

A Prediction of Financial Distress for Listed Companies of the New tertiary board Based on Factor Analysis and Logistic Regression

Zhuoxi Yu

Jilin University of Finance and Economics, School of
Management Science and Information Engineering
Jilin Province Key Laboratory of Internet Finance
Changchun, China
yzx8170561@163.com

Milan Parmar

Jilin University of Finance and Economics, School of
Management Science and Information Engineering
Jilin Province Key Laboratory of Internet Finance
Changchun, China
milaanparmar9@hotmail.com

Qi Sun

Jilin University of Finance and Economics, School of
Management Science and Information Engineering
Jilin Province Key Laboratory of Internet Finance
Changchun, China
949667392@qq.com

Tiansong Zhang

Jilin University of Finance and Economics, School of
Management Science and Information Engineering
Jilin Province Key Laboratory of Internet Finance
Changchun, China
1416931552@qq.com

Abstract—Under the complex and changeable marketing environment, all walks of life are facing various risks and challenges. The enterprise financial crisis is more likely to happen. The company's financial situation is a standard judgment for investors, creditors and suppliers to make a decision. And it can affect the development of enterprises in the future directly. According to the features of the selected data, we established the binary Logistic regression equation of financial early warning model combined with factor analysis method. This model can analyze the financial status of listed companies in the new tertiary board market. Through the establishment of the model, stakeholders of listed companies may identify the signal of financial crisis and get ready ahead. The government regulators can monitor the quality of listed companies and the risk in the stock market. It also can reduce the volatility of the capital market, promote the benign development of listed companies.

Keywords—New tertiary board; Factor analysis; Logistic regression; Financial early warning model

I. INTRODUCTION

The new tertiary board market is originally to provide an off-exchange trading platform for shares in companies which are circulated in the main market. The assignments are handled by brokerages. In January 2006, two technology companies Riel Beijing century technology co., LTD and Beijing zhongke software co., LTD in Zhongguancun Science Park entered the third board market, other high-tech enterprises in the park followed. The companies with non-public shares in third board market first appeared. It is fundamentally different with the original one. In order to distinguish, the new market is known as "new tertiary board" in the industry [1]. "New tertiary board" market refers to financing market of unlisted enterprises in the science and technology industrial park. It serves enterprises in national high-tech zone and focus on enterprises of "high tech" and "innovative"[2]. China Securities Regulatory

Commission classified " new tertiary board" as one of the priorities for the whole year in 2011. It issued many times in order to promote the further development of the market. In January 2013, "National small and medium-sized enterprise stock transfer system co., LTD." was formally established, along with business rules trying out. In December, the State Council issued "Decision on issues related to the share transfer system of small and medium sized enterprises in China ", it set clear established rules about plate transformation mechanism. With the number of listed companies in the new tertiary board rise, the market took on an air of prosperity. It is widely distributed in most kinds of industries, not just limited to the high-tech enterprises. Market transactions are exceptionally prosperous.

New tertiary board market is the innovation of China's capital market on a world scale; it is only eight years since its emergence. It emphasizes more on companies' growth and innovation. So establishing early warning analysis on financial status which determine its growth appears to be very important. There are many research literatures about new tertiary board at present, mainly focus on market positioning, the legal system, development patterns, financing system or so, not too much on its growth especially on its early warning analysis of financial status [3]. Therefore, making financial warning analysis of listed companies is of great practical significance. Research on financial early warning model at home and abroad has been very mature. It can be classified as single variable model, multiple variable models, conditioning probability analysis model and artificial nerve network analysis model by its complexity [4]. Some early warning models are widely used so far, including the model based on principal component analysis and logistic regression [5], the model based on multiple discriminant analysis and neural network technology [6], and the model based on BP neural network [7], etc. These models

have their own advantages and disadvantages. Considering the applicability and operability, this paper used the factor analysis to reduce the dimensions of financial indicators. Then use the factor score as explanatory variables to set up Logistic model for early warning analysis and prediction.

II. THE INTRODUCTION OF MODEL AND METHOD

A. Factor analysis

Factor analysis is a multivariate statistical analysis method. It starts from (kicks off with) the internal dependencies of the correlation matrix of the indicators. Then summarize some variables with intricate relationship into fewer comprehensive factors. Its basic idea is to group variables with high degree of correlation in a group. Each group of variable represents a basic structure - common factor. Variables in different groups have low or no correlation. That is, factor analysis can obtain a series of relatively independent new variable group in term of linear combination of the original signals. Although the number of variables that have been reassembled is less than original variables, it contains the bulk of the information.

The basic idea of factor analysis modeling is: suppose the number of the sample is n , each sample has p observed variable. For the comparison among p variable, we standardized data firstly, which makes the mean of the standardized data (of each variable) is equal to 0, and the standard deviation is 1. The mean of standardized vector which is composed of p variable is 0. Covariance matrix is equal to the correlation coefficient matrix. So the form of factor model is as follows.

$$\underset{(p \times 1)}{X} = \underset{(p \times m)}{A} \underset{(m \times 1)}{F} + \underset{(p \times 1)}{\varepsilon} \quad (1)$$

Among them, "x" stands for p -dimensional observation vectors. A is the factor load matrix. F is common factor. ε is special factor. Factor analysis is a way of dimension reduction, so m is usually less than P . Elements in A are called the factor loading. The bigger the absolute value of the element is, the more the relevance of element in F are. Estimation of matrix A mainly used the principal component analysis to decompose. The main idea is to calculate eigenvalues from the covariance matrix of X decomposition, and to orthogonalize feature vectors. The square root of each eigenvalue multiply by their orthogonalization eigenvectors. Such matrix is called a factor load matrix. The average vector of F is 0. Covariance matrix is an identity matrix. That is, each component is independent to each other. In addition, assume that the mean vector of ε is 0. Covariance matrix is a diagonal matrix. F and covariance matrix are also independent.

Generally the process of factor analysis is: First, standardize the original data. The goal is to eliminate the dimension influence between variables. This makes the possibility of comparative between them. Second, factorize the relevant data. Extracting the common factor which is bigger than the threshold set before. Third, calculate each factor score according to the factor score matrix. It is also the score of composite indicator. Each factor is a linear combination of the different original data.

B. Logistic regression model

Logistic regression is one of the generalized linear regression models. It is often used in data mining, disease diagnosis, and economic forecasting, etc. The curve shows an S or S nonlinear form. Usually, the dependent variable of the Logistic regression is classified variable (Also can be in a form of numerical probability value), the independent variables of the Logistic regression are at least one numerical variables. There are multiple forms of Logistic regression. It can be divided into binary Logistic regression and multivariate Logistic regression according to the morphology of the dependent variable. The difference between them is that the former dependent variable is dichotomous, the latter is multi-classification variable. This article use the method based on the binary Logistic regression model to evaluate listed companies' profit in the new tertiary board.

A hypothesis of data distribution characters is needed before establishing a binary Logistic regression. First, the dependent variable must be dichotomous (equal to zero or equal to 1), or be a probability value (between 0 and 1). Second, the dependent variable and residual are both obedient to the binomial distribution. So when we analyze the data, estimation method is not to obtain the parameter by the least square method like regression model. It is to solve the equation estimation problems by maximum likelihood estimation. Third, logit transformation of the independent variable and dependent variable is a linear or linearized relationship. Forth, each observed object is independent. Its general form is as follows:

$$\ln \frac{p}{1-p} = \alpha + \sum \beta_i x_i \quad (2)$$

Its equivalent form is as follows:

$$p = \frac{e^{\alpha + \sum \beta_i x_i}}{1 + e^{\alpha + \sum \beta_i x_i}} \quad (3)$$

In the binary Logistic regression, the dependent variable is generally stored by 0-1 variables. In actual forecasting, the prediction of dependent variable is between 0 and 1. We usually set a threshold value, the predicted value greater than the threshold value is set to 1 and less than the threshold value is set to 0. In this study we take 0.5 as the threshold of probability, and using maximum likelihood method to estimate model parameters. Then put the financial indicators data into the regression equation. If the predicted value is greater than 0.5, this sample is judged to be a company of financial crisis, else it's a normal company.

The basic idea for setting up Logistic model in this paper is as follows. First, we take the four score which get from the financial indicators of listed companies in 2013 according to the factor analysis method as the independent variable. Net profit (greater than 0 is 0, less than 0 is 1) were regressed. The method of parameter estimation is the maximum likelihood estimation; second, evaluate the parameter estimation; third, using Logistic regression model to predict the net profits of

2013 and 2014, and take 0.5 as a threshold. The predicted value greater than 0.5 is classified to 1, less than or equal to 0.5 is classified to 0, and then compared with the real value [8].

III. THE EMPIRICAL ANALYSIS BASED ON FACTOR ANALYSIS AND LOGISTIC REGRESSION

In this paper, firstly we describe the basic characteristics of the financial indicators of 2013 and 2014. Then using the factor analysis to reduce the dimension of financial indicators of 2013, and to extract the relevant composite indicator. By bringing the composite indicator as explanatory variables and net profit as dependent variable into the Logistic regression model. Dependent variable is categorical variable in Logistic regression. We need to convert net profit into categorical variable. The financial data of 2014 are chosen as the test set which is used to verify the model prediction effect.

A. Financial index selection and statistical description

This article selects index which reflect debt-paying ability, including current ratio, quick ratio, cash ratio, asset-liability ratio, equity ratio, interest coverage. And select index which reflect operational capacity, including accounts receivable turnover, inventory turnover, total asset turnover, operating cash/gross revenue. And select index which reflect profitability, including return on equity, earnings per share, return on total assets, net profit margin on sales. And select index which reflect growth capability, including the rate of growth of gross operating income, net profit growth [9]. Corporate profit is target variable. Select 6175 samples of 2013 and 2014 which from the Wind database. Index connotation is as follows.

TABLE I. THE MEANING OF FINANCIAL INDICATORS

Variable	Index	Variable	Index
X_1	current ratio	X_9	total asset turnover
X_2	quick ratio	X_{10}	operating cash/gross revenue
X_3	cash ratio	X_{11}	return on equity
X_4	asset-liability ratio	X_{12}	earnings per share
X_5	equity ratio	X_{13}	return on total assets
X_6	interest coverage	X_{14}	net profit margin on sales
X_7	accounts receivable turnover	X_{15}	the rate of growth of gross operating income
X_8	inventory turnover	X_{16}	net profit growth

Descriptive statistics of financial data was performed by SPSS 20.0. In terms of the average, other index of the same varieties changes identically besides X_7 and X_8 . In terms of standard deviation, we find that volatility of 2014 is bigger than 2013. This indicates its instability. In terms of deviation degree, most indicators show right skewed. Compared with the financial data of two years, the skewness of X_{11} , X_{12} and X_{16} have changed from left to right. In terms of kurtosis, the financial data of two years has little difference. The data distribution pattern has an obvious peak and fat tails. Table II is financial data description statistics of 2013.

TABLE II. FINANCIAL DATA DESCRIPTION OF LISTED COMPANIES IN THE NEW TERTIARY BOARD IN 2013

variable	mean	standard deviation	skewness	kurtosis
X_1	3.42	29.90	45.78	2440.01
X_2	2.88	26.67	52.04	3276.07
X_3	1.02	11.56	41.10	1792.01
X_4	54.14	33.28	6.57	126.87
X_5	4.01	73.70	68.98	5110.81
X_6	35.45	3229.86	27.63	2317.54
X_7	5697.05	427350.94	78.45	6161.31
X_8	58.16	1258.40	53.89	3361.87
X_9	1.01	0.93	7.97	184.95
X_{10}	-45.71	1451.71	-36.52	1658.61
X_{11}	7.84	167.03	-32.56	2096.72
X_{12}	0.30	5.73	-13.88	2739.46
X_{13}	3.57	20.17	-10.09	200.79
X_{14}	-31.11	707.33	-38.98	1782.32
X_{15}	110.785	5211.75	73.58	5589.38
X_{16}	113.33	4038.05	-12.29	2498.83

The financial data description of listed companies in the new tertiary board in 2014 is shown in the following table III.

TABLE III. FINANCIAL DATA DESCRIPTION OF LISTED COMPANIES IN THE NEW TERTIARY BOARD IN 2014

variable	mean	standard deviation	skewness	kurtosis
X_1	5.23	96.87	44.89	2202.47
X_2	4.45	81.70	40.03	1651.29
X_3	2.39	63.45	50.74	2759.57
X_4	49.90	43.47	31.36	1647.19
X_5	2.51	50.91	75.94	5887.95
X_6	49.62	1032.06	19.14	1012.84
X_7	102.76	3930.79	66.96	4817.89
X_8	2687.90	204077.25	78.56	6172.94
X_9	1.05	0.96	9.60	255.04
X_{10}	-8.83	466.12	-30.00	1237.97
X_{11}	15.22	148.41	3.33	1256.03
X_{12}	0.42	3.49	31.74	2081.72
X_{13}	6.13	20.08	-13.95	458.97
X_{14}	-14.44	519.60	-40.59	2006.66
X_{15}	134.03	2263.36	53.77	3226.30
X_{16}	337.93	4514.12	27.80	1381.40

B. Analysis of the index factor

We use factor analysis to analyze the financial index of 2013 after normalization. Figure 1 shows that scree plot has leveled off after the fourth common factor. They covered most of the information which can describe the company's financial situation. And there is no information overlap between them. So four common factors were extracted.

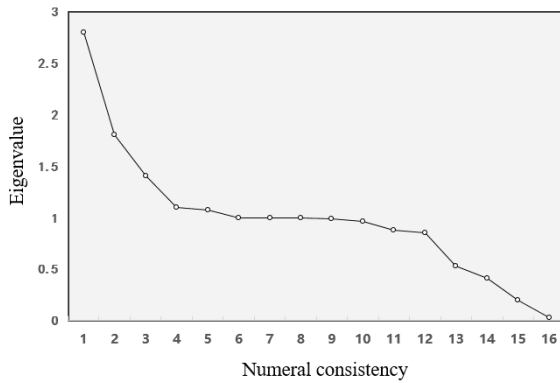


Fig.1. Scree Plot of factor analysis

On the basis of rotating factor score matrix shown in table IV, and accounting the score of four factors of 2013. We observed weight coefficient of each factor, it can be found that some of the original data of factor weights are big. Through sorting, we can obtain corresponding dimension indicators of each factor. The first factor includes current ratio, quick ratio, cash ratio, asset-liability ratio, equity ratio, interest coverage. It was named solvency factor. The second factor includes accounts receivable turnover, inventory turnover, total asset turnover, operating cash/gross revenue. It was named workability factor. The third factor includes return on equity, earnings per share, return on total assets, and net profit margin on sales. It was named profitability factor. The fourth factor includes the rate of growth of gross operating income, net profit growth. It was named growth ability factor.

TABLE IV. ROTATING FACTOR SCORE MATRIX OF FINANCIAL INDICATORS

variable	The factor score coefficient			
	1	2	3	4
current ratio	.984	-.056	.037	.029
quick ratio	.964	-.060	.038	.025
cash ratio	.936	-.057	.029	.035
asset-liability ratio	-.175	-.596	.321	.349
equity ratio	-.014	-.080	.111	-.360
interest coverage	.012	.259	.118	-.115
accounts receivable turnover	.002	.013	-.015	-.125
inventory turnover	.010	.026	-.012	.045
total asset turnover	-.042	.147	.083	.594
operating cash/gross revenue	-.009	.389	.744	-.006
return on equity	.011	.254	-.170	.640
earnings per share	.023	.443	-.336	-.055
return on total assets	.072	.789	-.349	-.003
net profit margin on sales	.001	.529	.651	-.025
the rate of growth of gross operating income	.004	.030	-.028	-.097
net profit growth	-.046	.057	-.032	.158

Through the division of the four dimensions, we can get four factor score expressions of 2013.

$$\text{Debt-paying ability:} = 0.984 * X_1 + 0.964 * X_2 + 0.936 * X_3 - 0.175X_4 - 0.014X_5 + 0.012X_6 + 0.002X_7 + 0.01X_8 - 0.042X_9 - 0.009X_{10} + 0.011X_{11} + 0.023X_{12} + 0.072X_{13} + 0.001X_{14} + 0.004X_{15} - 0.046X_{16}$$

$$\text{Operational capacity:} = 0.056 * X_1 - 0.06 * X_2 - 0.057 * X_3 - 0.596X_4 - 0.08X_5 + 0.259X_6 + 0.013X_7 + 0.026X_8 + 0.147X_9 + 0.389X_{10} + 0.254X_{11} + 0.443X_{12} + 0.789X_{13} + 0.529X_{14} + 0.03X_{15} + 0.057X_{16}$$

$$\text{Profitability:} = 0.037 * X_1 + 0.038 * X_2 + 0.029 * X_3 + 0.321X_4 + 0.111X_5 + 0.118X_6 - 0.015X_7 - 0.012X_8 + 0.083X_9 + 0.744X_{10} - 0.17X_{11} - 0.336X_{12} - 0.349X_{13} + 0.651X_{14} - 0.028X_{15} - 0.032X_{16}$$

$$\text{Growth capability:} = 0.029 * X_1 + 0.025 * X_2 + 0.035 * X_3 + 0.349X_4 - 0.36X_5 - 0.115X_6 - 0.125X_7 + 0.045X_8 + 0.594X_9 - 0.006X_{10} + 0.64X_{11} - 0.055X_{12} - 0.003X_{13} - 0.025X_{14} - 0.097X_{15} + 0.158X_{16}$$

C. The establishment of the Logistic regression model

First of all, the financial data were treated by means of Logistic regressions. Parameter estimation and testing were showed in table V.

TABLE V. THE RESULT OF LOGISTIC REGRESSION OF LISTED COMPANIES IN NEW TERTIARY BOARD IN 2013

Variable	B	The standard deviation	Wald value	Degree of freedom.	significance	Exp (B)
constant	-2.210***	0.060	1356.295	1	.000	0.110
debt-paying ability	-0.057	0.052	1.193	1	.275	0.945
operational capacity	-5.140***	0.169	924.845	1	.000	0.006
profitability	1.995***	0.140	201.669	1	.000	7.349
growth capability	-2.331***	0.121	373.483	1	.000	0.097

Cox - Snell R square is 0.6325, it showed that the model constructed possesses has better description ability for the profitability of listed companies in the new three boards. Estimated coefficient of debt-paying ability in four factors is not significant, it indicates that the prediction of a company's profitability is weak. The estimated coefficient of operation ability and growth ability is negative. It showed that the operation ability, growth ability are negative correlation with the company's profitability. Estimated coefficient of profitability is positive. It showed that profitability is the positive factor that affects whether a listed company profits in the new tertiary board. Equation expression can be got from the table above. F1-F4 are solvency factor, workability factor, profitability factor, growth ability factor.

$$\ln \frac{p}{1-p} = -2.21 - 0.057 * F1 - 5.14 * F2 + 1.995 * F3 - 2.331 * F4 \quad (4)$$

D. The results and analysis of the prediction of model

The main method to describe the effectiveness of a model is to analyze its predictive ability. This article put all the financial indicators of listed companies in new tertiary board in 2013 as the training data set to establish the Logistic regression model. The prediction test was made which showed the correctness of the model using the financial data of 2013, as it showed in table VI.

TABLE VI. LOGISTIC REGRESSION PREDICTION EFFECT OF LISTED COMPANIES IN NEW TERTIARY BOARD IN 2013

Observed		Predicted		
		Net profit		Percentage of correction
		0	1	
Net profit	0	4609	179	96.3
	1	530	857	61.8
Total percentage				88.5

From table VI we can know that Logistic model make predictions for a total of 6175 data. There are 4788 samples which their actual values are 0. There are 4609 right samples and 179 wrong samples in the prediction. The prediction accurate is 96.3%. There are 1387 samples which their actual values are 1, including 857 right samples and 530 wrong samples. It showed that the result of the model is preferable.

Put the financial date of 2014 as testing dataset to test the prediction effect. The result was showed in table VII. There are 5283 samples which their actual values are 0, including 4902 right samples and 381 wrong samples. The prediction accurate is 92.8%. There are 892 samples which their actual values are 1, including 696 right samples and 196 wrong samples. The prediction accurate is 78.0%. Compared with the prediction result of Logistic regression model which was using the data of 2013, the precision was improved. The total prediction accuracy which was 90.7% is better than the accuracy of 2013.

TABLE VII. LOGISTIC REGRESSION PREDICTION EFFECT OF LISTED COMPANIES IN NEW TERTIARY BOARD IN 2014

Observed		Predicted		
		Net profit		Percentage of correction
		0	1	
Net profit	0	4902	381	92.8
	1	196	696	78.0
Total percentage				90.7

IV. CONCLUSION

The financial early warning analysis can help listed companies to find the deficiencies in their managements and the problems existing in the operation. Companies can adjust business strategy in time so that they can make their companies develop continuously and healthily. It is also convenient for state operator to make control and management. They can devise new strategies to promote the orderly operation of the security market [10]. In addition, immediate analysis of the present business status may help the enterprises adjust their investment strategy. Investors shall find legal, effective invest channels. This can protect the legal rights of the creditor.

ACKNOWLEDGMENT

The research was supported by the National Social Science Foundation of China (No.16BTJ020); "Thirteenth Five-Year Plan" Science and Technology Research Project of the Education of Jilin Province (No. 2016103).

REFERENCES

- [1] A.J. Lin, and Z.Y. Lin, "Research on development model and countermeasures of the New three board," *Economic reform*, vol.30, 2012, pp. 111-114(In Chinese)
- [2] Y. Ding, "The new tertiary board listed company financial growth research," *Harbin University of Commerce*, 2015(In Chinese)
- [3] L. Jing, and Q.J. Meng. "Problems and system analysis in the development of the New tertiary boardmarket," *Finance economy*, vol.32, 2013, pp.58-60(In Chinese)
- [4] F. Li, Z.T. Du, and L.G. Li. "Enterprise financial early warning model: theory review and comment," *Management Review*, vol.23, 2011, pp.144-151(In Chinese).
- [5] Z.H. Lei. "Financial crisis warning accuracy test research of Logistic model," *Seeker*, vol.32, 2012, pp.54-56(In Chinese)
- [6] L. Zhang, S. Chen, and X. Zhang. "Financial crisis warning based on the multiple discriminant analysis and neural network technology," *Systems Engineering*, vol.23, 2005, pp.49-56.
- [7] S.E. Yang, and L. Huang. "Financial crisis warning model based on BP neural network," *System Engineering Theory and Practice*, vol.25, 2005, pp.12-18+26(In Chinese).
- [8] Z. Gao. "Financial crisis warning analysis based on Logit model—taking manufacturing listed companies for example," *China Journal of Commerce*, vol.22, 2013, pp.41+43(In Chinese).
- [9] Z.M. Qin. "The research about warning variables selection on financial crisis of China's listed companies," *Dongbei University of Finance and Economics*, 2012(In Chinese)
- [10] Feng Y L, and McClean S. "A data mining approach to the prediction of corporate failure," *Knowledge-based systems*, vol.14, 2001, pp.189-195.