Hitotsubashi Journal of Commerce and Management 38 (2003), pp.13-28.

CORPORATE FINANCIAL DISTRESS DIAGNOSIS IN CHINA: EMPIRICAL ANALYSIS USING CREDIT SCORING MODELS*

QI LIANG**

Abstract

Corporate performance is undoubtedly of great interests to the owners, managers, creditors and regulatory institutions. This study attempts to extend and improve upon the prior studies in China, particularly in its greater sample size and comparative analysis between MDA and logistic regression analysis in the financial distress prediction. Empirical results show that logistic regression analysis has relatively higher prediction accuracy and lower Type I & II errors. Together with its great flexibilities and efficient combination of data from both financial statements and capital market prices, logistic regression analysis is considered as the best technique to classify and predict financial distress of listed companies in nowadays China.

Key Words: Financial Distress; Multiple Discriminant Analysis; Logistic Regression Analysis; Listed Company
JEL classification: G33; C49

I. *Introduction*

Credit risk has become the leading risk measurement and management challenges of the late 1990s. Globally, institutions are taking on an increasing amount of credit risk. In the year of 2002, 47 companies listed on the two stock exchanges in China fell into financial distress and were being specially treated.[1] This amount has reached a record high since the first "ST" company appeared in 1998. Listed companies are the cornerstones of the stock market, especially when the market is at its developing stage. The firms' performances are no doubt of

---

* I am grateful to my academic adviser, Professor Yoshinori Shimizu for his constructive comments. I would like to thank Monbukagakusyou for its support of my post-doctoral research at Hitotsubashi University. Funding from the Project "Research on the Establishment of Credit Risk Management System in Chinese Commercial Banks" (70201001) supported by the NSFC and Project "Quantitative Measurement and Management of the Credit Risk of the Commercial Banks" (00CJY026) supported by the NPOPSS are also gratefully acknowledged.

** Qi LIANG is an Associate Professor of Finance at Nankai University, China and visiting Post-doctoral Researcher at Hitotsubashi University; liangqi@nankai.edu.cn

[1] The China Securities Regulatory Commission (CSRC) issued a file in 1998 and required the Shanghai and Shenzhen Stock Exchanges to make special treatment to the everyday trade of the negotiable stock of the listed company that experience financial distress and/or other abnormalities. Financial distress means that the company experiences loss in two consecutive years or the net assets per share is lower than its face value while some unusual events like failure to pass the audit process, suffering of natural disasters, involvement in big lawsuit and etc. are included in the other abnormalities.

great interests to the owners, managers, creditors and regulatory institutions. Wang et al. (2002) investigated the movements of negotiable share prices to the "ST" announcements using 72 selected "ST" companies in the 1998-2000 periods as original sample and found a significantly negative response of stock price to the announcements. As credit exposures have multiplied, the need for more accurate risk measurement and management techniques for credit risk has also increased.

Considering the pre-maturity of the financial markets and the availability of the economic and financial data in present China, this study utilizes traditional credit scoring model to predict the corporate financial distress and attempts to find the best approach by making a comparatively theoretical and empirical analysis on two practicable techniques, i.e., multiple discriminant analysis (MDA) and logistic regression analysis. The rest of the paper is organized as follows. Section II discusses prior research of the credit scoring models on the corporate distress prediction. Section III gives a short description of the two credit scoring techniques. Section IV presents the model sample and the originally designed 25 financial ratios. Section V analyzes the optimal predictors selected for the two techniques as well as the classification and prediction accuracies decided by the two model functions. Section VI concludes the whole paper.

II. *Prior Research on Credit Scoring Models*

In the credit risk measurement literature, credit-scoring models are among the mainstream approaches to model firm credit events of financial distress that are assumed to occur completely unexpected. In other words, all factors not limiting to economics and finance may lead to firm financial distress. Credit scoring models usually combine a set of quantifiable financial ratios of firm performance through alternative statistical methods to empirically search for optimal predictors that lead to the lowest firm financial distress misclassification rates. Then the credit scores assigned to firms can be used together with a large credit database, if exists, or some capital market risk equivalents such as the bond ratings to make inferences of probability of default of the firm and possibly loss given default on any credit portfolio that includes the claims on the firm.

The earliest study of the credit scoring models on the firm financial distress prediction might be dated back to the 1930s. Fitzpatrick (1932) compared the values of financial ratios between the failed and no-failed firms and found that the failed firms usually had poorer variables. In 1966 Beaver pioneered the univariate discriminant approach and found that the cash flow to total liability was the most prognostic financial ratio in the firm failure predictions. In 1968, Altman presented a Z-score model built on the MDA technique, which is first proposed by Fisher (1936) and is probably first applied by Durand (1941) in the economic and financial fields. The Z-score model function is composed of five variables chosen by MDA as doing the best overall job together in the prediction of corporate failure. Because of the relatively strong classification and prediction ability and the model simplicity, MDA has become the dominant approach in the firm financial distress prediction and alternative Z-score models have been developed in over 25 countries. Afterwards, Beaver (1968), Deakin (1972), Edmister (1972), Altman, Handelman & Narayanan (1977), Izan (1984) and Boritz (1991), surely not limiting to these studies, investigated various financial ratios under the framework of both linear and quadratic discriminant analysis and established respective decision rules. For example, Beaver (1968) tried to incorporate the movements of market share value as a possible prognostic factor in forecasting firm failure and found that the market anticipates firm failure at least a year prior to its actual occurrence. Izan (1984) constructed an industry-relative business failure classification model based on the MDA and investigated the indications of corporate distress in Australia. The prediction accuracy of the 5-variable linear discriminant function 2 and 3 years prior to the actual failure reach 82% and 75.5%. In addition, the failure rate data is also presented for the first time that can be useful for time series comparisons.

Logistic regression analysis has gain more and more popularity in recent decades in social science. In 1977 Martin used both logistic and discriminant analysis to predict bank failure on 23 failed bank sample in the 1975-1976 periods and achieved similar model results of the classification accuracy. Huffman & Ward (1996) applied logistic model to distress prediction of 171 defaulted high yield corporate bonds in the 1977-1991 periods and found that high yield issues that default are characterized by having higher asset growth rates, lower operating profit margins, larger levels of collateralizable assets, and larger changes in net working capital. In addition, the study also implies that investors should use default prediction models instead of a classic bankruptcy prediction model to aid forming high-risk bond portfolios. In a recent study, Laitinen & Laitinen (2000) attempted to improve functional form of decision rules by applying Taylor's series expansion to logistic regression and found that the firm cash ratio, cash flow ratio and shareholder's equity ratio are the major factors affecting the firm insolvency risk.

In China, the research on firm financial distress has been impeded by the non-availability of the public credit data. The lack of time series data has also limited the research to make any meaningful inference of the probability of default and loss given default that are the crucial inputs in quantitative credit risk measurement. Since 1990s, quite a lot credit data have been accumulated in the newly established stock markets. On the other hand, there has also appeared about 150 "ST" listed companies since 1998. All the financial information can be used in the financial distress prediction analysis of which the proxy measurement of the credit risk may become possible. Chen J. (1999) is among the earliest to apply MDA to financial distress prediction. A 6-variable discriminant function is built to empirically discriminate financial distress of 27 "ST" listed companies appeared in the year of 1998. Wang et al. (2001) used a modified discriminant approach named projection pursuit in order to better fit the non-normal and high dimensional distribution characteristics of financial data. Shi et al. (2001) studied the financial distress problem of listed companied by means of MDA with a bigger original sample of 128 firms in the 1999-2000 periods. Chen Y. (2000) applied both univariate and muitivariate ways to predict financial distress based on the 67 sample firms that had been specially treated by the end of 1999. The principle component analysis, regression analysis and discriminant analysis were all being used. Among them, the best classification accuracy 1 year prior to the financial distress reaches 93.1% and the best prediction accuracy 2 and 3 year prior to the financial distress reach 89.66% and 79.31% respectively. Zhang et al. (2001) attempted to improve the decision rule by incorporating principle component analysis within regression model and constructed a model using 8 variables. The accuracy 1, 2 and 3 years prior to the financial distress were 92.50%, 87.50% and 77.50%. All of the aforementioned studies have used the various financial ratios only from the financial statements and find the firm profitability indicators are the most important predictors in the financial distress diagnosis. Till now, the credit scoring models used to predict the financial distress of listed companies in China are primarily focused on the discriminant analysis.

III. *Credit Scoring Techniques*

1. Multiple Discriminant Analysis

MDA is used to model the value of a categorical dependent variable based on its relationship to more than one independent variable. In its most common form MDA tries to derive a linear combination of characteristics of these variables that best discriminates between the categories, based on the statistical decision rule of maximizing the between category variance while minimizing the within category variance among these variables. One advantage of MDA is the reduction of the analysis space dimensionality, i.e. from the number of independent variables to k-1 dimension(s), where k equals the number of original *a priori* categories. Since the financial distress prediction is concerned with only two categories of "ST" group and non-"ST" group, the analysis is transformed into its simplest one dimension and the discriminant function transforms the values of variables to a single discriminant score of Z, which is then used to classify and predict the financial performance of the original firms or/and out-of-the-sample ones. Following the treatment of Huberty (1994), MDA can be described mathematically as follows:

Consider n firms in the model sample, and a set of p independent variables (financial ratios), $X_1, X_2, \ldots, X_p$, and a binary category variable Z referring to firm financial performance. The predicted categorical measure Z_u (discriminant score) for firm u may be represented as

$$Z_u = b_0 + \sum_{i=1}^{p} b_i X_{iu}, \qquad i = 1, 2, \ldots, p \qquad u = 1, 2, \ldots, n \tag{1}$$

where b_i is the discriminant coefficient and b_0 is the constant. MDA assigns firm u to the financial distress category of g if the posterior probability of membership of firm u in category g is greater than that in the non-financial distress category of g'. That is

$$P(g \mid X_u) > P(g' \mid X_u). \qquad g \neq g' \tag{2}$$

Posterior probability is a likelihood of category membership conditioned on knowing X_u. Assuming that the independent variables follow muitivariate normal distribution and the two category covariance matrices are equal, then the posterior probability of membership of firm u in category g is given as

$$P(g \mid X_u) = \frac{q_g \cdot \exp(-\frac{1}{2} D_{ug}^2)}{\sum_{g'=1}^{k} q_{g'} \cdot \exp(-\frac{1}{2} D_{ug'}^2)}, \tag{3}$$

where q_g and $q_{g'}$ denote respectively the prior probabilities of membership in category g and g', "prior" in the sense that this is a probability of category membership before X_u is known. D_{ug} and $D_{ug'}$ are distance between the observation vector of firm u and the centroid of category g

and g'. Seeking for answer of equation 2 is equivalent to maximizing the numerators of equation 3, or its natural logarithm. This may be expressed as:

$$L_{ug}=\ln q_g-\frac{1}{2}D_{ug}^2=\ln q_g-\frac{1}{2}(X_u-\overline{X}_g)'S^{-1}(X_u-\overline{X}_g)$$
$$=\left[\overline{X}_g'S^{-1}\right]X_u+\left[-\frac{1}{2}\overline{X}_g'S^{-1}\overline{X}_g+\ln q_g\right], \qquad 4$$

where the terms in the first and second parentheses are discriminant coefficients and the constant. Thus, the maximum probability rule for the p-predictor normal, equal covariance matrices can be put forward as assigning firm u to category g if

$$L_{ug}>L_{ug'}. \qquad g\neq g' \qquad 5$$

The linear relationship between L_{ug} and X_u shows that MDA here adopts linear decision rule to classify the firm financial distress.

2. Logistic Regression Analysis

Logistic regression analysis is another method to predict a categorical variable from a set of continues or categorical independent variables. Traditional regression cannot be used directly when the dependent is a dichotomy because the assumptions of normality and homoscedasticity are being violated. However, after the transformation of the dependent into a logit variable which is the natural log of the odds of the dependent occurring or not, values of new dependent belong to the whole real number population and the logistic regression analysis may be applied. By the means of maximum likelihood estimation, logistic regression analysis estimates the probability of a certain event occurring. For example, each firm is assigned a probability of financial failure through the logistic function in the financial distress prediction. Following the treatment of Hosmer and Lemeshow (2000), logistic regression analysis can be described mathematically as follows:

Consider a collection of p financial ratios denoted by the vector $X=(X_1, X_2, \ldots, X_p)^T$. Let the conditional probability of firm financial distress be denoted by $P(Z=1|X)=p(X)$. The logistic model is given as:

$$\pi(X)=\frac{e^{\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_iX_i}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_iX_i}}, \qquad i=1, 2, \ldots, p \qquad 6$$

where$\beta=(\beta_1, \beta_2, \ldots, \beta_k)^T$ is the logit coefficient and β_0 is the constant. These unknown parameters are being estimated through the likelihood functions, described in equation 7, which maximizes the probability of obtaining the observed set of data.

$$l(\beta)=\prod_{i=1}^{p}\pi(X_i)^{z_i}[1-\pi(X_i)]^{1-z_i} \qquad i=1, 2, \ldots, p \qquad 7$$

In order to find the value of β that maximizes $l(\beta)$, the $l(\beta)$ should be differentiated with respect to β_0 and β and then the resulting expressions be set equal to zero. These $p+1$ equations, known as maximum likelihood equations, are:

$$\sum_{i=1}^{p}[z_i-\pi(X_i)]=0 \qquad i=1, 2, \ldots, p \qquad 8$$

and

$$\sum_{i=1}^{p} X_{ij}\left[Z_i - \pi(X_i)\right] = 0 \qquad i, j = 1, 2, \ldots, p \quad i \neq j \qquad 9$$

Since the expressions in equations 8 and 9 are nonlinear in β_0 and β, logistic regression analysis uses iterative method to estimate the logit coefficients and constant.

IV. *Sample Representation*

1. Model Sample

The model sample is composed of 138 firms with 69 in each of the two categories, i.e., financial distress group of "ST" firms and non-financial distress group of non-"ST" firms (see appendix A). The sample size represents the largest in the corporate financial distress prediction in China till now. There were total 93 new "ST" firms appeared in the 2000-2002 period, among which 22 were being specially treated due to other abnormalities reasons and these firms must be excluded from the model sample. Otherwise, corporate financial distress prediction will be negatively affected by unreal firm performances. Additionally, there are another two "ST" firms whose assets sizes are too big relative to the group average to be included. After deducting all these 24 firms, the financial distress group now has 69 valid "ST" firms (see table 1). The non-financial distress group consists of a paired sample of firms chosen on the stratified random basis of asset size and annual profit. The asset value of the financial distress group ranges from \$18.25 million to \$383.31 million with the average of \$98.55 million while the non-financial distress group ranges from \$18.50 million to \$380.65 million with the average of \$103.02 million.[2]

TABLE 1. "ST" FIRMS SELECTED IN THE FINANCIAL DISTRESS GROUP

	No. of "ST" firms entered into the financial distress category	No. of "ST" firms excluded from the financial distress category	Total
New "ST" in 2002	29	16	45
New "ST" in 2001	19	3	22
New "ST" in 2000	21	5	26
Total	69	24	93

2. Industry Affiliation

The industry affiliation of the financial distressed category is given in table 2. These firms come from 18 sectors including 1 agriculture, 43 various manufactures, 4 real estates developers and builders, 6 wholesalers and retailers, 5 service firms and 10 conglomerates. According to the classification stipulation newly issued by the CSRC, the industrial affiliations of listed companies are being divided into 22 sectors. As the 18 industry sectors of the model sample firms account for a huge part of the whole sectors, this study is deemed applicable

[2] All of the data are from the www.cninfo.com.cn which is a designated website of stock market information publication by the CSRC. The exchange rate of \$1=RMB8.27 is used in the asset size conversion.

TABLE 2. INDUSTRY REPRESENTATION OF THE FINANCIAL DISTRESS GROUP

Code No.	Industry Name	No. of "ST" Firms
1	Agriculture, Forestry, Fishing and Hunting	1
2	Food, Beverage	2
3	Textile, Apparel, Leather	2
4	Paper, Printing	2
5	Petroleum, Chemical Product, Plastics, Rubber	4
6	Electrical Equipment	6
7	Metal, Nonmetallic Mineral Product	8
8	Machinery, Equipment, Meter	5
9	Medicine, Biologic Products	3
10	Other Manufacturing	8
11	Construction	1
12	Transport, Storage	1
13	Information, Technology	3
14	Wholesale and Retail Trade	6
15	Real Estate	3
16	Social Services	3
17	Transmission, Culture	1
18	Conglomerate	10
	Total	69

across almost all firms only except financial institutions, mining companies, wood products firms and public utilities.

3. Financial Ratios

The design of the financial ratios is crucial to the financial distress prediction. Data pertinent to estimating corporate financial distress arise mainly from two sources: financial statements and market prices of firms' equity and debt. Presently, more use is still made of financial statements than capital market prices. Needless to say, the most accurate financial distress prediction derives from models employing both sources, but there sometimes is a limit to the information that can be extracted from statements or prices. This study makes new attempts in this field by incorporating stock market indicators of the negotiable shares into the financial ratio design. Based on the popularity in the previous literature, the systematic characteristics of Chinese listed companies and the potential relevancy to this study, 25 financial ratios are being chosen (see appendix B). Although many ratios are theoretical predictors of financial distress and even statistically relate to the likelihood of financial distress, they contribute only marginally to prediction accuracy. In order to prevent this problem, all the 25 financial ratios are being found to be contributive in the financial distress prediction from a number of previous papers. These studies include Beaver (1966), Altman (1968), Altman, Haldeman & Narayanan (1977), Izan (1984), Laitinen & Laitinen (2000), Chen J. (1999), Chen Y. (2000), Wang & Li (2000), Zhang, Zhu & Xu (2001) and Shi & Zhou (2001).

V. *Models Comparative Analysis*

1. Optimal Financial Ratios

This study applies stepwise method in MDA and logistic regression analysis to select optimal financial ratios in the financial distress prediction. By doing this, objective comparisons between the two credit scoring techniques can be achieved. Widely used in the discriminant and regression analysis, stepwise method provides a useful and effective data mining tool to screen a large number of variables. The stepwise procedures for inclusion or exclusion of variables in a model are based on a statistical algorithm that checks for the "importance" of variables, and either selects or deletes them on the basis of a fixed decision rule. The "importance" of a variable depends on a measure of statistical significance of the coefficient for the variables and the statistic used is decided by the assumptions of the model. In MDA, the significance is assessed via an F-test due to an assumed normal distribution of the model errors and the criterion in the step procedure for entering or removing variables into the function is based on group separations. For any given step, the next variable added is the one that increases group separation the most of all the remaining independents. When all the variables in the model meet the criterion to stay and none of the other variables meets the criterion to enter, the stepwise selection process stops. In the logistic regression analysis, the errors are assumed to follow a binary distribution and the significance is assessed via the likelihood ratio chi-square test. Thus, at any step the most statistically important variables in the procedure is the one that produces the greatest change in the log-likelihood relative to a model not containing the variable, i.e., the one that would result in the largest likelihood ratio statistic. The procedures start by estimating only constant forced into the model. Next, the procedures compute the chi-square statistics for all the variables not in the model and examine the largest of these statistics. If they are significant at the specified level, the variables are entered into the model. The stepwise selection process terminates if no further variable can be added to the model, or if the variable just entered into the model is the only variable removed in the subsequent elimination. Stepwise selection is characterized by one or more elimination process after each selection step, i.e., the variables already selected into the model do not necessarily stay. The significance level for adding or retaining variables in both MDA and logistic regression analysis are set to be equal of 0.05.

Table 3 shows the selected financial ratios 1, 2 and 3 years prior to the financial distress for the MDA and logistic regression analysis.[3] The order of the selected variables is being arranged according to its "contribution" to the respective functions.

TABLE 3. FINANCIAL RATIOS SELECTED FOR THE TWO TECHNIQUES

	MDA	Logistic Regression Analysis
1 year prior to "ST"	X_3, X_6, X_2, X_{15}	X_3
2 year prior to "ST"	X_6, X_{15}, X_{11}, X_{22}	X_6, X_{15}, X_{22}
3 year prior to "ST"	X_3, X_4, X_8, X_{20}	X_7, X_2, X_{21}

[3] SPSS version 11.5.01 is used to do the empirical works of MDA and logistic regression analysis.

In order to compare the optimal financial ratios selected for the two techniques, the 25 original ratios are being divided into seven general dimensions based on firm performance evaluation. They are profitability, liquidity, activity, solvency, potentiality, capital market and capital structure. The selected financial ratios for both techniques 1 and 2 years prior to the financial distress are quite similar. Furthermore, the financial ratios selected for logistic regression analysis of these two years are a subset of those selected for MDA. Among the financial ratios selected 1 year prior to the financial distress, X_3 (Net Income/Total Assets), an investment return indicator, is the most important. It is the only optimal financial ratio that enters the logistic function and it is also the most significant financial ratio that contributes the greatest to the discriminant function. Besides X_3, three other financial ratios of X_2 (Net Sales/ Total Assets), X_6 (EBIT/Total Assets) and X_{15} (Growth Rate of Total Assets) are also being selected for the MDA function. Since these selected financial ratios are mainly from investment return and prime business indicators, the firm profitability predictors can be considered as having the highest diagnostic capability in the financial distress classification. As for the financial ratios selected 2 years prior to the financial distress, X_6, X_{15}, X_{11} (Operating Cash Flow Ratio) and X_{22} (Equity Increasing Ratio) are being chosen in the MDA function. Optimal financial ratios selected for the logistic function are X_6, X_{15} and X_{22}, which is a subset of that selected for the MDA function too. All of these selected ratios show that in addition to the profitability indicators, the optimal predictors 2 years prior to the financial distress have enlarged to include indicators from firm potentiality analysis and capital market prices.

Contrast distinctively to the quite similar optimal financial ratios selected for the two techniques 1 and 2 years prior to the financial distress, totally different optimal financial ratios are selected for the two techniques 3 years prior to the financial distress. Two investments return indicators of X_3 and X_4 ((Net profit + Interest Expense)/Total Assets), one working capital indicator of X_8 (Current Ratio) and one long-term activity indicator of X_{20} (Total Asset Turnover) are selected for the MDA function while one prime business indicator of X_2, one capital investment indicator of X_7 (Net Profit/Equity) and one capital market indicator of X_{21} (MV of the Negotiable Shares/Total Debt) are selected for the logistic function. Possible explanations to this difference include: first, the "ST" listed companies caught in the financial distress are defined mainly as firms suffering two consecutive years of losses. Thus, all the sample firms are still making profits 3 years prior to the financial distress. This implies that the differences of the financial ratios between the financial distress group and the non-financial distress group are becoming less clear; second, there may exist real distinctions in different firms that can be measured by different financial ratios. Hence, the two empirical techniques use this information in alternative ways and thus the optimal financial ratios selected for them differ from each other; third, accidental elements originated from the model sample and/or pure statistical decision rule applied in the selection process cannot be underestimated. To find more satisfactory answers to this problem, this study tries a step further by drawing support from factor analysis. As a data reduction or structure detection method, factor analysis is useful when there are a great number of variables and there exist a high correlation among variables. From the 25 financial ratios, factor analysis extracts six factors, among which three factors, firm profitability, liquidity and potentiality, are primarily loaded on the financial ratios that are selected for both techniques 3 years prior to the financial distress. Factor analysis further shows that high correlations exist among some of the 25 financial ratios, especially within the profitability indicators. For example, the coefficient of correlation between X_3 and

X_4 reaches as high as 0.977. Since some financial ratios are indeed measuring the same economic and financial dimensions of the firm performance, high correlations can interfere the models results or the differences in seemingly alternative ratios can be so small that the selection among several ratios into the functions can be more or less random.

Moreover, for 1, 2 and 3 years prior to the financial distress, the number of the selected optimal financial ratios for MDA functions are all more than that for logistic functions. Among all the optimal financial ratios, X_3 and X_6 are being selected the highest of three times, followed by X_{15} of twice. X_3 enters both model functions 1 year prior to the financial distress while X_6, X_{15} and X_{22} enter both model functions 2 years prior to the financial distress. Noting that X_{21} and X_{22}, financial ratios containing information of forward-looking capital market prices, are selected for the logistic functions both 2 and 3 years prior to the financial distress.

2. Prediction Results

The MDA and logistic functions are used to classify and predict the financial performances of listed companies in the model sample. Cross-validation method is utilized to estimate the misclassification rate in order to modify the potential bias in the model. It involves a two-step process: first, one firm is deleted and the function is determined on the remaining *n*-1 firms. Then the function is used to classify the deleted firm into financial distress group or non-financial distress group. This process is carried out *n* times and the proportions of deleted firms not correctly classified are used as misclassification estimates. Given in table 4, the classification and prediction accuracies of both models are quite good based on the data 1, 2 and 3 years prior to the financial distress. The classification accuracies of both models 1 year prior to the financial distress exceed 98% and the prediction accuracies 2 and 3 years prior to the financial distress reach 80%.

TABLE 4. CLASSIFICATION AND PREDICTION ACCURACIES

		MDA			Logistic Regression		
Year	No. of Firms	Type I error	Type II error	Total	Type I error	Type II error	Total
1	138	0	0	138	1	1	136
	(%)	0.00	0.00	100.00	1.45	1.45	98.55
2	138	3	1	134	2	1	135
	(%)	4.35	1.45	97.10	2.90	1.45	97.83
3	138	9	16	113	7	9	122
	(%)	13.04	23.19	81.88	10.14	13.04	88.41

Accuracy comparisons show that MDA has relatively better financial distress classification ability. It successfully classifies all the firms in the model sample and the classification accuracy reaches perfect of 100%. But the prediction ability of MDA seems a little weaker as the accuracies decline to a certain extent. In particular, the prediction accuracy 3 years prior to the financial distress drops suddenly from the classification accuracy of 100% to only 81.88%. Accordingly, Type I error and Type II error increase sharply to 13.04% and 23.19% respectively. The poor prognostic capability of MDA may be possibly explained by: first, MDA incorporates a prior probability to account for the relative occurrence of observations in model population. However, due to the available credit database, it is almost impossible to built a satisfactory financial distress population to observe a prior probabilities. For example, McAllister & Mingo (1994) estimated that to develop very stable estimates of default probabilities, any institution would need some 20,000-30,000 "name" in its database. Very few institutions worldwide come even remotely close to approaching this number of potential clients. Consequently, instead of relating the estimates of the a prior probabilities to the population priors, most studies assume equal priors or priors derived from the model sample. MDA is very sensitive to this violation and the practice in fact limits its ability to make any meaningful inferences; second, MDA suffers from its strict assumptions such as muitivariate normal distribution and equal covariance matrices that are violated very often by the economic and financial data; third, the often-existed high correlations among some financial ratios lead to the problem of multicollinearity in the discriminant function that produces great negative affects to the model estimation procedures.

On the contrary, logistic regression analysis has much better prediction ability. Both prediction accuracies 2 and 3 years prior to the financial distress are higher than that of the MDA. The prediction accuracy 3 years prior to the financial distress even reaches 88.41% while the Type I and II errors are only 10.14% and 13.04%. These remarkable model results can be attributed to the flexibilities of logistic regression analysis that makes no assumptions about the distributions of the independent variables. In particular, through the logit transformation, the logistic regression analysis possesses much more advantages than MDA does, such as that it does not assume multivariate normality and equal covariance matrices. These assumptions are largely conforming to the distribution characteristics of the economic and financial data. Most of all, logistic regression analysis chooses optimal financial ratios that contain capital market prices that are formed by capital providers as they anticipate the future prospects of the firm. Therefore, it incorporates forward-looking information in its function and realizes the efficient combination of data from both financial statements and capital market prices at the same time.

VI. *Conclusion*

The main problems in constructing financial distress prediction models are the design of the financial ratios and the functional form between these variables. Till now, two major approaches have been developed in this field that are the empirical study for optimal predictors that lead to lowest misclassification rate on the one hand and modification for statistical methods that might improve the prediction accuracy on the other hand. However, all these techniques suffer from the problem of random variables selection ignoring the economic and financial features of variables. This is imputed to the fact that financial distress prediction lacks of any unified theory. Therefore, the best techniques are often being decided by comparing the classification and prediction accuracy of the model results. Since 1990s, the economic theories in the financial distress prediction fields have experienced outstanding development. Structural approach, for example, a sophisticated technique built on the option pricing theory, has gain popularity among both the practitioners and scientist, though many theoretical and practical problems around the approach are still in intense disputes. Unfortunately, these techniques cannot be implemented in nowadays China since they are all based on the data from capital market where advanced financial systems are a prerequisite.

This paper extends and improves upon the prior studies in China, primarily in its greater sample size and theoretical and empirical comparative analysis between MDA and logistic regression analysis in the financial distress prediction. Both MDA and logistic functions are constructed to comparatively analyze the selected optimal financial ratios and classification and prediction accuracy 1, 2 and 3 years prior to the firm financial distress. In summary, four conclusions can be made. First, quite similar financial ratios are being selected for both techniques 1 and 2 years prior to the firm financial distress and the optimal financial ratios selected for the logistic model are the subset of that for the MDA. Second, contrast to the similar selected financial ratios mentioned above, totally different financial ratios 3 years prior to the firm financial distress are being selected for the two techniques. Factor analysis shows that these different financial ratios are primarily loaded on the same factors and the high correlations among some financial ratios may help explain this random selection. Third, MDA has better classification ability but worse prediction capability. This might illustrate that the problems connected with strict model assumptions such as normality are not weakening the classification ability of MDA, but its prediction capability. Fourth, logistic regression analysis has relatively better prediction capability. The overall prediction accuracy is higher than that of MDA, and both Type I error and Type II error are lower than that of MDA. This makes it a useful tool in developing firm performance warning systems that are beneficial to both internal and external institutions. To the firm owners and managers, the earlier they can forecast firm financial performances, the better they can discover any hidden problems and be fully prepared to act. To the firm investors, the earlier and more accurate of the performance prognostic, the more investment returns can be guaranteed. Since logistic regression analysis shows higher prediction accuracy and lower Type I & II errors, together with its great flexibility and efficient combination of data from both financial statements and capital market prices, it is considered at present time as the best technique to classify and predict the financial distress of listed companies in China.

References

1. Altman, E.I., Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, Vol.23(1968), pp.589-609.
2. Altman, E.I., Haldeman, R., and Narayanan P., ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking and Finance*, June 1977, pp.29-54.
3. Altman, E.I., G. Marco, and F. Varetto, Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (the Italian Experience). *Journal of Banking & Finance* 18 (1994), pp.505-529.
4. Altman, E.I., and A. Saunders, Credit Risk Measurement Development over the last Twenty Years. *Journal of Banking & Finance* 21 (1998), pp.1721-1742.
5. Back, B., T. Laitinen, K. Sere and M. Wezel, Choosing Bankruptcy Predictors: Using Discriminant Analysis, Logit Analysis, and Genetic Algorithms. Technical Report, No. 40. Turku Center for Computer Science. 1996.
6. Beaver, W., "Financial Ratios as Predictors of Failure, Empirical Research in Accounting: Selected Studies." *Journal of Accounting Research* 5, supplement (1966), pp.71-111.

7. Boritz, J., The "Going Concern" Assumption: Accounting and Auditing Implications. Research Report. Canadian Institute of Chartered Accountants. 1991.
8. Cortina M.J., Big Things have Small Beginnings: An Assortment of "Minor" Methodological Misunderstandings. *Journal of Management* 28 (2002), pp.339-362.
9. Deakin, E., A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research* 1, Spring (1972), pp.167-179.
10. Durand, D., Risk Elements in Consumer Installment Financing. *Studies in Consumer Installment Financing*. NBER (1941), pp.105-142.
11. Edmister, R., An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. *Journal of Financial and Quantitative Analysis* 2, March (1972), pp.1477-1493.
12. Fisher, R.A., The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*. 7 (1936), pp.179-188.
13. Hosmer, W.D., and S. Lemeshow, *Applied Logistic Regression*. John Wiley & Sons, Inc. 2000.
14. Huberty, C.J., *Applied Discriminant Analysis*. John Wiley & Sons, Inc. 1994.
15. Huffman, P.S., and J.D. Ward, The Prediction of Default for High Yield Bond Issues. *Review of Financial Economics* 5 (1996), pp.75-89.
16. Izan, H.Y., Corporate Distress in Australia. *Journal of Banking & Finance* 8 (1984), pp. 303-320.
17. Karels, G., and A. Prakash, Multivariate Normality and Forecasting of Business Bankruptcy. *Journal of Business Finance Accounting* Winter (1987), pp.573-593.
18. Giesecke, K., Credit Risk Modeling and Valuation: An Introduction. Humboldt-Universitat zu Berlin. 2002.
19. Kealhofer, S., and R.J. Bohn, 2001. Portfolio Management of Default Risk. KMV Corporation, San Francisco, CA.
20. Laitinen, K.E., and T. Laitinen, Bankruptcy Prediction: Application of the Taylor expansion in logistic regression. *International Review of Financial Analysis* 9 (2000), pp. 327-349.
21. McAllister, P. and J. Mingo, Commercial Loan Risk Management, Credit-scoring and Pricing: The Need for a New Shared Data Base. *Journal of Commercial Bank Lending*. (1994), pp.6-20.
22. Martin, D., Early Warning of Bank Failure. *Journal of Banking & Finance* 1 (1977), pp. 249-276.
23. McQuown, A.J., A Comment on Market vs. Accounting-based Measures of Default Risk. KMV Corporation, San Francisco, CA. 1993.
24. Platt, D.H., and B.M. Platt, A Note on the Use of Industry-relative Ratios in Bankruptcy Prediction. *Journal of Banking & Finance* 15 (1991), pp.1183-1194.
25. Richardson, F., and L. Davidson, An Exploration into Bankruptcy Discriminant Model Sensitivity. *Journal of Business Finance Accounting* 2 (1983), pp.195-207.
26. Suh, C., A. Rajagopalan, X. Li, and K. Rajan, The Application of Principle Component Analysis to Materials Science Data. Working Paper, Rensselaer Polytechnic Institute.
27. Varetto, F., Genetic Algorithms Application in the Analysis of Insolvency Risk. *Journal of Banking & Finance* 22 (1998), pp.1421-1439.
28. Vasicek, A.O., 1999. Credit Valuation. KMV Corporation, San Francisco, CA.

References in Chinese Journals

29. Chen J., Empirical Analysis of Financial Distress of Listed Companies. [China] *The Journal of Accounting Research*. No.4, 1999.
30. Chen Y., Empirical Analysis of "ST" Firms in the Chinese Stock Market. [China] *Economic Science*. No.6 (2000), pp.57-67.
31. Gao, P.Y., Discriminant Research on the Firm Failure. [China] *The Journal of Statistical Research*. No.10, 2000.
32. Liang, Q., PCA & Multivariate Discriminant Model and Prediction of Corporate Distress. [China] *Studies of Finance & Economics*. No.5, 2003.
33. Liu, W., The Research on the Early Warning System of Unemployment in Shenzhen. [China] *The Journal of Quantitative & Technological Economics Research*. No.2, 2001.
34. Shi, X.Q. and X.Y. Zhou, The Application of Canonical Discriminant Analysis in Credit Risk Evaluation of Enterprise. [China] *The Study of Finance and Economics*. No.10 (2001), pp.53-57.
35. Wang, Ch.F. and W.H. Li, The Evaluation of Credit Risk of Commercial Banks: A Projection Pursuit Model. [China] *Journal of Industrial Engineering/Engineering Management*. No.2 (2000), pp.43-46.
36. Wang, Zh., L. Liu and Ch. Chen, The Information and Influence Factors of the "ST". [China] *Journal of Financial Research*. No.9 (2002), pp.61-71.
37. Zhang, A.M., Ch.Sh. Zhu and D.J. Xu, Principle Component Model and its Empirical Test in the Financial Distress Analysis. [China] *Journal of Financial Research*. No.3 (2001), pp.10-25.

APPENDIX A. MODEL SAMPLE: "ST" & NON-"ST" FIRMS & YEAR OF FAILURE

Year	"ST" Listed Companies		Non-"ST" Listed Companies	
2002	000013	000723	000035	600531
	000405	000738	000538	600535
	000412	000826	000619	600536
	000514	000851	000700	600587
	000533	000893	000920	600614
	000544	600057	600087	600641
	000560	600082	600120	600651
	000566	600622	600126	600706
	000567	600669	600145	600711
	000571	600786	600345	600729
	000572	600792	600355	600742
	000586	600842	600371	600756
	000603	600858	600397	600825
	000678	600870	600428	600880
	000699		600456	
2001	000007	600167	000034	600360
	000033	600703	000406	600367
	000047	600734	000503	600381
	000416	600743	000518	600518
	000430	600745	000540	600558
	000675	600753	000839	600621
	000788	600768	600202	600636
	000838	600848	600310	600739
	600109	600853	600321	600770
	600150		600330	600771
2000	000003	600097	000005	000935
	000010	600137	000023	600110
	000011	600629	000046	600157
	000025	600658	000404	600180
	000038	600683	000525	600697
	000411	600696	000583	600704
	000504	600721	000612	600724
	000546	600762	000615	600857
	000592	600806	000716	600883
	000689	600845	000876	600889
	001696			

Appendix B. Financial Ratios Presentation

Type of the Ratio	Code	Ratio	Studies
Profitability			
Primary Business Index	X_1	Gross Profit	CY, W, Zh
	X_2	Net Sales/Total Assets	A
Investment Return Index	X_3	Net Income/Total Assets	C, CY, B, Zh
	X_4	(Net profit + Interest Expense)/Total Assets	W
	X_5	Net Profit/Net Assets at year end	C, Sh
	X_6	EBIT/Total Assets	I
Capital Investment Index	X_7	Net Profit/Equity	CY, W, Zh
Liquidity			
Working Capital Index	X_8	Current Ratio	C, CY, W, B, Zh, Sh, I
	X_9	Quick Ratio	W
	X_{10}	Working Capital/Total Assets	C, B, A, Sh
Cash Flow Index	X_{11}	Operating Cash Flow Ratio	A, A-H-N
	X_{12}	Total Cash Flow/Total Debt	B, L-L
Solvency			
Debt Ratio Index	X_{13}	Total Debt/Total Assets	C, CY, W, B, A-H-N, Sh
Interest Payment Index	X_{14}	EBIT/Interest Payments	I
Potentiality	X_{15}	Growth Rate of Total Assets	Zh
	X_{16}	Accumulation Rate of Equity	Zh
Activity			
Short-term Index	X_{17}	Inventory Turnover	W
	X_{18}	Receivable Turnover	W
Long-term Index	X_{19}	Fixed Assets Turnover	W
	X_{20}	Total Assets Turnover	C, CY, W, Zh
Capital Market	X_{21}	MV of the Negotiable Shares/Total Debt	A, A-H-N, I
	X_{22}	Equity at the year beginning/Equity at the year end	W
Capital Structure	X_{23}	Retained Earning/Total Assets	A, A-H-N, Sh
	X_{24}	Equity/Total Assets	Zh, L-L
	X_{25}	Total Debt/Equity	I

Legend:

C	Chen, J. (1999)	CY	Chen, Y. (2000)	W	Wang, C.F. etc. (2000)
Zh	Zhang, A.M. etc. (2001)	Sh	Shi, X.Q. etc. (2001)	B	Beaver (1966)
A	Altman (1968)	A-H-N	Altman, Haldeman, Narayanan (1977)		
I	Izan (1984)	L-L	Laitinen & Laitinen (2000)		