Construction of Credit Risk Evaluation System for Small- and Medium-Sized Enterprises

Based on Principal Component Analysis and Logistic Model

Qiguang An
School of Finance
Shandong University of Finance and Economics
Ji'nan, China

Yuyang Zhao
School of Finance
Shandong University of Finance and Economics
Ji'nan, China

Abstract—**With the fast development of the domestic financial sector, the financing of SMEs has become the significant part of domestic credit, but due to the SMEs' stabilization of funds, technology and other aspects, SMEs are facing financial difficulties or loan default problems. This paper presents the financial and non-financial indicators combined with the evaluation system, using principal component analysis (PCA) and logistic regression (LR), a credit risk identification model is established. And then through the SPSS software to calculate the SMEs compliance probability, then analyze the impacts of SMEs compliance. Finally, some suggestions are put forward to improve the credit database, adjust the model according to the industry attributes, and strengthen the model test.**

Keywords—SMEs; PCA; credit risk; logistic regression

I. INTRODUCTION

A. Research Background

SMEs occupy a very important position in the national economy. The government introduced plenty of supportive policies to help SMEs, but the problem of "information asymmetry" in economics in the process of SME financing is always cannot get effective solution, namely commercial Banks unable to timely and accurate to receive the credit condition of SME actual, and the Banks "toxic assets", which seriously affect the normal operation of bank bad debts, or the bank won't for some SMEs to provide the funds needed for the operation. In general, in the absence of information, SMEs have financing difficulties, so the current society is in urgent need of a comprehensive evaluation system to measure the financial risk of SMEs.

B. Research Framework

The observation of the 60 small business financial statement analysis, put forward the financial and non-financial indicators of combining the evaluation system, using PCA and LR method to establish the credit risk identification model, and then through SPSS software to calculate the probability of default of SMEs, the analysis of the SMEs, many influence factors, finally from perfecting the credit database, attributes according to the industry adjustment model, put forward suggestion to strengthen the model test.

C. Literature Review

To construct credit risk evaluation system of SMEs, well-known scholars at home and abroad use different theoretical models and research methods to discuss it. Credit risk is often difficult to distinguish among various risks, so it is necessary to construct the credit evaluation system with the help of some mathematical statistics method. The earliest statistical method was the discriminant analysis proposed by Fisher [1] (1936), and Altman [2] (1968)'s "z-score" model was the typical application of this method in credit risk measurement. Currently, most scholars studying such problems at home and abroad use logistic regression analysis.

Earlier abroad began to use LR model to construct the SME credit risk evaluation index system is made by Martin [3]. He defined 58 distressed banks from about 5700 member banks of the Federal Reserve from 1970 to 1977, and selected 8 indexes such as total asset profit margin from 25 financial indexes to predict the probability of bankruptcy and default of the company, and established a LR model.

The earliest beginning to study SME credit risk evaluation index system of peace is Ren Yong-ping and Mei Qiang [4](2001), two scholars summed up the construction of system of principles, on the relationship between the enterprise overall credit and partial credit for processing, do a combination of the quantitative index and qualitative index, evaluation index and evaluation purposes. Wu Shi-nong and Lu Xian-yi [5] (2001) selected 70 samples of listed companies according to 1:1 ratio, selected 6 indicators from 21 financial indicators, and constructed three prediction models, among which Logit model was the best, with a prediction accuracy of 93.53%.Yu Li-yong and Zhan Jie-hui [6](2004) took the loan data of commercial Banks as samples and obtained a good prediction effect. This paper attempted to study the default probability with Logistic regression model, in order to provide a modeling method for quantitative measurement of credit risk. Liu Yi-wen [7](2010) constructed the high-tech SME credit evaluation system. He used the PCA method to make an empirical analysis of 31 high-tech SMEs in China, then determines the weight of

each factor in the index system, and analyzes and summarizes the main factors influencing the credit risk of high-tech SMEs in China.

In addition to discriminant analysis and logit model, there are also non-parametric methods and neural networks. There is a lot more advanced modern credit risk measurement model, such as the KMV [8] model based on option theory, and some statistical methods, such as neural network technology based on artificial intelligence [9], but these models and methods of operation process is complex and need large amounts of data, at present in our country is difficult to get practical application.

Recently, scholars more from the perspective of supply chain finance. Liu Yuan-liang and Gao Shu-li [10](2013) first puts forward the integrated industry, credit evaluation system of enterprises and the core elements such as using PCA and LR method to establish the credit risk identification model, and then by comparing the financial supply chain financing mode and traditional mode of small business credit probability, shows that the former can improve the small business credit, finally from perfecting the credit database, attributes according to the industry adjustment model, put forward suggestion to strengthen the model test. Fan Fang-zhi, Su Guo-qiang and Wang Xiao-yan [11](2017) established a three-party game model for Banks, core enterprises and SMEs, and analyzed the risk sharing principle of each participant in supply chain finance.

Based on the above literature research, this paper makes a more detailed division of the evaluation indicators, and better uses PCA and LR to build a credit risk evaluation system for SMEs.

II. CREDIT RISKS OF SMEs

A. Characteristics of Credit Risk

Credit risk has four main characteristics:

- Asymmetry: the expected return and loss are asymmetric. When a certain credit risk acts on a certain subject, the subject has asymmetric expected earnings and expected losses.

- Accumulative nature: the accumulative nature of credit risks refers to the characteristics of credit risks, such as chain reaction, vicious circle, continuous accumulation, and sudden outbreak of the financial crisis exceeded a certain critical point.

- Systematicness: caused by the uncertain macroeconomic factors. Driven by this, credit risk may evolve into systemic risk.

- Endogenous: credit risk is not completely driven by objective factors, but has the characteristics of subjectivity.

B. Causes of Credit Risk of SMEs

For the consideration of their own development, SMEs often need a large amount of capital to maintain a healthy business process, and the first thing enterprises seek is the form of direct financing. However, it is common for SMEs to go bankrupt due to poor management, so commercial Banks need to put the credit risk of SMEs in an important position, specifically for the following reasons:

First, the enterprise's own reasons, such as: 1) the enterprise management is not standard, some employees are not strong sense of responsibility, weak credit awareness, using the information asymmetry between the enterprise and the bank to cheat the bank loan, not timely return the loan, resulting in the loan overdue, thus bringing credit risk to the bank. 2) the operating situation is not stable. Many SMEs in China have backward production equipment, insufficient product innovation ability and limited market development ability, leading to the instability of upstream and downstream customers. When examining the loan qualification of SMEs, the future capital flow and repayment ability of enterprises are difficult to predict and monitor. 3) enterprise financial information is not standard. SMEs live in the absence of a reasonable financial system, resulting in financial statements information is not standard, opaque, accounting treatment is relatively arbitrary, income and profit with the owner's desire to transfer, cannot truly reflect the current financial situation of enterprises.

Second, the micro-factors inside the bank mainly include the disadvantages of the bank's own management level and management system, the incomplete risk control inside the bank, and the backward risk management technology of the bank itself.

III. CONSTRUCTION OF CREDIT RISK EVALUATION INDEX SYSTEM OF SMEs

A. General Construction Principles

Based on the index system of relevant literatures, this paper summarizes several general principles of system construction:

The combination of financial and non-financial indicators: most relevant literatures only use financial indicators to study the financial situation of SMEs, but ignore the influence of non-financial factors on the enterprises' credit risk. Therefore, this paper is committed to the use of financial and non-financial indicators combined with the index system.

The combination of qualitative and quantitative indicators: generally, the multivariate statistical model is constructed mainly with quantitative indicators.

The reason is that quantity can directly reflect the essential characteristics of things, but from the comprehensive consideration of system financial indicators, the selection of qualitative indicators should be considered, and qualitative indicators can be quantified or converted into quantitative indicators.

Comprehensiveness: comprehensiveness here refers not only to the completeness of financial indicators, but also to the comprehensiveness of the research involved in the system, including not only the operational factors of SMEs

themselves, but also industry factors and even the whole macroeconomic environment.

Hierarchy: in order to make the index system cover board aspects and show the composition of the evaluation index system in a scientific, reasonable and comprehensive way, we have established three levels of index respectively and carried out classification and summary step by step.

B. Selection and Analysis of Credit Risk Assessment Indicators for SMEs

According to the above system construction principles, the author selected 21 indicators as the main component analysis objects based on the consideration of various influencing factors, as shown in "Table I".

TABLE I.   CREDIT RISK ASSESSMENT INDICATORS FOR SMEs

| The target layer | Domain layer | Index layer | | | Index number |
		The first layer	The second floor	The third layer	
credit risk evaluation system for SMEs	SMEs themselves	financial index	solvency index	current ratio	X_1
				quick ratio	X_2
				asset-liability ratio	X_3
				interest coverage	X_4
			profitability index	operating profit margin	X_5
				cost profit margin	X_6
				return on total assets	X_7
			operating capacity index	inventory turnover	X_8
				accounts receivable turnover	X_9
				total asset turnover	X_{10}
				fixed assets turnover	X_{11}
			growth ability	main business growth rate	X_{12}
				operating profit growth rate	X_{13}
				net profit growth rate	X_{14}
		nonfinancial indicators	growth ability	leadership qualities	X_{15}
				the enterprise scale	X_{16}
			technical innovation ability	technological innovation achievements	X_{17}
	Industry factors	nonfinancial indicators	industrial production growth rate	the growth rate of production in the industry	X_{18}
			industry scale	the size of the industry	X_{19}
	Macroeconomic factors	nonfinancial indicators	GDP	gross domestic product	X_{20}
			legal policy environment	legal policy environment	X_{21}

1) Selection of financial indicators

It starts from the general principles of financial statement analysis, and different financial indicators are selected according to the four basic aspects of the enterprise's solvency, operating capacity, profitability and growth capacity, so as to reflect the health of the enterprise's financial status and the possibility of default.

2) Selection of non-financial indicators

Non-financial factor analysis is an analysis and research on the background of the debtor's financial factors and the background factors that affect the future development trend of its financial factors. It is mutually confirmed and supplemented with the financial factor analysis method, both of which are indispensable.

Compared with previous studies, the index system designed in this paper is special in that it places more emphasis on non-financial indicators than previous studies.

IV.   CREDIT RISK LR MODEL CONSTRUCTION OF SMEs IN CHINA

A. Establishment of Logistic Empirical Model

Given in this paper, which involves many variables proposed model, at the same time, considering the multivariate model leads to multicollinearity problem, before the establishment of logistic empirical model, need to dimension of index system, the principal component analysis, comprehensive analysis of 21 indicators, summed up the number of principal component index, so as to further establish the logistic empirical model.

The establishment process of logistic empirical model is as follows: the dependent variable of logistic regression is binomial classification variable, and P is the probability that SMEs abide by the loan repayment agreement. In this paper, the default probability is greater than or equal to 50% as a

sign of good credit standing. If the probability of an enterprise is less than 50%, this paper considers the enterprise as a high-risk enterprise with poor credit status.

B. Data Sources

This paper obtained the 2018 annual financial statement data of 60 sample companies from Wind database and other databases, including 20 ST or *ST listed companies and 40 companies with good credit, and calculated and processed the financial data to get the corresponding indicators. Since different variables have different units, to remove the possible influence of diverse dimensions and orders of magnitude between variables, this paper adopts Winsorize to process extreme values, and then carries out descriptive statistical analysis of each original data variable.

C. Principal Component Analysis Test

Before empirical analysis, KMO test and Bartlett test are used to examine the correlation between model variables.

And the inspection result is: KMO value is 0.612, it can be considered that there exist many mutual divisors among variables, which is fit for PCA; while the statistical value of Bartlett's sphericity test is 805.328, and the probability value of this statistic is significant and suitable for factor analysis.

D. Analysis of Empirical Results

1) PCA

PCA is a multivariate statistical analysis mean proposed by Karl Pearson. Its basic principle is to calculate the correlation of variables, and group them according to the correlation size, so that only variables in the same group can have high correlation.

So as to avoid the multicollinearity between model variables, this paper extracts 6 indicators through PCA of 21 indicators, and uses SPSS to obtain the total variance interpretation table, as shown in "Table II".

TABLE II.   TOTAL VARIANCE EXPLANATION TABLE

Composition	Initial eigenvalue			Sum of the extracted loads		
	total	percentage of variance	Cumulative %	total	percentage of variance	Cumulative %
1	3.636	30.198	30.198	3.636	30.198	30.198
2	2.629	24.604	54.802	2.629	24.604	54.802
3	2.249	12.492	67.295	2.249	12.492	67.295
4	1.818	10.097	77.392	1.818	10.097	77.392
5	1.503	6.351	83.743	1.503	6.351	83.743
6	.995	3.530	87.273	.995	3.530	87.273

SPSS was used to obtain the rotated component matrix, and the practical significance of principal component factors obtained by each indicator was obtained through factor loading, as shown in "Table III".

TABLE III.   ROTATED COMPOSITION MATRIX

	Resend Scale					
	1	2	3	4	5	6
VAR0001	.917	.114	.042	.242	.079	.015
VAR0002	.912	.147	.042	.211	.068	.066
VAR0003	-.754	.000	-.041	-.213	.072	.242
VAR0004	-.799	.143	.111	.127	-.107	-.004
VAR0005	.172	-.005	.080	.893	-.099	.037
VAR0006	.233	-.046	.018	.834	-.019	.130
VAR0007	.161	.038	-.155	.735	.498	.152
VAR0008	.348	.759	.031	-.119	.030	.143
VAR0009	.176	.936	.019	-.060	.078	.022
VAR00010	-.116	.859	.128	.234	-.040	-.103
VAR00011	-.117	.886	-.072	-.073	.173	-.043
VAR00012	.017	.078	.788	.108	.133	.164
VAR00013	-.016	.110	.830	.140	-.095	-.080
VAR00014	-.101	-.040	.872	-.224	.039	.067
VAR00015	-.057	-.055	.889	.258	.003	.051
VAR00016	-.069	.046	.876	-.008	-.104	.083
VAR00017	.150	-.142	.711	-.087	-.109	.333
VAR00018	.098	.209	-.047	-.147	.813	.004
VAR00019	-.215	-.033	-.365	-.024	.888	-.080
VAR00020	.009	.025	.081	.034	.067	-.834
VAR00021	.001	-.074	.033	.015	.014	.831

The economic significance of principal components is determined by the comprehensive significance of the indexes with the largest weight in each linear combination.

As can be seen from "Table III", FAC1_1, FAC2_1, FAC3_1, FAC4_1, FAC5_1 and FAC6_1 have relatively high factor loading values on different indicators, so the factors can be named accordingly. Current ratio, asset-

liability ratio, quick ratio and equity multiplier have high loading on FAC1_1. Therefore, FAC1_1 mainly reflects the solvency of the enterprise, which can be called as debt repayment factor. Total asset turnover, accounts receivable turnover, inventory turnover and fixed asset turnover have high loading on FAC2_1. Therefore, FAC2_1 mainly reflects the operating capacity of enterprises, which can be called as operating factor. Main business growth rate, net profit growth rate, operating profit growth rate, technological innovation results, enterprise size, leader quality these 6 variables have a high load on FAC3_1. Therefore, FAC3_1 mainly reflects the enterprise's development prospect and growth ability, which can be called as the growth ability. The three variables of operating profit rate, cost profit rate and return on total assets have a high load on FAC4_1. Therefore, FAC4_1 mainly reflects the ability of enterprises in terms of profitability, which can be called as profit factor. The two variables of industry production growth rate and industry

size have a high load on FAC5_1, so FAC5_1 mainly reflects the industry situation of the enterprise, which can be called as industry factor. The two variables of GDP and legal and policy environment have a high load on factor FAC6_1. Therefore, factor FAC6_1 mainly reflects the macroeconomic environment in which enterprises are located, which can be called as macro environment factor.

2) LR model

Based on principal component analysis, 6 principal components were analyzed by logistic regression model. Through the combination of the six principal components, it is found that when F_1, F_2, F_3, F_4 and F_6 are taken as independent variables, each variable has a strong significance and is suitable for the comprehensive index of the credit evaluation system. See "Table IV" for specific variable names.

TABLE IV.   THE VARIABLES IN THE EQUATION

	Steps	Score	Degrees of freedom	Significant
Variable	Debt service factor	.037	1	.007
	Operating factor	.147	1	.031
	Macroenvironmental factor	.215	1	.012
	The profit factor	1.238	1	.013
	Growth factor	.333	1	.042
	constant	-0.97	1	.025

In this paper, SPSS was used for calculation. After five iterative calculations, the final model retained five variables, namely F_1, F_2, F_3, F_4 and F_6. Therefore, at a significant

level of 5%, the relevant variables passed the test. Expression of regression model:

$$\ln\left(\frac{P}{1-P}\right) = -0.970 + 0.037F_1 + 0.147F_2 + 0.333F_3 + 1.238F_4 + 0.215F_6$$

The credit risk evaluation model is established according to the existing data, and the data of SMEs are brought into the model to predict the credit of enterprises. The P value obtained from the formula is the result of enterprise credit risk evaluation, $P \in (0,1)$. The higher the P value is, the higher the possibility of enterprise default is. The closer the p value is to 0, the worse the enterprise credit will be. The closer the p value is to 1, the better the enterprise credit will be.

E. Result Analysis

As for the sorting of samples' principal component scores, the method used in this paper is to make linear combination

of 6 principal components, the variance contribution rate of each principal component as weight.

Take the sample data of 60 enterprises into the regression model, and the prediction accuracy of the test model is shown in "Table V". For defaulting enterprises, the accuracy rate of enterprise prediction is 80%, while for non-defaulting enterprises, the accuracy rate of enterprise prediction is 90%, and the overall accuracy rate is 86.6%, indicating that the prediction accuracy rate of regression model is high. Therefore, the model can accurately evaluate the credit risk of SMEs in financial business.

TABLE V.   RESULT TEST

		Category	Regression model discriminant results		Total
			default	No default	
Back to the band	count	default	16	4	20
		No default	4	36	40
	The proportion	default	80%	20%	100%
		No default	10%	90%	100%
The overall accuracy rate was 86.6%					

Calculate the P value of the company's credit risk in 2018, in descending order. Some results are shown in "Table VI":

TABLE VI.   RANKING OF CREDIT RISK P VALUE

Corporate name	FAC1_1	FAC2_1	FAC3_1	FAC4_1	FAC6_1	Composite score	P
Netsun	3.09981	4.06700	0.43029	-0.88267	0.67442	1.887643939	0.79688
Sinobioway Biomedicine	3.49420	-0.56927	1.04248	2.17978	0.51183	1.544931379	0.76645
Aisidi	-0.97492	5.16727	1.61730	-0.41166	-1.02031	0.962802864	0.76218
Huaming Equipement	0.89271	-0.70205	0.09662	1.13729	0.87279	0.422205890	0.74326
Juli Culture	0.21415	1.73975	-1.32410	1.22981	1.76337	0.613530687	0.73887
Jiaying Pharmaceutical	0.45588	-0.29114	1.32595	1.33458	-0.80145	0.430497063	0.72884
Zhongheng Electric	3.74870	-0.83249	0.38569	-0.50204	-0.61491	0.884845335	0.71830
Changjiang Runfa	0.33104	0.82861	-0.89071	0.05201	2.22890	0.407637405	0.71053

P value represents the compliance probability of financing enterprises. P value approaches 1 means that the credit level of financing enterprises is high. P value approaches 0 means that the credit level of financing enterprises is low.

V.   THE APPLICATION OF THE NEW INDEX SYSTEM

A. Government: Establishing and Perfecting the Credit Legal System

The state should accelerate the construction of credit investigation legislation. The operation of relatively mature economic system is based on perfect social credit legislation. To have an effective credit investigation system, the relevant laws and regulations are necessary. We should accelerate the legislative construction, constantly improve the laws and regulations based on the regulations on the administration of credit investigation, and promote the promulgation of relevant laws as soon as possible. Financial institutions should be urged to use credit data legally and reasonably, to ensure the authenticity and effectiveness of credit information, and to protect the legitimate rights and interests of the people being investigated. In order to improve the cost of breach of trust, the whole society should set up the mechanism of breach of trust, so that market actors are not only limited to the immediate temporary benefits, but also take a long-term view and pay more attention to credit. Furthermore, financial institutions can determine whether to issue loans and reduce non-performing assets by inquiring the credit status of the borrowers.

B. Commercial Banks: Improving Internal Constraint Management Mechanism and Making Full Use of Credit Data

Constantly improve the internal management mechanism about east. At present, there is a phenomenon that commercial Banks in China "attach importance to application and neglect management", which leads to potential risks. They blindly use the credit evaluation system and ignore the losses caused by the neglect of internal management. Commercial Banks should not only attach importance to the use of credit evaluation system, but also strengthen internal management and constantly improve the internal management constraint mechanism while mastering the application of the system. First of all, the management should be aware of this clearly, to master and credit system

internal management at the same time. Secondly, the credit investigation business department should regularly conduct detailed investigation on the credit investigation business system, continuously improve and perfect the system loopholes and improve the security of the system. Finally, the credit investigation department should strictly supervise the legal use of the system by commercial Banks and implement various rules and regulations.

In addition, I actively communicated with the credit information center of the People's Bank of China, regularly fed back the problems and difficulties encountered in the use of the credit information management system, sought guidance and help from the People's Bank of China, and constantly improved my technical and management defects. Regularly assign professionals to the central bank to learn advanced experience and technology, always keep up with the advanced level.

C. Enterprises: Increasing the Skills Training of Relevant Personnel

In this paper, PCA is adopted as the analysis method of index selection, and logistic model is used to establish the model. The mastery of the above analysis methods and models influences the popularization and application of the new index system. To make this index system play a better role in the process of evaluating credit risks of SMEs and avoid the problems of using it incorrectly, the evaluators should be well trained in relevant knowledge to ensure that the new index system can be effectively popularized and applied.

VI.   CONCLUSION

Scientific and accurate evaluation of enterprise credit risk is the key to the healthy operation of SME financing. Based on the principal component analysis of the financial and non-financial indexes of 60 listed SMEs, this paper selects the index factors that can best reflect the credit risk of enterprises, establishes the credit risk evaluation system of SMEs, simulates and verifies the construction of the credit risk evaluation model in the form of empirical research, and draws the following conclusions:

The credit risk model LR has a high accuracy in the overall credit risk prediction of China's listed SMEs.

For SMEs, their credit risk is determined not only by the company's financial situation, but also by macroeconomic

indicators such as the GDP of the province, the growth rate of industrial production and the legal and policy environment.

Among the financial indicators of a company, operating profit margin, cost profit margin and return on total assets are the most important indicators that determine the financial condition of the company.

REFERENCES

[1]   Fisher R A. The Use of Multiple Measurements in Taxonomic Problems [J]. Annual Eugenics vol 7 1936: 179-188.

[2]   Altman E I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy [J]. Journal of Finance, September 1968: 589-609.

[3]   ARTIN D. Early Warning of Bank Failure:A Logit Regression Approach [J]. Journal of Banking and Finance, 1977(3):249-27.

[4]   Ren Yong-ping, Mei Qiang. Discussion on Credit Evaluation Index System of SMEs [J]. Discussion on Modern Economy, 2001(4):60-62.

[5]   Wu Shi-nong, Lu xian-yi. Prediction Model of Financial Distress of Listed Companies in China [J]. Economic Research, 2001(6):46-56.

[6]   Yu li-yong, Zhan Jie-hui. Prediction of Default Probability Based on Logistic Regression Analysis [J]. Financial Research, 2004(9):15-23.

[7]   Liu Yi-wen, Xia En-jun, He Ming. Credit Risk Assessment of High-tech SMEs Based on Financial Indicators [J]. Management and Technology of SMEs, 2010, 2010(4):59-60.

[8]   Zeng Shi-hong, Wang fang. Credit Risk Evaluation of Listed Manufacturing Companies Based on KMV Model [J]. Prediction. 2013(02).

[9]   Wu Yan. Construction of Credit Evaluation System of Small and Medium-sized Enterprises Based on Neural Network [J]. Science and Education Collection (the following ten-day issue). 2013(06).

[10]   Liu Yuan-liang, Gao Shu-li. Credit Risk Identification of Small Enterprises in the Mode of Supply Chain Finance — An Empirical Study Based on Credit Data in Beijing [J]. New finance, 2013(1):45-49.

[11]   Fan fang-zhi, Su Guo-qiang, Wang Xiao-yan. Credit Risk Assessment and Risk Management of SMEs under the Supply Chain Finance Model [J]. Journal of Central University of Finance and Economics, 2017(12).