International Journal of Security and Its Applications
Vol. 11, No. 6 (2017), pp.35-52
http://dx.doi.org/10.14257/ijsia.2017.11.6.04

Credit Risk Rating System of Small Enterprises Based on the Index Importance*

Chi Guotai[1], Zhao Zhichong[1**] and Mohammad Zoynul Abedin[1, 2]

[1]*Faculty of Management and Economics, Dalian University of Technology, Dalian 116024, China*
[2]*Department of Finance and Banking, Hajee Mohammad Danesh Science and Technology University, Dinajpur 5200, Bangladesh*
zhaozhichong0829@163.com

Abstract

The main criteria to establish the credit risk evaluation index system is the indicators default identification ability. There is mutual influence between indices, a single index which has the default identification ability, but if put this indicator into the index system, and it will no longer have the default identification ability because of the impact of other indicators. This study therefore deletes the indicators of repeated information using colinearity diagnostics, and determines the order of indicators into the index evaluation system by calculating the score statistic of every indicator. We established credit risk evaluation index system of small businesses, including 14 indicators, such as cash ratio, the corporate credit situation nearly 3 years, by extracting the related data of 28 regional commercial bank branches of China, and the judgment accuracy of default and non default samples is 99.0%.

Keywords*: Credit risk; credit rating; Index system; Default identification ability; index importance; small enterprises*

1. Introduction

The essence of the credit rating is the default risk rating, which evaluates the possibility of repayment of one debt. There are not perfect credit risk index systems of small business for commercial banks because of small enterprises' high risk, small amount, and untrue financial data, *etc.* However, perfect an index seems and however high frequency uses in actual, the index is useless in credit rating if it cannot effectively distinguish default and non default customers. Therefore, the main criteria to establish the credit risk evaluation index system is the indicator have default identification ability.

A credit risk evaluation index system is composed of a number of indicators. There is mutual influence between indices, a single index has the default identification ability, but if put this indicator into the index system, and it will no longer have the default identification ability because of the impact of other indicators. So the establishment of credit risk evaluation index system should not only ensure the single index has default identification ability, but also ensure every index has default identification ability considering the interaction between indicators [1].

There are two difficulties in constructing the credit risk evaluation index system, on the first hand, if added a newly index into system which will lead to the original indicators become not effectively distinguish default and non default customers because of the mutual influence between indices. Second, In the process of considering the interaction between indices, the order of the index entry system has great influence on the final

* Authors contributions are equal
**Corresponding Author

ISSN: 1738-9976 IJSIA

selection result. Therefore, the existing index screening methods ignored the interaction between indicators, which only disicriminant a single indicator has default identification ability or not, which cannot ensure all the indices in indicator system have default identification ability

In order to overcome the above shortcomings, this paper established a credit risk evaluation index system of small enterprises based on the index importance. Using a Chinese regional commercial bank's 3045 small enterprises loan for empirical study, and the result shows the credit risk evaluation index system including 14 indicators, such as cash ratio, the corporate credit situation nearly 3 years, and the judgment accuracy of default and non default samples is 99.0%. The proposed model performs well on the evaluation customers default state.

The rest of the paper is organized as follows: Section 2 is the literature review. Section 3 introduces the methodology of this paper. Section 4 presents the empirical study. We conclude the paper in Section 5.

2. Literature Review

There are many credit risk evaluation index systems. Standard & Poor's Ratings [2] from two aspects that the business situation and financial factors to build an enterprise credit risk evaluation index system. Fitch Ratings [3] built a credit rating system mainly from the corporate structure, corporate profitability and corporate strategy aspects. Psillaki *et al.* [4] constructed the enterprise credit risk evaluation index system including 16 indicators such as the pretax profit, the employee number, which can use to discriminate the enterprise's default state. Maik *et al.* [5] set up a business credit risk rating index system including enterprise prospect, annual product sales volume. Industrial and Commercial Bank of China [6] established a credit evaluation index system for small business customers, including asset liability ratio, the basic situation of legal representative, the industry sentiment index.

In the indicator screening and evaluation methods, there are many researches, for example, Guotai *et al.* [7] using multiple hybrid screening index which reflects the information repeated of index, and using logistic regression discriminant the key factors which can significant distinguish the default and non-default customers. Shi and Chi [8] have established the index system of credit evaluation for framer loans using the method of Significant discriminant. Fan and Zhu [9] established the credit rating system for SMEs through the correlation analysis delete redundant index, and by discrimination analysis retain the indicators which have big identify ability. Abedin *et al.* [10] established Chinese small business credit risk evaluation system including 81 indicators, which using filter analysis to delete duplicate index, and employing significant discriminant method screened indicator which cannot significant distinguish the default state.

Although the existing researches have made great progress in constructing credit risk evaluation index system, there are still some drawbacks. First of all, the existing index screening methods ignored the interaction between indicators, which only disicriminant a single indicator has default identification ability or not, which cannot ensure all the indices in indicator system have default identification ability under the comprehensive consideration of the index mutual influence. Secondly, the existing studies did not consider the order of indicators into the index evaluation system, which will result in the big information content indices are deleted by random determining the order.

3. Methodology of Establishment Credit Risk Evaluation Index System

3.1. Data Standardization

Data standardization means change the original data of indicators into the [0,1] range, which will not only eliminate the impact of the dimensional difference, but also change the qualitative indicator that cannot be directly used into the quantitative indicator.

The quantitative indices include positive indices, negative indices and interval indices. The positive indices are the values bigger, the better; the negative indices are the values smaller, the better. And, the interval indices are indicators which are logical only when their values lie in certain intervals. Let x_{ij} denote the standardization score of the i^{th} indicator and the j^{th} customer. Let v_{ij} denote the original data of the i^{th} indicator and the j^{th} customer. Let m denote the number of customers. Let q_1 denote the left boundary of the ideal interval. Let q_2 denote the right boundary of the ideal interval. The standardization equations of the positive, the negative and the interval indices are shown as Eq. (1) to (3) respectively.

$$x_{ij} = \frac{v_{ij} - \min\limits_{1 \le j \le m}(v_{ij})}{\max\limits_{1 \le j \le m}(v_{ij}) - \min\limits_{1 \le j \le m}(v_{ij})} \tag{1}$$

$$x_{ij} = \frac{\max\limits_{1 \le j \le m}(v_{ij}) - v_{ij}}{\max\limits_{1 \le i \le m}(v_{ij}) - \min\limits_{1 \le j \le m}(v_{ij})} \tag{2}$$

$$x_{ij} = \begin{cases} 1 - \dfrac{q_1 - v_{ij}}{\max(q_1 - \min\limits_{1 \le j \le m}(v_{ij}), \max\limits_{1 \le i \le m}(v_{ij}) - q_2)}, v_{ij} < q_1 & (a) \\ 1 - \dfrac{v_{ij} - q_2}{\max(q_1 - \min\limits_{1 \le j \le m}(v_{ij}), \max\limits_{1 \le i \le m}(v_{ij}) - q_2)}, v_{ij} > q_2 & (b) \\ 1, \qquad q_1 \le v_{ij} \le q_2 & (c) \end{cases} \tag{3}$$

Table 1. The Standard of Qualitative Index

(1) No.	(2) Index	(3) Options properties	(4) Mark
1	Credit card record	1) Credit card no default record	1
2		2) there is default record of credit card or data missing	0
3	Stay in his post time	1) more than 8 years	1
4		2) more than 5 years and less than 8 years	0.7
5		3) more than 2 years and less than 5 years	0.4
6		4) less than 2 years or data missing	0
…	…	…	…
8	Enterprise tax record	1) tax history more than 3 years and no tax arrears records	1
9		2) tax history less than 3 years and no tax arrears records	0.75
10		3) onetime tax arrears record, and paid in full later	0.5
11		4) no tax record	0.25
12		5) more than two time tax arrears records or data missing	0

By rational analysis and expert investigation for qualitative indices, the scoring standard of qualitative indices can be obtained, which are shown in Column 2 to 3 of Table 1.

3.2. The First Index Screening Based on Colinearity Diagnostics

The objective of the first index screening is deleting the repeated information indices by co-linearity diagnostics, which can avoid the redundancy of credit risk evaluation index system. The steps of colinearity diagnostics, (i) Building regression equation, *e.g.*, x_{ij} denote the standardization score of the i^{th} indicator and the j^{th} customer, n denote the number of indices, a_i denotes the estimated parameters, and the regression equation between the i^{th} indicator and other n-1 indicators is as follows.

$$x_{ij}=a_0+a_1x_{1j}+\ldots+a_{i-1}x_{i-1,j}+a_{i+1}x_{i+1,j}+\ldots+a_nx_{n,j}+u_j \tag{4}$$

The estimated values a_i can be obtained by using least squares estimation in Eq. (4). Substituting these parameters a_i into Eq. (4), the estimated value $\hat{x}_{ij}$ of x_{ij} can be obtained.

(ii) The variance inflation VIF reflects the correlation between the index and the rest of indices. The bigger the VIF is, the more serious the multiple co-linearity between indices. Let VIF_i denotes the variance inflation factor of the i^{th} indicator, R_i^2 denotes the determination coefficient of the i^{th} indicator, $\overline{x}_i$ denotes the mean value of the i^{th} indicator, $\hat{x}_{ij}$ denotes the least squares estimated value of the i^{th} indicator and the j^{th} customer, then

$$R_i^2=\sum_{j=1}^{m}\left(\hat{x}_{ij}-\overline{x}_i\right)^2 \Big/ \sum_{j=1}^{m}\left(x_{ij}-\overline{x}_i\right)^2 \tag{5}$$

$$VIF_i=1\big/\left(1-R_i^2\right) \tag{6}$$

The bigger determination coefficient R_j^2 is, the bigger VIF_i, which means the information of the i^{th} indicator reflection can be replaced by other n-1 indicators, so the i^{th} indicator should be deleted.

(iii) If the variance inflation factor VIF_j is greater than 10, it indicates that there is a multi colinearity between the i^{th} indicator and the rest of indicators, and delete the largest VIF indicator in all indicators whose VIF greater than 10.

(iv) Cycle of step (i) to step (iii) until all the variance inflation factor are less than 10, the colinearity diagnostics ended.

3.3. The Second Index Screening Based on the Index Importance

3.3.1. Determination of the Importance of Index

The objective is ranked the indices according to the importance of index, which determined the order of indices into the evaluation index system.

Let S_i denotes the score of the i^{th} indicator; x_{ij} denote the standardization data of the i^{th} indicator and the j^{th} customer; m denote the number of customers; y_j denotes the default state of the j^{th} customer, y_j=1 represent default, and y_j=0 represent non-default; $\overline{y}$ denotes the mean value of default state for m customers; $\overline{x}_i$ denotes the mean value of m customers for the i^{th} indicator; so the Score calculation formula of the i^{th} indicator is:

$$S_i=\frac{\left[\sum_{j=1}^{m}x_{ij}\left(y_j-\overline{y}\right)\right]^2}{\overline{y}\left(1-\overline{y}\right)\sum_{j=1}^{m}\left(x_{ij}-\overline{x}_i\right)^2} \tag{7}$$

The Eq. (7) refers that if the characteristics difference is smaller among different customers for the i^{th} indicator, and if default distribution is discrete for different customers, then the Score S_i is bigger, which represents the index is more important and has the priority into the index system to be selected.

3.3.2. The Second Index Screening

The objective of the second index screening is reserving the index that can significant distinguish default and non-default customers, which means the index has default identification ability under the premise of considering the interaction between indicators.

(1) The principle of screening index based on logistic regression model

Let $P(y=1)$ denotes the default probability of the j^{th} customer; z_j denotes the Latent variables; x_{ij} denotes the standardization score of the i^{th} indicator and the j^{th} customer; m denotes the number of customers; n denote the number of indices; α denotes the constant; β_i denotes the regression coefficient of the i^{th} indicator; ε denotes the random error term. The logistic regression model is as follows.

$$P(y=1)=\frac{1}{1+e^{-z_j}} \tag{8}$$

$$z_j=\alpha+\sum_{i=1}^{n}\beta_i x_{ij}+\varepsilon \tag{9}$$

The regression coefficient β and its standard error SE_β can be obtained by using maximum likelihood estimation in Eq. (9), this process can be realized by SPSS software.

Assuming W_i denotes the *Wald* statistic value of the i^{th} indicator; $\hat{\beta}_i$ denotes the estimated coefficient of the i^{th} indicator; $SE_{\beta i}$ denotes the standard error of coefficient β_i. Thus, the *Wald* statistic is given by

$$W_i=\left(\hat{\beta}_i \big/ SE_{\beta i}\right)^2 \tag{10}$$

The Eq. (10) refers that the coefficient of the i^{th} indicator β_i is or not equal to zero by constructing *Wald* statistic W_i, which can determine the indicator has or not significant effect on customers' default state y_j .

Given a significance level α=0.05, and get the threshold $\chi_{0.05}^2(1)$ =3.841 by checking the χ^2 distribution table under the condition of freedom equal to 1. If the value of *Wald* statistic W_i are more than 3.841, then the i^{th} indicator has significant effect on customers' default state.

Therefore, ranking the indices according the Score that calculated in 3.3.1, the *Score* S_i is bigger, which represents the index is more important and has the priority into the index system to be selected.

Taking the index which has the biggest Score as independent variable, and the default state as the dependent variable, the logistic regression model was established. We can test the index has default identification ability or not according to the steps 1 to 3 in (1) above. If the index has default identification ability, turn to Step 3, else delete the index, and repeated the process according the Score from big to small until the first remaining index has default identification ability.

Assuming the two indices as independent variables which are the remained index in Step 2 and the new index has the biggest Score in remaining indices, and the default state as the dependent variable, the logistic regression model was established. We can test the new indicator has default identification ability or not according to the steps 1 to 3 in (1) above, and delete the index if the index has not default identification ability. Then repeat this process by increasing the index which has biggest Score in remained indices. If the new index has default identification ability, we need test the index which remained in

Step 2 still has default identification ability or not, if has default identification ability, retained the index, else deleted the index.

The process of index screening can be attributed to the following three cases:

First, the coefficient of the new addition index is not significant, that is, the index has not default identification ability, delete it directly. Second, the coefficient of the new addition index is significant, and the retained index in index system is still significant, retain all the significant indices. Third, the coefficient of the new addition index is significant, but the retained index in index system is not significant, delete the index which is not significant. It should be noted that we directly delete the index not significant is reasonable. The first enter into the index system is more importance, but the importance is based on the degree of dispersion of index data, which does not mean the index has big default identification ability.

Cycle of Step 2 to Step 3 until every index in index system has significant default identification ability.

3.4. The Rationality Test of Index System Based on ROC Curve

(1) The prediction of the default probability. Take the index which retained by 3.2-3.3 as independent variable, the default probability P (y=1) prediction model was established according to the Eq. (8) – (9), which was showed in Eq. (11).

$$P(y=1)=\frac{1}{1+e^{-z_j}} \tag{11}$$

(2) The classification of model identify results. Contracted the calculated default probability $P(y=1)$ with the real default state of customers, if the default probability $P(y=1)\geq 0.5$, the customers are discriminated default, else $P(y=1)<0.5$, the customers are not default. The classification result by contracting predicted default state and real default state is shown in Table 2.

(3) The construction of ROC curve. According to the classification results in Table 2, the two variables are defined, which are the horizontal and vertical coordinates of the ROC curve. Vertical coordinate: Also known as the true positive rate (TPR), which is the ratio of predict the correct default sample TP accounted for the total sample (TP+FN), with the formula expressed as:

$$\text{TPR=TP/(TP+FN)} \tag{12}$$

Horizontal coordinate: Also known as the false positive rate (FPR), which is the ratio of wrong predict sample TP that non default customers are predicted default, accounted for the total sample (TP+FN), with the formula expressed as:

$$\text{FPR=FP/(FP+TN)} \tag{13}$$

Table 2. The Classified Result of Predict Model

Actual default state	Predicted default state	
	1 (default)	0 (non-default)
1 (default)	The number of actual default customers are judged to be correct True Positive(TP)	The number of actual default customers are judged to be wrong False Negative(FN)
0 (non-default)	The number of actual non-default customers are judged to be wrong False Positive(FP)	The number of actual non-default customers are judged to be correct True Negative(TN)

(4) The rationality test. Computing the area under curve AUC below the ROC curve

establishment in (3), which value is between 0.5–1. The AUC is more close to 1, the default identification effect is better. If AUC=1, means the predicted results are entirely consistent with actual state and this is the most ideal situation. If AUC＞0.9, the discriminant model has a high accuracy, and this means the predicted model passed the rationality test.

3.5. Credit Grade Division

This paper will be divided into AAA, AA, A, BBB, BB, B, CCC, CC and C nine credit grades, from AAA to C credit rating from high to low, representing customer credit station from good to worse.

(1) Calculate the customer's credit score S. The Eq. (11) represents the default probability P (y=1), the bigger the default probability, the worse the customers' credit status, and corresponding the lower the customer's credit score. We can get the credit score in Eq. (14).

$$S=100\times(1-P(y=1)) \tag{14}$$

(2) Primary division of credit rating. Customers are ordered according to the credit score from high to low. Customers are divided to nine grades; there is at least one default customer in each grade. We can get the loss given default (LGD) of each grade according to the data of customer's "receivable principal and interest" and "uncollected principal and interest".

(3) Adjusting the credit rating. Constantly adjust the boundaries of each credit rating, that is, to adjust the number of customers per level, until the credit rating to meet the Default Pyramid standard. Default Pyramid standard means the higher of credit rating, the lower the LGD of customers. In the course of the adjustment of the level, once a level of customer number changes, the number of customers adjacent level will change, the corresponding LGD will change. The actual Adjustment process is more complex, and does not belong to the scope of the study, this paper will not repeat them. The detail adjusting process is showed in invention patent which is authorized by Patent Office of the People's Republic of China.

4. Empirical Study

4.1. Audition of Indices and Data Obtain

(1) Audition of indices

Focus on the high-frequency indices from the typical international rating agencies such as Standard & Poor's, Moody's and Fitch Fitch, and the major domestic financial institutions such as China Construction Bank and Industrial and Commercial Bank of China, combined with classical literature at home and abroad, we built the credit risk audition index system of small enterprise, and selected 107 indices which are shown in Table 3. By observation principle, we delete the source of repayment and other indices which are unable to access the data, and eventual established 81 indices which covered 7 criteria layer like internal corporate financial factors, internal corporate non-financial factors, external macro environment, the basic situation of legal representative, the credit status of enterprise, the business reputation and mortgage collateral security factor. Data cannot be observed in column 8th in table 3 with "Unavailability delete" mark.

Table 3. Index Set of Small Businesses Credit Risk Evaluation Chosen Extensively

(1) No.	(2) Criterion Layer		(3) Indicator	(4)Indicator type	(5) Index source	(6) Final result of screening
1	Internal corporate financial factors	Solvency	Asset liability ratio	Negative	[3],[5-6], [8]	Deleted by significance
…			…	…	…	…
29		Profitability	Net assets return ratio	Positive	[3-6], [8], [10]	Deleted by significance
…			…	…	…	…
45		Operating Capacity	Accounts receivable turnover rate	Positive	[4-8], [11-12]	Deleted by significance
…			…	…	…	…
55		Growth Capacity	Operating income growth rate	Positive	[4],[7-9], [11], [13]	Deleted by significance
…			…	…	…	…
64	Enterprise external macroeconomic conditions		Consumer price index	Positive	[3-6], [10-13]	Reserved
…			…	…	…	…
73	Internal corporate non-financial factors		Entire period of actual operation	Qualitative	[6], [8-10], [12-13]	Reserved
…			…	…	…	…
86	Basic situation of the legal representative		education	Qualitative	[8], [10-12]	Reserved
…			…	…	…	…
99	Basic situation of credit enterprise		Category of registered capital	Qualitative	[3-6],[10],[12]	Reserved
…			…	…	…	…
102			Customer complaint rate	Qualitative	[12]	Unavailability delete
103	Enterprise commercial credit		Tax records	Qualitative	[3],[8-10],[13]	Deleted by colinearity diagnostics
…			…	…	…	…
106			Number of breach of contract	Qualitative	[3-5],[9],[10-12]	Deleted by significance
107	Pledged collateral factors		Mortgage / pledge / guarantee	Qualitative	[3-6],[9-13]	Reserved

(2) Database

Using a Chinese regional commercial bank's 3045 small enterprises loan for empirical study, involved nearly 20 years' data from branches of Beijing, Tianjin, Shanghai, Chongqing and other 28 cities, in which there are 2995 non-default small enterprise and 50 default small enterprise. The data can be obtained in accordance with the order of the 81 indicators in order to $X_1, X_2, \ldots, X_{81}$ annotation, showed in column *b* in Table 4. Table 4 constituted by 2 parts, the first part is the original data showed in column 1-3045, remember to matrix (v_{ij}); the second part is the standardization data showed in column 3046-6090, recorded for matrix (x_{ij}), the process of standardization are showed in 4.2.

Table 4. The Original Data and Standardized Data of Small Enterprises' 81 Indicators

(*a*) No.	(*b*) Indicator	Original data of 3045 customers v_{ij}			Standardized data of 3045 customers x_{ij}					
					50 default customers			2995 non-default customers		
		(1) customer 1	…	(3045) customer 3045	(3046) customer 1	…	(3095) customer 50	(3096) customer 51	…	(6090) customer 3045
1	X_1 Asset liability ratio	0.523	…	0.603	0.454	…	0.654	0.000	…	0.369
2	X_2 Net cash flow ratio from current liabilities operating activities	-0.054	…	0.136	0.472	…	0.461	0.000	…	0.496
3	X_3 Quick ratio	0.000	…	0.640	0.000	…	0.000	0.682	…	0.073
…	…	…	…	…	…	…	…	…	…	…
48	X_{48} Retained earnings growth rate	0.888	…	0.888	0.519	…	0.501	0.503	…	0.513
49	X_{49} Industry cycle indicator	134.750	…	123.300	0.695	…	0.700	0.695	…	0.579
…	…	…	…	…	…	…	…	…	…	…
76	X_{76} Enterprise credit in 3 years	Have credit record、no default	…	Have credit record、no default	0.000	…	0.000	0.000	…	1.000
…	…	…	…	…	…	…	…	…	…	…
81	X_{81} Score of pledged collateral	Self-built office building mortgage	…	No mortgage and pledge	0.669	…	0.649	0.100	…	0.570
82	Default or non-default y_i	1.000	…	0.000	1.000	…	1.000	0.000	…	0.000

4.2. Standardization of Indices Data

(1) Data standardized of quantitative indices. For the data matrix (v_{ij}) in 1-81 rows and 1-3045 columns in Table 4, each data v_{ij} represents the original data of the i^{th} index for the j^{th} customer. Among them, we can find a maximum and a minimum value from 3045 data in each row, that is max(v_{ij}) and min(v_{ij}) which needed in formula (1) – (3).

The original data can be standardized according to the types of indices in 7$^{\text{th}}$ column in Table 3, that means take the original data v_{ij}, maximum value max (v_{ij}) and minimum value min (v_{ij}) substituted into the corresponding types formula. The standardization data are showed in column 3046 – 6090 in Table 4. There are two interval type indicators in the 81 indices, respectively, the consumer price index and age. The best interval of the age index is [31, 45], the best range of consumer price index is [101, 105]. Take the original data v_{ij} in Eq. (3), the standardized data x_{ij} can be obtained.

(2) Data standardization of qualitative indices. According to the standardized method of qualitative indices in 3.1 (4), the indices value are changed into [0, 1] range. Take "X_{74} Working time holding the position" from row 74 in Table 4 as an example, the data in row 74 column 1 in Table 4 is No, so the standardizing data is zero according to the standard of scoring in row 6 column 3 in Table 1, which is showed in row 74 column

3046 in Table 4. Similarly, the other customers' standardization data are obtained and showed in row 74 in Table 4.

4.3. The Establishment of Credit Risk Evaluation Index System

4.3.1. The First Index Screening

We build a multiple linear regression model by taking asset-liability ratio X_1 in Table 4 line 1 as the dependent variable and the remaining 80 indicators in Table 4-81 line 2 as independent variables. So we put the data from line 1 column 3046-6090 of Table 4 into

Table 5. The Screening Process of the Multicolinearity

(1)No.	(2) Indicator	(3) VIF_j 81 indices	(4) VIF_j delete X_9	(5) VIF_j delete X_9 X_{29}	…	(10) VIF_j delete X_9 X_{29} X_{19} X_{30} X_{43} X_{72} X_{71}	(11) VIF_j delete X_9 X_{29} X_{19} X_{30} X_{43} X_{72} X_{71} X_6	(12) Result
1	X_1 Asset liability ratio	238.6	5.83	5.80	…	5.80	5.78	Reserved
…	…	…	…	…	…	…	…	…
6	X_6 Earnings before interest and taxes /Current liabilities	11.82	11.80	11.79	…	**10.51**	—	Deleted
7	X_7 Full capitalization rate	4.38	4.36	4.36	…	4.29	4.27	Reserved
9	X_9 Shareholder equity ratio	**242.1**	—	—	…	—	—	Deleted
…	…	…	…	…	…	…	…	…
29	X_{29} EBITDA	92.66	**92.66**	—	…	—	—	Deleted
…	…	…	…	…	…	…	…	…
70	X_{70} Corporate credit situation nearly 3 years	9.76	9.76	9.76	…	6.17	6.16	Reserved
…	…	…	…	…	…	…	…	…
81	X_{81} Score of pledged collateral	1.48	1.48	1.48	…	1.48	1.48	Reserved

the left side of Eq. (4), the data from line 2-81 column 3046-6090 into the right side of the equation (4), and estimate coefficients of Eq. (4) by using the least squares estimation method, and get a equation as follows:

$$X_1=-0.014-0.033\times X_2-0.001\times X_3+\ldots+0.000\times X_{81} \tag{10}$$

By putting the data from line 2-81 column 81-6090 of Table 4 in Eq. (10) respectively, we get 3045 estimated value of the X_1 which is the estimated value of 3045 samples of first index $\widehat{X}_1$, satisfying with: $\widehat{X}_1=\left(\widehat{x}_{11},\widehat{x}_{12},\cdots,\widehat{x}_{1,3045}\right)$. Determinate the average $\overline{x}_1$ of data from line 1 column 6090-3046 in Table 4, and put the standardized data x_{1j} of the first indicator X_1, the average $\overline{x}_1$ and estimated value $\widehat{X}_1$ into equation (5), we get the determination coefficient $R_1^2=0.996$ of first indicator X_1, and bring R_1^2 into equation (6), we get the variance inflation factor $VIF_1=238.58$, and list in Table 5 line 1 3 columns.

Similarly, by taking the second index X2 as dependent variable and $X_1, X_3, \ldots, X_{81}$ as the independent variable, we build a multivariate regression model, and determine variance inflation factor $VIF_2=72.59$ of index X2, list in line 2 column3 of Table 5. Determine the variance inflation factor of other 79 indicators, and list the results in corresponding line column 3 of Table 5. The variance inflation factor of variable X1, X2, ... , the X72 are all more than 10, which shows that there are serious multicolinearity between these variables. To eliminate multicolinearity we should remove a variable which has the biggest *VIF* and also is bigger than 10, that is to say, remove index X9 which is shareholders' equity ratio.

After deleting index X9, for the rest of the 80 indicators, repeat the above (1)-(3) process, delete index which is based of the largest *VIF* and more than 10, namely indicators X_{29}EBITDA. Similarly, proceeding colinearity deletion of other indicators, until all *VIF* of the index values is less than 10. For example, after deleting eight indices such as X9 from line 11of Table 5, there remain 73 indicators whose variance inflation factor *VIF* is less than 10, so there are no multicolinearity between 73 indicators.

4.3.2. The Order of Indices According to the Importance

Determine the average of the default state y_i of the last line columns 3046-6090 of Table 4, because when default, it is y=1, if not y=0, there are 50 samples of default in 3045 clients, that means there are 50 samples where y=1, y is 0 in the remaining 2995 samples, so the average of default states of 3045 small business is 0.0164.

Table 6. The Score and Rank of 73 Indices

(1) No.	(2) Indicator	(3) Score S_i	(4) Rank
1	X_1 Asset liability ratio	13.505	26
2	X_2 Net cash flow ratio from current liabilities operating activities	12.667	28
3	X_3 Quick ratio	4.528	45
…	…	…	…
64	X_{70} Corporate credit situation nearly 3 years	78.256	4
65	X_{73} Enterprises operating conditions	21.449	17
…	…	…	…
70	X_{78} Consumer price index	168.372	1
71	X_{79} City residents' disposable income	105.425	3
72	X_{80} Engel coefficient	133.408	2
73	X_{81} Score of pledged collateral	24.965	14

We determine $\overline{x}_i$ by calculating average of index from each row of column 6090-3046 in Table 4, and put average $\overline{x}_i$ of *i-th* index、 average $\overline{y}$ of default state and standardized value x_{ij} of each index into Eq. (7), we obtain the index Score S_i, the results listed in column 3 and 7 of Table 6. As is shown in column 4 and 8 of Table 6, the scores are ranked from big to small. The greater score, the more important the index is, and the more front place it takes, and when select index with resolution capability into system, the more priority this index has.

4.3.3. The Second Index Screening

We built a Logistic regression model by taking index X_{78}, ,which got the biggest ranking score points and took first place as the independent variables, this index is the fourth index from bottom of column 6 Table 6, and taking default state as the dependent variable, which vary from 1 to 0, so we put the data of column 6090-3046 from line 70 and the last line of Table 4 and y representing default state into Eq. (9), and use the maximum likelihood estimation method to determine the coefficient of β_i=-8.234, and the coefficient of the standard error $SE_{\beta i}$=2.344, of the model, the results are listed in columns 3-4 of line 1 Table 7.

We calculate the Wald statistic by putting the coefficient β_i=-8.234 and the coefficient of standard error $SE_{\beta i}$=2.344 into Eq. (10),

W_1=(-8.234/2.344)^2=12.341, the results are listed in line 1 column 5 of Table 7

From distribution table of χ^2, we can see that critical value of χ^2 is $\chi_{0.05}{}^2(1)$=3.841 when the degrees of freedom equal to 1 and the significance level is α=0.05. Therefore, W_1=12.341 > 3.841, which means indicator X_{78}, consumer price index, has a significant influence, so it is preliminary reserved in the index system.

Taking two indicators were reserved in step (1) as independent variables, indicator X_{78}, consumer price index, and indicator X_{80}, Engel's coefficient, which gets the highest score, and default state as dependent variable, we build a binary Logistic regression model, then repeat the calculation process of model coefficient β_i and Wald statistics in step (1), the results are listed in columns 3-5 of Table 7.

Similarly, from the relationship between Wald statistic value and critical value$\chi_{0.05}{}^2(1)$=3.841 of χ^2, we can see that X_{78} and X_{80} have significant influence on the default state, so X_{78} and X_{80} are reserved. Taking two indicators X_{78} and X_{80} which were reserved in step (2) and X_{79}, consumer price index, which gets the highest score among the indicators except indicators X_{78} and X_{80}, as independent variables, and default state as dependent variable, we build a binary Logistic regression model, then repeat the calculation process of model coefficient β_i and Wald statistics in step (1), the results are listed in columns 3-5 of Table 7.

However, from the relationship between Wald statistic value and critical value $\chi_{0.05}{}^2(1)$=3.841 of χ^2, we can see that X_{78}, X_{80} and X_{79} have significant influence on the default state. From Table 7 after step 3, we can see that Wald statistic of index of X_{79} is 36.723, which is far greater than the critical value 3.841; the index has a significant influence on default state, so keep the index initially. But because of the influence of index X_{79}, Wald statistics of X_{80} changes to 0.146, which is less than the critical value of 3.841, so the index X_{80} no longer has ability of default identification, therefore delete indicator X_{80}, and reserve X_{78}, X_{79}.

Table 7. The Screening Process of the Binary Logistic Analysis

(1) Screen times	(2) Indicator	(3) Coefficient β_i	(4) error $SE_{\beta i}$	(5) Wald	(6) significant	(7) Constant α	(8) Default identify ability
The first step	X_{78} Consumer price index	-8.234	2.344	12.341	0.000	3.912	√
The second step	X_{78} Consumer price index	-8.306	2.674	9.645	0.002	14.526	√
	X_{80} Engel coefficient	-15.122	1.915	62.384	0.000		√
The third step	X_{78}Consumer price index	-7.057	1.911	13.640	0.000	7.053	√
	X_{80} Engel coefficient	1.194	3.130	**0.146**	0.703		×
	X_{79} City residents' disposable income	-13.443	2.218	36.723	0.000		√
…	…	…	…	…	…	…	…
The 73rd step	X_{78} Consumer price index	-9.120	2.757	10.939	0.001	17.718	√
	…	…	…	…	…		…
	X_{50} Whether the audit	1.072	0.933	1.322	0.250		×
The index system	X_{78} Consumer price index	-9.902	2.782	12.673	0.000	18.432	√
	X_{79} City residents' disposable income	-17.807	2.811	40.132	0.000		√
	X_{70} Corporate credit situation nearly 3 years	-2.342	0.679	11.895	0.001		√
	X_{49} entire period of actual operation	-2.816	0.546	26.608	0.000		√
	X_{55} Product sales scope	-3.026	0.751	16.221	0.000		√
	X_{32} Net cash flows from operating activities	-5.482	1.373	15.946	0.000		√
	X_{81} Score of pledged collateral	-1.868	0.673	7.695	0.006		√
	X_{10} super quick ratio	-6.031	2.431	6.153	0.013		√
	X_{69} Category of registered capital	2.632	0.710	13.733	0.000		√
	X_7 Full capitalization rate	-1.084	0.542	3.996	0.046		√
	X_{48} Retained earnings growth rate	3.864	1.530	6.379	0.012		√
	X_{40} Working capital allocation ratio	3.830	1.117	11.748	0.001		√
	X_{58} Education	-1.678	0.630	7.090	0.008		√
	X_{14} Cash ratio	-2.339	0.829	7.959	0.005		√

On the basis of step (3), keep index X_{78}, and X_{79}, and repeat the process above, the other indicators are filtered, the results are shown as in the corresponding column of Table 7 after step 4-73. After 74 step of screening, we determine not only individual indicators' identification ability, but also indicators' identification ability under condition of influencing on each other, eventually established a small business credit risk evaluation index system including 14 indices such as the consumer price index, which is shown in final screening results in Table 7.

4.4. The Rationality Test

The logistic regression model is constructed according to the 14 indices which are selected in 4.3, the index and coefficients of indices comes from column 2-3 in Table 7.

$$z_j=18.432-9.902X_{78}-17.807X_{79}-2.3427X_{70}-2.816X_{49}-3.026X_{55}-5.482X_{32}-1.868X_{81}-6.031X_{10}+2.632X_{69}-1084X_7+3.864X_{48}+3.830X_{40}-1678X_{58}-2.339X_1 \quad (15)$$

Take the data x_{ij} in column 3046-6090 in Table 4 in Eq. (15), and predicted the default probability $P_j(y=1)$ of the j^{th} customer combining the Eq. (11). Contracted the calculated default probability $P(y=1)$ with the real default state of customers, if the default probability $P(y=1)\geq 0.5$, the customers are discriminated default, else $P(y=1)<0.5$, the customers are not default. The classification result by contracting predicted default state and real default state is shown in Table 8. This process can be achieved by SPSS.

From Table 8, the number of actual default customers judged to be correct is 26, the number of actual non-default customers judged to be correct is 2998, so the judging accuracy rate is (26+2988)/3045=99.0%. Therefore, the credit risk evaluation index system constituted by 14 indices has strong default identification ability. The accuracy of the model is as high as 99.0%.

Table 8. Classification Using Logistic Regression

Actual default state	Classification results		
	(1) Default	(2) Non-default	(3) Total
(1) Default	**26(TP)**	24(FN)	50
(2) Non-default	7(FP)	**2988(TN)**	2995
(3) Total	33	3012	3045

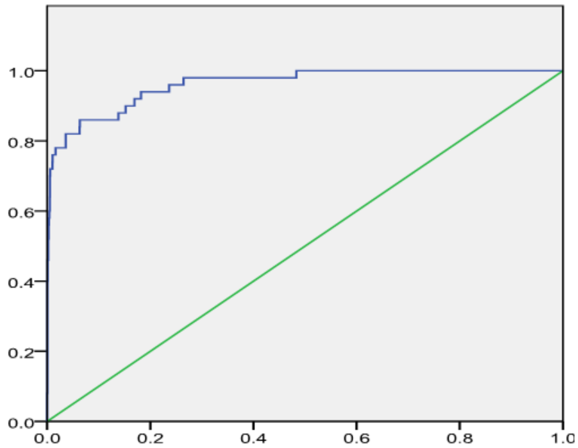

Figure 1. ROC Curve

The critical value is set 0.5, the classification results are showed in Table 8, according which the value of horizontal and vertical coordinates are calculated and the ROC curve is painted. That is take the data in row 1-2 column 1-2 in Table 8 into Eq. (12)-(13), the vertical coordinates value TPR=26/(26+24)=52.0%, and the horizontal coordinates value FPR=7/(7+2988)=0.23%. Take different critical values, we can get different horizontal and vertical coordinates values, and can draw a ROC curve, as shown in Figure 1. The process can be realized by SPSS. The area under curve is calculated, the result is AUC=0.962＞0.9, which indicates the prediction model passed the rationality test. This means the credit risk evaluation index system constituted by 14 indices has strong default identification ability.

Table 9. The Credit Rank Result of 3045 Small Enterprises

(1) No.	(2) default probability P_j	(3) Credit score S_j	(4) Receivable principal and interest	(5) Uncollected principal and interest	(6) Credit rating	(7) Sample number	(8) *LGD*	(9) Score interval
1	0.00	100	5191667	0	AAA	1549	0.421%	$99.95 \le S \le 100$
…	…	…	…	…				
1550	0.0005	99.95	10688862	0	AA	1410	0.621%	$90.22 \le S < 99.95$
…	…	…	…	…				
2960	0.1006	89.94	1859951	0	A	17	1.227%	$85.80 \le S < 90.22$
…	…	…	…	…				
2977	0.1442	85.58	2113150	0	BBB	16	3.343%	$74.70 \le S < 85.80$
…	…	…	…	…				
2993	0.2580	74.20	1067601	0	BB	12	6.938%	$62.33 \le S < 74.70$
…	…	…	…	…				
3005	0.3767	62.33	105280000	0	B	6	13.745%	$54.70 \le S < 62.33$
…	…	…	…	…				
3011	0.4705	52.95	5344191	0	CCC	4	31.491%	$49.66 \le S < 54.70$
…	…	…	…	…				
3015	0.5034	49.66	5636352.5	5033596	CC	17	51.781%	$32.72 \le S < 49.66$
…	…	…	…	…				
3032	0.6811	31.89	54419002	49272507	C	14	69.313%	$0 \le S < 32.72$
…	…	…	…	…				
3045	1	0.00	78905	78905				

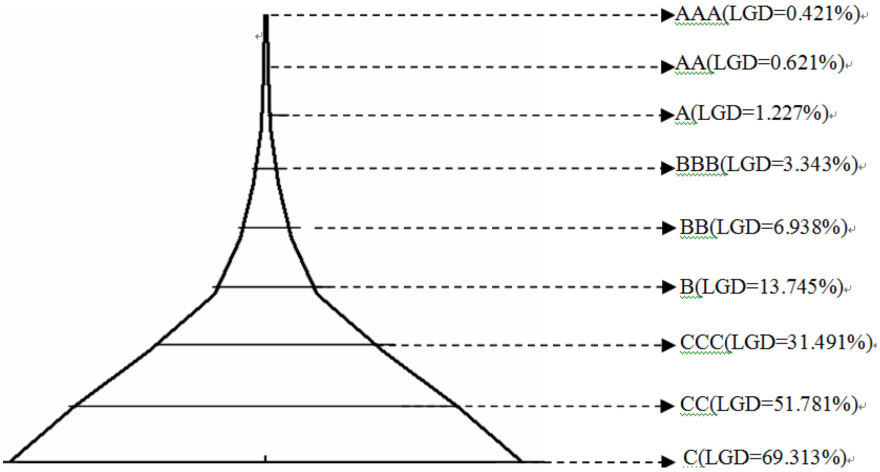

Figure 2. The LGD Distributions of 3045 Small Enterprises

4.5. The Division of Credit Rating

4.5.1. Calculate the Customer's Credit Score

Take the standardized data in Table 4 in the Eq. (15), we can get the latent variable value z_j, then put the z_j into formula (11), the default probability $P(y=1)$ are calculated, showed in column 2 of Table 9. Taking the default probability $P(y=1)$ in Eq. (14), the credit score are obtained, which are showed in column 3 of Table 9. The data in column 4-5 of Table 9 are origin data which obtained from database of commercial bank credit management system.

4.5.2. The Credit Rating of Small Enterprises

According to the ideas of credit rating in above 3.5, through the division patent method, the 3045 small enterprises classified credit rating, the results as shown in 6-9 columns of Table 9. Column 8 in Table 9 indicates that the credit rating result meets the default pyramid standard. As the credit rating from AAA, AA,..., C gradually decreases, the LGD of each grade from 0.421%, 0.621%,..., 69.313% gradually increases. We represent the credit rating and LGD, shown in column 6 and 8 in Table 9 in Figure 2, where the credit rating is exhibited decreasingly from up to down, and the length of each line represents the size of the LGD corresponding to the credit rating.

5. Conclusion

It is hard for commercial banks to accurately evaluate the credit risk of small business loans, because of mall enterprises' high risk, small amount, and untrue financial data, *etc.* The main criteria to establish the credit risk evaluation index system is the indicators have default identification ability. There is mutual influence between indices, a single index has the default identification ability, but if put this indicator into the index system, and it will no longer have the default identification ability because of the impact of other indicators. The prerequisite for study of the mutual influence among indices is determined the index importance order. This study deletes the indicators of repeated information using colinearity diagnostics, and determines the order of indicators into the index evaluation system by calculating the score statistic of every indicator. After increase an indicator in system, we will test if all the indicators in system have default identification ability by *Wald* statistic, and select the indicators into the system which has default identification ability.

The proposed approach has been verified using the data of 28 regional commercial bank branches of China, the results of our empirical analysis show that the credit risk evaluation index system of small businesses, including 14 indicators, such as cash ratio, the corporate credit situation nearly 3 years, and the judgment accuracy of default and non default samples is 99.0%. The proposed model performs well on the evaluation customers default state.

Acknowledgments

This work was supported by the National Social Science Foundation Project [16BTJ017], Social Science Planning Fund of Liaoning Province [L16BJY016], the Credit Risk Rating System and Loan Pricing Project for Small Enterprises for the Bank of Dalian [2012-01], and the Credit risks evaluation and loan pricing for petty loan funded for the head office of post savings Bank of Dalian [2009-07].

References

[1] G. Chi, M. Z. Abedin and F. E. Moula, "Modeling Credit approval data with neural networks: an experimental investigation and optimization", Journal of Business Economics and Management, doi: 10.3846/16111699.2017.1280844, vol. 18, no. 2, **(2017)**, pp. 224-240.
[2] "Standard & Poor's. Standard & Poor's guide to credit rating essentials", Standard &Poor's, **(2011)**.
[3] "Fitch Ratings. Fitch Ratings global corporate finance 2012 transition and default Study", Credit market research - Fitch Ratings, **(2013)**.
[4] M. Psillaki, I. E. Tsolas and D. Margaritis, "Evaluation of credit risk based on firm performance", European Journal of Operational Research, vol. 201, no. 3, **(2010)**, pp. 873-881.
[5] M. Dierkes, C. Erner, T. Langer and L. Norden, "Business credit information sharing and default risk of private firms", Journal of Banking & Finance, vol. 37, no. 8, **(2013)**, pp. 2867-2878.
[6] "Industrial and Commercial Bank of China. The notice about printing the evaluation method of the small enterprises credit rating of China Industrial and Commercial Bank means", Industrial and Commercial Bank of China, no. 78, **(2005)**, pp. 5-7, 13-21.

[7] G. Chi, M. Z. Abedin and F. E. Moula, "Chinese Small Business Credit Scoring: Application of Multiple Hybrids Neural Network", International Journal of Database Theory and Application, http://dx.doi.org/10.14257/ijdta.2017.10.2.01, vol. 10, no. 2, **(2017)**, pp. 1-22.
[8] B. Shi and G. Chi, "A Model for Recognizing Key Factors and Applications Thereof to Engineering", Mathematical Problems in Engineering, Article ID 862132, vol. 2014, **(2014)**, pp. 1-9
[9] F. Bonai and Z. Wenbin, "Theory Selection and Empirical Analysis of Small and Medium-sized Enterprise Credit Evaluation Index", Scientific Research Management, vol. 06, **(2003)**, pp. 83-88.
[10] M. Z. Abedin, G. Chi and F. E. Moula, "Credit Default Prediction of Chinese Small Business: A Neural Network Methodology", European Journal of Economics, Finance and Administrative Sciences, http://dx.doi.org/10.14257/ijdta.2017.10.2.01, no. 77, **(2015)**, pp. 32-50.
[11] W. Chi and C. C. Hsu, "A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model", Expert Systems with Applications, vol. 39, no. 3, **(2012)**, pp. 2650-2661.
[12] R. Hwang, H. Chung and C. Chu, "Predicting issuer credit ratings using a semiparametric method", Journal of Empirical Finance, vol. 17, no. 1, **(2010)**, pp. 120-137.
[13] T. Harris, "Quantitative credit risk assessment using support vector machines: Broad versus Narrow default definitions", Expert Systems with Applications. vol. 40, no. 12, **(2013)** pp. 4404-4413.

Authors

Chi Guotai, Professor of Finance, Doctoral supervisor, Faculty of Management and Economics, Dalian University of Technology. His research interest includes bank assets and liabilities management, credit risks rating, loan pricing, the trading risks management of future goods and so on.

Zhao Zhichong, she is currently working toward the PhD degree in Dalian University of Technology, China. Her research interest includes credit risk rating, complex system evaluation, and bank assets and liabilities management.

Mohammad Zoynul Abedin, Investment Theory Doctoral Student, Faculty of Management and Economics, Dalian University of Technology, China. His research interest includes credit risk rating, complex system evaluation.