ISSN : 0974 - 7435
Volume 10 Issue 9
2014
BioTechnology
An Indian Journal
FULL PAPER
BTAIJ, 10(9), 2014 [3823-3830]

Comparison of financial distress prediction model on real estate listed companies

Rong Li, Wang Jiuling
Jiangxi University of Finance and Economics, Jiangxi Nanchang 330013, (CHINA)
E-mail : rongli91@tom.com

ABSTRACT

Take real estate listed companies,for example, we studied the internal connection between the companies marked by STand their financial indexesin the three years before the occurrence of financial distress, wemodel the financial crisis of early warning by statistical (Logistic Model) and artificial intelligence (BP neural networks, support vector machine model). Finally we made a comparison to the predict theaccuracy of the classification and the applicability of the three models. Empirical analysis shows that theaccuracy of support vector machine model is the highest, logisticmodel reflected the lowest accuracy, and the BP neural network stood in the middle. Based on the characteristics of the three types of models, wemade some researches ontheir application.

KEYWORDS

Listed real estate companies; Financial crisis; Warning model; Accuracy comparison.

INTRODUCTION

Financial crisis warning has been a hot issue since itcame into being in the field of financial theory and practice, especially under the fierce competitive market environment. With the rapid development of China's economy, the real estate companies,due to higher business risk, aremore prone to fall into financial crises. The real estate industry asthe pillar industry of the national economy, directly influences the other industries, and even affects the development of China's capital market. Therefore, the establishment of a reasonable model to correctly predict the real estate company's financial crisis has a practical significance.

For the increasingly high price of houses in the past years, the government has graduallyintroduced the "eight policy of regulating house price" and other special regulation, these macroeconomic policies aimingto contain the accumulated bubble of real estate industry. Affected by these policies, the demand for China's real estate is gradually shrinking, and housing prices began to slowly decline. However, real estate companies have to face to the dilemma on the aspect of cash flow, which lead to more and more financial crises happening to real estate companies. Study listed real estate company's financial risk profile can not only provide reliable information for the majority of investors, also can help the operators recognize their own risks, to some certain extent provide assistance to improve risk control and governance.

LITERATURE REVIEW OF FINANCIAL CRISISEARLYWARNING RESEARCH METHODS

Empirical research methods of financial crisis early warning at home and abroad mainly experienced through a process that from the class method by statistical models to artificial intelligence class model.

The first statistical model as[2] Fitzpatrick (1932), he choose single financial ratios as the variable to study the significant correlation withbankruptcy[5]. Martin (1977), first adopted Logistic model to predict bankruptcy in the banking sector. He selected 58 banks bankruptedas a research object from about 5700 members of the Federal Reserve Bankbetween 1970-1977, Then he picked 25 financial variables as alternative indexes. Finally, his model contains eight explanatory variables such as total assetsnet profit rate, respectively compared with the resultin Z, ZETA, Logistic model. He verified the accuracy of Logistic models is significantly higher than the other two models. Since entering tothat stage of artificial intelligence,[6]Odom and Sharda (1990),[7]Pang Qing Yue and Liu Xinyun (2011),[3]Gong Xiaofeng (2012) use the neural network model,[0]BaoXinzhong (2013) uses cluster analysis method,[4]Lu Yongyan (2012) uses Cox proportional hazards model,[8]Wang Qian (2013) discusses the support vector machine model,[9]YeHuanzhuo, Yang Qing (2013) applied adaptive Bayesian networkto the common early warning model for listed companies' financial crisis. However, the models related to the real estate industry for the financial crisis early warning are not always observed, especially those emphasized on how to choose warning variables based on the perspective of the particularity of the real estate industry as well ashow to study the comparative result on various types of predictive model accuracy. Due to the special nature of various industries, financial ratios are far different in industries. For example,there are clear differences on equity ratio between thesales industry and the real estate industry. If using a common early warning model, its prediction accuracy will be seriously reduced.

Therefore, the financial crisis early warning model for the real estate company must consider the disclosure of financial data from the goal company. We intend to use China's real estate companies as the research objects, respectively using Logistic model, BP artificial neural network model and support vector machine model to establish an appropriate financial crisis early warning model, and making a comparison between the three models.

SAMPLE SELECTION AND CONSTRUCTION OF INDEXSYSTEM

Sample selection

The companies which actually bankrupted in China's capital market do not meet the sample size, because they are not sufficient to achieve the goal and the bankrupt companies do not have universal applicability. Therefore, this article considered those stocks which were branded "ST" in front of their names symbolizing the financial crisis as the sample in real estate company. The empirical analysis fetched the raw data from the financial listed companies statements and financial indexes in "RESSET database" developed byBeijing Poly source RESSET Technology Co., Ltd. Some missing data were complemented by the reference derived from "Wind Database" and "Sina Finance Database". Considering that we plan to use the relevant financial data three years before topredict whether this companies will be Special Treated in the next year, we named the ST year as T years, 1-3 years ago are recorded as T-1, T-2, T-3 years. We divided the research sample into two types, learning and testing samples. The two types of samples consist of ST and non-ST companies.

Since the number of ST companies newly added in one year islimited. We, inorder to obtain adequate financial crisis samples, selected a total of 49real estate companies warned by ST in 2007-2013, andwe defined them as the company in financial distress. According to the principle of 1:1 matchedsample, we selectthe real estate companies that had never been STin 2007-2013 as the representative of non-financial distress. In the operation,we removed the company which disclosed incomplete data to reserve relevant information as much as possible. Besides, to avoid the influence of outliers in the data model accuracy the indexesexceeded the averages 50 times are also excluded.

Eventually we adopted the data from 67 real estate companies as the final sample,including 37 ST companies and 30 non-ST companies. More precisely, the learning sample consists of 30 companies, 24 non-ST companies, and the test samplesinclude 7 ST companies, 6 non-ST companies.

Index selection

We try toselect the financial indexes which able to represent the financial distress in the various aspects reflected the company's financial situation. The basic principle of selection are following :(1) It was commonly used in previous studies, and proved to be effective; (2) The selected indexes can be obtained from the corporate balance sheet, income statement and cash flow statement or can be effectively calculated; (3) Select the relativeindexes in order to eliminate the influence brought by company scale; (4) Consider the index of cash flow.

In accordance with the principle of selecting indexes, we mainly took the financial indexes from five areas. Such as corporate solvency, profitability, operational capacity, growth ability, cash flow capacity.

(1) Enterprise Solvency: Current ratio (X1), quick ratio (X2), cash flow debt ratio (X3), asset-liability ratio (X4);

(2) Corporate profitability: return on net assets (X5), sales margin (X6), sales gross margin (X7), earnings per share (X8);

(3) Business operations capabilities: Accounts receivable turnover ratio (X9), inventory turnover (X10), mobile asset turnover (X11), total asset turnover (X12);

(4) The ability of business growth: revenue growth (X13), net profit growth (X14) Total assets growth rate (X15), the net asset growth (X16), earnings per share growth rate (X17);

(5) The ability of corporate cash flow: Cash flow ratio (X18), cash sales of goods labor income / operating income (X19), the total assets of the cash recovery rate (X20), capital expenditure / depreciation and amortization (X21).

Data preprocessing

Because we couldn't judge that whether the financial indexes of the real estate company belong to the same distribution, we couldn't not use the two-tailed T test method to estimate thesignificance of financial indexes. We conducted a non-parametric testmethod, Kruskal-walis H test, to present the significant differences in financial indexes of among T-3year, T-2year and T-1year when the level of significance was 5%.

After analyzing bySPSS19,we can get only 7 significant financial indexes in T-3 year. They are current ratio, return on equity, sales margin, gross margin, earnings per share, inventory turnover and net asset growth. There are 9 significant financial indexes in T-2 year, adding2 extra indexes to the basis of T-3year, which are the growth rate of total assets and the capital expenditures / depreciation and amortization. Thenumber of significant financial indexes has reached to 11 in T-1year,another 2 indexes are added based on T-2year's significant indexes, which are revenue growth ratio and cash flow ratio. Because these significant indexes are not the same in each year, weshould build 3 different models based on the different significant indexesin these three years to predict which year the financial crisis will most likely occurred for the sample company.

MODEL SELECTION

Logistic model

Logistic regression is a probabilistic nonlinear regression model,and the linear expression is as following:

$$P = \frac{\exp(Y)}{1+\exp(Y)} = \frac{e^{(\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_mx_m)}}{1+e^{(\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_mx_m)}}$$

Due to the financial indexes overlaps and the articulation of financial data among the calculated variables in real estate companies, it is likely to cause the problem ofmulticollinearity in the modeling indexes and insignificance of the regression coefficients, so we will use principal component analysis to reduce the dimension of the selected indexes, and then to establish the model.

Principal component analysis:

We conduct a KMO and Bartlett coefficient test to the financial data extracted from the significant indexes for the T-3 year, T-2 year, T-1year to examine whether the selected financial indexes are suitable for factor analysis.

The KMO and Bartlett coefficient test results on T-3year'sfinancial indexes variable is 0.691, greater than 0.5, and it's suitable for the principal component analysis. Extract the first3 componentswhose eigenvalues is greater than 1, we can see the cumulative contribution rate is 73.993%, therefore the main component F1, F2, F3 can well represent the overall index and also have a higher credibility

Principal component is calculated as follows:

$$F1 = 0.729X1 + 0.386X5 + 0.350X6 + 0.558X7 + 0.069X9 - 0.092X14 + 0.835X15 + 0.778X16$$

$$F2 = 0.089X1 + 0.808X5 + 0.845X6 + 0.118X7 + 0.041X9 + 0.707X14 \quad 0.013X15 + 0.434X16$$

$$F3 = 0.153X1 - 0.019X5 - 0.049X6 - 0.460X7 + 0.932X9 + 0.058X14 + 0.026X15 - 0.059X16$$

Component F1 mainly represents business development capabilities, F2 mainlyrepresents profitability, and F3 mainlyrepresents the operating capacity. The T-3 year financial early warning model after the dealt with Logistic regression is:

$$P = \exp((0.473 - 1.294F1 - 1.409F2 + 0.350F3))/(1 + \exp((0.473 - 1.294F1 - 1.409F2 + 0.350F3)))$$

The KMO and Bartlett coefficient test results on T-2year's financial indexes variable is 0.408,less than 0.5, so the indexes in T-2year is not suitable for the principal component analysis. In another word, we do not reduce the dimension for the financial indexes in T-2 year. The T-2 year financial early warning model after the dealt withLogistic regression is:

$$P = \exp^{\uparrow}((1.277 + 0.046X1 - 0.052X5 + 0.002X6 - 0.065X7 - 0.327X8 + 0.029X10 - 0.026X15 + 0.01X16 + 0.048X21)$$

The KMO and Bartlett coefficient test results on T-1year's financial indexes variable is 0.541, greater than 0.5, we can make the principle component analysis. Extract the first 4 components whose eigenvalue is greater than 1, the total contribution rate is 66.723%, therefore the main component F1, F2, F3, F4 can also represent the overall index and have a good credibility.

Component F1 mainly represents corporate profitability, F2 mainly consider corporate cash flow capabilities, F3 mainly emphasize enterprise operating capacity, F4'smain consideration is the development of capabilities. The T-1 year financial early warning model after the dealt with Logistic regression is:

$$P = \exp^{\uparrow}((0.522 - 1.7F1 + 0.286F2 + 0.12F3 - 1.211F4))/(1 + \exp^{\uparrow}((0.522 - 1.7F1 + 0.286F2 + 0.12F3 - 1.211F4)$$

The Result of Logistic financial forecasting models are shown in TABLE 1.

TABLE 1 : Result of Logistic financial forecasting models

Observation		Prediction								
		T-3 year			T-2 year			T-1 year		
		ST or not		Accuracy of prediction	ST or not		Accuracy of prediction	ST or not		Accuracy of prediction
		0	1		0	1		0	1	
ST or not	0	2	4	33.3	4	2	66.6	5	1	83.3
	1	2	5	71.4	2	5	71.4	2	5	71.4
Total Accuracy				52.4			69.0			77.4

BP neural network model

BP neural network model is aartificial intelligence model which has stronger self-learning ability than the traditional statisticalmodel. It is nonlinear with high fault tolerance features. BP neural network is a multi-layer neural network based on back-propagation (Back-Propagation). It is the core issue to determine the optimal weights between the layers of neurons BP neural network by reiterative process.

Adopted the pre-processed data, weuse a neural network toolbox in Matlab (R2011a) to construct athree-layer feedforward BP neural network which contains an input layer, a hidden layer and an output layer. The STcompanyis recorded as 1 while the opposite is recordedas 0, and they are the training target data. At the same time, we use thenormalized financial indexes asthe training data because the normalization process can be unitizethe distribution of the sample, making it easy to calculate. Using the tansig transfer function between the input layer and the hidden layer, using the Purelin transfer function between the hidden layer and the output layer, and about the training function, we choose the way of trainlm.

Using the results in earlier principal component analysis, 8 financial indexes in T-3 years are regarded as the 8 neurons in the input layer joining the neural network training,in the T-3-year model: M = 8, N = 1 (1 in output represents the financial crisis will occur, output in 0 represents the good financial prospect). After we repeated tests, we found that when the number of neurons in the hidden layer is 25, the training group (learning samples) can be converged after the 61th iterations, the training error SSE fall below.001, and then we use the prediction group (test samples) to validate the predictive ability of the model, but the results shows only 53.6 %, in detail,3 of 7 ST companies 3 years later have been mistaken for healthy,3 of 6 healthy companies 3 years later has wrongly recognized as ST. Similarly, we established two 9-25-1, 11-25-1 three-layer BP neural network based on the financial indexes in T-2 year and T-1 year, the results show the prediction accuracy of T-2 year 77.4%, the prediction accuracy of T-1 model also reach 77.4%. Specific predictions are shown in TABLE 2.

TABLE 2 : Result of BP neural network model

Observation		Prediction								
		T-3 year			T-2 year			T-1year		
		ST or not		Accuracy of prediction	ST or not		Accuracy of prediction	ST or not		Accuracy of prediction
		0	1		0	1		0	1	
ST or not	0	3	3	50.0	5	1	83.3	5	1	83.3
	1	3	4	57.1	2	5	71.4	2	5	71.4
Total Accuracy				53.6			77.4			77.4

The results also confirm that the closer to ST year, the more predictive the data are, and the more predictive valuable on the corresponding model.

Support vector machine model

Support vector machine is called SVM for short. It is also a kind of artificial intelligence model. It is more accuratethan BP neural network when estimating in the study of a small sample statistical prediction problems. The basic idea is to classify the two types of samples by using a line of distinction, which making the class interval maximum. The sample points on the classified line are called support vectors.

We use a libsvm tool package of Matlab version under the environment of Matlab R2011a, establishinga support vector machine model based on RBF kernel function. Still,1 represents ST companies, 0 marks non-ST companies. After several parameter optimization, we determine the parameters C = 1000, gama = 0.00001.

Using thecode model = svmtrain (train_label, train_data, '- t 2 -c 1000 -g 0.00001') with the financial indexes in T-3, T-2, T-1-relatedas the training data in the model, we obtained three groups of support vector machine model. Only 1 company is misclassified among the 13 company samples of the test sample group (the first 7 are STcompanies; the left 6 are non-STcompanies). Specific predictions are shown in TABLE 3.

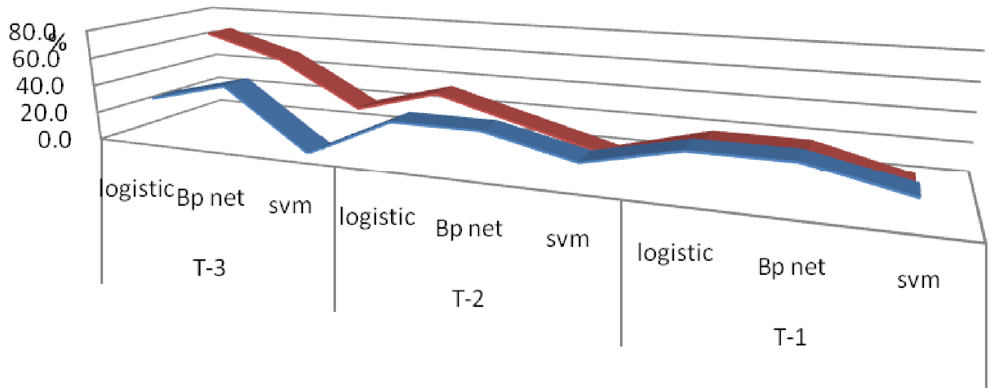
TABLE 3 : Result of support vector machine model predictions

Observation		Prediction								
		T-3 year			T-2 year			T-1 year		
		ST or not		Accuracy of prediction	ST or not		Accuracy of prediction	ST or not		Accuracy of prediction
		0	1		0	1		0	1	
ST or not	0	5	1	83.3	6	0	100.0	6	0	100.0
	1	0	7	100.0	1	6	85.7	1	6	85.7
Total Accuracy				91.7			92.9			92.9

The same to previous two models, the closer to the ST Year, the financial indexes express the stronger explanatory ability, the corresponding model are more predictive.

COMPARISON OF THE PREDICTION IN MODELS

Synthesize the above forecasting results of the three models, the effect of China's real estate listed companies' financial early warning model can be summarized in Figure 1. We can make a horizontal and vertical comparison about the model's false positives rate, according to the past practice, the financial crisis, we called the mistake of misjudging the ST Company as the Type I error, andthe other is called the Type II error.



	T-3			T-2			T-1		
	logistic	Bp net	svm	logistic	Bp net	svm	logistic	Bp net	svm
Type I error	28.6	42.8	0.0	28.6	28.6	14.3	28.6	28.6	14.3
Type II error	66.7	50.0	16.7	33.3	16.7	0.0	16.7	16.7	0.0

Figure 1 : Comparison of 3 model's prediction

The three prediction models all represent that the T-1year model is more predictive than the T-2 yearand T-3 year in the temporal dimension. Compared to other industries the real estate enterprises has some characteristics, such as longer cost recovery period, higher risk of operating costs. Therefore,it is no doubt that there is significance tofind the signal of future financial crisis in the early stages. We found that SVM warning model is the best after thehorizontal comparison of the three prediction models, therefore, support vector machine model is undoubtedly the best choice to predict if the financial crisisoccurs.

CONCLUSIONS AND OUTLOOK

(1) This article selected 21 indexes related to reflect the real estate company's profitability, solvency, operational capacity, growth ability, cash flow ability, it may be influenced by the particularity of real estate industry or the shortage of learning samples, which causes that we found only 10 significant indexes after hypothesis testing. By comparingthe found significant indexes of T-3 and T-1,we found that cash flow indexes influence more when the date is getting closer to the financial crisis date, which is consistent with the common reasons of broken capital chain when facing to the financial crisis, that is to say insufficient cash flow accelerate the corporate financial distress. Therefore, the short term cash flow forecasting index has symbolic significance especially for the corporate managers, who should be very wary of irregular changes in cash flow problems. In addition, these financial indexes, such ascurrent ratio, return on equity, sales margin, gross margin, earnings per share, inventory turnover and net asset

growth are all significant in the three-year forecast period, and it is worth paying long-term close attention for the users of financial statements.

(2) Although the forecast accuracy of Logistic model in T-3year and T-2 year is not high, this model is simple, easy to understand and easy to operate after all. So it is still acceptable to forecast the next year's financial condition for the small investors. But for the professionals in investment consulting institutions or real estate managers who are willing to handle the future development or even to estimate the trend of entire real estate, they need more accurate data in order to make strategic decisions based on the prediction or adjust strategies followedbythe industry trends. The results derived from neural network model and support vector machine model have more reference significance. Particularly, support vector machine model, due to the highest forecast accuracy and the least affected by sample size, it is the best choice to adopt support vector machine model under permitting conditions.

(3) The model does not take other non-financial data indexes into account, mainly due tothe causes like non-financial data is difficult to quantify, its sensitivity is not strong and its objectivity is not as good as the financial data and other reasons.

REFERENCES

[1] Bao Xinzhong,Yang Yi; Early Warning Based clustering - Rough Set - Neural Networks Enterprise Financial Crisis. Journal of Systems Management, 358-365 **(2013)**.

[2] F.J.Fitzpatrick; A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Firms, Certified Public Accountant, **2**, 589-605 **(1932)**.

[3] Gong Xiaofeng; Improvement and Exploration of Enterprise Financial Crisis Warning Variables Based on BP Neural Network. Research of Finance and Economics, 111-116 **(2012)**.

[4] Lu Yongyan; The industry Differences Research under Corporate Financial Distress risk, Macroeconomic Economic Research, **3**, **(2012)**.

[5] D.Martin; Early Warning of Bank Failures: A Logit Regression Approach, Journal of Banking and Finance, **9**, **(1977)**.

[6] M.D.Odom, R.Sharda; A Neural Network Model for Bankruptcy Prediction, IEEE INNS IJCNN, **2**, 163-168 **(1990)**.

[7] Pang Qing Yue, Liu Xinyun; Financial Crisis Prediction Method Based on Ant Colony Neural Network. Statistics and Decision, **3**, 554-561 **(2011)**.

[8] Wang Qian, Shi Wenchao, Xi Feng; An Empirical Study of Financial Crisis Warning Based on Rough Sets-Support Vector Machine. Statistics and Decision, **20**, 156-158 **(2013)**.

[9] Ye Huanzhuo, Yang Qing, Wang Yonghua; Intelligent Financial Crisis Warning and Reason Diagnosis Method Research- Based on Analysis of Adaptive Bayesian Network Model SABNM. Audit and Economic Research, 74-82 **(2013)**.