

Evaluating Predictive Ability of Classification Models with Ordered Multiple Categories

Oong-Hyun Sung¹⁾

Abstract

This study is concerned with the evaluation of predictive ability of classification models with ordered multiple categories. If categories can be ordered or ranked, the spread of misclassification should be considered to evaluate the performance of the classification models using loss rate since the apparent error rate can not measure the spread of misclassification. Since loss rate is known to underestimate the true loss rate, the bootstrap method were used to estimate the true loss rate. Thus, this study suggests the method to evaluate the predictive power of the classification models using loss rate and the bootstrap estimate of the true loss rate.

1. Introduction

In developing predictive classification models, it is useful to employ three stages of the analysis. The first stage, concerned solely with the original samples, is used to estimate a classification function for a purpose of classification. There are two important points to be checked before the estimation of the classification function. First, the distribution assumptions concerning the data needed to be checked. The normal classification model assumes multivariate normality of the independent variables, while rank transformation classification models and the logit model require no assumptions on the form of the distribution of the independent variables.

The second stage is concerned with the performance of a classification model. One way of evaluating its performance is to calculating the apparent error rate, which is the proportion of observed misclassification made by the classification function on its own sample. If each category is to be treated equally, the use of the apparent error rate is useful to measure the performance of the classification model. However, if categories can be ordered or ranked, the spread of misclassification should be considered to measure the performance. Therefore, an alternative misclassification rate such as a loss rate, which can measure the spread of misclassification, should be employed when dealing with the classification case with ordered multiple categories.

The third stage is concerned with predictive power of the classification model. As with

1) Associate Professor, Department of Statistics, Hanshin University, Osan, Kyounggido, 447-791, Korea

any inferential technique based on samples, it is well known that the misclassification rates obtained in the second stage tend to underestimate the true misclassification rate. This is because the prediction is based on the same data used to derive the classification functions.

Recently, resampling method such as bootstrap method has been used to estimate the predictive power of the classification model. Efron (1979,1983), and Efron and Gong (1983) compared the bootstrap method with the cross-validation method under normal distribution in the linear classification model. He concluded that the cross-validation method gives a nearly unbiased estimate of the true error rate, but often unacceptably high variability, particularly if the sample size is small. However, the bootstrap method produced an almost unbiased estimate with a small variance.

For the empirical study with ordered multiple categories, the classification of the bond ratings are considered. Pogue and Soldofsky (1969), and Kaplan and Urwitz (1979) used regression model and Pinches and Mingo (1975) used a linear and quadratic classification model. They employed holdout method to evaluate the predictive power of the models with small holdout samples. One drawback of this method is that there are problems connected with the size of the test sample—if it is large, a good assessment of the performance will be obtained, but if it is small, its performance is highly variable. The problems found in the previous works are: (1) the violation of a multivariate normality in the normal classification models, (2) the report of overall misclassification rate, and (3) inaccurate measure of predictive ability.

Therefore, this study deals with the common problems appearing in the classification analysis. First, this study compares the efficiency between the normal classification models (linear, quadratic), rank transformation classification models (linear, quadratic) and the logit model. Second, this study introduces a loss function in order to measure the spread of misclassification and evaluates the performance with the loss rate as well as the apparent error rate. The loss rate can be defined as the average of the loss made due to misclassification. Third, this study is concerned with the estimation of the true loss rate using the bootstrap method to measure the predictive power of the classification models.

2. Misclassification Rate

Consider classification models with g ordered multiple categories. The data consist of random sample of size n_i from each category i and size N for entire sample. The model defines disjoint classification regions $R_1(x), R_2(x), \dots, R_g(x)$ such that a future observation X_o is classified to category i if $X_o \in R_i(x)$. The apparent error rate (APER) is defined as the proportion of observed classification error made by the estimated classification function in the sample. It is useful to define a misclassification variable $M_{ij}(X_o|X)$ which indicates the misclassification of an single observation X_o from category i into category j , given the entire

sample. Thus

$$M_{ij}(X_o|X) = \begin{cases} 0 & \text{if } X_o \in R_i(x) \\ 1 & \text{if } X_o \in R_j(x) \end{cases}$$

Let $APER_{ij}$ be an estimate of conditional probability of misclassifying an observation from category i into category j . The apparent error rate of category i , $APER_i$, is an estimate of the marginal probability of misclassifying an observation from category i . It can be expressed as the proportion of observed errors made by the classification functions on its own sample for category i .

$$APER_i = \sum_{j \neq i}^g APER_{ij} = \frac{1}{n_i} \sum_{t=1}^{n_i} \sum_{j \neq i}^g M_{ij}(x_{it}). \quad i = 1, 2, \dots, g. \quad (2.1)$$

The overall apparent error rate, $APER$, is an estimate of the overall probability of misclassification and is obtained by multiplying each $APER_i$ by its sample proportion $s_i = n_i/N$ and summing it over i :

$$APER = \sum_{i=1}^g s_i APER_i = \frac{1}{N} \sum_{i=1}^g \sum_{t=1}^{n_i} \sum_{j \neq i}^g M_{ij}(x_{it}). \quad (2.2)$$

The apparent error rate is intuitively appealing and easy to calculate. It is used as an estimate of the true error rate. Unfortunately this will typically lead to an optimistic(under)estimate of the true error rate for future observation. Efron (1983) defined the true error rate (TER) which is the probability of incorrectly classifying a randomly selected future observation X_o in the binary classification case. Extending his theory the true error rate for category i using (2.1) can be defined as:

$$\begin{aligned} TER_i &= \sum_{j \neq i}^g TER_{ij} \\ &= \sum_{j \neq i}^g E\{M_{ij}(X_o|X) : F_i\} \\ &= \sum_{j \neq i}^g P\{X_o \in R_j(x) : F_i\}, \quad i = 1, 2, \dots, g. \end{aligned} \quad (2.3)$$

In the expression the data X and the classification regions $R_j(x)$ are regarded as fixed. The symbol $E\{M_{ij}(X_o|X) : F_i\}$ indicates expectation over a single new observation X_o from unknown distribution F_i . $P\{X_o \in R_j(x) : F_i\}$ indicates probability of misclassifying a single new observation from F_i into category j . The overall true error rate is

$$TER = \sum_{i=1}^g p_i TRE_i, \quad (2.4)$$

where p_i is a prior probability for category i .

3. Loss Rate and Excess Loss

For the case of ordered multiple categories the application of the apparent error rate, which assigns equal loss of misclassification, seems to be inappropriate since the apparent error rate can not measure the spread of the misclassification. An alternative method should be used to measure the spread of misclassification. In ordered categories it is assumed that the loss of misclassification increases as the difference between the observed order and the estimated order becomes wider. Therefore, a general loss function should be introduced when a new observation X_o from category i is misclassified as category j as follows.

$$L_{ij}^{\alpha}(X_o|X) = \begin{cases} |i-j|^{\alpha} & \text{if } X_o \in R_j(x) \\ 0 & \text{if } X_o \in R_i(x) \end{cases} \quad (3.1)$$

where α is a given positive number. Since α is usually unknown, the assignment of α must be highly subjective matter. When $\alpha=2$, $L_{ij}^2(X_o|X)$ is called the squared loss function, When $\alpha=1$, $L_{ij}^1(X_o|X)$ is called the absolute loss function, and when $\alpha=0$, $L_{ij}^0(X_o|X)$ is called the constant loss function which is the same as the misclassification variable $M_{ij}(X_o|X)$.

The loss rate of category i can be defined as the average of the loss made in the sample of category i due to misclassification.

$$LR_i^{\alpha} = \frac{1}{n_i} \sum_{t=1}^{n_i} \sum_{j \neq i}^g L_{ij}^{\alpha}(x_{it}), \quad i = 1, 2, \dots, g. \quad (3.2)$$

For example, the larger the difference of the constant loss rate and the absolute loss rate becomes, the wider the spread of the misclassification gets. Therefore, the loss rate should be considered to measure the predictive ability of the classification models. And overall loss rate can be defined as

$$LR^{\alpha} = \sum_{i=1}^g p_i LR_i^{\alpha}. \quad (3.3)$$

However, like the apparent error rate the sample based loss rate usually tends to be smaller than the true loss rate because the sample data have been used both to construct and to evaluate the classification functions. The true loss rate can be defined as the expected value of the loss function, which measures the loss of a randomly selected future observation X_o from an unknown distribution. Let TLR_{ij}^a be the true loss rate of misclassifying an observation from category i into category j . The true loss rate for category i , TLR_i^a , is

$$\begin{aligned} TLR_i &= \sum_{j \neq i}^g TLR_{ij}^a \\ &= \sum_{j \neq i}^g E\{L_{ij}^a(X_o|X) : F_i\} \\ &= \sum_{j \neq i}^g |i - j|^a P\{X_o \in R_j(x) : F_i\}, i = 1, 2, \dots, g. \end{aligned} \quad (3.4)$$

and the true overall loss rate is

$$TLR^a = \sum_{i=1}^g p_i TLR_i^a. \quad (3.5)$$

Since loss rate tends to be biased from the true loss rate, excess loss R_i^a for category i can be defined as

$$\begin{aligned} R_i^a &= TLR_i^a - LR_i^a \\ &= \sum_{j \neq i}^g TLR_{ij}^a - \sum_{j \neq i}^g LR_{ij}^a \\ &= \sum_{j \neq i}^g E\{L_{ij}^a(X_o|X) : F_i\} - \sum_{j \neq i}^g E\{L_{ij}^a(X_o|X) : \hat{F}_i\} \\ &= \sum_{j \neq i}^g |i - j|^a P\{X_o \in R_j(x) : F_i\} \\ &\quad - \sum_{j \neq i}^g |i - j|^a P\{X_o \in R_j(x) : \hat{F}_i\}, \end{aligned} \quad (3.6)$$

and the overall excess loss rate R^a is

$$R^a = \sum_{i=1}^g p_i R_i^a. \quad (3.7)$$

The notation \hat{F}_i is empirical probability distribution which puts mass $1/n_i$ at each observation in category i . The symbol $P\{X_o \in R_j(x) : \hat{F}_i\}$ indicates probability of misclassifying an observation X_o from the original sample of category i into category j . Therefore, the second term of the right side of (3.6) is

$$\sum_{j \neq i}^g |i - j|^\alpha P\{X_o \in R_j(x) : \hat{F}_i\} = \frac{1}{n_i} \sum_{t=1}^{n_i} \sum_{j \neq i}^g L_{ij}^\alpha(x_{it}) \quad (3.8)$$

the loss rate for category i .

4. Bootstrap Estimation of the Excess Loss

If the population distributions were known or more data were available, they could be employed to estimate the true loss rate and to obtain a better estimate. However, the assumption for the moment is that the population distributions are unknown, and the data available for the estimate are those for the original sample. Thus the bootstrap estimation method can be used to assess the statistical accuracy of the estimate given by the sample. The bootstrap procedures for the estimation of the expected excess loss are presented in four steps:

Step 1: An empirical probability distribution, \hat{F}_i , of the data for category i is constructed by putting the mass $1/n_i$ on x_{it} . Bootstrap samples are generated by independent random sampling with replacement from the empirical probability distributions. The bootstrap samples for category i can be denoted by

$$X_i^* = (X_{i1}^*, X_{i2}^*, \dots, X_{in_i}^*) \sim_{iid} \hat{F}_i, \quad i = 1, 2, \dots, g. \quad (4.1)$$

Let r_{it} be the resampling proportion of x_{it} selected in the bootstrap sample of category i and \hat{F}_i^* be the empirical probability distribution of the bootstrap sample of category i which putting mass r_{it} on x_{it} .

Step 2: Given the bootstrap samples, the bootstrap classification functions are constructed.

Step 3: A bootstrap estimate of the excess loss is calculated. The bootstrap estimate of the excess loss for category i is

$$\begin{aligned} B^*(R_i^\alpha) &= TLR_i^{*\alpha} - LR_i^{*\alpha} \\ &= \sum_{j \neq i}^g TLR_{ij}^{*\alpha} - \sum_{j \neq i}^g LR_{ij}^{*\alpha} \\ &= \sum_{j \neq i}^g E\{L_{ij}^\alpha(X_o | X^*) : \hat{F}_i\} - \sum_{j \neq i}^g E\{L_{ij}^\alpha(X_o | X^*) : \hat{F}_i^*\} \\ &= \sum_{j \neq i}^g |i - j|^\alpha P\{X_o \in R_j(X^*) : \hat{F}_i\} \\ &\quad - \sum_{j \neq i}^g |i - j|^\alpha P\{X_o \in R_j(X^*) : \hat{F}_i^*\}, \end{aligned} \quad (4.2)$$

and the bootstrap estimate of the overall excess loss is

$$B^*(R^a) = \sum_{i=1}^g p_i B^*(R_i^a). \quad (4.3)$$

The first term of the right side of (4.2) is the expectation of the loss function over an X_o from \hat{F}_i , given the entire bootstrap sample. This is equal to the loss rate made in the original sample of category i by the bootstrap classification functions:

$$\sum_{j \neq i}^g |i-j|^a P\{X_o \in R_j(X^*): \hat{F}_i\} = \frac{1}{n_i} \sum_{t=1}^{n_i} \sum_{j \neq i}^g L_{ij}^a(x_{it}|X^*) \quad (4.4)$$

The second term of the right side of (4.2) indicates the expectation of the loss rate over an X_o from \hat{F}_i^* , given the entire bootstrap sample. This is the loss rate made in the bootstrap sample of category i by the bootstrap classification functions:

$$\sum_{j \neq i}^g |i-j|^a P\{X_o \in R_j(X^*): \hat{F}_i^*\} = \sum_{t=1}^{n_i} \sum_{j \neq i}^g L_{ij}^a(x_{it}|X^*) r_{it}. \quad (4.5)$$

Therefore, a bootstrap estimate of the excess loss for category i can be rewritten as

$$B^*(R_i^a) = \sum_{t=1}^{n_i} \sum_{j \neq i}^g L_{ij}^a(x_{it}|X^*) \left(\frac{1}{n_i} - r_{it} \right). \quad (4.6)$$

Step 4: The steps from 1 to 3 are repeated some large number $NBOOT$ times, obtaining independent bootstrap replication. A bootstrap estimate of expected excess loss for category i is

$$\bar{B}(R_i^a) = \frac{1}{NBOOT} \sum_{b=1}^{NBOOT} B_b^*(R_i^a) \quad (4.7)$$

approximately by the averaging $NBOOT$ replications. The bootstrap estimate of expected overall excess loss is

$$\bar{B}(R^a) = \sum_{i=1}^g p_i \bar{B}(R_i^a). \quad (4.8)$$

The bootstrap estimate of the expected excess loss $\bar{B}(R^a)$ can be used as an estimate of

the expectation of the true excess loss $E(R^a)$. Thus, the bootstrap estimate of the true loss rate can be calculated by adding the bootstrap estimate of the expected excess loss to the loss rate as defined by (3.2). The bootstrap estimate of the true loss rate for category i is

$$B(TLR_i^a) = LR_i^a + \overline{B}(R_i^a) \quad (4.9)$$

and the bootstrap estimate of the true overall loss rate is

$$B(TLR^a) = \sum_{i=1}^g p_i B(TLR_i^a). \quad (4.10)$$

When $a=0$, $B(TLR^0)$ is called a bootstrap estimate of the true overall constant loss rate, and when $a=1$, $B(TLR^1)$ is called a bootstrap estimate of the true overall absolute loss rate. Efron (1979), Singh (1981), and Bickel and Freedman (1981) showed that as few as 100 bootstrap replications may be required to get a reliable estimate of the bias. Based on their theory, the bootstrap estimate of the true loss rate converges in probability to the true loss rate as the sample size and the number of replications approach infinity:

$$B(TLR_i^a) \xrightarrow{p} TLR_i^a \quad \text{as } n_i \text{ and } NBOOT \rightarrow \infty. \quad (4.11)$$

These conditions also imply

$$B(TLR^a) \xrightarrow{p} TLR^a. \quad (4.12)$$

5. Evaluating Predictive Ability of Bond Rating Classification

Bond ratings are based, in part, on available statistics depicting a firm's operating and financial conditions. In addition to quantifiable data, the rater's qualitative judgement concerning the future ability of a firm to make the scheduled interest and principal or sinking fund payment in time also influences the bond ratings. Under the commercial bank regulations issued by the Controller of the Currency, bonds rated in the top four categories by Moody's- Aaa, Aa, A, Baa- generally considered eligible for bank investment. The Baa rating, bordering between investment and speculative categories, is the lowest which qualifies for investment. Therefore, the classification of bonds into investment and speculative categories as well as the classification of bonds into individual rating categories with a limited number of independent variables is of interest.

One way of evaluating a model's performance is to calculate the apparent error rate. However, if the categories can be ordered or ranked, the use of a loss rate is more

meaningful for the measuring the misclassification rate. Since the loss rate is typically biased like the apparent error rate, the bootstrap method can be employed in order to obtain an approximately unbiased and stable estimate of the true loss rate. In this study, the predictive ability of five possible models of the bond ratings classifications are compared: (1) a linear classification model (LCM), (2) a quadratic classification model (QCM), (3) a rank transformation linear classification model (RLCM), (4) a rank transformation quadratic classification model (RQCM), (5) logit model (LTM).

In order to obtain the sample, the data was collected on 130 outstanding industrial corporate bonds for the top six ratings from Moody's industrial manual in 1987 (source: 응용다변량분석 p.268). Industrial corporate bonds include those issued by manufactures, retails, and the like. Utilities, banks, other financial firms and transportation companies were eliminated because they had different financial measures. Table 1 presents the distributions of issues by rating for both population and sample and shows sample proportion and prior probabilities associated with their ratings.

Table 1 : Distribution of Bonds by Rating

Rating		Aaa	Aa	A	Baa	Ba	B	Total
Sample	size	8	20	35	22	18	27	130
	%	6.2	15.4	26.9	16.9	13.8	20.8	100
Population	size	16	41	120	80	44	108	409
	%	3.9	10.0	29.3	19.6	10.8	26.4	100

Bond ratings have been proved in previous research to be based, in part, on several variables depicting a firm's operating and financial conditions. Since the purpose is to compare and evaluate predictive classification models and not to discover new rating determinants, the selection of the independent variables were based on the previous studies. Therefore, the following proxies for the selected independent variables of the firms were obtained from Moody's Industrial Manual, Moody's Handbook of Common Stocks, and Standard and Poor's Stock Report index: (1) X_1 : size variable = total asset in 100 million dollars, (2) X_2 : leverage variable = long term debt / total capital, (3) X_3 : profitability variable = net income / total assets, (4) X_4 : instability variable = coefficient of variation of net income, and (5) X_5 : stock ranking(six point scale). Financial ratios from X_1 - X_3 were computed using a five-year sample mean of the annual ratios from 1982 to 1986. The sample mean and standard deviation of the instability variable were also computed from 1982 to 1986, and the stock rankings in 1987 were obtained. The assumption of multivariate normality was examined using the chi-square probability plot. After the natural log transformation were applied to size and instability variable, a reasonable approximation of the multivariate normality was obtained.

5.1 Empirical Results in Binary Classification

The purpose of the section is to analyze binary classification into investment and speculative categories with the use of five classification models with prior probabilities

in Table 1, including the loss rate in the original sample and the bootstrap estimate of the true loss rate for future sample with 100 bootstrap replications. The constant loss rates (same as apparent error rates) and the estimate of the true loss rates obtained from five classification models.

Table 2: Constant and bootstrap estimate of true loss rate

Category	LCM	QCM	RLCM	RQCM	LTM
Investment	8.2 % (8.3)	2.4 % (3.6)	11.8 % (13.7)	8.2 % (8.2)	5.9 % (5.9)
Speculative	8.9 (12.6)	46.7 (50.3)	2.2 (4.4)	8.9 (10.1)	4.4 (8.0)
Overall	8.5 (9.9)	18.9 (21.0)	8.2 (10.2)	8.5 (8.9)	5.3 (6.7)

note: () = bootstrap estimate of true loss rate.

The LTM yields an overall constant loss rate of 5.3 % which is the smallest overall constant loss rate compared with others classification models. An estimate of the true overall constant loss rate of the LTM was also superior to the results obtained from the other classification models. One interesting point from Table 2 is the pattern of constant loss rates. There is a serious variation in the constant loss rates between categories. The QCM and RLCM yielded a much lower constant loss rate for one category and a much higher constant loss rate for the other category, while the LCM, RQCM, and LTM produced similar constant loss rates for both categories.

5.2 Empirical Results in Ordered Multiple Classification

Ordered multiple classification models should be evaluated with loss rate because a model can produce high correct classification rates, but that the spread of the misclassification may be serious. Therefore, the absolute loss rate as well as the constant loss rate should be examined for a comparison of predictive power of classification models. The summary of the constant loss rates and the absolute loss rates are presented in Table 3.

The classification results are generally less satisfactory than those of the binary classification in Table 2. All classification models classified most Aaa, A, and B bond correctly and had trouble with Ba bonds. The QCM produced a better result by reducing the

constant overall loss rate of the LCM by 3.9%, but its spread of misclassification was wider than the LCM. The RLCM yielded an overall constant loss rate of 29.5% which is slightly better than the LCM and its spread of misclassification was relatively narrow. The RQCM results were better than the RLCM, but the spread of misclassification for certain category was serious. For example, the difference between the constant loss rate and the absolute loss rate for category A were surprisingly 11.4%. Conover and Iman's (1980) simulation studies have shown that the RLCM is likely to be better than the LCM, but not as good as the RQCM when the data are non-normal. The results of this study support their findings. The LTM resulted in the lowest constant overall loss rate of 24.2 % and the lowest absolute loss rate of 25.2 % when compared with other classification models in the data we examined. Therefore, the LTM seems to be more successful than the other classification models and can be used as a prediction model because of its smaller constant loss rate and narrower spread of the misclassification.

To evaluate classification models' prediction ability for future bonds, 100 bootstrap replications were used to calculate the bootstrap estimate of the true loss rate. The bootstrap estimate of the true constant loss rate and the bootstrap true loss rate based on 100 bootstrap replications are presented in Table 4. Comparing Table 3 and 4, we can see that both the constant overall loss rate and the absolute overall loss rate are biased optimistically. The bootstrap estimate of expected excess loss for each category can be obtained by the difference between values of Table 3 and Table 4. For example, the bootstrap estimate of expected excess absolute loss of category A for LCM model was 4.1% by subtracting 25.7% from 29.8%. The linear types of classification models tended to produce a smaller difference between the loss rate and the true loss rate when compared with the quadratic types of the classification models.

The LCM produced a constant overall loss rate of 36%, on the average, on future data compared with 30.6% on the original data. The QCM produced the constant overall loss rate of 26.7% on the original data but yielded an estimate of the true constant overall loss rate of 37% which was even higher than the LCM. The RLCM produced almost same predictive ability as LCM and yielded a better prediction ability than the RQCM. The RQCM performed poorly in prediction, yielding an estimate of the true constant overall loss rate of 39% and an estimate of the true absolute overall loss rate of 49.4%. The LTM produced the best predictive ability and gave an estimate of the true constant overall loss rate of 30.1% and an estimate of the true absolute overall loss rate of 32.2%.

Table 3: Constant and Absolute Loss Rates

Category	loss rate	LCM	QCM	RLCM	RQCM	LTM
Aaa	constant	0 %	0 %	0 %	0 %	0 %
	absolute	0	0	0	0	0
Aa	constant	45	45	35	20	45
	absolute	45	50	35	25	45
A	constant	25.7	11.4	25.7	25.7	11.4
	absolute	25.7	11.4	25.7	37.1	11.4
Baa	constant	31.8	27.3	36.4	45.5	31.8
	absolute	36.4	36.4	40.9	54.5	31.8
Ba	constant	77.8	88.9	77.8	50	66.7
	absolute	83.3	105.6	83.3	50	66.7
B	constant	14.8	14.8	11.1	11.1	11.1
	absolute	22.2	25.9	14.8	11.1	14.8
overall	constant	30.6	26.7	29.5	25.0	24.2
	absolute	34.0	33.7	32.0	29.0	25.2

Table 4: Bootstrap Estimates of True Loss Rates

Category	Bootstrap	LCM	QCM	RLCM	RQCM	LTM
Aaa	constant	5.3 %	13.3 %	8.1 %	17.3 %	5.9 %
	absolute	5.3	20.8	8.1	24.4	2.4
Aa	constant	49.8	57.7	42.9	32.4	52.9
	absolute	49.8	64.5	42.9	38.3	52.9
A	constant	29.8	20.0	31.1	34.1	16.8
	absolute	29.8	20.7	31.6	48.2	17.3
Baa	constant	39.0	42.4	44.6	59.2	40
	absolute	45.3	59.1	51.0	72.8	41.9
Ba	constant	84	97.3	82.8	66.4	71.7
	absolute	90.3	114.6	88.5	69.8	71.9
B	constant	20.5	22.7	17.0	21.9	17.0
	absolute	31.7	38.9	24.0	30.8	24.0
overall	constant	36.0	37.0	35.9	39.0	30.1
	absolute	40.9	47.5	39.8	49.4	32.2

6. Conclusions

After developing classification rules, the natural next step is to evaluate the performance of the classification models. It is usually assumed that the loss of misclassification of one category to other category is equal. This may be inappropriate in practice when dealing with ordered categories. Therefore, the loss function should be employed to deal with different

losses of misclassification as well as the absolute loss rate should be examined. The main point of this study is the estimation of the true loss rate of classification models when the sample size is not large. To obtain a more accurate estimate of the true loss rate, the bootstrap method was used.

Based on the results of predictive power of classification models, the linear types of classification models were found to be more useful for the determination of bond ratings than the quadratic types of classification models. In case of ordered multiple categories, the quadratic types of classification models were not appropriate in predicting bond ratings for future samples because of possible wider spread of the prediction. The logit model clearly dominated all classification models in terms of the classification results of the original sample and prediction ability for future samples. It produced the narrowest spread of misclassification.

Reference

- [1] Conover, W.J. and Iman, R.L. (1980), The Rank Transformation as a Method of Discrimination with Some Examples, *Communications in Statistics -Theory and Method*, A9(5), 456-487
- [2] Bickel, P.B. and Freedman, D.A. (1981), Some Asymptotic Theory for the Bootstrap, *The Annals of Statistics*, 9, 1196-1217
- [3] Efron, B. (1979), Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics*, 7, 1-26
- [4] Efron, B. (1983), Estimating the Error Rate of a Prediction Rule : Improvement on Cross-Validation, *Journal of the American Statistical Association*, 78, 316-331
- [5] Efron, B. and Gong, G. (1983), A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation, *The American Statistician*, 37, 36-48
- [6] Kalplan, R. and Urwitz, G. (1979), Statistical Models of Bond Ratings: A Methodological Inquiry, *Journal of Business*, 52, 231-261
- [7] Pinches, G. and Mingo, K. (1975), The Role of Subordination and Industrial Bond Ratings, *Journal of Finance*, 30, 201-206
- [8] Pogue, T. and Soldofsky, R. (1969), What's In A Bond Rating ?, *Journal of Financial and Quantitative Analysis*, 4, 201-228
- [9] Singh, K. (1981), On The Asymptotic Accuracy of Efron's Bootstrap, *The Annals of Statistics*, 9, 1187-1195
- [10] 성웅현 (1997), 응용다변량분석, 탐진