Logistic Regression-Based Enterprise Credit Evaluation Model

Ziyuan Li

***Abstract*—In recent years, enterprises in Beijing have developed rapidly. As one of the economic centers in China, the integrity system of enterprises has become a hot topic. An objective and effective evaluation model for enterprises in Beijing can not only help consumers avoid consumption risks and provide reference for the government to formulate relevant policies, but also provide a reference for enterprises in choosing partners. The research objective of this paper is to construct a logistic regression-based enterprise credit evaluation model in Beijing. In the empirical study, 4648 Beijing enterprises were taken as samples, and the K-S test and Mann-whitey U test were used to test the correlation between test indicators and their correlation with explanatory variable y, and the qualified 6 test indicators were selected from the 15 indicators. KMO and Bartlett tests were used to test the six indicators to see whether they were suitable for principal component analysis. After principal component analysis of 6 indicators, the principal component factors obtained were taken into the binary logistic regression model as input variables, and the stepwise forward method based on maximum likelihood estimation was adopted to estimate the model parameters, so as to determine the final variables entering the model. Finally, test samples are used to test the accuracy of the model. The test results show that the model can predict the integrity of enterprises in Beijing.**

***Index Terms*—Logistic regression, Beijing enterprises, factor analysis, parameter significance test.**

I. Introduction

Beijing, as the political and economic center of China, the credit status of enterprises in Beijing plays a fundamental role in innovation activities. Therefore, this paper takes enterprises in Beijing as the subject of this research. Targeting at the development status of enterprise credit system in Beijing, an enterprise credit evaluation model was constructed based on logistic regression algorithm.

Although Beijing has its enterprise credit information system in use, there are still three main problems. First, the rating method lacks objective basis, so the evaluation of enterprise risk is not accurate enough. Second, the enterprise credit evaluation standard is not uniform. The credit evaluation methods and systems are independent without mutual communication of information. Third, the information of enterprise credit database is not enough, which leads to the inaccurate assessment of the rating agencies in Beijing (Xiang Lu, 2003) [1].

Therefore, the public needs an objective data model when evaluating the integrity of enterprises. For example, a total of 6,336 enterprises of Beijing were collected as samples in this study, and then their dishonesty was studied. From Table I, Beijing enterprises are divided into 20 industries, and the dishonesty of each industry is calculated.

From Table I, it can be seen that apart from culture, education, printing, publishing and advertising media, all industries have broken faith, especially the total rate of breaking faith reaches 1.5%, which is still a large probability.

Set up an enterprise credit evaluation system not only can constraint corporate behavior, but also help consumers choose according to quantitatively the enterprise good faith make reasonable consumption, help company to better avoid risks with choosing partner (Addo, 2018) [2]. This study combines literature search, data collection, statistical analysis, empirical analysis, case analysis and other methods to build, test the logistic regressions-based evaluation model of enterprise integrity in Beijing, and analyze the actual cases in Beijing. The technical route studied in this paper is shown in Fig. 1.

II. Literature Review

Western countries' credit rating developed very early, especially the United States as a country with a developed credit system, whose credit rating agency is the earliest and most sound in the world (Jia Mei, 2014) [3].

The initial credit rating of the United States was mainly based on statistical analysis, which in short was mainly based on qualitative analysis (Zou Ding, 2013) [4]. This type of rating model focuses more on the subjective factors of the raters. More representative selection methods of rating indicators include minimal incoherence method, maximum generalized difference method, principal component analysis method and cluster analysis method (Yu Shun, 2014) [5]. The financial risk assessment model based on statistical analysis can be divided into three types. They are univariate model, multivariate model and probabilistic model (Jia Wenrong, 2014) [6].

The univariate model was first proposed by Professor William Beaver of the University of Chicago in 1966 (Diakite, 2015) [7]. Beaver uses the financial reporting data of all industrially listed companies contained in the Moody's Industry Manual as a research sample (Beaver, 1966) [8]. He compared the similarities and differences between successful and failed enterprises from 1954 to 1964, and finally concluded that the lowest financial ratio of misjudgment rate was debt guarantee ratio = cash flow ÷ total debt (Beaver, 1966) [8]. Moreover, different financial ratios are used depending on the situation (Beaver, 1966) [8].

Manuscript received July 9, 2021; revised October 9, 2021.

Ziyuan Li is with Columbia Business School, China (e-mail: ziyuanli2020@outlook.com).

doi:10.18178/joebm.2021.9.4.662

His financial ratios include debt to asset ratio = total liabilities ÷ total assets, return on assets = net income ÷ total assets and security ratio = liquidity to asset-liability ratio (Beaver, 1966) [8].

TABLE I: INDUSTRY CLASSIFICATION OF ENTERPRISE DISHONESTY IN BEIJING

	Faithless announcement						
Beijing enterprise group category	**yes**	**no**	**Faithless rate**	**Beijing enterprise group category**	**yes**	**no**	**Faithless rate**
Telecommunications, radio, television and satellite transmission services	5	285	1.72%	Agricultural food and cosmetics	6	226	2.59%
Internet and related businesses	7	549	1.26%	Engineering and technology	2	168	1.18%
Foundation and new materials industry	19	425	4.28%	Manufacturing of industrial goods and trade of industrial goods	6	209	2.79%
Enterprise Technology Center	5	345	1.43%	Petroleum chemical industry	2	166	1.19%
Cars and transportation equipment	4	304	1.30%	Knitted garment paper leather	2	118	1.67%
The biological and pharmaceutical industries	5	363	1.36%	Financial investment management and certification	3	75	3.85%
Electronic information industry	4	686	0.58%	Environmental protection and energy saving	5	114	4.20%
Equipment industry	2	434	0.46%	Advertising media	0	86	0.00%
Software and information technology services industry	12	1304	0.91%	Packaging logistics	1	40	2.44%
Urban industry	5	175	2.78%	All enterprises (total)	95	6241	1.50%
Culture and education printed and published	0	169	0.00%	In total		6336	

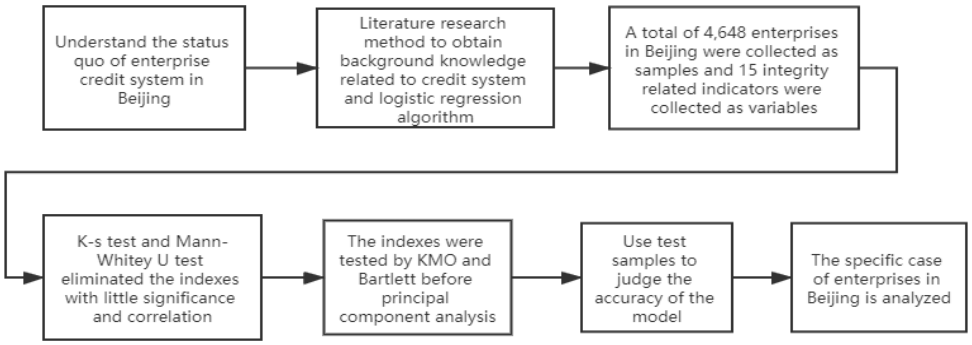

Fig. 1. Technical route of research.

Later, Altman proposed in 1968 that univariate analysis model may lead to errors in the prediction of enterprise failure risk (Diakite, 2015) [7]. He also gave an example that if a company has relatively low profitability and solvency, it will be regarded as a risk of failure (Diakite, 2015) [7]. However, if the firm has above-average liquid assets, the conclusion is different. Therefore, Altman started to study the multivariate analysis model from 1968 (Diakite, 2015) [7]. He divided the 66 companies in the study sample into two groups (Diakite, 2015) [7]. Group 1 is 33 manufacturing companies that went bankrupt between 1946 and 1965, while group 2 is 33 companies that did not go bankrupt (Diakite, 2015) [7]. He collected the companies' balance sheets and income statements, then compared 22 ratios. Among these 22 ratios, he found 5 ratios that could best predict the risk of enterprise failure (Diakite, 2015) [7]. These 5 ratios are working capital/total assets, retained profits/total assets, earnings before interest and tax/total assets, market price of equity/book value of liabilities, and sales/total assets (Diakite, 2015) [7]. Then he came to the conclusion that when z-Score value is 2.675, it is the best critical value to identify whether an enterprise has the risk of failure (Diakite, 2015) [7].

In addition to univariate analysis model and multivariate analysis model, probabilistic model is also used (Jia Wenrong, 2014) [6]. For example, logit is a conditional probability model. Martin, Henderson, Perry&Cronan used logIT analysis model to predict bond rating in 1984 (Jia Wenrong, 2014) [6]. They do this by selecting financial ratios and setting a warning line for credit risk. Peel and Keasey used logit model in 1987 and 1990 respectively to predict whether an enterprise would go bankrupt in several financial reports before bankruptcy (Jia Mei, 2014) [3]. Meyer also proposed the linear probability model in 1970 (Jia Mei, 2014) [3].

III. ENTERPRISE CREDIT INFORMATION COLLECTION

Enterprise credit information has two sources, one is internal non-public information, another is the public information collected by the Beijing Enterprise Credit Information Network, which is run by the Beijing Municipal Administration for Industry and Commerce.

The index selection of the raw data follows the following three principles. Firstly, the number of indexes in the model should be controlled. Too many indexes will reduce the correlation between indexes and lead to poor fitting degree of the model. Do not select too little, or the prediction accuracy of the model will be reduced. Second, the pertinence of index selection. The selected indicators should be linked to the enterprise credibility as far as possible. Third, the stability of index selection. Indexes should objectively reflect the integrity of the enterprise, as shown in Table II.

TABLE II: CREDIT INFORMATION COLLECTION OF MODEL SAMPLES

	Information type	samples	Data source		Information type	samples	Data source
1	Company name	6336	non-public	13	Major cases of illegal taxation	6336	non-public
2	The trademark	6336	non-public	14	Blacklist of government procurement	6336	non-public
3	Software copyright	6336	non-public	15	Commercial litigation	6336	non-public
4	Copyright table	6336	non-public	16	Copyright infringement	6336	non-public
5	The referee documents	6336	non-public	17	List of intellectual property Demonstration pilot projects	528	2019 List of National Intellectual Property Advantage Enterprises List of National Intellectual Property Demonstration Enterprises in 2019 2016 National Intellectual Property Demonstration Enterprise
6	The court announcement	6336	non-public	18	Seriously illegal enterprise	57462	Public disclosure of serious illegal enterprise information in 2019
7	Perform announcement	6336	non-public	19	Judicial assistance equity freeze	1486	List of companies subject to the 2018 judicial assistance Equity Freeze List of enterprises subject to equity freeze in 2019 List of enterprises subject to equity freeze in 2020
8	Faithless announcement	6336	non-public	20	List of Class A tax paying enterprises	50448	2017 Beijing Good faith Creation Enterprise review announcement list 2017 Beijing Good faith enterprises to create the final public list
9	The hearing announcement	6336	non-public	21	List of Honest enterprises	792	2017 Beijing Good faith Creation Enterprise review announcement list List of Honest enterprises in Beijing in 2017
10	Negative public opinions of operation	6336	non-public	22	List of operating exceptions	133612	List of abnormal enterprises in Beijing from 2019 to 2020.4.4 List of Enterprises subject to Administrative punishment in Beijing in 2019
11	Negative financial public opinion	6336	non-public	23	The administrative punishment	105904	List of Enterprises subject to Administrative punishment in Beijing in 2019
12	Suspected illegal or illegal conduct	6336	non-public				

The modeling preprocessing on the data of these 6,336 enterprises in Beijing mainly including two parts:

(1) Elimination of invalid information: The indicators of some enterprises cannot all be collected. For example, there are a total of 15 indicators, but some enterprises can only find 5 indicators information, the remaining 10 indicators cannot be found. The incomplete indicator information is not conducive to the accuracy of modeling (Table III).

(2) Elimination of extreme outliers: Outliers will greatly affect the fitting degree of the model curve, and eventually lead to a great decline in accuracy.

TABLE III: ELIMINATION OF DATA

Filter condition	quantity
In addition to the announcement of dishonesty, the rest of the data are 0 enterprises	299
More than three indicators are empty enterprises	1122
Judgment documents, executive announcement, court announcement, opening announcement more than 50 enterprises	267

IV. CONSTRUCTION OF CREDIT RISK MODEL BASED ON LOGISTIC REGRESSION

A. Binary LOGISTIC Regression Model

The probabilistic model is divided into linear probabilistic model and nonlinear probabilistic model. For the linear model, when the value of independent variable is too large or too small, the value of dependent variable, namely the probability value of the occurrence of events, will exceed 1, and the judgment will lose some accuracy (Shi Mengwei,2010) [9]. Therefore, this study decided to use the logistic regression model in the nonlinear probabilistic model to predict the degree of integrity of enterprises in Beijing, which can well avoid the shortcomings of the linear model. For example, the dependent variables of the Logistic regression model can be discrete or continuous.

Logistic regression models are divided into multiple regression and binary regression. Where, multiple regression method requires explanatory variables, that is, dependent variables are continuously spaced data (Shi Mengwei,2010) [9]. But there are only two options in predicting whether the company is an illegal one. 1 means that the enterprise is classified as an illegal enterprise, 0 means that the enterprise is classified as a normal enterprise. Therefore, the binary regression method was used in this study. Because binary logistic regression requires that dependent variables are dichotomous variables and can only be represented by 0 and 1, which exactly conforms to the type of explanatory variables in this study.

B. Credit Risk Model Based on Logistic Regression (Fig. 2)

Enterprise credit y is a dichotomous variable, namely only 0 and 1, $y \in \{0,1\}$. $p_i = P(y_i = 1 | x_i)$ is the violation rate of enterprises in Beijing, and $P(y_i = 0 | x_i) = 1 - p$ is the normal probability of enterprises.

The regression model of enterprise credit assessment in Beijing can be expressed by the following formula:

$$\ln(\frac{p}{1-p}) = g(x) = w_0 + w_1 x_1 + \ldots + w_n x_n \quad (1)$$

It can be deduced that the expression of the QTH influence factor of Y is:

$$\ln(\frac{p_i}{1-p_i}) = w_0 + \sum_{q=1}^{q} w_q x_{iq} \quad (2)$$

After converting the above formula into the nonlinear form, it can be obtained that:

$$p_i = \frac{1}{1 + \ln(-(w_0 + \sum_{q=1}^{q} w_q x_{iq}))} \quad (3)$$

According to the above formula, the integrity of a specific Beijing enterprise can be evaluated, among which w_0 and w_q are evaluated according to the likelihood estimation method. The specific construction process of the risk model is shown in the Fig. 2 below:

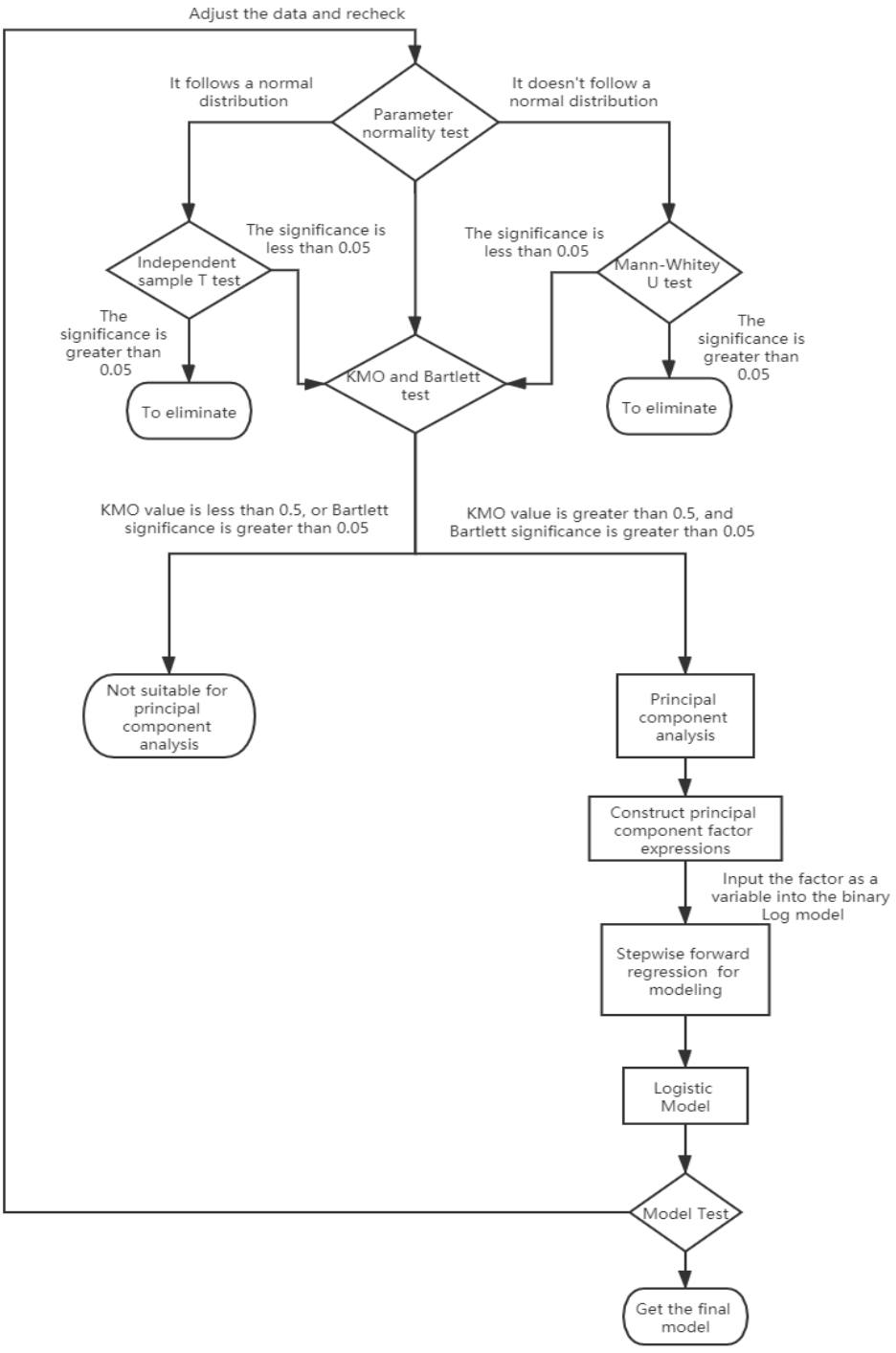

Fig. 2. Specific construction process of the risk model.

In this study, enterprises published on the list of dishonest enterprises are classified as illegal enterprises, otherwise, they are classified as normal enterprises. A total of 4,648 enterprises were collected as samples in this study. The total samples are divided into modeling samples and testing samples. Among them, there are 4,600 modeling samples, 138 of which are illegal enterprises and 4,462 normal enterprises. The test samples are randomly selected from the total samples. There is a total of 48 enterprises, among which 39 are illegal enterprises and 9 are normal enterprises.

C. Selection of Model Indicators: K-S Test

The Kolmogorov-Smirnov test is used to test the normality of each financial index. When the indicator set is normally distributed and the population variance of the two groups of samples is equal, independent sample T-test is used. For indexes that do not conform to normal distribution, Mann-Whitey U test in non-parametric test is used. The purpose of this step is to test the difference between the classified data and the quantitative data.

TABLE IV: SINGLE SAMPLE K-S TEST RESULTS

	Trademark	Software copyright	Copyright table	The referee documents	The court announcement	Perform announcement
Case quantity	4600.00	4600.00	4600.00	4600.00	4600.00	4600.00
The normal parameters a, b (average)	36.81	21.21	3.38	9.27	0.88	0.37
The standard deviation	173.47	41.26	105.66	53.57	2.83	6.06
The most extreme difference (absolute)	0.42	0.30	0.49	0.43	0.41	0.48
positive	0.34	0.21	0.47	0.32	0.41	0.48
negative	-0.42	-0.30	-0.49	-0.43	-0.38	-0.48
Test statistics	0.42	0.30	0.49	0.43	0.41	0.48
Asymptotic significance (double tails)	.000c	.000c	.000c	.000c	.000c	.000c

TABLE V: SINGLE SAMPLE K-S TEST RESULTS

	The hearing announcement	Pilot demonstration of intellectual property rights	List of tax-paying enterprises	List of operating exceptions	The administrative punishment
Case quantity	4600.00	4600.00	4600.00	4600.00	4600.00
The normal parameters a, b (average)	1.62	0.01	0.48	0.02	0.13
The standard deviation	7.72	0.11	0.50	0.14	0.57
The most extreme difference (absolute)	0.42	0.53	0.35	0.54	0.51
positive	0.33	0.53	0.35	0.54	0.51
negative	-0.42	-0.46	-0.33	-0.45	-0.41
Test statistics	0.42	0.53	0.35	0.54	0.51
Asymptotic significance (double tails)	.000c	.000c	.000c	.000c	.000c

TABLE VI: SINGLE SAMPLE K-S TEST RESULTS

	The judicial freeze	List of Honest enterprises	Serious illegal
Case quantity	4600.00	4600.00	4600.00
The normal parameters a, b (average)	0.01	0.05	0.00
The standard deviation	0.07	0.22	0.02
The most extreme difference (absolute)	0.52	0.54	0.51
positive	0.52	0.54	0.51
negative	-0.47	-0.41	-0.49
Test statistics	0.52	0.54	0.51
Asymptotic significance (double tails)	.000c	.000c	.000c

It can be seen in the Table IV, Table V, Table VI trademark, copyright, copyright table, the written judgment, court announcement, announcement, the hearing announcement, intellectual property pilot demonstration, list of corporate tax business exception list, administrative penalties, judicial freeze, the good faith enterprise list and serious illegal don't obey the normal distribution, so the next step is to use the Mann - Whitey U method for further testing.

D. Selection of Model Indicators: Mann-Whitey U Test

The Mann-Whitey U test can compare whether there is a significant difference between the mean values of explanatory variables Y and independent variables.

In order to analyze the ability of each independent variable index to distinguish illegal enterprises from normal enterprises, Mann-Whitey U, a non-parametric test, was used in this study to analyze independent variables that do not obey normal distribution.

According to the results, the six indexes of judgment document, court announcement, executive announcement, court hearing announcement, list of honest enterprises and list of tax-paying enterprises can be significantly divided into illegal enterprises group and normal enterprises group.

E. Extraction of Principal Component Factors: KMO and Bartlett Tests

KMO and Bartlett tests are used to determine whether

indicator variables are suitable for principal component analysis. KMO value is used to judge the correlation between independent variables. The value is in the interval. The closer the KMO value is to 1, the greater the correlation between indicators will be, and vice versa.

If the KMO statistic is lower than 0.5, it indicates that the correlation between indicators is low and it is not suitable for factor analysis. Because only when there is a high correlation between the indicators, the common factor can be extracted.

The Bartlett test statistic is derived from the determinant of the correlation coefficient matrix. If $p > 0.05$, the correlation coefficient matrix is judged to be the identity matrix, the null hypothesis cannot be rejected, and it is not suitable for factor analysis. If $p < 0.05$,the null hypothesis should be rejected.

TABLE VII: KMO AND BARTLETT TEST RESULTS

KMO and Bartlett tests		Bartlett sphericity test	
Number of KMO sampling appropriateness	0.67	The approximate chi-square	5161.51
		Degrees of freedom	15.00
		significant	0.00

As can be seen from Table VII, the value of KMO is 0.635, which is greater than 0.6. Therefore, factor analysis can be continued. According to the significance P value of Bartlett test is 0.000, it can be concluded that the null hypothesis should be rejected. That is, there is a good correlation between independent variables in the sample, so factor analysis can be performed.

F. The Score Coefficient of Principal Component Factor

Only factors with factor eigenvalue over 1 were selected for analysis. The cumulative value of the second principal component factor is 55.149%, which means the first two factors reach the cumulative distribution of 54.09%. It can well explain the difference of the original variables in Table VIII.

TABLE VIII: TOTAL VARIANCE INTERPRETATION

composition	Initial eigenvalue	Extract the sum of squares of loads	
	sum	Percentage of variance	Cumulative %
1.00	2.25	37.41	37.41
2.00	1.00	16.68	54.09
3.00	0.99	16.53	70.62
4.00	0.88	14.59	85.21
5.00	0.62	10.28	95.49
6.00	0.27	4.51	100.00

TABLE IX: COMPONENT MATRIX

	component	
	1.00	2.00
The referee documents	0.84	-0.01
The court announcement	0.47	0.06
Perform announcement	0.71	-0.04
The hearing announcement	0.89	0.00
List of tax-paying enterprises	-0.13	-0.24
List of Honest enterprises	-0.02	0.97

It can be seen that for the first principal component factor, the biggest contribution is the judgment document and court notice. For the second principal component factor, the biggest contribution is the tax enterprise list and the court notice.

According to the component matrix in Table IX, the two principal component factors can be obtained as follows

$$F_1 = 0.84x_1 + 0.469x_2 + 0.71x_3 + 0.893x_4 - 0.127x_5 - 0.021x_6 \quad (4)$$

$$F_2 = -0.014x_1 + 0.064x_2 - 0.036x_3 - 0.004x_4 - 0.238x_5 + 0.969x_6 \quad (5)$$

V. RESULTS AND VALIDATION OF THE MODEL

After screening the modeling indicators and obtaining the principal component factor, the principal component factor was introduced into the binary Logistic regression model, and finally the empirical parameters and results were obtained. The empirical process of this study includes a total of 4,547 samples, and the data is complete without any missing data. The forward method was used to incorporate F1 and F2 variables in the logistic regression model. Table X shows that the Wald value of F1 and F2 is 129.163 and 12.355 respectively.

It can be concluded that Wald test results are significant, that is, there is a significant correlation between these two independent variables and dependent variables. Then hosmer-Lemeshow was used to test the goodness of fit of the model, and the results were shown in the Table XI.

The chi-square statistics of the model correspond to P =4.925, and the significance is 0.766. The significance is greater than 0.05. Therefore, if the null hypothesis is rejected, it can be proved that the overall fitting degree of the model to the sample observation data is relatively good.

TABLE X: VARIABLES IN THE EQUATION

	B	Standard error	wald	Degrees of freedom	significant	Exp(B)	95% confidence interval for EXP(B)	
							The lower limit	
step 1a	F1	0.08	0.01	221.23	1.00	0.00	1.08	1.07
	constant	-4.81	0.16	951.57	1.00	0.00	0.01	
Step 2b	F1	0.07	0.01	129.16	1.00	0.00	1.07	1.06
	F2	-1.22	0.35	12.36	1.00	0.00	0.29	0.15
	constant	-4.96	0.17	886.62	1.00	0.00	0.01	

TABLE XI: HOSMER-LEMESHOE TEST

step	chi-square	Degrees of freedom	significant
1.00	8.10	7.00	0.32
2.00	4.93	8.00	0.77

Next, it can be seen from the Table XII of the model abstract that after two steps of regression, the likelihood ratio test statistic of the model is 815.385, the fitting coefficient Cox-Snell R square reaches 0.088, and The Negoko R square reaches 0.373.

It can be explained that the two independent variables in the current model can well explain the integrity of the dependent variable enterprise, and the model has a good fitting degree.

TABLE XII: MODEL SUMMARY

step	Minus 2 logarithmic likelihood	Cox snell R squared	Nagorco R squared
1.00	834.178a	0.08	0.36
2.00	815.385a	0.09	0.37

Therefore, the credit evaluation model of enterprises in Beijing can be obtained as follows:

$$P = \frac{1}{1+e^{-(-4.955+0.068F_1-1.224F_2)}} \quad (6)$$

This study divides the sample data of Enterprises in Beijing into two parts: test sample and model sample. In order to test the prediction effect of this model, 39 normal enterprises and 9 offending enterprises in the test sample were inserted into the established Logistic model, and 0.5 was regarded as the threshold for offending enterprises. Then calculate the corresponding default rate, and then compare with whether the company is an illegal enterprise, get Table XIII.

TABLE XIII: TEST RESULT

The actual violations	Actual non-violation	Predicts irregularities	Forecast non-violation
9	39	7	38
Accuracy of illegal prediction		0.777778	
Non-violation prediction accuracy		0.974359	
Overall accuracy		0.9375	

VI. CONCLUSION

Based on the previous research knowledge of scholars, this paper establishes a mathematical model that can predict the credit status of enterprises in Beijing. This mathematical model is based on logistic regression algorithm and k-S test. Mann-whitey U test completes the screening of 15 sample indicators, and the remaining 6 valid indicators are finally left. Then, 6 valid indicators were put into the binary Logistic model and the forward method was used for regression. Finally, Wald test and Hosmer-Lemeshow test are combined to test the fitting degree of the model.

After the real and effective mathematical model is obtained, the prediction accuracy of the model is preliminarily tested with test samples. As can be seen from the test results of the model, this model has a high prediction accuracy, with its total accuracy up to 94%. It can well provide support for government decision-making, help consumers to choose reliable businesses, and help enterprises to choose more trustworthy partners.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

In the whole process, the author herself collects, organizes, thinks, revises and finally completes this article.

ACKNOWLEDGEMENT

This thesis is completed under the guidance of Zhenji Zhang and Daqing Gong. From the topic selection, literature review, further polishing of research direction and determination of research methods, they gave me careful suggestions and timely guidance, which helped me to find the direction of my paper and conduct specific research. Their guidance made me have a clearer understanding of the research process and a deeper understanding of the research field.

I would like to express my gratitude to Beijing Jiaotong University for its cultivation in the past years and to every teacher who helped me with my professional course guidance. My progress and growth are inseparable from their instruction. I would also like to thank the owner of the literature and research ideas cited in this paper. It is the hard work of predecessors that has laid a good knowledge foundation for the smooth progress of my research. Finally, I would like to thank my classmates and my family, whose care and companionship have always been my driving force.

REFERENCES

[1] X. Lu, "Research on enterprise credit in Beijing," Ph.D. dissertation, Capital University of Economics and Business, 2003.
[2] P. M. Addo, D. Guegan, and B. Hassani, "Credit risk analysis using machine and deep learning models," *Risks*, vol. 6, no. 2, p. 38, 2018.
[3] M. Jia, "Research on SME credit rating index system," Ph.D. dissertation, Minzu University of China, 2013.
[4] D. Zou, "Research on the construction of credit rating index system for small and micro enterprises," Ph.D. dissertation, Lanzhou University, 2013.
[5] S. Yu, "Exploration on the construction of enterprise credit evaluation system and system platform," Ph.D. dissertation, Suzhou University, 2014.
[6] W. Jia, "Research on the Construction of SME Credit Rating Index System in China," Ph.D. dissertation, Tianjin University of Finance and Economics, 2012.
[7] S. M. Diakite. (2015). Corporate failure prediction: Z-score model and PBT model investigated in different external situations (Order No. 3729820). [Online]. Available: https://ezproxy.rit.edu/login?url=https://search-proquest-com.ezproxy.rit.edu/docview/1734473907?accountid=108
[8] W. Beaver, "Financial ratios as predictors of failure," *Journal of Accounting Research*, vol. 4, pp. 71-111, 1966.
[9] M. Shi, "Research on enterprise financial credit evaluation based on logistic regression model," Ph.D. dissertation, North China Electric Power University (Hebei), 2010.

Ziyuan Li was born on 12 Nov., 1998. Ziyuan Li did her undergraduate course in Beijing Jiaotong University with a bachelor degree of management information system (May, 2020); now she is making her master study in Columbia Business School. During her undergraduate study, she was selected to Global Scholar Program and had one-year experience of studying abroad in Rochester Institute of Technology.