Acta Oeconomica, Vol. 65 (S2), pp. 3–16 (2015)
DOI: 10.1556/032.65.2015.S2.2

STUDY ON EARLY WARNING OF ENTERPRISE FINANCIAL DISTRESS – BASED ON PARTIAL LEAST-SQUARES LOGISTIC REGRESSION

Kun XU[1] – Qilan ZHAO[2] – Xinzhong BAO[3]

[1]*School of Management, Beijing Union University, Chaoyang District, Beijing, 100101. E-mail: gltxukun@buu.edu.cn*
[2]*Economic and management school, Beijing Jiaotong University, Haidian District, Beijing, 100044. E-mail: zhaoqilan@126.com*
[3]*School of Management, Beijing Union University, Chaoyang District, Beijing, 100101. E-mail: xinzhongbao@vip.sina.com (corresponding author)*

Establishment of an effective early warning system can make the company operators make relevant decisions as soon as possible when finding the crisis, improve the operating results and financial condition of enterprise, and can also make investors avoid or reduce investment losses. This paper applies the partial least-squares logistic regression model for the analysis on early warning of enterprise financial distress in consideration of quite sensitive characteristics of common logistic model for the multicollinearity. The data of real estate industry listed companies in China are used to compare and analyze the early warning of financial distress by using the logistic model and the partial least-squares logistic model, respectively. The study results show that compared with the common logistic regression model, the applicability of partial least-squares logistic model is stronger due to its eliminating multicollinearity problem among various early warning indicators.

Keywords: early warning, financial distress, partial least-squares logistic regression, logistic model

1. INTRODUCTION

Financial situation is the earliest standard to assess the enterprise risk, because financial crisis often leads to the credit crisis, and the financial distress also indicates the higher credit risk. With the rise and development of the enterprise, as early as the initial stage in the last century, early warning of risk has been explored and studied in the world. After nearly a century development, the research on financial early-warning method has achieved more mature results (Mérő 2014). Initially, simple univariate analysis and multivariate discriminant analysis (MDA) were used for the study on the early warning of financial distress. Fitzpatrick

(1932) first found in the study by the univariate analysis method that the net profit ratio of rights and interests and the ratio of property right had higher prediction of energy. Subsequently, MDA which uses multiple variables to distinguish the categories of study objects was applied in the early warning of financial distress field (Virág and Kristóf 2014). In 1968, Altman firstly used the MDA to analyze 33 couples of companies, and built the famous Z-SCORE model. After MDA, the Logistic regression and Probit regression models were introduced into the early-warning model of financial distress.

Logistic regression method can overcome the boundedness of the hypothesis that the variable belongs to the normal distribution and the samples have the same covariance matrix in MDA model. Ohlson (1980) adopted the logistic regression method to construct the financial early-warning model of enterprise. Later, it was found that the size of company, the capital structure, operating performance and financing capacity of company had very high contribution for the discriminant accuracy of financial early warning, and the model prediction accuracy is more than 96%. Zmijewski (1984) used the Probit model for analysis for the first time. Since 1990s, non-statistical method such as neural network, etc. began to be used to establish the early-warning model of financial distress. The research results of Desai et al. (1996) showed that compared with linear discriminant analysis and logistic regression model, the prediction accuracy of neural network model was higher. Jae Kwon Bae (2012) analyzed the annual data of 1888 manufacturing companies collected by Korea Credit Guarantee Fund, built a support vector machine (SVM) model based on radial basis function, and compared the discriminant accuracy of RSVM and artificial intelligence technology. In the traditional study on the financial distress prediction by using the data mining theory, a single method is often adopted. The use of single model would cause the overvalued or undervalued predictive ability of model due to the each kind of method with more or less fixed defects. For that reason, Sung-Hwan Min, Jumin Lee et al. (2006) combined the genetic algorithm and SVW to predict bankruptcy, improved the performance of SVW in two aspects of feature subset selection and parameters optimization. In addition, there were other scholars such as Zhongsheng Hua (2007), Melek Acar Boyacioglu (2009), Yang Zijiang (2011), and so on who combined the SVW and other methods. Esteban Alfaro and Noelia García et al. (2008) applied the integrated learning algorithm Adaboost for the prediction of company failure, as well as considered the effect of quantitative index and qualitative index. Compared with the neural network, this method reduced 30% generalization error; moreover, Sun Jie, Jia Ming et al. (2011) also used this method. Sungbin Cho and Jinhwa Kim et al. (2009) optimized the financial early-warning binary classification problem by combining the multivariate discriminant analysis, logistic regression, decision-making tree and neural network. Through the empirical test, this model is superior to discriminant analysis and feed forward neural network.

In the application of various models, the logistic model has been widely used in the empirical study on financial early warning of various countries due to its relatively loose application condition. But the independent variable as the compositional data has the unit-sum constraint and multicollinearity, and the ordinary logistic model is extremely sensitive to multicollinearity.

Hence, the traditional logistic model can cause the expansion of standard error, and then would meet bottleneck in the process of modeling. The conference about partial least-squares method and the related methods was launched in 1999, which promoted its research and application in the fields of management, engineering and life science, and so on. In 2005, Vinzi and Tenenhaus (2005) uses PLS Logistic Regression model to achieve the treatment in advance for the initial explanatory variable to remove the effect of unit-sum constraint by Log-ratio conversion for the principal component data. That can convert the principal component data into the latent variables that can take values freely rather than non-component data and solve the problem of common logistic regression modeling. Hair et al. (2014) also studied the application of this method.

The data of real estate industry listed companies in China are used to evaluate the early warning of financial distress by using the logistic model and the partial least-squares logistic model, respectively.

2. COMPARATIVE ANALYSIS OF TWO KINDS OF MODELS

2.1. Logistic regression model and its characteristics

Logistic function was initially put forward and applied to demography by P.F. Verhuist in Belgium. Logistic regression, namely, logistic regression analysis, is a multiple quantitative analysis method to analyze and predict discrete-type dependent variable according to single or multiple continuous or discrete variables. It is mainly used to study the relationship between the probability of various states for the dependent variables and the independent variable values, and probabilistically predict the events affected by multiple factors. Logistic regression model is the further expansion and extension for the multivariate linear model. Its nonlinear characteristics replace the linear ones, the dependent variables are converted from the quantitative variables into qualitative variables. Simultaneously, it gradually improves the many assumption problems in the univariate and multiple discriminant method, and is more consistent with the actual economic situation. Therefore, it has been widely used in various fields. The essence of the model is a multivariate analysis method that is based on the statistical pattern classification, selects the relevant needed indexes to carry out 0–1 judgment. Logistic regression has several merits as follows.

1) Robustness: the model selects feature indexes, uses statistical methods to accurately describe the company's financial condition, and has the robustness in the process of classification.

2) Simplicity: during the process of selecting indicators, SPSS software is used to make the correlation analysis for the data. And then the feature space is reduced for feature index optimization selection, and keep primitiveness of data as far as possible. Less data input can get more accurate results, which can avoid dimensionality disaster.

3) Rationality: previously, the financial crisis early-warning model mostly required data obey the normal distribution, but the reality often largely deviates from the hypothesis proposition. Logistic regression model evades this defect, and reduces the restrictions for sample hypothesis conditions, so as to make the results more in line with the enterprise status.

Logistic regression is the most widely used multivariate analysis method, when the regression analysis is conducted on the explained variable of the binary classification. Scholars use multivariate logistic regression model to analyze the financial crisis of listed companies. Actually, they calculate and analyze the probability of the enterprise in financial distress by using logistic model according to the financial and non-financial parameters of company. Logistic regression model is to overcome the problems existing in the linear discriminant function, and its content is as follows.

Assume that Y represents financial situation of a company, when the financial crisis happens, $Y = 1$, when not happening, $Y = 0$. Y value depends on another unobservable variable M. It is assumed that there is a simple linear relationship between M and the used prediction value X $M = f(x) = \alpha + \beta x_i$ M as a critical value, its value determines whether the Y event occurs or not. If $M > 0$ that is equivalent to $Y = 1$.

So the probability that some prediction results occur (namely, Y = 1) is $P(Y=1) = P(M>0) = P(\alpha > -\beta x_i)$. Supposing the probability distribution function of α, the assumed probability distribution function of Logistic regression model is $F(t) = \frac{e^t}{1+e^t}$. According to $F(-t) = 1 - F(t)$, $P(Y=1) = P(M>0) = P(\alpha > -\beta x_i)$ can be transformed into $P(Y=1) = P(M>0) = P(\alpha > -\beta x_i) = 1 - P(\alpha \le -\beta x_i) = 1 - F(-\beta x_i) = F(\beta x_i)$. And then the probability that the first sample company occurs the financial crisis is $P(Y=1) = e^{\beta x_i} / 1 + e^{\beta x_i}$ and $\beta x_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$ in addition, $P(Y=0) = \frac{1}{1+e^{\beta x_i}}$. The ratio of the probability whether the event occurs or not is $p_i / 1 - p_i = e^{\beta x_i}$, taking the logarithm for both sides of the equation can obtain the linear equation of regression model.

$\ln(p/1-p) = \beta x_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$ is its Logit form, namely, $p = 1/1 + e^{-(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik})}$. Because p_i may be equal to 0 or 1, then the probability ratio would be equal to 0 or infinity, and its logarithm has not practical significance. Hence, it is difficult for the parameters of the Logit model to commonly adopt ordinary least-square method (OLS) to estimate. However, the maximum likelihood method is used to estimate parameters, and likelihood function formula of the model logarithmic function is:

$$l(\beta) = \min \log L(\beta) = \log(\prod_{i=1}^{m} h_\beta(x_i)^y (1 - h_\beta(x_i)^{1-y})$$

$$\text{Here } h_\beta(x) = g(\beta^T x) = \frac{1}{1 + e^{-\beta^T x}}, y = 0 \ \ or \ \ 1$$

Where β is the parameter to be estimated. The maximum likelihood estimate value of β can be resolved by iteration according to the sample data. Based on the research conclusions, optimal split point of probability is 0.5 as prediction, the company of $P > 0.5$ can be judged as the one with financial crisis; on the contrary, it is the healthy company.

2.2. Model and characteristics of partial least-squares logistic regression

It is known from the above that $N = \ln(p/1-p) = \beta x_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$. Where N is still the linear regression in relation to *x,* N is still monotonous to *N.* Hence, ordinary logistic regression model did not completely solve the problem of linear compensation. In 1982, PLS Logistic model analysis method was proposed for the first time, the quadratic term and cross term were introduced into logistic function, which makes the compensatory effect among the variables not only rely on each independent variable but also depend on other variables. Therefore, it contains all of the original independent variables, and also completely solves the linear compensation problem.

Change the equation $$N = \ln \frac{p}{1-p} = \beta x_i = \beta_0 + \beta_1 x_{i1} + \beta_3 x_{i3}.$$

The partial differential is used to find the effect of variables on the explained variables, for example, the influence of x_{i1} on *N.*

From the above formula, there are only three independent variables, but the formula contains nine explanatory variables such as x_{i1}, x_{i2}, x_{i3}, x_{i1}^2, x_{i2}^2, x_{i3}^3, $_{i1}x_{i2}$, $x_{i1}x_{i3}$, $x_{i2}x_{i3}$ and so on. They give full consideration to the relationship among the independent variables.

Partial least-squares logistic regression analysis is approximately equal to sum of the multivariate linear regression analysis, canonical correlation analysis and principal component analysis. Because the model is still using the formula of logistic regression model, the difference is the change of the index. And the index includes main component parts, and the correlation part between explanatory variables and explained variables. The research core of the partial least-squares analysis method is to build the model for the data containing multiple dependent variable and the analyzed data. The central hypothesis is that the observed data are generated by the system driven by a small amount of potential components, so the selection of data is vital to reflect the accuracy of model analysis.

3. EMPIRICAL ANALYSIS ON FINANCIAL CRISIS EARLY-WARNING MODEL OF ENTERPRISE

3.1. Early-warning indicators and sample selection

3.1.1. Standards of financial distress

Different scholars have different definitions for Financial Distress. For listed companies, the poor financial situation will bring numerous negative impact to the company. Take the stock trading for an example. In China, "Shanghai Stock Exchange stock listing rules" regulates that when the financial condition or other conditions of a listed company are abnormal, this makes the investors difficult to judge the company's prospect, and their rights may be damaged. Then the stock exchange will be conducted on Special Treatment (ST). The stock with the special treatment will be labeled as "ST" before abbreviation of stock, and daily price limits of stock quotes is 5%. If the following two conditions occur, the shares of listed companies will be specially treated. One is the net profit of two fiscal years of listed company audited being negative. The second is the audited net assets per share of a listed company in the most recent fiscal year being lower than par value of stock. Therefore, the ST of listed companies is regarded as the mark to be caught in financial distress in the related research in China. This paper also regards the ST of companies as the standard of being caught in financial distress.

3.1.2. The choice of financial early-warning indicators

Through the comparison of ST financial data and non-ST financial data, it is found that there exists mainly great difference of the cash flow between the two, which is regarded as the focus of the analysis. The financial indicators are used in this paper are profit ability, debt paying ability, operation ability and cash flow. The specific early-warning index systems are as follows.

1) Profitability indicators: return on net assets, net profit rate, gross profit rate, net profit, earnings per share, operating income, and main business income per share.
2) Operation ability indicators: receivable turnover ratio, receivable turnover days, inventory turnover ratio, inventory turnover days, current assets turnover ratio, and current assets turnover days.
3) Debt paying ability: current ratio, quick ratio, cash ratio, interest payment multiple, equity ratio, asset-liability ratio.
4) Cash flow indicator: ratio of net cash flow of operating cash to sales revenue, ratio of return of assets' operating cash flow, the ratio of operating cash flows to net income, cash flow rate, the ratio of operating cash flows to the debt.

3.1.3. Sample selection

Over the past decades, the real estate industry got rapid development in China. But with the slowdown of Chinese economic development and the adjustment of macroeconomic policy, the real estate industry is facing more serious challenges. This paper chooses the listed companies in Chinese real estate industry as the research object to validate the application of the two models. This article selects 132 listed real estate companies as the sample data, including 18 ST companies and 114 non-ST companies, and the financial data of ST and the non-ST companies is divided into two groups for comparison by enterprise annual in 2014 of the stock data center in Sina finance and economics official website.

3.2. Descriptive statistics analysis of sample data

The SPSS statistical software is used in this paper, compared with other statistical software, the SPSS has efficiency to display various management and analysis data as well as the advantage that dialog is used to show various kinds of performance options. And it can directly read the Excel software and DBF data software, and the results output are clear and intuitionistic. Moreover, it is suitable for non-professional statisticians. Before analyzing the data, the SPSS software can be used for data description to pretreat the data. The related indicators such as the minimum, maximum, total amount, mean and standard deviation (reflect the discrete degree among individuals in the group) of many objects can be obtained through the description of the SPSS software function. *Table 1* and *Table 2* show the descriptive statistical results of financial data for an ST company and a non-ST company, respectively.

Table 1. Financial data description of non-ST companies

	N	Min.	Max.	Sum	Mean	Std. deviation
Net profit ratio (%)	18	–121	30	–21	–1.18	37.120
Gross profit rate (%)	18	20	62	704	39.11	13.363
Net profits (million yuan)	18	–127	438	2,157	119.86	171.891
Earnings per share (yuan)	18	0	1	3	0.16	0.244
Operating income (million yuan)	18	11	3,554	15,143	841.27	947.175
Main business income per share (yuan)	18	0	3	24	1.35	1.086
Inventory turnover ratio (time)	18	0	1	5	0.26	0.282
Inventory turnover days (day)	18	276	9,114	47,865	2,659.17	2,446.667
Current asset turnover (time)	18	0	1	5	0.26	0.168
Current asset turnover (day)	18	587	7,547	40,869	2,270.50	1,847.253
Current assets turnover days (%)	18	1	16	47	2.62	3.347
Quick ratio (%)	18	0	13	20	1.12	2.956
Cash ratio (%)	18	1	486	907	50.41	112.257
Multiple of interest payment	18		3,551		–1,374.46	7,977.432
Stockholders equity ratio (%)	18	–20	86	666	36.99	23.917
Asset-liability ratio (%)	18	14	120	1,134	63.01	23.917
Ratio of net flow of operating cash (%)	18	–4	1	–5	–0.29	1.154
Rate of return of assets' operating cash flow (%)	18	0	0	0	0.01	0.158
Ratio of net flows to debt of operating cash (%)	18	–2	0	–1	–.07	0.537
Cash flow ratio (%)	18	–205	44	–270	–15.00	64.702
Valid N (listwise)	18					

It can be seen from the above data that when the company got into trouble, the financial condition would sharply deteriorate. Therefore, there is the mark to follow to predict whether the company gets into financial crisis by using the radix of financial indicators. Hence, the experience equation mentioned above can be used to predict the results of the dependent variable according to the values of independent variables as well as find the factors which have the more significant influence on the dependent variables.

Table 2. Financial data description of non-ST companies

	N	Minimum	Maximum	Sum	Mean	Std. deviation
Return on net assets (%)	113	–6	41	1,184	10.47	8.047
Net profit ratio (%)	113	–122	62	1,383	12.24	17.269
Gross profit ratio (%)	113	–15	77	4,186	37.04	14.463
Net profit (million yuan)	114	–97	12,551	67,286	590.23	1,490.139
Earnings per share (yuan)	114	0	2	45	0.39	0.418
Business income (million yuan)	114	0	103,116	531,949	4,666.22	12,074.699
Main business income per share (yuan)	114	0	15	361	3.17	2.919
Receivable turnover (time)	107	1	25,737	58,036	542.40	2,642.885
Receivable turnover days (day)	107	0	337	3,751	35.05	63.084
Inventory turnover ratio (time)	110	0	16	59	0.53	1.605
Inventory turnover days (day)	110	22	61,017	275,412	2,503.75	6,063.924
Turnover of current assets (time)	114	0	1	39	0.34	0.224
Current assets turnover days (day)	113	287	36,000	214,837	1,901.21	3,716.462
Current ratio (%)	114	0	28	249	2.18	2.531
Quick ratio (%)	114	0	21	92	0.81	1.939
Cash ratio (%)	114	0	2,044	5,881	51.59	191.323
Multiple of interest payment	114	–20,7165	72,098	6,995	61.36	22,108.640
Stockholders equity ratio (%)	114	–340	93	3,846	33.74	39.394
Asset-liability ratio (%)	114	7	440	7,554	66.26	39.394
Ratio of net flow of operating cash (%)	113	–5	2	–7	–0.06	0.657
Rate of return of assets' operating cash flow (%)	114	0	0	0	0	0.083
Ratio of net flows to net profit of operating cash (%)	111	–44	263	249	2.24	26.526
Ratio of net flows to debt of operating cash (%)	114	0	0	0	0	0.160
Cash flow ratio (%)	114	–82	51	–49	–0.43	20.519
Valid N (listwise)	103					

3.3. Empirical analysis process of logistic regression model

The SPSS software is used to obtain the treating processes and prediction results of the logistic regression model, as shown in *Tables 3* and *4*, respectively.

It can be seen from the above table that the accuracy of binary classification logistic regression model for predicting data is 82.6%, among them the accuracy of non-ST company is 90.4%, and the accuracy of ST company is 33.3%. Results obtained by the logistic regression are still not ideal, and there exists great difference.

Table 3. Omnibus tests of model coefficients

		Chi-square	df	Sig.
Step 1	Step	76.584	1	.000
	Block	76.584	1	.000
	Model	76.584	1	.000
Step 2	Step	5.650	1	.017
	Block	82.234	2	.000
	Model	82.234	2	.000
Step 3	Step	8.453	1	.004
	Block	90.687	3	.000
	Model	90.687	3	.000
Step 4a	Step	-.818	1	.366
	Block	89.869	2	.000
	Model	89.869	2	.000
Step 5	Step	5.924	1	.015
	Block	95.794	3	.000
	Model	95.794	3	.000

a. The cut value is .500

Table 4. Classification table

Observed			Predicted		
			Y		Percentage correct
			0 (Non-ST)	1 (ST)	
Step 1	Y	0	103	11	90.4
		1	12	6	33.3
		Overall percentage			82.6

3.4. Empirical analysis process of partial least-squares logistic regression model

According to the above description for the model, partial least-squares logistic analysis process can be divided into the following steps:

(1) Extract the potential factors: the first pair of constituent data in groups of dependent and independent variables are extracted, respectively, and their correlation would be the greatest. Assume that the first pair of constituents in the two groups extracted are t_1 and w_1, t_1

is the linear matrix of independent variable set $X = (x_1 \quad x_2 \quad \dots \quad x_k)^T$, namely,

w_1 is the linear matrix of independent variable set, namely, $t_1 = u_{11}x_1 + \dots + u_{1k}x_k = u_1^T X$.

In order to meet the needs of the regression analysis, that requires:

1) The variation information of its corresponding variable set are extracted from both t_1 group and w_1 group.

2) The degree of correlation of t_1 and w_1 reaches the maximum.

The R software is used to read data, and the model is established by the partial least-squares analysis to get effect analysis of the model. As shown in *Figure 1*, the difference of root mean square error approximately trends to be stable as taking to ten factors, that is to say, it can well represent the selected 24 indicators.

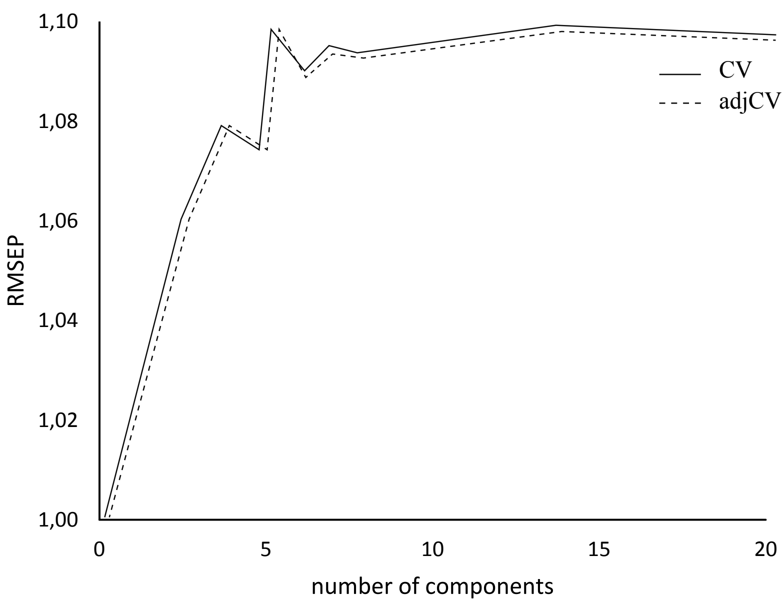

Figure 1. Cross-validation figure

(2) Fitting degree of evaluation model

It can be seen from the analysis of model effect that the influence on dependent variables tends to be stable since the 10th independent variable in 24 independent variables and the fitting degree of model is obtained.

Table 5. Explanatory variables detection

	1 comps	2 comps	3 comps	4 comps	5 comps	6 comps	7 comps	8 comps
X	10.00	23.29	28.60	33.05	38.85	43.04	50.32	55.35
Y	12.40	17.20	19.93	21.13	21.40	21.54	21.57	21.59
	9 comps	10 comps	11 comps	12 comps	13 comps	14 comps	15 comps	16 comps
X	59.14	61.67	63.93	66.64	70.58	73.51	76.59	79.91
Y	21.61	21.61	21.61	21.61	21.61	21.61	21.61	21.61
	17 comps	18 comps	19 comps	20 comps	21 comps	22 comps	23 comps	24 comps
X	83.93	85.82	89.29	91.26	93.14	95.37	97.44	100.00
Y	21.61	21.61	21.61	21.61	21.61	21.61	21.61	21.61

AIC 87.5956

AIC information criterion, namely, Akaike information criterion, is a kind of standard to measure fitting property of statistical model. From AIC = 87.5956, it can be seen that the fitting degree of the model is better.

(3) Extraction of factor and test results

The extracted factors are as follows.

1) x_1, x_4, x_8, x_{20}; 2) x_1, x_{24}; 3) x_{22}; 4) x_{12}, x_{17}; 5) x_{15}, x_{14}, x_{16}; 6) x_2, x_{23}; 7) x_8, x_{18}, x_7; 8) x_9; 9) x_{10}, x_{21}, x_{19}; 10) x_{11}, x_6, x_{13}

R software is used for the partial least-squares logistic regression analysis of financial data of listed companies, the results are shown in *Table 6*.

Table 6. Inspection results of partial least-squares logistic regression

Observed		Predicted Y: 0	Predicted Y: 1
Y	0	100	14
	1	3	15
	Overall percentage		

It can be seen from the above test results that the number of ST company of erroneous judgment is three, and the mistake rate was 16.7%. The number of non-ST company of erroneous judgment is fourteen, and the mistake rate was 12.3%, so the comprehensive accuracy is 87.1%. It can be observed that the prediction effect of partial least-squares logistic regression analysis for data is better, with the error rate of 12.9%, and has better prediction ability on the financial status of listed companies.

4. CONCLUSIONS

In view of the high-dimensional and high-correlated characteristics of financial data in listed company, logistic regression model can effectively reduce dimensions and eliminate collinearity, but it does not weigh the correlation between the dependent variables and independent variables, so as to result in the loss of important factors of interpretation variables which have the close relationship with the financial condition of enterprise when extracting the principal components. Partial least-squares logistic regression analysis model has the better degree of fitting and higher prediction ability for the business operation failure. The core assumption of the partial least squares is that the observed data are generated by the system driven by a small amount of potential components (not directly observed or measured variables) or process. Therefore, the key points of study are how to find the potential components which could represent the original data on the maximum limit and with the biggest covariance as well. It can be seen from the above that the partial least-squares logistic model has the stronger applicability than that of the common logistic model in the financial early warning of enterprise. The degree of fitting of PLS model established in this paper is 87.60%, hence the prediction results are more ideal.

ACKNOWLEDGEMENTS

This research is partially supported by the National Social Science Foundation of China (No. 14BGL034), the Key Program of Beijing Social Science Foundation of China (No. 15JGA003) and the Funding Project for Academic Human Resources Development in Beijing Union University.

REFERENCES

Alfaro, E., García N., Gamez, M. and Elizondo, D. (2008). Bankruptcy forecasting: an empirical comparison of AdaBoost and neural networks [J]. *Decision Support Systems*, 45, 110–122.

Altman, E.I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy [J]. *The Journal of Finance,* 4, 589–609.

Boyacioglu, M.A., Kara, Y. and Baykan, O.K. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: a comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey [J]. *Expert Systems with Applications,* 36, 3355–3366.

Cho, S., Kim, J. and Bae, J.K. (2009). An integrative model with subject weight based on neural network learning for bankruptcy prediction [J]. *Expert Systems with Applications,* 36, 403–410.

De Andrés, J., Lorca, P., De Cos-Juez, F.J. and Sánchez-Lasheras, F. (2011). Bankruptcy forecasting: a hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS) [J]. *Expert Systems with Applications,* 38, 1866–1875.

Desai V.S., Crook, J.N. and Overstreet G.A. Jr. (1996). A comparison of neural networks and linear scoring models in the credit union environment [J]. *European Journal of Operational Research,* 95(1), 24–37.

Fitzpatrick, P.J. (1932). A comparison of ratios of successful industrial enterprises with those of failed firms [J]. *Certified Public Accountant*, 10, 598–605, 11, 656–662, 12, 727–731.

Hair, J.F. Jr., Sarstedt, M. Hopkins, L. and Kuppelwieser, G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research [J]. *European Business Review*, (2), 106–121.

Hua, Z., Wang, Y. and Xiaoyan, X. (2007). Predicting corporate financial distress based on integration of support vector machine and logistic regression [J]. *Expert Systems with Applications,* 33, 434–440.

Jae, K.B. (2012). Predicting financial distress of the South Korean manufacturing Industries [J]. *Expert Systems with Applications*, 39(10), 9159–9165.

Mérő, K. (2014). Macroprudential warnings in the Euro zone and Hungary. *Acta Oeconomica*, 64(4), 397–417.

Min, Sung-Hwan, Lee Jumin and Han Ingoo (2006). Hybrid genetic algorithms and support vector machines for bankruptcy prediction [J]. *Expert Systems with Applications*, 31, 652–660.

Ohlson, J.A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research,* 18 (1), 109–131.

Sun, Jie, Ming-yue, J. Li, H. (2011). Adaboost ensemble for financial distress prediction: An empirical comparison with data from Chinese listed companies [J]. *Expert Systems with Applications*, 38, 9305–9312.

Tenenhaus, M., Vinzi, V.E., Chatelin, Y. and Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis*, 48(1), 159–205.

Virág, M. and Kristóf, T. (2014). Is there a trade-off between the predictive power and the interpretability of bankruptcy models? The case of the first Hungarian bankruptcy prediction model. *Acta Oeconomica*, 64(4), 419–440.

Yang, Z., You, W.J. and Guoli, J. (2011). Using partial least squares and support vector machines for bankruptcy prediction [J]. *Expert Systems with Applications*, 38, 8336–8342.

Zmijewski, M.E. (1984). Methodological issues related to the estimation of financial distress prediction models [J]. *Journal of Accounting Research,* 22, 59–82.