ECONOMIC RESEARCH FORUM
منتدى البحوث الاقتصادية

2012

working paper series

PREDICTING BANK FINANCIAL STRENGTH RATINGS IN AN EMERGING ECONOMY: THE CASE OF TURKEY

Hulisi Öğüt, M. Mete Doğanay,
Nildağ Başak Ceylan, Ramazan Aktaş

Working Paper No. 740

PREDICTING BANK FINANCIAL STRENGTH RATINGS IN AN EMERGING ECONOMY: THE CASE OF TURKEY

Hulisi Öğüt, M. Mete Doğanay, Nildağ Başak Ceylan, Ramazan Aktaş

Working Paper 740

December 2012

Send correspondence to:
Hulisi Öğüt
Department of Business Administration, TOBB University of Economics
hogut@etu.edu.tr

First published in 2012 by
The Economic Research Forum (ERF)
21 Al-Sad Al-Aaly Street
Dokki, Giza
Egypt
www.erf.org.eg

Abstract

The recent global financial crisis shows us that the rating of bank's financial strength can be very misleading. As the credibility of the credit rating agencies has been shaken, the objectivity of the credit rating agencies has been questioned. Based on this observation, we investigate whether the forecast of the rating of bank's financial strength using the publicly available data is consistent with those of the credit rating agency. The data of Turkish banks is used for this investigation. Furthermore, we identify the variables that play an important role in assigning these ratings. For this purpose, we used quantitative proxies for some qualitative factors that are used by Moody's. The important factors in these ratings are profitability (measured by return on equity), efficient use of resources, and funding of businesses and households instead of government.

JEL Classification: G2, G3

Keywords: Rating Agencies; Bank financial strength rating; financial and operational ratios; rating prediction; multivariate statistical model; data mining technique

ملخص

الأزمة المالية العالمية الأخيرة يبين لنا أن تصنيف القوة المالية لأى بنك يمكن أن تكون مضللة جدا. وكما اهتزت مصداقية وكالات التصنيف الائتماني، تم التشكيك في موضوعية وكالات التصنيف الائتماني. وبناء على هذه الملاحظة، نقوم بالتحقيق في ما إذا كانت توقعات التصنيف للقوة المالية لأى بنك باستخدام البيانات المتاحة للجمهور تتسق مع تلك من وكالات التصنيف الائتماني. وتستخدم البيانات من البنوك التركية لهذا التحقيق. وعلاوة على ذلك، فإننا نقوم بتحديد المتغيرات التي تلعب دورا هاما في تحديد هذه التصنيفات. لهذا الغرض، استخدمنا التقدير الكمي لبعض العوامل النوعية التي يتم استخدامها من قبل وكالة موديز. العوامل الهامة في هذه الدرجات تكون الربحية (يقاس العائد على حقوق المساهمين) والكفاءة في استخدام الموارد، والتمويل من الشركات والأسر بدلا من الحكومة.

1. Introduction

Bank financial strength ratings are conducted by rating agencies such as Moody's and it defined as "[the] Moody's opinion of a bank's intrinsic safety and soundness" (Moody's 2006). Unlike other types of rating, Moody's states that bank financial strength rating measures a bank's ability to avoid default rather than measuring the ability of a bank to make timely payments. In other words, bank financial strength ratings provide information about the financial strength/weakness of a bank.

Moody's gives these ratings by combining letters between A and E, and (+) (-) signs. For example, C+ corresponds to a rating. Moody's takes into consideration some quantitative and qualitative factors when it determines these ratings. These factors are grouped into five broad categories: franchise value, risk positioning, regulatory environment, operating environment, and financial fundamentals. Some of these factors are general factors, which apply to all banks within an environment such as a country or a region whereas others are specific factors, which apply to individual banks. Franchise value is defined by Moody's as "the solidarity of a bank's market standing in a given geographical market or business niche". Franchise value encompasses sub-factors such as market share and sustainability, geographical diversification, earnings stability, earnings diversification, and vulnerability to event risk (risk that an event can destroy a bank's franchise value). Risk positioning is a measure of a bank's attitude towards risk and its ability to manage risk. This factor encompasses sub-factors such as corporate governance, controls, financial reporting transparency, credit risk concentration, liquidity management, and market risk appetite. Regulatory environment and operating environment are general factors and they are not related to individual banks. These two factors define the environment in which the bank is operating. Financial fundamentals encompass sub-factors such as profitability, liquidity, capital adequacy, efficiency, and asset quality. Moody's assesses the sub-factors and factors and assigns a rating to a bank according to a score based on the assessments.

Bank financial strength ratings have gained widespread popularity especially after the recent financial turmoil. Rating agencies were criticized because of their ratings and failure to predict the bankruptcy of the banks. Being motivated by these developments and the scarcity of studies related to bank financial strength ratings in the literature, our aim in this paper is to develop models to determine the significant factors that have an impact on bank financial strength ratings. Rather than developing alternative methodologies used by the rating agencies, our purpose in this paper is to determine the model that predicts the bank financial strength ratings best and the factors that are important in determining the financial strength ratings of banks. For this purpose, we used quantitative proxies for some qualitative factors that are used by Moody's. Because of this approach, our study differs from other studies that used only accounting and financial data. In addition, environmental factors can also be important in determining the ratings. But it is very hard to quantify environmental factors and the rater's judgment plays an important role in these factors. Other studies that used proxies for environmental factors found that these proxies did not have any explanatory power. For this reason, we did not consider environmental factors in our research. Furthermore, we can also obtain data for Turkish banks related to the proxies other than financial and accounting ones. Thus, we restricted our sample to only Turkish banks operating in the same economic and political environment.

The paper proceeds as follows: The second section provides a brief literature review. We explain the methodologies in the third section. The data are explained in the fourth section. Section five presents and discusses the empirical results of the models. The paper is concluded in Section six.

2. Related Literature

Ratings can be viewed as a classification problem as the cases (banks, firms, governments, etc.) are grouped based on their ratings. Classification models have long been applied to financial problems such as financial failure, audit reports, financial information manipulation, stock price manipulation, etc. These models have also been developed to predict classification or used to understand the determinants of the classifications. In 1995, Moody's started bank financial strength ratings. Poon et al. (1999) performed a pioneering study on this subject. They developed an ordered multiple logistic regression model to predict bank financial strength ratings. The data are gathered from 130 banks and more than 30 countries. Bank specific financial data and ratios that covered profitability, asset management and risk measures are used as explanatory variables. Poon et al. performed a factor analysis in order to reduce the number of variables by grouping them into factors. They also used an aggregate measure (between 0 and 100) representing political, economic, and financial risks of the country in which the bank was operating. This measure was obtained from the International Country Risk Guide. Short-term and long-term debt ratings of the banks were two other explanatory variables that were used in the model. The analyses showed that loan provision was the most important factor for the explanation of bank financial strength rating, followed by risk, and then profitability. Country risk measure was not a significant factor explaining the ratings. It was also found that models that included short-term and long-term debt ratings had better predictive powers. Boyacıoğlu and Kara's paper (2007) also predicted Moody's bank financial strength ratings. They used a binary dependent variable in their model using data from Turkish Banks. Ratings covering the period 2001-2005 were used. Independent variables were 20 bank specific financial ratios grouped by factor analysis. Models were developed for discriminant analysis, logistic regression, and neural networks. In the holdout sample they did not find any significant difference in the prediction power of the models.

In the 2000s different types of ratings were predicted and different models were used to predict these ratings. Bennell et al. (2006) developed artificial neural network and ordered probit models to predict sovereign credit ratings. They used macroeconomic indicators as explanatory variables and found that an artificial neural network model was superior to an ordered probit. Credit ratings attracted the attention of the researches after the publication of Basel Accords. Researchers developed different models to predict credit ratings of the companies. Doumpos and Pasiouras (2005) developed a multi-criteria classification model (a value function technique named UTADIS) to predict the ratings assigned by a regional agency in the UK. They used financial ratios as the evaluation criteria. They performed hold out sample tests by using out-of-sample and out-of-time data (firms other then the ones in model development and for a different time period). Kumar and Bhattacharya (2006) also attempted to predict credit ratings assigned by Moody's. They also used financial ratios as classification variables and developed a full connected and back-propagation artificial neural network model with three layers of neurons, one each for input (financial ratios), output (ratings), and a hidden layer. They state that they did not use a multiple hidden layer model because of the simplicity of the model. They used an appropriate portion of companies for training and another portion for testing. Researchers widely used artificial intelligence methods in predicting different types of ratings in the 2000s. Huang et al. (2004) used a back propagation neural network method and support vector machines. Support vector machines are a learning machine technique that automatically extracts knowledge from a data set, to predict credit ratings. Kim (2005) used the adaptive learning network, which is an artificial intelligence technique, to predict bond ratings by using publicly available information. Cao et al. (2006) used support vector machines in predicting the ratings of the bonds issued in the USA. Authors compared the classification accuracy of vector support machines method with those of the neural networks, ordered probit, and the logistics regression. Analyses showed that support vector machines had the best performance in predicting bond ratings.

3. Methodology

3.1 Logistic regression

Logistic regression (logit) model is used for estimating the probability and group membership of independent variables by making a logistic transformation of a linear combination of dependent variables. In order to find the parameters of the logistic function, logarithmic sums of predicted probabilities is minimized. Logistic regression is only suitable when the dependent variable takes binary outcome. When there are more than binary outcomes and there are ordered relationships among dependent variables, an ordered logit model should be used. This is the case in our paper since a higher ratings number given by credit rating firms indicates the improved financial status of a bank. In an ordered logit model, cumulative probabilities of class membership are used to derive the non-cumulative probabilities of class membership. After calculating the probability of class membership, the instance is assigned into the class having the highest probability. For an n-type ordered categorical variable, the non-cumulative probabilities of class membership is defined as

$$P(Y=1)=P(Y\le 1)$$
$$P(Y=2)=P(Y\le 2)-P(Y\le 1)$$
$$\vdots$$
$$P(Y=n)=1-P(Y\le n-1)$$

where $P(Y\le i)=\dfrac{1}{1+e^{-(c_i-(a_1x_1+a_2x_2+\ldots+a_nx_n))}}$, $a_1,a_2\ldots a_n$ are the parameters and $x_1,x_2\ldots x_n$ are the inputs. Note that for each *i*, c_i is different however $a_1,a_2\ldots a_n$ are the same.

3.2 Multiple Discriminate Analysis (MDA)

MDA is a method for combining independent variables in linear forms for the purpose of classification. For this aim, independent variables are grouped based on the values of dependent variables and they are tested whether there are significant differences in terms of their means. Specifically, the test of the ratio of the between-groups variance to within-group variance is computed in order to identify best discriminators among these independent variables. If this ratio is high, it is concluded that group means of the variables are significantly different. Then, using a linear combination of the best discriminators of independent variables generates discriminate functions for each group. These functions maximize degree of separation between two groups. MDA assumes that the independent variables come from a multivariate normal distribution and covariance/variance matrices for every group are homogenous.

4. Data

For our analysis, we collected 26 ratios of the banks as independent variables and their financial strength ratings as dependent variables from 2003 to 2009. Note that all ratios are not financial as some of them are proxies of qualitative data. These ratios are given in Table 1. Descriptive statistics of these ratios are presented in Table A.1. Since we have too many ratios as explanatory variables, we perform principal component analysis (PCA) on the data. Six factors having eigenvalue score greater than 1 have been determined. Descriptive statistics of these factors and the eigenvalue scores of them are presented in Table A. and Table A.2 respectively. Then, we perform a varimax rotation technique in order to get rotated factor loadings. The principal component analysis for these rotated factor loadings is presented in Table A.3. We found that variables are grouped in the following factors: X18, X19, X20, X21 and X22 in the first factor; X9, X12, X13 and X14 in the second factor; X4, X16 and X17 in the third factor; X2, X7 and X24 in the fourth factor; X23 in the fifth factor and X3 and X15 in the sixth factor. We grouped these variables together if the correlation between variable and the factor is greater than $|0.65|$. When we examine the factors we see that variables grouped in the first factor are related to the franchise value. Variables grouped in the second factor are related to the profitability (return on equity) and how efficiently the bank used its resources. Variables grouped in the third factor are related to revenue structure and non-performing loans. Variables grouped in the fourth factor are related to capital adequacy and liquidity. The variable in the fifth factor is related to deposit concentration. Variables grouped in the sixth factor are related to asset structure, especially the ratio of loans in assets and the percentage of deposits as loans.

As there is an ordered relationship between the rating of the banks and these variables are categorical, we transformed these variables into numeric form. For this reason, we assigned the lowest rating in our dataset as 1. For the other ratings, we used an increment of 1 as the financial strength rating if the bank improves by one grade. The number of the ratings and their rating frequencies in each year are reported in Table 2. Moody's provided these ratings and we determine the factors affecting these ratings in this paper. We also divided the dataset into two equal parts: test and training data. In order to get a homogenous split, we divide the data in each year equally into two parts. Furthermore, we try to have a homogenous distribution of financial strength ratings in each year for training and test data. Note that the data belonging to the earlier periods concentrates on low ratings while the data belonging to the later periods concentrates on higher ratings. For this reason, we did not divide the data based on years (for example earlier years for training data later years for test data) since such a split makes training and test data more heterogeneous.

5. Results

We used SAS for the implementation of MDA and ordered logistic regression classifiers. We used the same set of training and test data in both datasets across techniques in order to compare the performances of the classifiers. Six attributes (factor scores) are used as input variables. The output attribute takes the value of 1 to 6 based on the rating of the banks.

5.1 Ordered Logistic Regression

The first multivariate statistical technique that we use is ordered logistics regression (logit) model. We used these six variables for the determination of the classes. Table 3 shows the regression output of ordered logistics. Standard errors are shown in parentheses and ** and * indicates at the 5% and 10% significance level respectively. We find that as the factor 2 score increases, the probability of getting higher rating increases. As there is a positive correlation between X9, X12, X13, X21, X14 and factor 2, we expect that an increase in the X9, X12, X13 and X14 scores also increase the factor 2. We also find that as the factor 6 score increases, the probability of getting a higher rating decreases. Thus, as there is negative correlation between X3, X15 and factor 6, we expect that increase in the X3 and X15 scores increase the probability of getting a higher rating. Since only two attributes are statistically different from 0 at the 5 % significance level, we choose these variables (Factor 2 and Factor 6) for the prediction of classes and find the accuracy rates of logistic classifier on the test data as 60.47% (26/43). We also provide confusion matrix of test data in Table 4. Note that cut variables in Table 3 represent the constant in the cumulative probability functions. For the sake of completeness, we also perform a probit analysis using our data as well. We found the accuracy rate of the probit classifier on test data as 60.47% (24/43). Since our estimation results are similar to that of ordered logistic regression, we did not report the regression output of the probit model in our paper.

5.2 Multiple Discriminate Analysis

The second multivariate statistical model used is multiple discriminate analysis (MDA). MDA generates n linear functions belonging to n classes (i.e. rating). The linear multiple discriminate model is presented below.

$$Z_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 ... + \beta_k X_k$$

In this model, Z_i represents the class score for ith classes, β_k represents the weight for input variable X_k. After N scores for each class are computed, instance is assigned to the class having the highest score. We used a stepwise discriminate method to select input variables in the prediction of the ratings. An stepwise discriminate method selects factors 2, 4 and 6 as the significant variables at the 5% level. The class weights for factors considered in the model are presented in Table 5. We find the accuracy of MDA classifier as 53.49% (23/43). We also provide confusion matrix for the test data in Table 6.

The interpretation of the linear discriminate score is not straightforward as the ordered logistic classifier. However, we can say that an increase in the factor 2 score increases the rating of the banks as the weight of factor 2 score increases consistently with the increase in the rating score. As there is positive correlation between X9, X12 X13 and X14 and factor 2, increases in these variables increase the probability of a higher rating. For the same reasoning, increase in factor 4 score increases the score for a higher rating except for the class 1 function. Since the correlation between X2 and Factor 4 is positive and X7, X24 and factor 4 is negative, a higher X2 variable and a lower X7 and X24 variable increases the rating of the banks. For factor 6, a higher score decreases the probability of a higher rating except for class 1 and 4 scores. As the correlation between X3, X15 and factor 6 is negative, we can say that higher values of these variables increase the rating of banks in general.

5.3 Discussion of the results

We discuss the results from two views, methodological and financial. When we compare classifier matrices in terms of hit rate for each rating, we can say that the number of correct classification for each rating is similar to each other although there are minor differences. This shows the robustness of the performance of our classifiers. Furthermore, we reported the classifiers and their performances on the test data in Table 7 for purpose of comparison. Based on the results of the analyses, we found that the ordered logistic regression have done a better job compared to other techniques in predicting rating of the banks as the performance of the ordered logistic regression in terms of total classification accuracy is superior to those of other techniques. A possible explanation for this result is that the ordinal relationship is taken into account in the logistic regression, while multiple discriminate analysis did not take into account the ordinal relationship and treated a dependent variable as any categorical data.

When we look at the results from a financial perspective, we see that an increase in the variables X3, X9, X12, X13, X14, and X15 increases the probability of getting higher rating. The variable X3 is total loans/total assets. As the amount of loans in the assets increase the rating of the bank increases. In Turkey, there was a special case for the banks prior to 2002 when major bank restructuring took place. Most of the banks placed their resources into government debt securities. Government debt securities' yields were very high and they did not have default risk. But the main function of the banks is to provide funds for households and businesses. After 2002, the yield of the government debt securities have decreased gradually and as a result banks decreased government debt securities portfolios and increased their loans. Our inference is that the rating agency perceives that as a bank increases its loan portfolio, it is acting more like a commercial bank and placing its funds more efficiently. As the yields of the government debt securities decrease, placing the funds as loans increases the revenue of the bank. The X9 variable is return to equity and it is the main profitability ratio.

The rating agency is assigning a higher rating to those banks whose return on equity (profitability) is higher. The variables X12, X13, X14 are efficiency ratios. When a bank uses its resources (human capital and other capital) efficiently, it receives a higher rating. The variable X15 is total loans/total deposits ratio. As more and more deposits, which are the main funds of Turkish banks, are placed as loans banks receive a higher rating. The explanation is the same as the one for X3 since the rating agency wants the banks to act like a commercial bank (fund households and businesses) instead of channeling the funds it acquired to the government.

6. Conclusion

The aim of this research is twofold. One of them is to forecast the ratings of the banks by using financial and operational variables; another one is to determine the variables that play an important role in assigning the ratings. For this purpose, we use two multivariate techniques (MDA and logit model). We found that the ordered logistic classifier has done a better job compared to others in forecasting the financial strength rating of the banks, as the performance of this method in test data in terms of total classification accuracy is superior to those of data mining techniques. We also find relevant input variables for the prediction of financial strength rating of the banks in ordered logistic regression, which is the best model. According to the results the most important factors are efficiency, profitability, and the proportion of loans in the assets. The rating agency assigns a higher rating to those banks that generate high net income for shareholders, use resources efficiently, and channel funds as loans to households and businesses. According to our inference rating agencies find it less profitable for banks to place a high proportion of their funds (mainly deposits) in government debt securities. The last result indicates that the rating of a bank is higher if its risk is shared with different groups. These results may guide banks in order to get a higher rating and improve in terms of financial strength. Prediction performance of the classifier techniques suggested that our predictions are consistent with those of Moody's financial strength rating in general.

In our paper, we only used data from Turkish banks since we can use proxies for efficiency and franchise value for these banks. Furthermore, although it is possible to find financial ratios for the banks all over the world, it is very difficult to find ratios such as net interest revenue/number of branches and net interest revenue/number of employees that are found to be important explanatory variables in our analyses. Also it is our inference that the rating agencies take into account country specific factors when they assign a rating. For this reason, we believe that country specific research provides more insight. For example, more banks financed the government instead of households and businesses in Turkey prior to 2002. According to our analysis, these types of banks cannot get higher ratings. We must note that we cannot find suitable proxies for some qualitative factors. So, the judgment of the raters also plays an important role in determining the ratings. Even with these restrictions, we believe that the performances of classifiers are quite high (up to 61% prediction accuracy) when we consider the dependent variable takes six different values. Thus, we can say that our predictions are consistent with those of Moody's financial strength rating in general.

References

Bennell, Julia A., Crabbe, D., Thomas, S., ap Gwilym, O. 2006. "Modeling Sovereign Credit ratings: Neural networks versus ordered probit." *Expert Systems with Applications* 30: 415-425 .

Boyacıoğlu, Melek Acar, Kara Y. 2007. "Türk Bankacılık Sektöründe Finansal Güç Derecelerinin Tahmininde Yapay Sinir Ağları ve Çok Değişkenli İstatistiksel Analiz Tekniklerinin Performanslarının Karşılaştırılması." *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* 22:197-217 (published in Turkish)

Cao, Lijuan, Guan, L.K., Jingqing, Z. 2006. "Bond rating using support vector machine." *Intelligent Data Analysis*10: 285-296.

Doumpos, Michael, Pasiouras, F. 2005. "Developing and Testing Models for Replicating Credit Ratings: A Multicriteria Approach." *Computational Economics* 25: 327-341.

Kim, Kee S. 2005. "Predicting bond ratings using publicly available information." *Expert Systems with Applications* 29: 75-81.

Kumar, Kuldeep, Bhattacharya, S. 2006. "Artificial neural network vs. linear discriminant analysis in credit ratings forecast: A comparative study of prediction performances." *Review of Accounting and Finance* 5:216-227.

Martin Linda .J., Henderson, G.V. 1983. "On bond ratings and pension obligations: A note." *Journal of Financial and Quantitative Analysis* 18: 463-470.

Moody's, Bank Financial Strength Ratings: Revised Methodology, 2006.

Poon, Winnie P.H, Firth, M., Fung, H-G. 1999. "A multivariate analysis of the determinants of Moody's bank financial strenght ratings." *Journal of International Financial Markets, Institutions and Money* 9:267-283.

Table 1: The Ratios Used as Explanatory Variables

Total equity/Total assets	X2
Total loans/Total assets	X3
Non-performing loans/Total loans	X4
Non-current assets/Total assets	X5
Liquid assets/Total assets	X6
Liquid assets in foreign currency/Total liabilities in foreign currency	X7
Net period income/Assets	X8
Net income/Equity	X9
Interest revenues/Interest expenses	X10
Total deposits/Total assets	X11
Net interest revenue (loss)/Number of branches	X12
Net interest revenue (loss)/Total assets	X13
Net interest revenue (loss)/Number of employees	X14
Total loans/Total deposits	X15
Net interest revenue / Total revenue from operations	X16
Non-interest revenue / Total assets	X17
Assets/Total assets of the sector	X18
Loans/Total loans of the sector	X19
Deposits/Total deposits of the sector	X20
Number of branches/Total branches of the sector	X21
Number of employees/Total number of employees of the sector	X22
Personal deposits/Total deposits	X23
Foreign branches/Total branches	X24
Specialized loans/Total loans	X25
Assets in foreign currency/Liabilities in foreign currency	X26

Table 2: The Number of Ratings and Their Frequencies

Year	Number of Rating	E (1)	E+ (2)	D-(3)	D(4)	D+ (5)	C- (6)
2003	11	2	2	1	1	5	
2004	13	2	2		2	7	
2005	12	1	2		2	7	
2006	15		3	2	2	8	
2007	13			1	2	5	5
2008	12				2	4	6
2009	10					4	6

Table 3: The Output of Ordered Logistic Regression

	-1	-2
Factor 1	-0.0165	
	-0.1495	
Factor 2	1.116**	1.097**
	-0.2918	-0.2756
Factor 3	-0.04684	
	-0.3536	
Factor 4	0.1843	
	-0.193	
Factor 5	-0.118	
	-0.2459	
Factor 6	-1.186**	-1.190**
	-0.3211	-0.3099
cut1	-4.324**	-4.101**
	-0.8903	-0.8079
cut2	-2.542**	-2.522**
	-0.5821	-0.5672
cut3	-2.022**	-2.030**
	-0.5179	-0.5036
cut4	-1.110**	-1.132**
	-0.4494	-0.4308
cut5	2.835**	2.863**
	-0.6972	-0.6934
N	43	43
lnL	-46.62	-47.41

Table 4: The Confusion Matrix of Logistic Classifier for Test Data

Actual\Predicted	1	2	3	4	5	6
1	1	0	0	0	1	0
2	0	3	0	0	1	0
3	1	0	0	0	1	0
4	0	1	0	0	5	0
5	1	0	0	0	16	2
6	0	0	0	0	4	6

Table 5: The Input Variables and Their Weight in MDA Model

Class/Variable	1	2	3	4	5	6
Constant	-6.65	-4.37	-3.85	-2.32	-0.72	-4.08
Factor 2	-1.29	-0.98	-0.57	-0.12	0.03	1.34
Factor 6	0.13	2.27	1.36	-0.30	-0.13	-2.03
Factor 4	-1.35	0.29	0.16	0.17	-0.01	-0.24

Table 6: The Confusion Matrix of MDA Classifier for Test Data

Actual\Predicted	1	2	3	4	5	6
1	0	1	0	0	1	0
2	0	2	0	0	2	0
3	0	1	0	0	1	0
4	0	1	0	0	5	0
5	0	2	0	0	15	2
6	0	0	0	0	4	6

Table 7: The Summary of Accuracy Rates of Classifiers

Classifier	Accuracy rates (%)
Ordered Logistic Regression	60.47
Multiple Discriminant Analysis	53.49

Appendices

Table A.1: The Descriptive Statistics of Variable

Variable	Mean	Median	Minimum	Maximum	Standard Deviation
Factor 1	0.0000	-1.2153	-2.8402	3.8931	2.1254
Factor 2	0.0000	-0.1363	-5.3089	3.7592	1.7356
Factor 3	0.0000	0.5263	-12.9510	3.4412	1.7531
Factor 4	0.0000	0.1487	-11.9400	1.2229	1.4228
Factor 5	0.0000	-0.0763	-2.7698	3.1677	1.2935
Factor 6	0.0000	-0.2815	-2.4330	3.0908	1.2212

Table A.2: The Principal Component Analysis of the Variables

Factors	Eigen Value	% explained	% cumulated
1	6,625131	26,50	26,50
2	5,884382	23,54	50,04
3	2,971074	11,88	61,92
4	2,698196	10,79	72,72
5	1,416465	5,67	78,38
6	1,066103	4,26	82,65
7	0,86506	3,46	86,11

Table A.3: The Factors Eigen Value Scores

Variable -	Factor 1 Correlation	Factor 2 Correlation	Factor 3 Correlation	Factor 4 Correlation	Factor 5 Correlation	Factor 6 Correlation
X2	0.1542	0.1859	0.2553	0.8009	-0.1569	-0.3267
X3	-0.521	0.3033	0.0575	0.1984	0.0873	-0.6928
X4	-0.0423	-0.0546	-0.8197	-0.3129	0.1147	0.1385
X5	0.3039	-0.5436	-0.2032	0.0067	-0.4408	-0.1695
X6	0.1728	-0.0518	-0.0925	-0.3601	-0.3622	0.5315
X7	-0.0949	-0.0516	0.2165	-0.9151	0.0378	0.1343
X8	0.0562	0.5111	-0.7212	0.3106	0.0405	-0.0577
X9	0.2422	0.8064	0.1129	-0.0402	0.1467	-0.1522
X10	-0.2671	0.5794	0.247	0.34	-0.2046	-0.1055
X11	0.1126	-0.2286	-0.0977	-0.4963	0.1893	0.637
X12	0.2185	0.8729	-0.0398	0.1131	0.0157	-0.2672
X13	-0.1284	0.7109	0.1659	0.3009	0.0799	0.0633
X14	0.2497	0.8708	-0.0865	0.1029	0.0549	-0.2203
X15	-0.4833	0.3106	0.0804	0.2073	0.0758	-0.7005
X16	-0.0811	0.4193	0.8156	-0.0084	0.1521	0.091
X17	-0.0639	-0.1202	-0.9337	-0.0765	-0.142	0.1263
X18	0.9673	0.1002	-0.0045	0.0182	0.0452	0.1693
X19	0.914	0.0809	-0.0402	-0.0004	-0.2633	-0.1638
X20	0.9369	0.0799	0.0054	0.0121	0.1365	0.2785
X21	0.8949	0.0905	0.0083	0.0509	0.2393	0.2421
X22	0.9362	0.0676	-0.0292	0.0221	0.1779	0.1534
X23	0.1436	0.152	0.0245	-0.0568	0.8622	-0.1148
X24	0.0183	-0.2489	-0.1895	-0.6578	-0.1483	0.0957
X25	0.3564	-0.0154	0.0693	-0.0075	0.5988	0.5972
X26	0.6036	-0.0811	0.3835	0.4084	-0.2455	0.1222
Variance Explained	5.7834	4.17	3.2116	2.9749	1.8951	2.6265