Financial prediction: Application of Logistic Regression with factor Analysis

Han dongmei, Ma Liqun, Yu changrui
Department of Information management and engineering
Shanghai University of Finance and Economics
Shanghai,China
hdm19610320@yahoo.com.cn

***Abstract*—Logistic regression is a very common method in financial prediction. In order to establish the more effective model, the paper introduces factor analysis into logistic regression to overcome multiple co-linearity among variables and meanwhile retain useful information of original variables. The result shows that this model has better predictive capability. This suggests that for policymakers and others interested in prediction system, this may be a useful tool for forecast firm failure.**

Keywords- financial prediction; Logistic regression; Factor analysis

I. INTRODUCTION

There are many different definitions about financial crisis, and a most common viewpoint is that financial crisis is the phenomenon that one enterprise is incapable to pay for due debt or expense, including the failure of fund management, bankruptcy, and as well as between the two kind of situations. And financial prediction is the possibility of predicting financial distress by establishing mathematical models on the basis of the existing financial ratios. Through such prediction, the investors and the creditors may avoid the risks of investments and lending money and the suppliers may formulate the more reasonable credit policies. Speaking of enterprises themselves, the effective financial crisis prediction is helpful to look for the root of financial distress, take the measure and prevent the further worsening situation.

In our stock market, the stocks named ST mostly refer to the kind "continuously two years loss or net assets of each one are lower than the face value", namely that the worsening of financial ratios is the main reason to make the company ST. Therefore, domestic researches generally make the ST Company as the standard of falling into financial crisis. This paper will also uses this standard.

The logistic regression model transforms the questions to calculating the probability of falling into financial crisis of a company in the certain period according to their financial condition. If the probability value is bigger than a certain supposed value, then we judge this company will fall into financial crisis. Because logistic regression does not request normal distribution of data, thus it is steadier than Discriminant analysis. It estimates parameters according to the Maximum likelihood estimators using sample data and then obtains the probability of taking certain value of response variable. The equation of logistic model is as follows.

$$\ln\left(\frac{p}{1-p}\right) = a_0 + a_1 F_1 + \cdots + a_n F_n$$

$$\Rightarrow p = \frac{\exp(a_0 + a_1 F_1 + \cdots + a_n F_n)}{1 + \exp(a_0 + a_1 F_1 + \cdots + a_n F_n)}$$

Here y= (1,0) refers to the times of occurrences of certain status. y=1 means it will happen and otherwise it won't happen. p=P(y=1) is assumed to the probability of occurrences.

Logistic regression model is also sensitive to multiple co-linearity, which is similar to other regression models. When relativity of variables is very high, very smaller change of samples will be able to bring the sweeping change of coefficient estimation, which will reduce the effect of prediction. But almost all financial ratios are mutually related and the degrees of relativity are often very high. In order to overcome the influence of multiple co-linearity, a very simple method is to delete certain variables from the model, but this may lose very useful financial information. Considering all these questions, this article first carried on the factor analysis on the financial ratios and then selected certain factor variables according to contribution rates to carry on logistic regression.

II. SAMPLE SELECTION

The data used in the paper are composed of basic and the test samples. The basic sample, consisting of 72 firms, is used to develop a financial distress prediction model, while the test sample, consisting of 28, is used to evaluate the predictability of the model developed. Among the 72 firms, there are 36 that were special treated in 2006 and among the 36, there are 18. And other firm is normal and used to match with ST companies.

After the initial groups are defined and firm selected, balance sheet and income statement are collected. Because of the large number of variables found to be significant indicators of corporate problems in past studies, a list of 32 potentially helpful variables(ratios) was complied for evaluation. The variables are classified into five standard ratio categories,

978-1-4244-2108-4/08/$25.00

including liquidity, profitability, leverage, and activity. The ratios are chosen on the basis of their popularity in the literature and their potential relevancy to the study.

We selected 32 financial ratios, and they are listed in the following table.

Financial Ratios			
X1	Return on equity	X17	Growth rate of revenue
X2	Return on assets	X18	Growth rate of total assets
X3	Operating earning ratio	X19	Growth rate of net assets
X4	ratio of net profit margin	X20	Growth rate of net earning
X5	ratio of gross profit margin	X21	Growth rate of EPS
X6	Current ratio	X22	Growth rate of NAVPS
X7	Quick ratio	X23	Cash/current debt
X8	liability/asset ratio	X24	cash/total debt
X9	equity/debt ratio	X25	cash reversion ratio
X10	net tangible asset/debt	X26	operating cash flow to profit ratio
X11	total assets turnover ratio	X27	cash flow per share
X12	current ratio	X28	net operating cash flow to sales ratio
X13	AR turnover ratio	X29	Operating/ total profit ratio
X14	Inventory turnover ratio	X30	Operating cost ratio
X15	net assets turnover ratio	X31	Company expenses ratio
X16	Fixed asset turnover ratio	X32	Management fee rate

III. Empirical Analysis

A. Factor Analysis

The factor analysis is a kind of multiple statistics methods that concentrates original variables into a few variables with the least information loss, and makes them have higher explanation functions. The factor variables are not from original variables directly and simply but are some new factors through new synthesis that can affect original variables, and simultaneously each variable is independent, therefore this can effectively overcome multiple co-linearity among original variables.

Using SPSS software, we can obtain Initial Eigenvalues and Extraction Sums of Squared Loadings.（Table1）

Table1

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.191	19.347	19.347	6.191	19.347	19.347
2	4.364	13.636	32.983	4.364	13.636	32.983
3	3.359	10.497	43.480	3.359	10.497	43.480
4	2.782	8.692	52.173	2.782	8.692	52.173
5	2.570	8.032	60.204	2.570	8.032	60.204
6	1.929	6.027	66.232	1.929	6.027	66.232
7	1.740	5.438	71.669	1.740	5.438	71.669
8	1.675	5.235	76.904	1.675	5.235	76.904
9	1.276	3.989	80.893	1.276	3.989	80.893
10	1.189	3.715	84.608	1.189	3.715	84.608
11	1.058	3.307	87.915	1.058	3.307	87.915
12	1.003	3.135	91.050	1.003	3.135	91.050

From table 1, we can see， Contribution rate of the first 12 Eigenvalues has reached 91.05%, namely they have already contained 91.05% of information of original variables. Therefore, we can replace the original variables by the 12 factors. In order to explain the 12 factors, we need to obtain the factor loading of 32 financial ratios to the 12 factors, namely the correlation coefficient of each factor to primitive financial ratios. This article used varimax orthogonal rotation to get the factor loadings .We can see the rotated component matrix in the following table.

Table2

Rotated Component Matrix(a)

	1	2	3	4	5	6	7	8	9	10	11	12
X1	0	0	0	0	0	0	0	0	1	-0	0	0
X3	0	1	0	0	0	0	0	0	0	0	-0	0
X7	0	0	1	0	0	0	0	-0	0	0	0	0
X8	0	0	0	0	0	1	0	0	0	-0	0	-0
X12	0	0	0	0	1	0	0	0	0	-0	0	-0
X15	0	0	0	0	0	0	0	0	0	1	0	0.2
X16	0	0	0	0	0	0	0	0	0	0	1	0
X17	0	0	0	1	0	0	0	0	0	-0	0	-0
X21	0	0	0	0	0	0	0	1	-0	-0	0	0
X25	0	0	0	0	0	0	1	0	-0	-0	-0	-0
X29	0	0	0	0	0	0	0	0	0	-0	0	0.9
X32	1	0	0	0	0	0	0	0	-0	-0	-0	-0

From the component matrix, we can see：

(1)The control variable of the first principal component is X31 and X32, which are company expense and management fee ratio. That is to say it mainly summarizes the corporate expense disbursement, and we take X32 as its explanatory variable.

(2) The control variable of the second principal component is X3 and X30, which are operating earning ratio and operating cost ratio. That is to say it mainly summarizes a company's operating condition, and we take X3 as its explanatory variable.

(3) The control variable of the third principal component is X7, which is Quick ratio. That is to say it mainly summarizes a company's ability to pay for short-term debt, and we take X7 as its explanatory variable.

(4) The control variable of the fourth principal component is X17 and X18, which are Growth rate of revenue and total assets, That is to say it mainly summarizes a company's growth situation, and we take X17 as its explanatory variable.

(5) The control variable of the fifth principal component is X12 and X13, which are current ratio and Accounts receivable turnover ratio. That is to say it mainly summarizes a company's operating ability, and we take X12 as its explanatory variable.

(6) The control variable of the sixth principal component is X2 and X8, which are Return on assets and liability/asset ratio, That is to say it mainly summarizes a company's ability to pay for a long-term debt, and we take X8 as its explanatory variable.

(7) The control variable of the seven principal component is X25, which is cash reversion ratio, That is to say it mainly summarizes a company's cash recycling ability, and we take X25 as its explanatory variable.

(8) The control variable of the eighth principal component are X20 and X21, which are Growth rate of net earning and EPS. That is to say it mainly summarizes a company's ability to develop, and we take X21 as its explanatory variable.

(9) The control variable of the ninth principal component is X1 and X9, which are Return on equity and equity/debt ratio. That is to say it mainly summarizes a company's ability to pay for debt, and we take X1 as its explanatory variable.

(10) The control variable of the tenth principal component is X5, which is ratio of gross profit margin. That is to say it mainly summarizes a company's ability to make profits, and we take X5 as its explanatory variable.

(11) The control variable of the eleventh principal component is X16, which is Fixed asset turnover ratio. That is to say it mainly summarizes a company's fixed assets situation, and we take X16 as its explanatory variable.

(12) The control variable of twelfth principal component is X29, which is operating profit to total profit ratio. That is to say it mainly summarizes company's ability to make profits from operation, and we take X29 as its explanatory variable.

The 12 principal components obtained according to the above factor analysis will been took as final variables in the following logistic model. They are mutually independent and represent 12 aspects separately, which basically conforms to the incompatible standard.

B. *Further Test*

In order to prove the method further, we select the financial ratios in 2004 from 76 companies as the sample. After carrying on the factor analysis, chooses establishes logistic model with Backward Stepwise method.

Variables in the Equation

Step 4	B	S.E.	Wald	df	Sig.	Exp (B)
F1	1.722	218.614	0.000	1	0.994	5.596
F2	-13.798	673.210	0.000	1	0.984	0.000
F4	2.477	314.328	0.000	1	0.994	11.911
F5	-0.925	91.913	0.000	1	0.992	0.397
F6	-11.745	299.878	0.002	1	0.969	0.000
F9	-0.403	246.685	0.000	1	0.999	0.668
Constant	-47.106	2,792.23	0.000	1	0.987	0.000

$\Rightarrow$

$$p = \frac{\exp(-47.106 + 1.722F_1 - 13.798F_2 - 0.925F_5 - 11.745F_6 - 0.403F_9)}{1 + \exp(-47.106 + 1.722F_1 - 13.798F_2 - 0.925F_5 - 11.745F_6 - 0.403F_9)}$$

		Predicted value		Rate of accuracy
		0	1	
Actual value	0	13	1	92.86%
	1	1	13	92.86%

From this result, we can see even though the data is from the year t-2,the rate of accuracy can still reach 92.86%, which is a stronger evidence to prove the better capacity of the prediction technique.

IV. Conclusions

The paper discusses the corporate financial prediction by introducing factor analysis into logistic regression. We choose 32 financial ratios ,nearly covering all financial aspects of a company, which helps the model reflect the corporate finance situations comprehensively, and simultaneously the ratios are the newest data from 2006 with complete effectiveness, which enables the model more prefect. The result, supported by a re-sampling study, shows that comparing the conventional method and simple logistic regression, the method may be a better valid alternative with better prediction capacity.

Certainly, due to a series of questions we can not decide, such as the false data, staff qualities and so on, the model is sure to have certain insufficiency and the limitation we can not overcome now. And the model needs further study.

References

[1] Juliana Yim, Heatber Mitcbell.A comparison of corporate distress prediction models in Brazil.Nove Economic BELO Horizonte_15(1),2005

[2] Stewart Jones, David A. hensher. Predicting Firm Financial Distress A Mixed Logit Model. The accounting Review, Oct 2004.

[3] Harlan D. Platt and Marjorie B. Platt. Predicting corporate financial distress Reflections on choice based sample bias. Journal of Economics and Finance; Summer 2002;26,2.

[4] Cheng Xiaojuan. The Model and Empirical Study on Pre-warning of Financial Crisis for Listed Company. Management Science and Engineering

[5] Edward I. Altman. Predicting financial distress of companies: revisiting the z-score and zeta models. Journal of Finance. Sep.2001

[6] Oslar, C, and G. Hong, "Rapidly Rising Corporate Debt: Are firms now vulnerable to an economic slowdown?" current issues in Economics & Finance, Federal Reserve Bank of New York, June 2000.