Comparison of the Prediction Effect between the Logistic Regressive Model and SVM Model

Li Sijia, Tan Lan*, Zhuang Yu, Yu Xiuliang
Zhejiang University City College
Hangzhou, P.R. China
tanl@zucc.edu.cn

Abstract—**Financial crises forewarning has important practical significance both for the investors and for the lenders. This paper uses the financial forewarning models, including the Logistic Regressive model and SVM model, to verify the feasibility of the short-term forecast for the financial situation of enterprises. And the paper also gives comparisons between these two models. The results of the study suggest that these two models are both feasible, and the SVM model can achieve better forecasting effects than the Logistic Regressive model.**

Keywords— financial crisis forewarning， Credit Risk Management，Logistic Regressive Model，SVM（Supporting Vector Machine）Model， prediction effect

I. INTRODUCTION

The effect of global financial crisis has not been wholly eliminated, and the economy has not fully recovered from the shock. China's banking sector has a lot to learn from the wave of the collapse of the banking industry in the United States. Commercial bank credit risk is the most important issue. Improvement of commercial bank risk management system must rely on the further research of financial situation of enterprises. With the development of China's capital market, Chinese scholars begin to introduce the western research, and based on these, a lot of theoretical discussion and empirical researches appears about the financial difficulties faced by Chinese companies, and many meaningful research results have been got. Researches involves the criterions for judging whether enterprises are in financial difficulties, the measuring and calculating of cost caused by financial difficulties, the building and improvement of financial early warning management system, and interaction between the behavior of enterprises in trouble and the behavior of the market, etc [1].

The company's financial situation is one of the focal questions which entrepreneurs, investors, creditors and auditors often pay close attention to. Operating well, companies can achieve healthy financial situation, enhance their reputation in the market, promote investors confidence, and expand financing channels. On the contrary, when meeting with financial difficulties, or even when approaching insolvency, the company not only brings great losses to investors, but also incurs great difficulty in the following production and management [2].

In fact, the company's financial crisis is a gradual process. Financial crises are not only with aura, but also predictable [3]. By establishing an effective early warning model of financial crisis, companies can obtain warning signals in advance, can take corresponding measures to improve the company's operations and financial condition as early as possible, investors or lender can elude or reduce risk, auditors can give a more accurate report on the financial situation of the enterprise. Meanwhile, the early warning model also has important practical significance for national securities regulatory authorities. By the model, they can monitor the quality of listed companies and reduce the risk of stock market.

The paper uses two econometric models, the Logistic Regressive model as the representation of the traditional financial early warning models, and SVM model (support vector machine model) as the representation of the emerging model of the financial early warning models, to forecast financial situation for some sample firms, summarizes the features of the two models, and then gives a comparison of their prediction effect. Finally, the paper gives some advices on how to approve the financial early warning models so as to give a better prediction.

II. OVERVIEW OF THE MODELS

A. *Logistic Regression Model*

Logistic binary response model is a nonlinear discriminant statistical method. In the binary response model, Y usually represents an individual or an experimental unit. Y has two possible values (Y=1, if the company is in financial crisis, and Y=0, if not). X represents the independent variable vector, $X= (X_1, X_2, X_3, \cdots, X_n)$, which, can be used to explain the possibility of the event of Y=1. P is used to represent a certain probability of occurrence of a particular situation. In this article, p stands for the probability of the occurrence of financial crisis for sample enterprises.

P can be obtained by the following equation:

$$\text{Logit}(p) = \text{Ln}\,[p/(1-p)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

And among them, $\beta=(\beta_1, \beta_2, \beta_3, .., \beta_n)$, is a set of corresponding regression coefficients of X, and β_0 is the model intercept. Both β_0 and β are parameters to be estimated.

B. *The Basic Principle of SVM Model*

SVM (Support Vector Machine, SVM) is a new machine learning method proposed by Vapnik [2]. The method follows the structural risk minimization principle and the limited sample assumptions, and can overcome some of the weakness of traditional machine learning (such as neural networks). For example, the SVM can overcome the convergence and other high dimensional disasters problems. So SVM has good learning ability and has great potential to be further generalized. Support Vector Machine (SVM) is

978-1-4244-6928-4/10/$26.00 ©2010 IEEE

one of the youngest branches in the statistical learning theory. It can be expressed in the form similar to neural networks. Statistical learning theory is considered as the best theory for prediction study and statistical estimation with small sample. It views the learning as a general process of function estimation based on empirical data. SVM follows the principle of structural risk minimization, which provides a new cognitive perspective for learning machine. The traditional neural network firstly fixes confidence risk, and then minimizes the empirical risk. While SVM adopts the opposite methods, fixing the empirical risk and minimizing the risk of Confidence, and then the input space is mapped to a high-dimensional inner product space, so the risk is only related with the number of samples inputted, independent of the input dimension, thus avoiding the "curse of dimensionality." By solving a linear constrained quadratic programming problem, SVM can get a global optimal solution, with no local minimum problem, and the fast algorithm ensures the convergence speed. The structural parameters are determined automatically in the above process from the samples, which means SVM can overcome the shortcomings of the traditional neural network, so can be used as a general-purpose learning machine. Financial crisis early-warning is a process of pattern recognition. SVM is suitable for pattern recognition, so suitable to be used in financial crisis earning-warning.

The basic idea of SVM can be summarized as: nonlinear separable problem in the input space can be transformed nonlinearly to a linear separable high-dimensional space, and in this high-dimensional space, optimal linear classification surface can be obtained. This nonlinear transformation process is realized by defining appropriate inner product function. So the training process of SVM becomes the process of solving the middle layer supporting vector node.

III. THE EMPIRICAL STUDY

A. *Description of the Date*

The scholars have different points of view about how to define financial crisis. The overwhelming majority of scholars domestically view special treatment (ST) due to abnormal financial position as logo of being in financial crisis. As this standard is consistent with the reality in China, and for the convenience of comparison of different studies, we also adopted this standard of definition for financial crisis.

More concrete criteria for selecting samples are summarized as following: (1) the sample companies should all belong to the manufacturing industry; (2) all the data are selected during the time spans between 2005 and 2008. All the date is collected from the securities Star database. Following the principles introduced above, we selected 20 ST companies (got special treatment for the first time during the time period between 2005 and 2008), and 20 non-ST companies as our sample. Since the year of (t-3) is the major turning point of financial position for ST companies, that is, one year before the company begins to lose money; the use of the data of sample companies in the year of (t-3) is of great practical significance. Therefore, the time period of the selected financial data ranges from 2002 to 2005.

To test the forecasting ability and stability of the extrapolation model, this paper divides the total sample into the training samples and test samples. Training samples are mainly used to build prediction model, the test samples to test the prediction model extrapolation capacity and stability. In this paper, sample of the company during 2005-2007 is used as a training sample; a sample of the company in 2008 is used as the test sample.

B. *Selection of Financial Indicators*

There exists no systematic and effective way to select variables for financial crisis forecasting model. Scholars always choose variables partly according to their individual preference, and partly relying on the statistical test [4]. As the focus of this paper is the comparison of the forecasting result between two forecasting models, so we make some appropriate simplification in the choice of predictor variables. Since Altman's Zeta model is acceptable to most scholars [5], we make some simple adjustment based on this model to achieve our goal.

We select 4 financial indicates from the model, including the ratio of working capital to total assets, ratio of EBIT to total assets, ratio of equity to debt, and ratio of income to the total assets, discard the indicate of ratio of retained earnings to total assets, and two financial indicates are complemented: the ratio of working financial capital to total financial assets, and the ratio of operating cash flow (net cash flow from operating activities) to the total assets. The working financial capital=cash + net short-term investments + short-term notes receivable - short-term loans - notes payable - long-term debt due within 1 year. The working financial capital reflects the company's cash liquidity. We choose this indicator due to the vulnerability of working capital to the earnings management and the quality of current assets, so as to complement the ration of working capital to total assets [6]. The ratio of operating cash flow to the total assets reflects situation of cash flow, on one hand, this indicator can reflect the company's ability to obtain operating cash, and on the other hand, it can also reflect the quality of the company's profitability and debt liquidity [7].

T test results of these indicators are shown in Table Ⅰ. As can be seen from Table Ⅰ, all financial indicators have reached the 1.96 significance level, indicating significant difference of the selected financial indicators between ST companies in the year of t-3 and the non-ST companies.

C. *Estimation Results by the Logistic Regressive Model*

By using the data from training samples, we give an empirical analysis based on the logistic regressive model introduce in section two. The estimation results by the Logistic Regressive model are reported in table Ⅱ. Two insignificant variables are dropped. The remaining four indicators all have significant influence on the possibility of incurring financial risks for companies.

TABLE I. T TEST OF SELECTED FINANCIAL INDICATORS

Ratio of	companies	Average value	Standard error	T value
working capital to total assets	Non-ST	0.232	0.205	2.690
	ST	0.169	0.198	
EBIT to total assets	Non-ST	0.066	0.043	8.996
	ST	0.033	0.044	
equity to debt	Non-ST	2.183	2.608	3.279
	ST	1.512	1.683	
income to the total assets	Non-ST	0.600	0.359	4.777
	ST	0.459	0.323	
operating cash flow to the total assets	Non-ST	0.054	0.073	6.839
	ST	0.009	0.089	
Working financial capital to total financial assets	Non-ST	0.060	0.223	9.621
	ST	−0.111	0.166	

And then we give a prediction for the financial risk occurrence for these training samples based on the coefficient estimated. When compared with the real-world data in 2008, we find that one non-ST company is predicted as a ST company inaccurately, and one ST company has not been recognized successfully. The total accuracy ratio is 96.67%.

D. Estimation Results by SVM model

The estimation results by SVM model are reported in table Ⅲ. Seen as in table Ⅲ, the prediction accuracy of our SVM for all of the 20 ST company and non-ST company is 100%. Error emerges neither in the first class nor in the second class. This indicates that SVM has had a great capacity for learning, so SVM can generate precise forecast for financial risk.

TABLE II. ESTIMATION RESULT BY THE LOGISTIC REGRESSIVE MODEL

Variables	coefficient	Standard error	Wald	P value
intercept	−0.133	0.232	0.328	0.567
income to the total assets	−0.706	0.377	3.505	0.061
Working financial capital to total financial assets	−3.869	0.571	45.883	0.000
operating cash flow to the total assets	−5.678	1.577	12.966	0.000
EBIT to total assets	−13.001	2.999	18.789	0.000

TABLE III. ESTIMATION RESULT BY SVM

items	Training samples			Test samples		
	ST	***Non -ST***	***In sum***	***ST***	***Non -ST***	***In sum***
In reality	20	20	40	20	20	40
In prediction	20	20	40	20	20	40
accurancy(%)	100	100	100	100	100	100
Error in first class (%)	0	0	-	0	0	-
Error in second class (%)	0	0	-	0	0	-

IV. CONCLUSIONS

SVM (Support Vector Machine Model) has more superior characteristic in performance comparing with the Logistic Regressive model. For our data, the recognition effect of SVM is better than the Logistic Regressive model. The prediction accuracy of SVM is 100%, while for the Logistic Regressive model, the accuracy is only 96.67%.

Superior performance of SVM is determined by its good theoretical basis. SVM firstly fixes empirical risk, and then minimize the confidence risk, by this process, SVM can achieve better performance. By contrast, the Logistic Regressive model may have several local minima, and the optimization process will converge to one of these minima. The quality of the solution depends on many factors, especially on the initial value of the network weights. To get good convergence performance, appropriate choices must be made about the initial parameters. And different choice of parameters means different results. To remedy this defect, SVM hypothesizes that the initialization value of hidden layer is distributed uniformly in the space over a weighted sphere, so can work out steady results.

ACKNOWLEDGMENT

This paper is supported by the construct program of the key laboratory in Hangzhou.

This paper is supported by the Student Research Project 2010 of Zhejiang University City College (number: X2010531145).

REFERENCES

[1] Gordon, “Towards A Theory of Financial Distress,” The Journal of Finance, vol.234, pp.221-243, May, 1971.

[2] Foster George, “Credit Risk Management,” proceedings of IEEE, vol.1, pp. 156-159, 2005.

[3] Wruck, “Financial Distress: Reorganization and Organization Efficiency,” Journal of Financial Economics, vol. 25, pp.39-54,1990.

[4] Koss, Fundamentals of Corporate Finance. 1rd ed. McGraw-Hill Inc, 1999.

[5] Morris, Early Warning Indicators of Corporate Failure, 1rd, ed. Ashgate Publishing Ltd, 1996.

[6] Ross, Corporate Management in Crisis, 1rd ed, Prentice Hall, 1998.

[7] Fitzpatrick P J, “A Comparison of Ratios of Successful Industries Enterprises with Those of Failed Film”, Certified Public Accountant, vol.54, pp88-98, Jun, 1999.

.