Credit risk assessment in bank based on SMEs using PCA

Zhou Huang¹, Zifeng Yan²*, Qingzhu Zhao³ and Keke Ma⁴

¹Department of Communication and Information Engineering, Xi’an University of Posts and Telecommunications, Xi’an, Shaanxi, 710061, China
²Department of Communication and Information Engineering, Xi’an University of Posts and Telecommunications, Xi’an, Shaanxi, 710061, China
³Department of Electrical Engineering, Xi’an University of Posts and Telecommunications, Xi’an, Shaanxi, 710061, China
⁴Department of Communication and Information Engineering, Xi’an University of Posts and Telecommunications, Xi’an, Shaanxi, 710061, China
⁵Corresponding author’s e-mail: yanzifeng1228@stu.xupt.edu.com

Abstract. In recent years, small and medium-sized enterprises (SMEs) have gradually become the backbone of China’s national economic development due to their strong innovation ability and high operational efficiency. However, such enterprises have poor anti risk ability in the early stage of development and need banks to lend to them. In order to help banks quickly and accurately evaluate the ability of small and medium-sized enterprises to determine the amount of loans, this paper proposes a comprehensive ability evaluation model based on PCA algorithm. Considering the strong impact of the new COVID-19 on SMEs in 2020, we introduce this sudden factor into the model to improve the adaptability of the model and help banks to accurately evaluate the enterprise capacity under the COVID-19.

1. Introduction
With the rapid development of the global economy, China’s small and medium-sized enterprises (SMEs) have become an important driving force in the development of China’s national economy because of their flexibility, adaptability and wide range of operation. We find that SMEs contribute more than 50% of the tax revenue, more than 60% of GDP, more than 70% of technological innovation, more than 80% of urban employment, and more than 90% of the number of enterprises. However, in reality, due to the relatively small size of SMEs and the lack of mortgage assets, they usually need to apply for loans from banks. Considering the instability of the early development of SMEs, SMEs have to apply for loans, there is a large potential credit risk in the lending bank, so the bank should evaluate the risk according to the strength and reputation of the small and medium-sized enterprises, and determine whether to lend or not according to the factors such as credit risk, loan amount, interest rate and term. In this context, this paper puts forward an enterprise comprehensive ability evaluation model based on PCA algorithm, which is helpful for banks to quickly and accurately obtain the comprehensive ability evaluation of an enterprise in a large number of data of many small and medium-sized enterprises. However, due to the outbreak of new coronavirus this year, small and medium-sized enterprises are short of working capital and have lost certain repayment ability, which further increases the risk of bank lending. Therefore, this paper improves the model, introduces the epidemic impact degree, improves the adaptability of the model, and
puts forward a better way for banks to solve the credit risk assessment under the special background of this year.

2. Evaluation model of enterprise comprehensive ability

2.1. Model preparation
We need to re-establish the ability of SMES.

| Table 1. Factors influencing enterprise competency value |
|--------------------------------------------------------|
| 1 Order abandonment rate \( \alpha = \frac{z}{n} \) \( \alpha \) is the invalidation rate, \( z \) is the invalidation invoice, \( n \) is the whole invoice. The larger the proportion of this invalid invoice, the lower the value of the company's comprehensive ability |
| 2 Negative enterprise invoice rate \( \omega = \frac{f}{n - z} \) Negative invoice represents returned goods, \( \omega \) is the rate of returned goods, \( f \) is the amount of returned goods |
| 3 The proportion of the returned amount \( \delta = \frac{\bar{m}}{m} \) \( \delta \) is the percentage of returned goods, is the amount of returned goods, \( m \) is the valid invoice amount |
| 4 Enterprise flow difference \( \varepsilon = n_1 - n_2 \) \( \varepsilon \) is the general ledger of the company, \( n_1 \) is the write-off amount, \( n_2 \) is the incoming amount. Reflect enterprise assets, whether to have reimbursement ability |
| 5 Average tax value of enterprises \( \eta = \frac{\sum x_i}{n - z} \) \( \eta \) is the average tax value, \( x_i \) is the amount of tax, \( y_i \) is the total tax value. Reflect the tax types of different companies |

The principal component analysis (PCA) method used in this paper is a multivariate statistical analysis method to reduce the dimension of multi-dimensional data through data compression. Its advantages are simple operation, no parameter limitation and wide application range. It is often used in face recognition, image compression, feature extraction and other fields [2]. We apply the idea of dimension reduction to the model building, extract the principal component of the impact on the evaluation of enterprise capacity value, and establish the quantitative equation for the evaluation of enterprise comprehensive capacity value.

2.2. Modeling
The establishment process of the model is as follows:
Step 1: For the data normalization processing, eliminate the dimensional influence, use the min max standardization method, and use the following formula for linear change,

\[ y_i = \frac{x_i - \min \{x_j\}}{\max \{x_j\} - \min \{x_j\}} \]  

That is, after treatment \( y_i \in [0,1] \) And dimensionless.

Step 2: For the filtered data, construct the data matrix and calculate the sample correlation coefficient matrix, which is expressed as:

\[ r_{ij} = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} \times x_{kj}), (i, j = 1, 2, \ldots, 7) \]  

Step 3: Get the eigenvalues and corresponding eigenvectors of the correlation coefficient matrix \( R \) by Jacobian method

Step 4: Calculate the principal component contribution rate and cumulative contribution rate:

\[ C = \frac{1}{n} \sum_{i=1}^{I} \frac{\lambda_i}{\sum_{i=1}^{I} \lambda_i}, (i, j = 1, 2, \ldots, 7) \]  

First \( k \) principal components are selected to make the cumulative contribution rate reach 85% to 95%

\[ C_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{I} \lambda_i}, (i, j = 1, 2, \ldots, 7) \]  

The contribution rate of the \( k \) principal components is as \( \gamma_k \).

Step 5: Calculate the principal component load:

\[ R_{ij} = P(z_i, x_j) = \sqrt{\lambda_i}, (i, j = 1, 2, \ldots, 9) \]  

Among them:

\[ z_i = R_{i1} x_1 + R_{i2} x_2 + \ldots + R_{i9} x_9 \]  

New variable index \( z_i \). They are the \( k \) principal components of the original variable.

Step 6: Calculate the principal component coefficient:

\[ P = \frac{X}{\sqrt{\lambda}} \]  

Among them, \( P \) The principal component represents the coefficient, \( X \) Represents the value in the principal component coefficient, \( \lambda \) Represents the characteristic value.

Through the calculation of the above steps, we can finally get the enterprise comprehensive capacity evaluation model:

\[ y_i = \gamma_1 z_1 + \gamma_2 z_2 + \ldots + \gamma_k z_k \]  

3. Results & discussion

3.1. Model Validation

We selected data collected by a bank on 302 enterprises with no credit record as our model test and we divided the size of the research enterprises as another factor affecting the evaluation of comprehensive capacity, among which the medium-sized enterprises accounted for 48.21%, the small enterprises accounted for 51.79%. Through SPSS calculation, we finally determined three principal components, the contribution rate of the first principal component is 74.31%, that of the second principal component is 12.73%, and that of the third principal component is 7.69%. The cumulative contribution rate of the three principal components has reached 94.73% and obtained a series of model data as shown in the table below
Table 2. Three principal component eigenvalues

|          | Primary principal component | Secondary principal component | Third principal component |
|----------|-----------------------------|-------------------------------|---------------------------|
| $\lambda$ | 7.446                       | 3.081                         | 2.436                     |
| The eigenvalue |                               |                               |                           |

Table 3. Three principal component contribution rate

|               | Primary principal component | Secondary principal component | Third principal component |
|---------------|-----------------------------|-------------------------------|---------------------------|
| Contribution /% | 0.7431                      | 0.1273                        | 0.0796                    |

Table 4. Three principal component coefficients

|                                | Primary principal component | Secondary principal component | Third principal component |
|--------------------------------|-----------------------------|-------------------------------|---------------------------|
| Order abandonment rate         | -2.3422                     | 0.0753                        | 1.0723                    |
| Negative enterprise invoice rate | -2.9323                     | -0.4163                       | -0.5295                   |
| The proportion of the returned amount | -3.6731                     | -0.1551                       | -0.4853                   |
| Enterprise flow difference     | 2.4097                      | 2.3216                        | -0.9349                   |
| Average tax value of enterprises | 2.6822                      | 0.1448                        | 1.6656                    |
| Enterprise size                | 3.8557                      | -1.9704                       | -0.7881                   |

The first, second and third principal components calculated by MATLAB are as follows:

$$z_1 = -2.34x_1 - 2.92x_2 3.67x_3 + 2.41x_4 + 2.68x_5 + 3.85x_6$$
(8)

$$z_2 = 0.07x_1 - 0.41x_2 0.15x_3 + 2.32x_4 + 0.14x_5 - 1.97x_6$$
(9)

$$z_3 = 1.07x_1 - 0.53x_2 - 0.48x_3 - 0.93x_4 + 1.66x_5 - 0.79x_6$$
(10)

The first principal component is affected the Enterprise size and the Order abandonment rate, the second principal component is affected by the Average tax value of enterprises and the Negative enterprise invoice rate, and the third principal component is greatly affected by the Enterprise flow difference.

For the three principal components, we take the numerical ratio of the cumulative contribution rate as the weight of each principal component, quantify the reputation risk, and obtain the comprehensive capacity value of each company. The quantitative equation is as follows:

$$Y_i = 0.7844z_1 + 0.1343z_2 + 0.084z_3$$
(11)

Therefore, we get the quantitative model of the comprehensive capacity of SMEs by modeling the data collected by a bank. The capability value calculated by the model accords with the real data of the enterprise.

3.2. Discussion

At the beginning of 2020, COVID-19 epidemic broke out, which was rampant in the world [3], which had a great impact on China and even the global economy. Most of China's small and medium-sized enterprises are still in the early stage of development, with weak foundation and poor anti risk ability. The unexpected impact caused by the epidemic situation has seriously hindered their business development and even threatened their survival [4]. In this paper, considering the impact of the epidemic situation of small and medium-sized enterprises in the context of the epidemic, we improve the model, and put forward effective suggestions for bank credit strategy.
Considering that the enterprises are affected by the new epidemic situation every quarter, and the impact of each quarter is different, we introduce the impact degree of the epidemic situation in each quarter, and calculate the quarterly comprehensive capacity value after the impact \( G_i \). Namely:

\[
G_i = Y_j \cdot H_j
\]  

(12)

among \( H_j \) is the impact degree of each quarter, and then the annual average comprehensive capacity value affected by the epidemic situation is calculated

\[
Y_j = \frac{G_1 + G_2 + G_3 + G_4}{4}
\]  

(13)

Thus, the improved model of enterprise comprehensive capacity under the influence of epidemic situation is obtained.

4. Conclusion

In this paper, we provide an evaluation method for the comprehensive ability of SMES. At the same time, combined with the impact of the new crown epidemic, we improved the model to better adapt to the changing environment. This model has good practical significance for reducing bank credit risk.

Acknowledgments

This research was funded in part by the Shanxi Provincial Key Research and Development Project (2020ZDXM-GY-091), and National Key Technologies Research and Development Program (2020YFB1309403) and the Scientific Research Plan Projects of the Shanxi Education Department (no. 17JK0702).

Reference

[1] Fang Yunlong. STUDY ON BANK CREDIT RISK management under new situation [J]. China’s collective economy, 2020(30) : 100-101.

[2] Jolliffe I T, Cadima J. Principal component analysis: a review and recent developments[J]. Philosophical Transactions of the Royal Society Mathematical Physical & Engineering Sciences, 2016,374(2065):20150202.

[3] Xintian Xu, Ping Chen, Jingfang Wang, Jianiian Feng, Hui Zhou, Xuan Li, Wu Zhong, Pei Hao. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission[J]. Science China Life Sciences, 2020,63(3).

[4] Yan shu-shan, Li Shan-yao. The impact of new crown pneumonia on small and medium-sized enterprises and its countermeasures [j]. Research in Business Economics, 2020(21) : 134-138.