中央大学理工学研究所論文集　第9号　2004年
Journal of the Institute of Science and Engineering. Chuo University

Prediction of Bankruptcy of Small to Medium Scale Companies via Semi-Definite Programming

Hiroshi KONNO*, Toshinari KAMAKURA*, Norio WATANABE*, Kazuyuki KOSHIZUKA*

1 Introduction.

The purpose of this project is to propose a new and practical method for estimating the failure probability of a large number of small to medium scale companies using semi-definite programming approach.

Calculation of failure probability plays an essential role for determining an appropriate level of interest rate of the money to be loaned to individual companies. Also, it can be used for failure discriminant analysis, i.e., for classifying companies into failing group and ongoing group during the next period.

Estimation of failure probability has a long history since the great depression in the 1930's. One of the most popular methods is to use the rating scores announced by reliable rating institutions such as S&P and Moody's. Unfortunately, however reliable rating scores are not available for small to medium scale companies, because it is very time consuming and expensive to acquire it.

There exists a number of methods for predicting the failure probability of companies using their financial data. Among successful methods are those based upon rating transition matrix. Also, a number of stochastic models of the evolution of the net capital have been proposed for estimating the failure probability [3].

2 Formulation of the Problem

2.1 Semi-Definite Logit Model

The method to be proposed in this project is based upon another well known approach using a logit model. Let $\boldsymbol{x}^i = (x_1^i, x_2^i, \ldots, x_n^i)$ be the vector of financial attributes associated with company i $(i = 1, 2, \ldots, m)$. Let M_1 and M_0 be, respectively the set of indices associated with failed and ongoing companies. We want to estimate the probability $y = f(\boldsymbol{x})$ of a company whose value of financial attributes is $\boldsymbol{x}$. Let

$$y_i^* = \begin{cases} 1, & i \in M_1 \\ 0, & i \in M_0. \end{cases}$$

Let us consider a logit function

$$f(\boldsymbol{x}) = \frac{\exp(\alpha_0 + \alpha_1 x_1 + \cdots + \alpha_n x_n)}{1 + \exp(\alpha_0 + \alpha_1 x_1 + \cdots + \alpha_n x_n)}, \tag{1}$$

which best fits the observed data $(\boldsymbol{x}^i, y_i^*)$, $i = 1, 2, \ldots, m$ using maximum likelihood method.

* Faculty of Science and Engineering, Chuo University

Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ and let us define

$$z_1(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{x}) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n, \tag{2}$$

which will be called a failure intensity function. Obviously, $f(\boldsymbol{x})$ tends to 0 as $z(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{x}) \to -\infty$ and $f(\boldsymbol{x})$ tends to 1 as $z(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{x}) \to +\infty$.

This method is known to lead to a reasonably good result by choosing appropriate set of financial attributes[4]. However, this model cannot take into account the correlation among financial attributes and the nonlinear dependence.

The simplest nonlinear extension of the logit model is the quadratic logit model where failure intensity function is a quadratic function of $\boldsymbol{x}$ [2]. Let $\boldsymbol{B} = (\beta_{ij}) \in R^{n\times n}$ be a real symmetric matrix and let us define the failure intensity function as follows.

$$z_2(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{B}, \boldsymbol{x}) = \alpha_0 + \sum_{j=1}^{n} \alpha_j x_j + \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} \beta_{jk} x_j x_k. \tag{3}$$

Then the likelihood function associated with this intensity function is given by

$$L(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{B}) = \prod_{i\in M_1} \frac{\exp z_2(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{B}, \boldsymbol{x}^i)}{1 + \exp z_2(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{B}, \boldsymbol{x}^i)} = \prod_{i\in M_0} \frac{1}{1 + \exp z_2(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{B}, \boldsymbol{x}^i)}. \tag{4}$$

To maximize L $(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{B})$, we maximize its logarithm which is a concave function.

This model achieves a better fitting to the learning data. However, it often results in the overfitting of the model to the data and tends to produce poor prediction performance. This is due to the fact that the set

$$S = \{\boldsymbol{x} \in R^n | z_2(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{B}, \boldsymbol{x}) \leq q\}$$

can exhibit a very complicated shape which contradicts common observation that financial data of the majority of successful companies with smaller failure probability are located in some convex region.

To account for this observation, we impose a condition that the set S is convex, i.e., either ellipsoid or paraboloid, not hyperboloid. This is equivalent to assume that $\boldsymbol{B}$ is either positive or negative semi-definite.

This assumption has several advantages over linear and general quadratic logit model. First, it will significantly reduce the chance of overfitting by restricting the shape of equi-intensity surface. Second, this model can account for mid-value property, i.e., the property that the failure probability is smaller when certain attribute attains its value in some interval.

The resulting maximum likelihood estimation problem becomes maximization of a concave function subject to semi-definite constraint:

$$(P)\left|\begin{array}{ll} \text{maximize} & \ln L(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{B}) \\ \text{subject to} & \boldsymbol{B} \succeq \boldsymbol{0} \end{array}\right. \tag{5}$$

where $\boldsymbol{B} \succeq \boldsymbol{0}$ means that $\boldsymbol{B}$ is positive semi-definite.

2.2 Failure Discriminant Analysis

The primary objective of our study is to provide an efficient and transparent method for estimating the failure probability of thousands of small to medium scale companies for which elaborate rating scores are not available.

To convince the validity of our approach, we will apply it to the failure discriminant analysis using the standard cross validation method of data mining analysis.

Let U be the set of financial data of small to medium scale companies. We used $U_1 \subset U$ as the set of training data and used $U_2 \subset U \backslash U_1$ for testing the quality of training. Let

$$U_3 = \{\boldsymbol{y}^1, \boldsymbol{y}^2, \cdots, \boldsymbol{y}^l\}$$

be the randomly chosen subset of U where l is about one half of the total number of data in U.

Let us specify the threshold probability $\alpha \in (0, 1)$ and classify $\boldsymbol{y}^i$'s into two groups as follows.

Failure group F: Those companies with $f(\boldsymbol{y}^i) \geq \alpha$
Ongoing group O: Those companies with $f(\boldsymbol{y}^i) < \alpha$

where $f(\cdot)$ is the estimated failure probability function.

Let $P_F(\alpha)$ be the percentage of companies in F which actually failed during the next period. Also, let $P_O(\alpha)$ be the percentage of companies in O which did not fail during the next period. When α is small enough, then $P_F(\alpha)$ is close to 1, but $P_0(\alpha)$ is close to zero and vice versa.

2.3 Computational Results

We conducted numerical experiments using the financial data of up to 15,000 small to medium scale companies [1]of the production industry in years 1998, 1999 and 2000. About 10% of these companies failed during the next 12 months.

Numerical experiments were conducted on a personal computer with CPU: Peutium IV 853MHz, Operating System: Vine Linux 2.1 CR and RAM: 1025MB. Also, we used NUOPT Version 5.0 (Mathematical Systems, Inc.) to solve a linearly constrained concave maximization problem.

Of crucial importance in this kind of analysis is the choice of appropriate financial attributes. We first generated 105 attributes representing such factors as safety, liquidity, capital efficiency, operating efficiency, asset efficiency, productivity, growth factor and the size of company. The basic strategy for choosing the "best" set of attributes is to find those which maximizes the likelihood function. This process is time consuming. In fact, it takes about one day to compute the best set of attributes. However, this procedure leads to a very good performance in prediction. Also, once the best set of attributes are determined, we can use them over and over again, so that this effort is well compensated.

The best set of attributes generated by this procedure was

Linear logit model: 9 attributes
Semi-definite logit model: 10 attributes.

Among these attributes, 4 were common.

We will present here the performance of the algorithm. Figure 1 shows the computation time for solving (P) when $m = 7800$. We see that the computation time increases more or less exponentially, as commonly observed in a wide class of outer approximation algorithms.

Figure 2 shows the computation time as a function of m, the number of companies. We see from this that the computation time increases more or less linearly. Therefore, we will be able to solve the problem even when m is as large as ten thousand.

Figure 3 shows the so-called efficient frontier based upon linear and semi-definite logit models using the best set of attributes. We see that semi-definite logit model outperforms linear logit model. Let α^* be the level of α such that $P_F(\alpha) = P_O(\alpha)$ and let $P^* = P_F(\alpha^*) = P_O(\alpha^*)$. We see from Figure 3 that $P^* = 0.8652$ for semi-definite logit model which is significantly better than the earlier results reported in [1]. Finally, Figure 4 shows the hitting ratio, i.e., the percentage of correct prediction as a function of α.

[1] Those companies whose capital is less than 300 million yen and number of employees is less than 300.

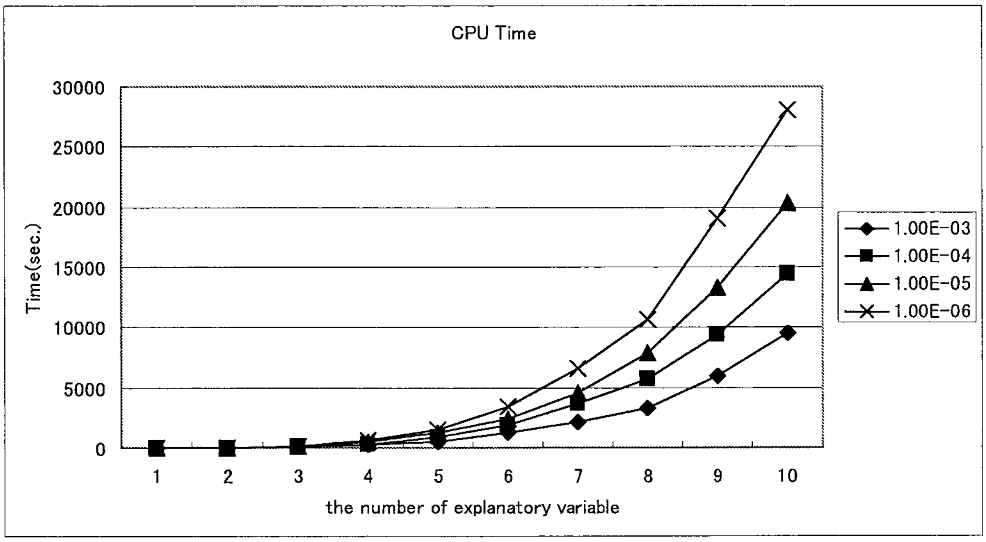

Fig. 1 Computation time

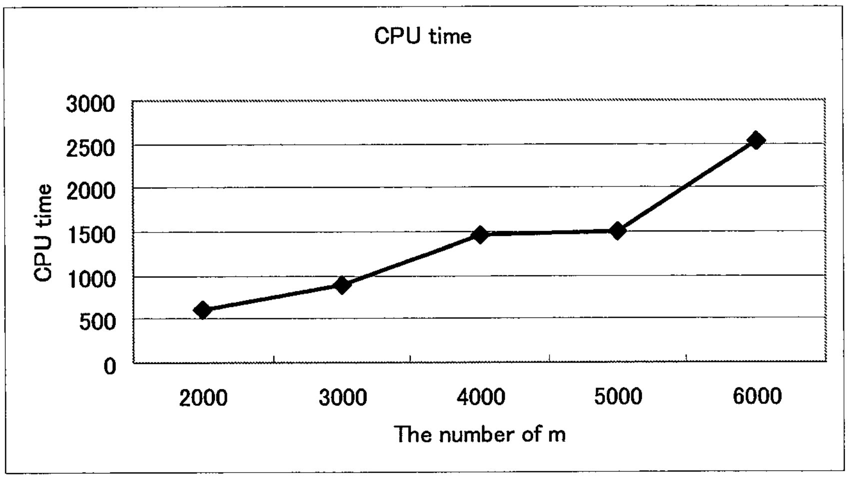

Fig. 2 The computation time as a function of m

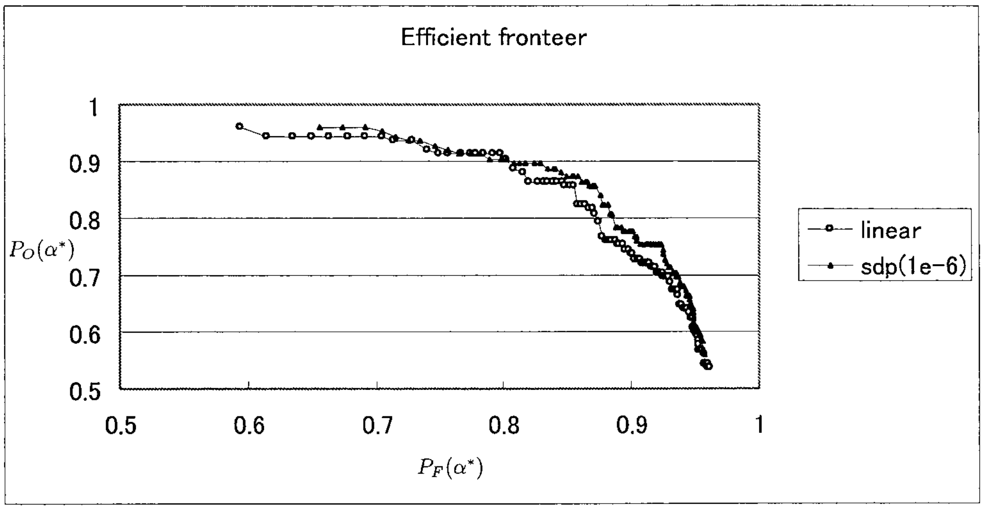

Fig. 3 Efficient frontier

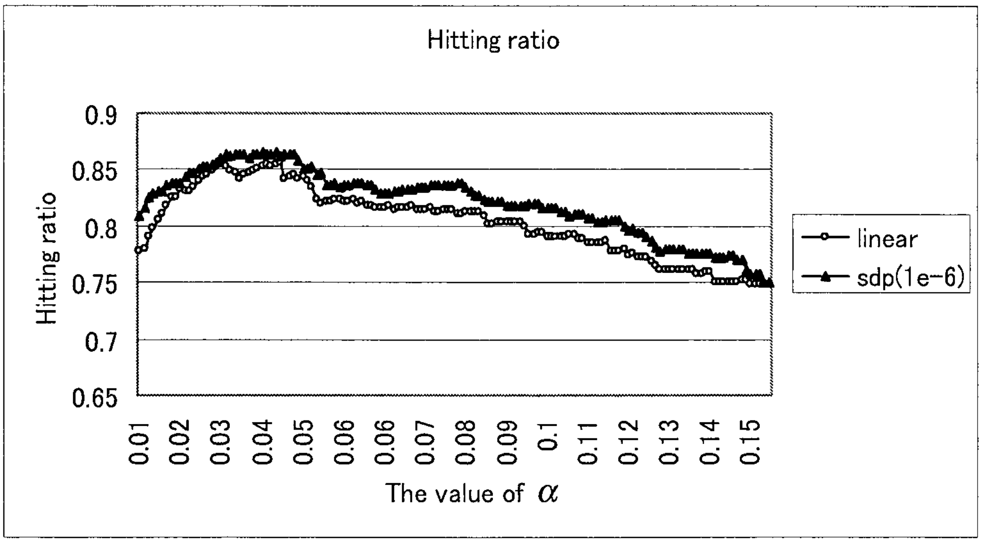

Fig. 4 Hitting ratio

Let us note that the efficient frontiers are associated with the best set of attributes i.e., 9 and 10 attributes for linear and semi-definite logit models respectively. However, it is sometimes too demanding to request small companies to prepare the complete set of data necessary to calculate 9 to 10 attributes. Therefore, one has to be satisfied with smaller number of attributes to calculate the failure probability of small companies. The difference of linear and semi-definite logit models is more significant when we use smaller number of attributes.

3 Conclusions

We showed here that semi-definite logit model can lead to a better prediction of failure probability than linear logit model. We believe that it also leads to a better estimation of failure probability of individual companies.

The calculated failure probability of each company can be used as the basic data for determining the appropriate level of interest rate of the money to be loaned to each company.

Let us add that Japan Credit Rating Agency, Ltd. (JCR), one of the largest rating institutions in Japan has recently released JCREST Scoring System using the method presented in this paper.

Acknowledgements This research was conducted under financial support of Chuo University.

References

[1] Konno, H. and Wu, D., “Estimation of Failure Probability using Semi-Definite Programming”, (in Japanese) J. of SIAM, Japan, 12 (2002) 121-134.

[2] Laitinen, E. K. and Laitine, T., “Bunkruptcy Prediction Applications of the Taylor’s Expansion in Logistic Regression”, International Review of Financial Analysis, 9 (2000) 327-349

[3] Merton, R., C., “An Intertemporal Capital Asset Pricing Model”, Exonometrica 41 (1973) 867-887.

[4] Shirakawa, H., “Credit Risk Management by Scoring”, Communications of Operations Research of Japan, 46 (2001) 628-634.