Proposal of Improving Model for Default Probability Prediction with Logit Model on Non-Compensatory Rule

Sakamaki, Yoshikazu
Department of Value and Decision Science, Tokyo Institute of Technology, Tokyo, 1528550, Japan

ABSTRACT

Financial statement analysis is widely used for credit risk analysis. This method was developed at the end of 19’s for the purpose of surveying credit reliability of credit customers. However, the result of analysis with financial statements is likely to be controlled by the work experience of analysts. As the result, it is difficult to maintain consistency of risk measurement, and moreover, it is expensive and time consuming to perform the vast amount of precise evaluations required in a short time. Many researchers have studied financial statement analysis scientifically applying mathematical and statistical methods, especially with use of multivariate statistical analysis.

This report presents the history of preceding risk control studies and considers issues of the preceding default forecasting model based mainly on the binominal logit model on compensatory rule. Moreover we propose the improving scoring model on the non-compensatory rule and verify that our model is superior to the binominal logit model, and is useful in business using based on actual financial data.

1. INTRODUCTION

During the period of long economic stagnation, Japan has been faced with the task of reestablishing the reliability of its financial systems. The financial markets must be reformed in order to obtain high evaluations worldwide. The Financial Services Agency claims that they will reduce the bad debt ratio of Japanese main banks by 50% by the end of 2004, and resolve the issue of non-performing loans. Adding to this policy, they also suggest that they intend to build a much stronger financial system that can support the promotion of structural reform. To achieve our goal, all financial institutes, personal investors and business corporations must recognize the existence of credit risk and understand the techniques of risk control.

Financial statement analysis is conventionally used as the basic method of credit risk analysis. Credit risk analysis based on financial statements developed at the end of the 19’s for the purpose of surveying the credit reliability of loan customers. However, the result of financial statement analysis is likely to be controlled by the level of work experience of analysts. As a result, it is difficult to maintain the consistency of risk measurement on such a subjective basis. Moreover, it is expensive and time consuming to perform the vast amount of precise risk analysis in the limited time.

Since W. H. Beaver (1967)[1], many researchers have studied financial statements analysis scientifically using mathematical and statistical methods, especially with the use of multivariate statistical analysis. One of the most well known studies is the Z-score model proposed by E. I. Altman(1966)(1968).

This report presents the history of the statistical approach of risk control and considers issues of the preceding default forecasting model with the linearly linked binominal logit model on compensatory rule, which is conventionally used as the scoring model. In addition, an improved version of the scoring model with the binominal logit model on non-compensatory rule is presented.

2. PRECEDING STUDY

This chapter introduces studies of credit risk measurement with a statistical approach in the default forecasting model. As mentioned in Chapter 1, the study of default forecasting originates from Altman’s discriminant model (1968). The result of his study is well known as the Z-score model, and it remains one of the most widely used methods for default forecasting all over the world.

Altman’s z-score model is based on discriminant analysis. In this model, they assume that each firm has its own risk factor, assigning appropriate weight to each risk factor and adding all of the weighted factors. As the result, we obtain a linearly linked risk factor for a bankrupt company and a non-bankrupt company.

That is, let \( x_{mi} \) be the \( m^{th} \) risk factor for company \( i \) \((m=1,2,\ldots,M)\) and \( \beta_m \) be the parameter for variable \( x_{mi} \), and we define z-score as

\[
z_i = \sum_{m=1}^{M} \beta_m x_{mi}
\]

(1)

Here, the company is default if \( z_i < 0 \), and non-default if \( z_i > 0 \). Now, we define

\[
Z_{1^{st}}: \text{Average of z-score for default companies}
\]

\[
Z_{2^{nd}}: \text{Average of z-score for non-default companies}
\]

---(2)

and let variance of the linearly-linked risk factors through all the sample data be \( \text{Var}(Z) \). Then, unknown parameters are given by maximizing
\[ \eta^2 = \frac{(z^2 - \bar{z}^2)^2}{\text{Var}(z)} \]  

With the discriminant score calculated from the linearly linked risk factors weighted by the estimated parameters, we can estimate whether each debtor belongs to the default or non-default group.

The evaluation methods of credit risk based on the linear discriminant function is mathematically easy to handle and analogous to traditional credit risk evaluation analysis using a score card. That is the reason why the scoring model with discriminant analysis is used all over the world. However, when we apply discriminant analysis to credit risk control, we have to be very careful of the following statistical assumptions specific for credit risk measurement.

<1> Normality and variance equivalence
If the variables used in the discriminant analysis is not distributed normally or variance is not equivalent between the default and the non-default group, default distributed normally or variance is not equivalent if the variables used in the discriminant analysis is not correctly.

<2> Uniformity of default/non-default probability
When we use discriminant analysis, we cannot have any information about the ratio of the default and the non-default firms. The discriminant model is usually built with the assumption that the ratio of the default firms and the non-default firms is equal. However, the ratio of default/non-default firms generally differs among banks.

<3> Meaning of calculated weighting parameters
Unlike the parameters estimated in regression analysis, the absolute value of discriminant parameters is not decided uniquely, and only the relative ratio among parameters is decided uniquely. That is, the absolute score in discriminant analysis cannot also be decided uniquely. In discriminant analysis, only the distance from the discriminant point is statistically significant.

Discriminant analysis is widely used but, as shown above, it involves the issue of statistical difficulty. Moreover, we need to make clear when bankruptcy occurs and what the default probability is, in order to qualify the credit risk. Even if we can know the relative default score as the result of discriminant analysis (z-score), we cannot know the absolute default probability.

To solve such an issues, normal regression analysis has begun to be used. Next we consider linear regression analysis. That is, let

\[ y_i: \text{Bernoulli random variable (If firm } i \text{ is non-default, it takes 0. And if firm } i \text{ is defaulted, it takes 1.)} \]
\[ x_{ij}^{'}: j^{th} \text{ credit risk factor of firm } i \ (j=1,2,\ldots) \]
\[ \beta_j: \text{Estimated parameter for } j^{th} \text{ risk factor} \]

And the expected value of the dependent variable given by equation (4) becomes the estimated default probability. However, in the linear regression analysis, when the summation of all of the weighted risk factor becomes very large, the default probability exceeds 1. On the other hand, if the summation becomes very small, the default probability is less than 0. And it becomes contradictory to the definition of statistics. To solve this issue, the following method is used to adjust the expected default probability estimated through the above regression analysis. That is, when the expected default probability exceeds 1, we truncate the default probability to 1, and when the expected default probability is less than 0, we truncate the default probability to 0. However, it is not statistically natural to put the expected default probability between 0 and 1 compulsorily.

To resolve these issues, logistic regression analysis (binominal logit model on compensatory rule) began to be used (Martin;1977). In the logistic regression analysis, let the summation of all of the weighted risk factors be

\[ Z_i = \beta_0 + \beta_1 x_{ij} + \cdots + \beta_n x_{im} \]  

and the default probability of firm \( i \) (\( p_i \)) be

\[ p_i = \frac{1}{1 + \exp(-Z_i)} \]  

If the default probability of firm \( i \) is given by equation (6), the likelihood function for parameter vector \( \beta \) is given by

\[ L(\beta) = \prod_{i=1}^{I} p_i^{y_i} (1 - p_i)^{1-y_i} \]  

Here, \( y_i = 1 \) if firm \( i \) is defaulted and \( y_i = 0 \) if firm \( i \) is not defaulted.

Moreover, log-likelihood function is defined by

\[ l(\beta) = \sum_{i=1}^{I} \{ y_i \log p_i + (1 - y_i) \log(1 - p_i) \} \]  

Generally, parameters that maximizes equation (8) are estimated by the maximum likelihood estimation method.

3. MODEL IMPROVEMENT

3-1. Issues in the preceding model

Default probability estimate based on logistic regression analysis is very significant because it enables measurement of default probability on an absolute scale, which cannot be performed in the classical credit risk models such as that of Altman (1968). However, in the normal logit model based on the compensatory rule, even if only one risk factor deteriorates, the default probability becomes very high although the firm does not actually confront management distress.

Regarding this point, researchers have shown very interesting report in preceding studies of marketing.
research. In the study of marketing research, compensatory rule is used widely in the consumer’s choice behavior model today because compensatory rule is widely believed to be able to mimic consumer’s choice behavior on non-compensatory rule, when the following two conditions are satisfied.

That is,
(1) Attributes are related monotonically to consumer’s preferences.
(2) There is no error or uncertainty about these preferences. (Dawes and Corrigan (1974[4]).

Because these conditions are thought likely to be satisfied in most choice behavior cases, compensatory models are widely used even when it does not reflect the actual decision making process (Cattin and Wittink (1982)[5]).

Despite these reports, there is still reason for concern that compensatory models may not always mimic non-compensatory processes. Several researchers (Curry Louviere, and Augustine (1981)[6]; Einhorn, Kleinmuntz and Kleinmuntz (1979)[7]) have argued that studies demonstrating robustness in linear models have underemphasized a major influence on predictive accuracy: the correlation among attributes of each alternatives.

Newman (1977) was the first to examine explicitly the relationship of correlations among attributes and the fit of linear models, following brief discussions offered by Einhorn (1970)[8] and Goldberg (1971) [9]. He noted that though Wainer et al (1976)[10] conjecture that small departures from optimal coefficients in a linear model do not make no nevermind, will often hold in practice, it will not hold when the attributes describing alternatives are negatively correlated.

Negative correlations also can diminish model performance because they make predictions more sensitive to the particular choice strategy used Einhorn, Kleinmuntz, and Kleinmuntz (1979) illustrate this effect through the example of a negatively correlated environment containing two options, described on two attributes.

The first is good on attribute A and bad on B whereas the second is bad on A and good on B. The option that will be picked in this case depends on how a choice strategy weighs the two attributes. A lexicographic strategy that selects the option best on attribute A, for example, would choose the first option. A similar lexicographic strategy looking first at attribute B would pick the second. In contrast, a compensatory rule that weighs the two attributes equally would be indifferent. If the two attributes were correlated positively, if the first option were good on both attributes and the second bad on both, one would pick the first option regardless of how the attributes are weighted. Hence, in general, the less the redundancy among attributes in a choice environment (the more negative the average inter-attribute correlation), the greater the difference in predictions made by different models.

A more precise discussion regarding this issue was made by Curry and Faulds (1986), who described the theoretical relationship that should be present between two different linear weighting schemes given different correlational structures. They noted that when two models are compared in a maximally negative inter-correlational environment, even small deviations in their weight can induce major discrepancies in prediction, approaching a complete rank-order reversal in overall preferences for options.

In this study, we propose an improving default probability forecasting model based on these discussions.

Table 1 shows the basic statistical value and correlation matrix of trunk seven representative financial ratios of default firms and non-default firms on the BULK database system provided by Nikkei group. As you can see, in the firms confronting bankruptcy, most of the financial indexes deteriorate compared with non-default firms.

Table 1 Basic statistics and correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>Non-default</th>
<th>N</th>
<th>average</th>
<th>std</th>
<th>minimum</th>
<th>maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash Ratio</td>
<td>0.1391</td>
<td>2820.3794</td>
<td>6.5273</td>
<td>1.1875</td>
<td>48.7401</td>
<td>101.0401</td>
</tr>
<tr>
<td>Stock Ratio</td>
<td>0.4446</td>
<td>2820.3794</td>
<td>4.8509</td>
<td>0.8909</td>
<td>265.8250</td>
<td>101.0401</td>
</tr>
<tr>
<td>Deposit-Loan Ratio</td>
<td>0.5719</td>
<td>2820.3794</td>
<td>0.5720</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Ratios of Return on Sales</td>
<td>0.1391</td>
<td>2820.3794</td>
<td>1.0452</td>
<td>0.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

*Correlation is calculated with use both of default and non-default data.

As shown in Table 1, on many risk factors, especially firms confronting bankrupt have deteriorating financial ratios. That is, it can be predicted that all of the financial ratios should become worse at the same time when bankruptcy occurs. And there seems to be very negatively strong correlation among the major financial ratios.

Taking these points into consideration, applying the logit model on the non-compensatory rule in which default probability increases only when all of the risk factors used in the model deteriorate seems to be more appropriate than applying the model of the compensatory rule.
In this study, we propose an improving default probability forecasting model based on non-compensatory rule. Moreover, we apply the actual financial data to our model and reveal a higher default forecasting power than the preceding model.

3-2 Model Improvement

In this section, we show two ways of improving the default probability forecasting model: the conjunctive type and the disjunctive type.

<Conjunctive type>

In this type of model, borderline for each risk factor is decided, and we assume that firm fails only when all risk factors go over the borderline. Borderline varies among risk factors, and if we assume that correlation among error terms become zero implicitly and variation fluctuates based on logistic distribution at random, we can define the probability for firm i to fail as

\[ P(y_i = 1|\beta) = P(x_{ik} - \tau_k + \varepsilon_{ik} \geq 0) \]

\[ = \prod_{k=1}^{K} P(x_{ik} - \tau_k + \varepsilon_{ik} \geq 0) \]

\[ = \prod_{k=1}^{K} \frac{1}{1 + \exp(-\beta_k(x_{ik} - \tau_k))} \]

(9)

It means that the default probability for firm i is in proportion to the distance from the borderline of each risk factor.

<Disjunctive type>

In this type of model, we assume that firm fails if, at least, any one of the risk factors goes over the borderline. Just as mentioned in the conjunctive type model, assuming that borderline varies among firms and variation fluctuates based on logistic distribution, we can derive the following equation.

\[ P(y_i = 1|\beta) = P(\exists k \; x_{ik} - \tau_k + \varepsilon_{ik} \geq 0) \]

\[ = 1 - P(x_{ik} - \tau_k + \varepsilon_{ik} < 0, \forall k) \]

\[ = 1 - \prod_{k=1}^{K} P(x_{ik} - \tau_k + \varepsilon_{ik} < 0) \]

\[ = 1 - \prod_{k=1}^{K} \frac{1}{1 + \exp(-\beta_k(x_{ik} - \tau_k))} \]

(10)

Here,

- \( x_{ik} \): kth risk factor of firm i (i=1,2,...,I; k=1,2,...,K)
- \( \beta_k \): Scale parameter for kth risk factor (k=1,2,...,K)
- \( \tau_k \): Border line of kth risk factor (k=1,2,...,K)
- \( y_i \): Bernoulli random variable (If firm i is non-default, it takes 0. And if firm i is defaulted, it takes 1.)

3-3 Comparison model

As the comparison model, we use the normal logit model shown by equation (6).

3-4 Parameter estimation

Parameters are estimated by equation (11) based on the MLE (Maximum Likelihood Estimation method). If we use MLE for the parameter estimation of nonlinear model, parameter solution sometimes drops in the local minimum.

To avoid this issue, parameters are estimated by comparing all of the combination patterns of variables to decide the most suitable variable combination, and we generated 20 combination patterns of values at random in advance which was used as the initial value for each combination pattern of variables.

\[ L(\beta) = \prod_{i=1}^{I} p_i^{y_i}(1 - p_i)^{1 - y_i} \]

(11)

3-5 Calibration of model fitness

In this study, log-likelihood, AIC criteria, Kolmogorov-Smirnov distance, and divergence are used as the index of model fitness to compare the appropriateness of the normal linearly linked binominal logit model on the compensatory rule and improving logit model on the non-compensatory rule.

Here, the Kolmogorov-Smirnov test measures the divergence between the default/non-default distribution. This difference is known as the KS distance and it is used to assess the fitness degree of our model.

Divergence is the statistical index that shows the separation of the distribution of the default and non-default firm, which is calculated with the average and variance of the score.

It can be calculated by

\[ \text{Divergence} = \frac{2(\mu_A - \mu_B)^2}{V_A + V_B} \]

(12)

Where,

- \( \mu_A, \mu_B \): Expected value of score given to default/non-default firms
- \( V_A, V_B \): Variance of score given to default/non-default firms

4. CALIBRATION

This chapter verifies the appropriateness of the proposed model. Publicly disclosed financial data of companies sold by Nikkei Group (BULK system) is used as the calibration data.

4-1 Outline of calibration data

1. Data source: Account settlement data of the company (From April 2000 to March 2001)
2. Data volume: 2,914 companies
3. Default ratio:
   If the company defaulted in three years after the accounting date, the company is regarded as defaulted. (Default ratio: 84/2,914=2.88%)
Here, especially, in the case of normal logit model, which has a linear structure of risk factors, outliers in financial ratios often cause the problem of biased default probability estimation. In this study, we set the lower cap at the 1 percentile point and the upper cap at 99 percentile point.

### 4-2 Parameter estimation

Based on this actual financial data, we estimate model parameters for the normal logit model, conjunctive/disjunctive type of logit model on the non-compensatory rule.

Generally, the fitness of the model on the training data set is fine, but the most important question is how the model works in out-of-sample data. So that in estimating parameters, we have chosen 60% of the sample data at random and used for in-sample-data, and 40% for out-of-sample data. Moreover, we generated 20 combination patterns of values at random in advance and used as the initial value of the maximum likelihood estimation method for the search algorithm, and prevented from dropping to the local minimum. Based on the value of the Log-Likelihood and AIC criteria, the proposed model adopts the estimated parameters which have minimum Log-Likelihood and AIC as the optimum parameters. It is probable that we obtain the result by chance, so that we tested several datasets to examine the superiority of our model. We show one of the results of parameter estimates in Table 3 due to the limitation of spaces.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Normal Logit Model</th>
<th>P-value</th>
<th>Compensatory Model (Conjunctive)</th>
<th>P-value</th>
<th>Non-Compensatory Model (Disjunctive)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B_0 ) Intercept</td>
<td>0.0021</td>
<td>0.00001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( B_1 ) Capital adequacy ratio</td>
<td>-3.9091</td>
<td>0.0219</td>
<td>-8.21248</td>
<td>0.00036</td>
<td>-2.94515</td>
<td>0.01859</td>
</tr>
<tr>
<td>( t_1 ) Acid Ratio</td>
<td>-1.93705</td>
<td>0.03447</td>
<td>-4.76147</td>
<td>0.0274</td>
<td>6.9669</td>
<td>0.05509</td>
</tr>
<tr>
<td>( B_3 ) Ratio of check account to sales</td>
<td>-5.31485</td>
<td>0.03125</td>
<td>-3.63163</td>
<td>0.0917</td>
<td>-2.07471</td>
<td>0.07811</td>
</tr>
<tr>
<td>( B_5 ) Ratio of return on sales</td>
<td>-5.90596</td>
<td>0.00049</td>
<td>-1.85393</td>
<td>0.13607</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( B_7 ) Cash Ratio</td>
<td>-0.01782</td>
<td>0.0021</td>
<td>-0.01782</td>
<td>0.0021</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Log-Likelihood (-2logL)</td>
<td>357.25248</td>
<td>310.964</td>
<td>386.82998</td>
<td>369.25248</td>
<td>326.964</td>
<td>398.42998</td>
</tr>
<tr>
<td>AIC</td>
<td>367.25248</td>
<td>326.964</td>
<td>398.42998</td>
<td>369.25248</td>
<td>326.964</td>
<td>398.42998</td>
</tr>
<tr>
<td>Divergence</td>
<td>0.153</td>
<td>0.162</td>
<td>0.108</td>
<td>0.108</td>
<td>0.108</td>
<td>0.108</td>
</tr>
<tr>
<td>K-S Distance</td>
<td>71.66%</td>
<td>75.26%</td>
<td>69.69%</td>
<td>71.66%</td>
<td>75.26%</td>
<td>69.69%</td>
</tr>
</tbody>
</table>

Here, parameter estimates are obtained from in-sample data. And Log-Likelihood (-2logL), AIC, Divergence and K-S Distance are calculated based on the out-of-sample data.

### 5. RESULTS

#### 5-1 Findings

As we can see in the financial policy for the revision of capital adequacy ratio based on BASEL II, application of rating technology for credit risk control is attracting significant attention from all over the world. For example, in BASEL II, it is permitted for each bank to use an original rating model, so-called internal rating grades, to grasp the total credit risk. And to realize the suitable increase of capital adequacy ratio, a more elaborate statistical model that can grasp total credit risk in the bank on time is needed. The linearly linked binominal logit model on the compensatory rule is used as the conventional model. However, generally, financial ratios that have strongly negative correlation are often used as a risk factor in the model, and as mentioned in the previous chapter, when we apply the normal logit model to the risk factors that have mutually negative correlations, it is expected that there exist stochastic issues.

So, this study proposed a default probability forecasting model on the non-compensatory rule and resolved these issues. Simultaneously, based on the actual financial ratio data, we have shown that our model was superior to the preceding model.

#### 5-2 Considerations
5-3 Parameter Estimates

As mentioned in the previous chapter, when there exist strongly negative correlations among variables of the model, it is suggested that model performance reduces on the model of compensatory rule like normal logit model.

At first, to confirm the level of correlation among some financial indexes, we made correlation matrix and confirmed that there existed strongly negative correlations among some variables. (Show Table 1).

And after that, we proposed improving default probability forecasting model that enabled to improve the model performance. We proposed two types of models (Conjunctive type and disjunctive type), and as the result of model validation, we confirmed that we could improve model performance dramatically with use of conjunctive type model.

It can be thought as the major cause that most of the financial indexes tend to deteriorate at the same time when firm fails in addition to there exists strongly negative correlations among financial indexes.

6. SUMMARY AND FUTURE INVESTIGATIONS

In this study, we proposed improving logit model based on non-compensatory rule instead of traditional normal logit model based on compensatory rule. We applied actual financial data for our model and showed superiority of our model.

Only financial ratios are used as the calibration data. But generally, when we apply default forecast model for small and medium entertainments or small individual firms, financial data may be window-dressed or the data may be modified intentionally. As a result, reliability for the financial data is reduced in such companies. To remove this issue, it is advocated that not only financial ratios but also individual attributes should be included in the model. However, including and combing such individual attributes is problematic. In an actual business, models are built by considering individual attributes and financial ratios simultaneously in the same logit model or mixing both of the estimated scores with a certain ratio after parameters are estimated independently on both of individual attribute data and financial ratio data. We can expect to build much stronger model through these processes.

REFERENCE

AUTHORS BIOGRAPHIES

Yoshikazu, Sakamaki is a student of doctor course in Tokyo Institute of Technology, Department of Value and Decision Science, Japan. He received a B.Eng in the Department of Control and System Engineering from Tokyo Institute of Technology and his M.Eng in the Department of Industrial Engineering and Management from Tokyo Institute of Technology. His recent work has involved financial engineering, marketing research with use of statistical approach. His email address is megacity@f3.dion.ne.jp.