ICIC Express Letters
Part B: Applications
Volume **8**, Number **5**, May **2017**

pp. 835–842

A NEW OPTIMIZATION MODEL OF CREDIT EVALUATION FROM MICRO-ENTERPRISES AND ITS APPLICATION

Zhanjiang Li

College of Economics and Management
Inner Mongolia Agricultural University
No. 306, Zhaowuda Road, Saihan District, Hohhot 010018, P. R. China
lizhanjiang582@163.com

Received December 2016; accepted March 2017

Abstract. *Constructing a credit evaluation model that can significantly distinguish credit characteristics of micro-enterprises is of an urgent problem for micro-enterprises in China. This paper constructs a new optimization model which can separate default micro-enterprises and non-default micro-enterprises to measure the nonlinear weight of the evaluation indicator. The new optimization model uses projection point of default micro-enterprises to approach negative ideal value, and projection point of non-default micro-enterprises to reach positive ideal value. The accuracy and robustness test reveals the new credit evaluation model is stable and reliable. Empirical comparative analysis reveals that credit evaluation model of projection pursuit discriminant is superior to credit evaluation model of logistic regression.*
Keywords: Optimization model, Credit evaluation, Micro-enterprises, Projection pursuit

1. **Introduction.** China's micro-enterprises are hugely numerous and energetic, but shortage of funds has always been a difficult problem to these micro-enterprises. Because existing credit evaluation method of enterprises cannot better reflect the credit characteristics of micro-enterprises, even most of commercial banks have not established credit evaluation system of these micro-enterprises. Therefore, constructing a credit evaluation model that can significantly distinguish credit characteristics of micro-enterprises is of an urgent problem for China's micro-enterprises.

The existing research about credit evaluation model of enterprises has the following three aspects.

(1) Evaluation model based on statistical and econometric methods.

Wu (2016) uses the Monte Carlo simulation method to research the impact relationship between the recovery and the default probability [1]. Zhang (2015) uses Markov transition probability to research the credit rating issue in business loans [2]. An (2014) uses a logistic regression model to establish prediction model of financial crisis of small and medium enterprises [3]. Wosnitza and Leker (2013) use kernel logistic regression to establish credit evaluation model of small enterprises [4]. Gama and Geraldes (2012) apply logit model to establishing credit scoring model of small enterprises in Portugal [5].

(2) Evaluation model based on artificial intelligence.

Chen (2016) uses support vector machine regression model to construct corporate credit scoring model [6]. Li and Zhou (2015) use scale-free network method to construct infection model of credit risk [7]. Huang (2014) uses fuzzy support vector machine technology to establish assessment model of financing risk [8]. Blanco et al. (2013) apply the neural network model with multi-layer perceptron to establishing credit scoring model of small loans [9]. Oreski et al. (2012) combine genetic algorithms and neural networks to assess the credit risk of retail enterprises [10].

(3) Evaluation model based on non-parametric technology and optimization model.

Zhang and Zhang (2015) use generalized semi-parametric additive model to estimate default probability of corporate customers [11]. Kruppa et al. (2013) use random forest method and nearest neighbor method to estimate the default probability of of individual consumption credit loan [12]. Wekesa et al. (2012) use survival analysis to research default probability of male loan applicants and female loan applicants [13]. Zhang and Zhou (2011) apply projection pursuit model to establishing nonlinear optimization model of corporate credit index weight [14].

Although the existing credit evaluation model is able to measure credit risks of enterprises, calculating credit scores of micro-enterprises and finding a suitable credit evaluation model are still an issue to be deeply examined. Compared with previous studies, the advantages of this research lie in constructing a new optimization model which can significantly distinguish between default enterprises and non-default enterprises.

The contributions of this research have two aspects. First, this research constructs a new credit evaluation model which has significant ability to identify the credit status of micro-enterprises. Second, this research uses this new model to solve measure problem of credit scoring of micro-enterprises in China.

This paper consists of five sections including introduction, optimization model, accuracy test and robustness test, application research, and conclusions.

2. **Optimization Model.**

2.1. **Traditional projection pursuit model.** Although traditional projection pursuit model can transform the non-linear, non-normality, multidimensional data of high-dimensional space into comprehensive indicator of the low-dimensional space, it cannot significantly distinguish credit characteristics of micro-enterprises.

The building step of traditional projection pursuit model is shown below.

(1) Standardized scoring of indexes

We need to calculate standardized scoring for indexes. After standardization, scoring interval for indexes is $[0, 1]$.

Let: x_{ij} – the ith micro-enterprises sample observation value of the jth credit evaluation index; y_{ij} – the ith standardized micro-enterprises sample observation value of the jth credit evaluation index, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$; m – number of evaluation indexes; n – number of micro-enterprises sample.

For positive index, the research uses Formula (1) to calculate its standardized score; for negative index, the research uses Formula (2) to calculate its standardized score value.

$$Y_{ij} = \frac{X_{ij} - \min(X_{ij})}{\max(X_{ij}) - \min(X_{ij})} \tag{1}$$

$$Y_{ij} = \frac{\max(X_{ij}) - X_{ij}}{\max(X_{ij}) - \min(X_{ij})} \tag{2}$$

(2) Projection objective function

Let: Z_i – the projection score value; $P = (P_1, P_2, \ldots, P_m)$ – the projection direction vector; S – standard deviation of projection score value; D – local density of projection score value. Then the calculation of Z_i, S, local density D are defined in Formulas (3)-(5) [14].

$$Z_i = \sum_{j=1}^{m} P_j Y_{ij},\ i = 1, 2, \ldots, n \tag{3}$$

$$S = \left(\sum_{i=1}^{n} (z_i - \bar{z})^2 / (n-1) \right)^{1/2} \tag{4}$$

$$D = \sum_{i=1}^{n}\sum_{j=1}^{n}(0.1S - |z_i - z_j|) \times u\,(0.1S - |z_i - z_j|) \tag{5}$$

In Formula (5), u is equal to 0 if $0.1S < |z_i - z_j|$. Then u is equal to 1 if $0.1S \geq |z_i - z_j|$.

The objective function Q of projection pursuit is defined in Formula (6) [14].

$$Q = S \times D \tag{6}$$

(3) Optimization model of projection objective function

The best projection direction vector P should expose characteristic of the overall dispersion and local dense of projection value. This moment, the nonlinear optimization model of traditional projection pursuit is shown in Formula (7) [14].

$$\begin{cases} Q_{\max} = S \times D \\ \text{s.t. } \displaystyle\sum_{j=1}^{m} p_j^2 = 1 \end{cases} \tag{7}$$

2.2. **New optimization model.** This paper firstly constructs optimization model of projection pursuit discrimination which can separate default micro-enterprises from non-default micro-enterprises in order to measure the weight of the evaluation indicator.

(1) Establishing objective function of optimization model

Let: $z_1(i)$ – the ith projection value of non-default micro-enterprise sample; $\max(z)$ – maximum value of projection vector z which can reflect positive ideal solution of projection vector. $z_2(j)$ – the jth projection value of default micro-enterprise sample; $\min(z)$ – minimum value of projection vector z which can reflect negative ideal solution of projection vector. k – non-default micro-enterprise sample number. Then objective function of optimization model is defined in Formula (8).

$$Q^* = \sum_{i=1}^{k}|z_1(i) - \max(z)| \times \sum_{j=1}^{n-k}|z_2(j) - \min(z)| \tag{8}$$

In Formula (8), formula $\sum_{i=1}^{k}|z_1(i) - \max(z)|$ reflects the absolute distance sum between all non-default micro-enterprise samples and positive ideal solution. Formula $\sum_{j=1}^{n-k}|z_2(j) - \min(z)|$ reflects the absolute distance sum between all default micro-enterprise samples and negative ideal solution.

(2) Establishing optimization function

The optimization model of projection pursuit discriminant is shown in Formula (9).

$$\begin{cases} Q^*_{\min} = \displaystyle\sum_{i=1}^{k}|z_1(i) - \max(z)| \times \sum_{j=1}^{n-k}|z_2(j) - \min(z)| \\ \text{s.t. } \displaystyle\sum_{j=1}^{m} p_j^2 = 1 \end{cases} \tag{9}$$

Formula (9) is the nonlinear optimization model of projection pursuit discriminant, where formula $\sum_{j=1}^{m} p_j^2 = 1$ is a constraint condition. The optimal solution of formula (9) is not weight of the evaluation index, but the projection direction vector $P_* = [p_{*1}, p_{*2}, \ldots, p_{*m}]$.

The features of Formula (9) lie in two aspects. First, Formula (9) constructs a new optimization model of projection pursuit discriminant, using combination between projection pursuit model and technique for order preference by similarity to an ideal solution. Second, the new model reflects weight calculation ideas that the more default sample differs from non-default sample, the more important the evaluation index is, thus solving weight measurement problem of the credit evaluation index for micro-enterprise.

(3) Solution of the optimization model

Most of the existing researches use the genetic algorithm to calculate projection pursuit optimization model of Formula (7), but the calculation result of Formula (9) shows that using the genetic algorithm has poor stability. Therefore, this paper uses direct search tool which is the pattern search algorithm to calculate Formula (9). The numerous calculation results of pattern search prove that solutions of Formula (9) are stable.

2.3. **Credit scoring measurement.** Let: w_i – projection weight of the ith credit evaluation index; $W = [w_1, w_2, \ldots, w_m]$ – weight vector of all credit evaluation index; $P_* = [p_{*1}, p_{*2}, \ldots, p_{*m}]$ – projection direction vector by Formula (9). Then projection weight is defined in Formula (10).

$$w_i = p_{*i}^2 \tag{10}$$

w_i by Formula (10) is the standardized weight coefficient to meet $w_1 + \cdots + w_m = 1$ and $w_i > 0$.

Let: S_i – credit scoring of the ith micro-enterprise; w_j – projection weight of the jth index; y_{ij} – the ith standardized micro-enterprises sample observation value of the jth index, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$. Then credit scoring S_i is defined in Formula (11).

$$S_i = 100 \times \sum_{j=1}^{m} w_j y_{ij} \tag{11}$$

The effect of Formula (11) is to measure credit scoring of hundred mark system for each micro-enterprise.

3. **Accuracy Test and Robustness Test.** The purpose of accuracy test and robustness test is to prove that credit scoring model (11) is stable and reliable when new micro-enterprise samples cannot be increased.

The first error is reflected by default micro-enterprise mistaken for non-default micro-enterprise and the second error is reflected by non-default micro-enterprise mistaken for default micro-enterprise.

3.1. **Accuracy test.** Training samples are obtained by using 80% random samples of micro-enterprise samples. Then this paper uses training samples to construct credit scoring model (11) and calculates credit score value of each training sample.

Let: S_0 – critical point of determining default state; SS_{1i} – credit scoring of the ith default micro-enterprise, $i = 1, 2, \ldots, 0.8k$; k – the number of default micro-enterprises; SS_{0j} – credit scoring of the jth non-default micro-enterprise, $j = 1, 2, \ldots, 0.8(n - k)$; n – the number of micro-enterprises. Then S_0 is defined in Formula (12).

$$S_0 = \frac{\frac{1}{0.8 \times k} \sum_{i=1}^{0.8k} SS_{1i} + \frac{1}{0.8 \times (n-k)} \sum_{j=1}^{0.8(n-k)} SS_{0j}}{2} \tag{12}$$

Let: A – the accuracy rate of credit scoring model (11); a_1 – the first error frequency of training samples; a_2 – the second error frequency of training samples; A_0 – critical point of the accuracy rate. Then the accuracy rate A is defined in formula (13).

$$A = 1 - \frac{a_1 + a_2}{0.8n} \tag{13}$$

Based on the popular literature, the critical value A_0 is set to 0.8.

3.2. **Robustness test.** Testing samples are obtained by using remainders of training samples. Then this paper calculates credit score value of each testing sample by credit scoring model (11).

Let: B – the robustness degree of credit scoring model (11); b_1 – the first error frequency of testing samples; b_2 – the second error frequency of testing samples; B_0 – critical point of the robustness degree. Then the robustness degree B is defined in Formula (14).

$$B = 1 - \frac{b_1 + b_2}{0.2n} \tag{14}$$

Based on the popular literature, the critical value B_0 is set to 0.7.

4. Application Research.

4.1. **Evaluation index system and standardized data.** The index system including three criteria layers and twenty-two evaluation indexes is shown in *a*-*b* column of Table 1. Because establishment of the index system belongs to research result of another paper, some indexes are omitted in Table 1.

Table 1. Evaluation index system and weight

Serial number	(*a*) Criteria layer	(*b*) Index layer	(*c*) p_{*i}	(*d*) w_i	Default sample (1) No.1	…	(30) No.30	Non-default sample (31) No.31	…	(300) No.300
1	Internal financial factors	Cash ratio of main business income Y_1	0.198	0.039	0.033	…	0.272	0.189	…	0.003
…		…	…	…	…	…	…	…	…	…
11		Rate of capital business income Y_1	0.142	0.020	0.197	…	0.197	0.146	…	0.248
12	Internal non-financial factors	Working years of related industries Y_{12}	0.219	0.048	0.000	…	0.000	0.000	…	0.750
…		…	…	…	…	…	…	…	…	…
20		Legal dispute case Y_{20}	0.104	0.011	0.000	…	0.000	0.000	…	1.000
21	External macro environment	Industry sentiment index Y_{21}	0.411	0.169	0.559	…	0.576	0.838	…	0.742
22		Per capita disposable income of urban residents Y_{22}	0.296	0.088	0.000	…	0.000	0.010	…	0.004

Index data comes from loan data system of a city commercial bank in China. After deleting outliers in original data, there are 300 samples including 30 default samples and 270 non-default samples. According to Formula (1) and Formula (2), standardized score of each index is calculated and shown in 1-300 column of Table 1.

4.2. **Calculating weight.** Use 1-300 column of Table 1 as all sample sources. 80% default samples are randomly selected from 1-30 column of Table 1 and 80% non-default samples are randomly selected from 31-300 column of Table 1. Then training samples are obtained including 24 default samples and 216 non-default samples.

After standardized index data of training samples is imported into matlab software, we use matlab pattern search algorithm toolbox to calculate the optimal solution of Formula (9) and obtain projection direction vector P_*. The result of projection direction vector P_* is shown in column c of Table 1.

According to Formula (10), projection weight of each credit evaluation index is obtained and shown in column d of Table 1.

4.3. **Calculating credit scoring.** The final model for micro-enterprise can be obtained according to credit scoring model (11). The result of empirical model is defined in formula (15).

$$S = 100 \times (0.039 \times Y_1 + 0.020 \times Y_{11} + \cdots + 0.088 \times Y_{22}) \tag{15}$$

Credit scoring of training samples and testing samples can be obtained by Formula (15). Credit scoring of training samples is shown in column 2 of Table 2 and credit scoring of testing samples is shown in column 2 of Table 3.

Table 2. Training samples and accuracy test

Serial number	(1) Default state	(2) S_i	(3) S_0	(4) Determined state	(5) a_1	(6) a_2	(7) A
1	1	35.382	45.605	1	0	–	0.925
...	...	...		...			
24	1	29.908		1			
25	0	56.093		0	–	18	
...	...	...		...			
240	0	74.675		0			

Table 3. Testing samples and robustness test

Serial number	(1) Default state	(2) S_i	(3) S_0	(4) Determined state	(5) b_1	(6) b_2	(7) B
1	1	39.076	45.605	1	0	–	0.817
...	...	...		...			
6	1	24.725		1			
7	0	68.098		0	–	11	
...	...	...		...			
60	0	54.214		0			

4.4. **Testing accuracy and robustness.**

(1) Testing accuracy

Critical point S_0 of determining default state is shown in column 3 of Table 2 through Formula (12). Determined results of default state are shown in column 4 of Table 2. Numerical 1 denotes default and numerical 0 denotes non-default. Frequency statistics results of two types of errors a_1 and a_2 are shown in columns 5-6 of Table 2. Calculation results of accuracy rate A is shown in column 7 of Table 2 through Formula (13).

Calculation results of accuracy rate A proves that credit scoring model (11) can be determined as exactness.

(2) Testing robustness

Critical point S_0 of determining default state is shown in column 3 of Table 3 through Formula (12). Determined results of default state are shown in column 4 of Table 3. Frequency statistics results of two types of errors b_1 and b_2 are shown in columns 5-6 of Table 3. Calculation results of robustness degree B is shown in column 7 of Table 3 through Formula (14).

Calculation results of robustness degree B proves that credit scoring model (11) can be determined as robustness.

(3) Results of accuracy test and robustness test

The result of accuracy test and robustness test proves that credit scoring model (11) is stable and reliable.

4.5. **Comparative analysis.** This paper uses logistic regression model to act as empirical comparative object. Calculation results of logistic regression model are shown in row 2 of Table 4 including a_1-a_2, b_1-b_2, A and B. Row 1 of Table 4 are copied from Tables 2 and 3.

Table 4. Empirical comparison

Serial number	(1) Model	Training samples			Testing samples		
		(2) a_1	(3) a_2	(4) A	(5) b_1	(6) b_2	(7) B
1	Projection pursuit discrimination	0	18	0.925	0	11	0.817
2	Logistic regression	3	25	0.883	1	14	0.750

Column 4 of Table 4 shows that accuracy rate A of projection pursuit discrimination model is greater than accuracy rate A of logistic regression model. At the same time, column 7 of Table 4 shows that robustness degree B of projection pursuit discrimination model is greater than accuracy rate B of logistic regression model.

Combining accuracy rate and robustness degree, the result of Table 4 can prove that credit evaluation model of projection pursuit discrimination is surprior to credit evaluation model of logistic regression.

5. **Conclusions.** This paper constructs a new optimization model which can separate default micro-enterprises and non-default micro-enterprises. The accuracy and robustness test reveals the new optimization model is stable and reliable.

Empirical comparative analysis reveals that credit evaluation model of projection pursuit discriminant is superior to credit evaluation model of logistic regression.

In credit evaluation research of micro-enterprises, future research trend is to apply the credit scoring of micro-enterprises to researching credit ratings and risk decision of micro-enterprises.

Acknowledgment. The research is supported by the China Postdoctoral Science Foundation (the grant number is 2015M582754XB) and supported by Natural Science Foundation of Inner Mongolia Autonomous Region of China (the grant number is 2016MS0714). We thank China Postdoctoral Science Foundation and Natural Science Foundation of Inner Mongolia Autonomous Region of China for their financial support.

REFERENCES

[1] L. H. Wu, Endogenous recovery and credit risk measurement, *China Management Science*, vol.24, no.1, pp.1-10, 2016.

[2] L. Zhang, Multistage Markov transition probability estimate of loan risk credit rating, *Statistics and Decision*, no.2, pp.44-49, 2014.

[3] L. An, Credit evaluation research of small and medium enterprises based on logistic regression, *Shanghai Finance College*, no.23, pp.154-158, 2015.

[4] J. H. Wosnitza and J. Leker, Why credit risk markets are predestined for exhibiting log-periodic power law structures, *Physica A: Statistical Mechanics and Its Applications*, no.21, pp.154-158, 2013.

[5] A. P. M. Gama and H. S. A. Geraldes, Credit risk assessment and the impact of the new basel capital accord on small and medium-sized enterprises: An empirical analysis, *Management Research Review*, vol.35, no.8, pp.727-749, 2012.

[6] Y. Chen, Corporate credit scoring model based on RS-SVM, *Application Research of Computers*, 2016.

[7] Y. K. Li and Z. F. Zhou, Infection delayed effect of credit risk based on scale-free networks, *Systems Engineering*, vol.30, no.5, pp.575-583, 2015.

[8] Q. Huang, Financing risk evaluation of the start-up of small and medium high-tech enterprises, *Value Engineering*, no.33, pp.141-142, 2014.

[9] A. Blanco, R. Pino-Mejías, J. Lara et al., Credit scoring models for the microfinance industry using neural networks: Evidence from Peru, *Expert Systems with Applications*, vol.40, no.1, pp.356-364, 2013.

[10] S. Oreski, D. Oreski and G. Oreski, Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment, *Expert Systems with Applications*, vol.39, no.16, pp.12605-12617, 2012.

[11] J. Zhang and B. B. Zhang, Generalized semiparametric additive credit scoring model, *Mathematical Statistics and Management*, vol.35, no.3, pp.517-524, 2016.

[12] J. Kruppa, A. Schwarz, G. Arminger et al., Consumer credit risk: Individual probability estimates using machine learning, *Expert Systems with Applications*, vol.40, no.13, pp.5125-5131, 2013.

[13] O. A. Wekesa, M. Samuel and M. Peter, Modeling credit risk for personal loans using product-limit estimator, *International Journal of Financial Research*, vol.3, no.1, pp.22-32, 2012.

[14] M. Zhang and Z. F. Zhou, Corporate credit rating model based on projection pursuit and optimal segmentation, *Operations Research and Management*, vol.20, no.6, pp.226-231, 2011.