FD prediction using the Bayes classifier with MFA

Jigang Xie, Zhengding Qiu, Yanjun Han

(Institute of Information Science, Beijing Jiaotong University, Beijing 100044)

Abstract

A Bayes classifier with mixtures of factor analyzers (MFA) is presented to predict financial distress (FD) of Chinese listed firms. The advantages of the Bayes classifier are that it can do the FD prediction with not only addressed with statistical rigor but also followed the results of default probabilities. We apply two performance measures to compare the Bayes classifier with multivariate discriminant analysis (MDA) and logistic regression (LR). Experimental results show that the Bayes classifier, for three years prior to FD, is superior to MDA and LR, which are the standard benchmarks for the loan default prediction problem.

Keywords: Financial distress prediction; Bayes classifier; Mixtures of factor analyzers; Multivariate discriminant analysis; Logistic regression

1. Introduction

Corporate financial distress (FD) prediction has long been an important and widely studied topic. The use of financial ratios to predict corporate FD has been the major methodology for this research topic. Since the criterion variable is categorical, healthy or unhealthy, the problem is one of classification. Thus, multivariate discriminant analysis (MDA) and logistic regression (LR) models have been typically used for this purpose.

More recently, new methods for predicting FD are developed due to the advantage of computer and information science, such as neural networks, decision trees, rough sets and rule-based methods [1][2][3]. These methods have been proved to be successful in FD prediction. However, there is an open issue that should desirably be addressed by the researchers who are interesting in the new techniques. Even though a prediction of the default event is by itself very useful, an estimate of the default probability is very desirable. For portfolio credit risk estimation, this is essential in order to compute the loss level. Research studies on using these new techniques have not satisfied with the objective [1]. Therefore, since the seminal papers of Altman [4] and Ohlson [5], MDA and LR had been widely used in practice and in many academic studies. They have been standard benchmarks for the loan default prediction problem.

Bayesian decision theory is a fundamental statistical approach to pattern classification. It makes the assumption that the decision problem is posed in probabilistic terms, and that all of the relevant probability values are known. However, the data's probabilistic structure is often unknown in practice. This study uses mixtures of factor analyzers (MFA), which are effective in high-dimensional density estimation [6][7], to build the Bayes classifier. Then, we apply the Bayes classifier to financial data in Chinese listed firms, and compare the results with those of MDA and LR.

The rest of the paper is organized as follows. In section 2, we review Bayes classifier and MFA briefly. Section 3 describes data set in detail. In section 4, we apply MDA, LR and the Bayes classifier on financial data, and analysis the experimental results. The paper concludes with some discussions in section 5.

2. Bayes classifier with MFA

2.1. Bayes classifier

The Bayes classification rule can be summarized as follows. For a K-category problem, let

- ω_i = event that class i occurs, $i = 1, \ldots, K$.
- $\vec{x}$ = observed feature vector, class unknown.
- $p(\vec{x})$ = the probability that a certain observation is made.
- $p(\omega_i)$ = the priori probability.
- $p(\vec{x}|\omega_i)$ =the class-conditional probability.
- $p(\omega_i|\vec{x})$ =the posterior probability.

By observing the feature vector $\vec{x}$, the posteriori probability that the true class is ω_i is

$$p(\omega_i|\vec{x}) = \frac{p(\vec{x}|\omega_i)p(\omega_i)}{p(\vec{x})},\ i = 1, \ldots, K \tag{1}$$

Where

$$p(\vec{x}) = \sum_{i=1}^{K} p(\vec{x}|\omega_j)p(\omega_j) \tag{2}$$

The decision rule is to decide on ω_i if

$$p(\omega_i|\vec{x}) > p(\omega_j|\vec{x}),\ \forall j \neq i \tag{3}$$

Without loss of generality, we let $p(\omega_1) = p(\omega_2) = 1/2$ in this paper. Then only the class-conditional probabilities are unknown. MFA, which is introduced in the next section, is applied to estimate the $p(\vec{x}|\omega_i)$ in this study.

2.2. Mixtures of factor analyzers

In a typical factor analysis model, each observation $\vec{x}_j \in R^d$ is modeled as

$$\vec{x}_j = W\vec{z}_j + \vec{\mu} + \vec{\varepsilon}_j \ , \ j = 1, \ldots, n \tag{4}$$

where $\vec{z}_j$ is a q-dimensional ($q << d$) vector of latent or unobservable variables called factors and W is a $d \times q$ matrix of factor loadings. The $\vec{z}_j$ are assumed to be i.i.d. as $N(0, I_q)$, independently of the noise $\vec{\varepsilon}_j$, which are assumed to be i.i.d. as $N(0, D)$, where D is a diagonal matrix,

$$D = diag(\sigma_1^2, \ldots \sigma_d^2) \tag{5}$$

and where I_q denotes the $q \times q$ identity matrix. Thus, the conditional distributions of the variables $\vec{x}_j$ are

$$p(\vec{x}_j | \vec{z}_j) = N(W\vec{z}_j + \vec{\mu}, D) \tag{6}$$

Unconditionally, the $\vec{x}_j$ are i.i.d. according to a normal distribution with mean $\vec{\mu}$ and covariance matrix

$$\Sigma = WW^{\mathrm{T}} + D \tag{7}$$

Thus, $\vec{x}_j$ is distributed as:

$$p(\vec{x}_j) = N(\vec{\mu}, WW^{\mathrm{T}} + D) \tag{8}$$

If q is chosen sufficiently smaller than d, representation (7) imposes some constrains on the covariance matrix Σ and thus reduces the number of free parameters to be estimated.

Since the factor analysis defines a proper probability model, it can be extended to mixture models. Thus, the density of each observation $\vec{x}_j$ is a mixture of g normal densities in proportions $\pi_1, \ldots, \pi_g$, that is

$$f(\vec{x}_j; \psi) = \sum_{i=1}^{g} \pi_i \phi(\vec{x}_j; \vec{\mu}_i, \Sigma_i) \tag{9}$$

Where

$$\Sigma_i = W_i W_i^{\mathrm{T}} + D_i \, , \ i = 1, \ldots, g \tag{10}$$

The MFA model can be fitted by using the expectation maximization algorithm (see details in [7]).

3. Data set description

3.1. Background

As the Chinese stock market develops, the rules and regulations controlling the money-losing listed firms have become more and more detailed. On 16 March 1998, in order to reduce market risk, the State Securities Regulatory Commission issued a regulation entitled 'ST' Treatment for Listed Firms in case of Financial Abnormalities [1]. The regulation stipulates that whenever financial abnormalities or other abnormalities make it difficult for investors to predict the future of a listed firm, and when such difficulties might damage the interest of investors, the stock exchanges shall give ST to the said firm. 'Financial abnormalities' here means loss in the past two consecutive years, or the net asset per share is lower than the face value, or both. 'Other abnormalities' refers to the virtual termination of production and operation of the firm due to natural disasters and/or serious accidents, or the firm involved in lawsuits having to pay indemnities that exceed its net assets. As long as the two situations occur in a listed firm, the firm automatically falls into the so-called ST block. This study takes the ST received by listed firms due to their abnormal financial performance as the indicator of FD.

3.2. The data set

The sample selection criteria are as follows:

- ST firms: received ST due to financial abnormalities in year t, and their financial statement in fiscal year (t-2), (t-3) and (t-4) available;
- Non-ST (N-ST) firms: listed before Jan. 1 in year (t-4), not ST treated in year t, and their financial statement in fiscal years (t-2), (t-3) and (t-4) available.

This study does not include firms that are special treated due to 'other abnormalities' as 'other abnormalities' are very uncertain and unpredictable. We concentrate on firms that are ST treated due to financial abnormalities. The reason why we do not use the financial statements in fiscal year (t-1) to predict the corporate status in year t is that according to the annual report disclosure practice in Chinese stock exchanges, the two events, namely, the announcement of the financial statement of fiscal year (t-1) and the

[1] The new Share-listing Rules and Regulations promulgated by both Shenzhen and Shanghai stock exchanges on 29 April 2000 have greatly expanded the ST definition. Since this paper studier the firms that got ST treated in 1998-2004, the ST definition adopted on 16 March 1998 is used here.

decision whether they get ST treated in year t, almost take place simultaneously. Upon obtaining the annual report of a firm in fiscal year (t-1), we almost know for sure whether it will get ST treated due to its financial abnormalities. Therefore, it is far more meaningful to use the financial statements of fiscal year (t-2) disclosed in year (t-1) to predict whether a firm will becomes a ST firm in year t. That is, all the predicting information used in this paper are publicly available information disclosed one year before the firm getting ST treated. Thus, the influence of the timing of estimation information on the predictive power of the model is effectively controlled. It should be noted that the financial statement of fiscal year (t-4) is only used to compute the growth ratios of the fiscal year (t-3).

According to the above sample selection criteria, we have eventually acquired a sample composed of 610 listed firms, of which 235 are ST firms, and 375 N-ST firms. In order to measure the performance of the FD prediction models, the 610 selected firms are divided into training and test sets.

3.3. Selection of variables

As there is a lack of economic theoretic guidance, and as the innate causes of ST are different, it is hard for us to describe adequately FD with just few simple financial variables/ratios. Different financial ratios usually represent different aspects of financial features of listed firms. 41 financial ratios, which are categorized as growth, profitability, leverage, efficiency, and productivity, were identified initially.

All the 41 ratios for every firm were calculated and the stepwise selection was done among these ratios 2 and 3 years prior to ST. The set of ratios for MDA and LR was chosen using stepwise selection. The variables that were selected into the MDA and LR models are presented in Table 1 and 2.

Table 1. Variables selected for MDA.

(t-2)	(t-3)
Debt/Equity	Debt/Total Assets
Current Assets Turnover	Sales/Total Assets
Net Income/Net Assets	Net Income/Net Assets
Operating Income/Revenue	Earnings per Share
Earnings per Share	EBIT/Sales
Net Share per Share	Logarithm Net Assets

4. Experiment and analysis

In order to evaluate the three models, two performance measures are used. The first is the prediction accuracy, which has been used by most researchers. The second is the curve of the trade-off function, which can show how effectively a model separates the two populations for different decision cut-offs [8]. The parameters of MFA are obtained by 5-fold cross-validation on the training data.

Table 2. Variables selected for LR.

(t-2)	(t-3)
Quick Ratio	Debt/Total Assets
Cash/Current Liabilities	Net Income/Total Assets
Cash/Total Liabilities	Net Income/Net Assets
Debt/Equity	Earnings per Share
Long-Term	EBIT/Sales
Liabilities/Total Assets	Growth Rate of Total
Profit Margin Rate	Assets
Net Income/Net Assets	Cash Flow to Share
Operating	Logarithm Total Assets
Income/Revenue	
Earnings per Share	

The prediction accuracies of the models are listed in Table 3. It clearly shows that the range of performances differs for the models 2 and 3 years prior to ST. For the validation 2 years prior to ST, MDA and LR are superior to the Bayes classifiers with MFA. However, for the validation 3 years prior to ST, the results are reverse.

Table 3. The prediction accuracies of MDA, LR and the Bayes classifier with MFA

Year	MDA	LR	MFA
(t-2)	80.97%	80.11%	76.7%
(t-3)	51.7%	57.67%	65.91%

The curves of the trade-off function of the models are described in Fig. 1 and 2. Since a model has a better performance as the curve is situated closer to the two axises [8], Fig. 1 and 2 show that the relative performance order of the investigated models is of no difference with Table 3.

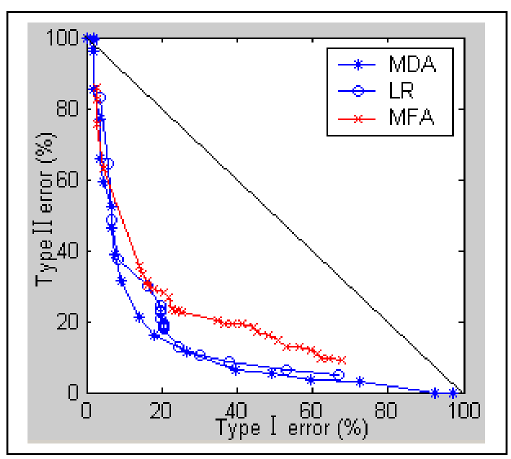

Fig. 1: The curves of the trade-off functions of (t-2) prediction models.

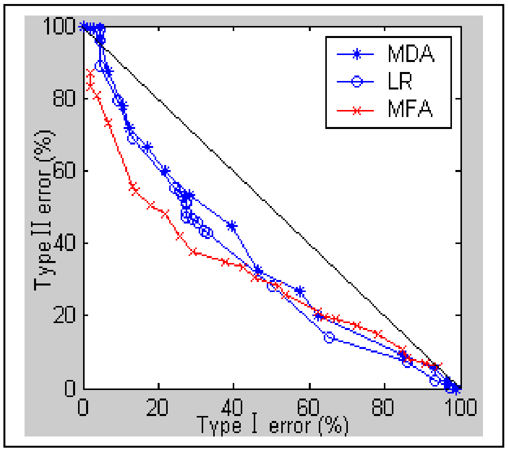

Fig. 2: The curves of the trade-off functions of (t-3) prediction models.

Both the performance measures show the Bayes classifier with MFA is efficient for FD in Chinese listed firms. We attribute this to factor analysis, which can be used not only for explaining correlations between variables in multivariate observations but also for dimension reduction. Thus, all the 41 financial (ratios) information could be used to build the prediction model. However, we see clearly that compared to some bankruptcy research findings [4][5], our prediction accuracies are on the lower side. We believe the reasons are two-fold.

The first reason is the different definition of FD. This study defines FD as being ST treated due to corporate financial abnormalities, that is, the firm suffers a loss for two consecutive years, whereas Ohlson, Altman and other researchers used bankruptcy as an indicator of FD. Apparently the latter is more serious than the former. Thus, compared to the difference between bankrupt and non-bankrupt firms, the difference between ST and N-ST firms is naturally smaller. This makes the ST prediction even more difficult than the bankruptcy prediction.

The second reason is the difference in the quality of financial statements. Since the Chinese capital market has only a very short development history, the capital market performance is far from perfect. Thus, the information content and quality of the financial statements of Chinese listed firms may not be as good as those of the financial statements in the more mature capital markets. Then both these factors lead to lower prediction accuracies.

5. Conclusion

In this paper, the Bayes classifier with MFA is presented to predict FD of Chinese listed firms. The advantages of the model are that it can do the FD prediction with not only addressed with statistical rigor but also followed the results of default probabilities. We apply two performance measures to compare the Bayes classifier with MDA and LR. Experimental results show that the Bayes classifier with MFA is efficient for FD prediction of Chinese listed firms.

Finally, it should be pointed out that the performances of all the models decline noticeably from 2 to 3 years prior to FD. We attribute this to the concept drift [9], i.e. the change of probability distribution over time, which is inevitably caused by the changing circumstance around the firms. An interesting attempt is to predict FD by using the relevant techniques in machine learning domain.

6. References

[1] A. Atiya, "Bankruptcy prediction for credit risk using neural networks: a survey and new results," *IEEE Transactions on Neural Networks*, vol. 12, pp. 929-935, 2001.

[2] M. Kim, "The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms," *Expert Systems with Applications.*

[3] A. Dimitras, S. Zanakis and C. Zopounidis, "A survey of business failures with an emphasis on prediction methods and industrial applications", *European Journal of Operational Research*, pp. 487-513, 1996.

[4] E. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *J. Finance*, vol. 13, pp. 589-609, 1968.

[5] J. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy," *J. Accounting Research*, pp. 109-131, 1980.

[6] G. McLachlan and D. Peel, "Mixtures of factor analyzers," In: Langley, P. (Ed.), *Proceedings of the Seventeenth International Conference on Machine Learning*. Morgan Kaufmann, San Francisco, pp. 599-606, 2000.

[7] P. Moerland, "Mixtures of latent variable models for density estimation and classification," [ftp://ftp.idiap.ch/pub/reports/2000/rr00-25.pdf]. *IDIAP Research Report*, Dalle Molle Institute for Perceptual Artificial Intelligence, Martigny, Valais, Switzerland, 2000.

[8] A. Siegel, "Going concern qualifications and bankruptcy prediction," *Working Paper*, presented at doctoral workshop Leuven on March 2, 1995.

[9] G. Widmer and M. Kubat, "Learning in the presence of concept drift and hidden contexts," *Machine Learning*, vol. 23: pp.69-101, 1996.