International Business Research; Vol. 6, No. 10; 2013
ISSN 1913-9004 E-ISSN 1913-9012
Published by Canadian Center of Science and Education

Credit Scoring Model for Iranian Banking Customers and Forecasting Creditworthiness of Borrowers

Mojtaba Dastoori[1] & Samira Mansouri[1]

[1] Financial Management, Department of management and insurance, Faculty of management, University of Tehran, Iran

Correspondence: Mojtaba Dastoori, Department of management and insurance, Faculty of management, University of Tehran, Iran. E-mail: dastoori@ut.ac.ir

Received: June 3, 2013 Accepted: August 2, 2013 Online Published: September 23, 2013

doi:10.5539/ibr.v6n10p25 URL: http://dx.doi.org/10.5539/ibr.v6n10p25

Abstract

Credit risk is one of important challenges facing banks and credit institutions in all economic systems. Accordingly, much research has focused on credit scoring of bank customers and to predict and classify solvent customers and insolvent customers. In this line, the present study has attempted to identify and employ financial ratios affecting the creditworthiness of banking customers using discriminant analysis and logistic regression to determine the reliability of different credit scoring models in predicting creditworthiness of banking customers. To do so, customers' credit files in one of the branches of commercial Iranian banks in Province were investigated and totally 54 creditable firms and 46 non-creditable firms were identified. The creditworthiness of these firms was determined through discriminant analysis and logic regression using financial information obtained from the sample firms from 2006 to 2011. The results of the study indicated that the two fitted models are reasonable reliable in predicting the creditworthiness of banking customers. However, the logic regression model had a higher discrimination power than the discriminant analysis model.

Keywords: credit risk, credit scoring, discriminant analysis, logistic regression, feature, sensitivity curve

1. Introduction

Clearly, borrowing/leverage has been one of the most important options to finance firms and much of financial resources needed by firms are provided through borrowing and leverage especially in countries with an underdeveloped capital market. Parallel to the economic development of different countries throughout the world and the emergence of financial businesses and their rapid expansion, the role of commercial banks play a vital role in absorbing micro resources and allocating them to financial businesses in the form of loans and banking credits.

As a result, loan portfolio seems to be considered as a major share of commercial banks assets. Commercial banks are mainly concerned with absorbing investors' deposits and granting them to the applicants in the form of loans. Consequently, banks are required to repay the original deposits and their interests to the depositors on the due date. However, these banks may be exposed to the borrowers' defaults. Accordingly, an examination of the creditworthiness of applicants to determine whether to grant them loans or not is of special importance.

When making investments or granting loans to their customers, banks and financial institutions have to take certain risks that will affect the repayment of the loans. Factors such as technological changes, changes made in the customers' types and their tastes, changes in the exchange rate, and the management replacement may put a part of banking resources at risk by increasing the risks associated with granting loans and credits (Saunders & Allen, 2009)

Banks, in their lifetime, face different risks such as liquidity risk, credit, commercial, and financial risks and inability to pay the exchange rate, the interest rate, and inflation; among which credit risk is of special importance since it is associated with the first and foremost role of banks in financial markets that is to absorb deposits and to grant loans. The possibility that the borrower is not able to repay his loans is termed credit risk or non-repayment risk (Sinky & Joseph, 1992). However, based on another definition under Basel Committee on Banking Supervision (2000), credit risk simply refers to the possibility that the borrower form the bank or its beneficiary is not be able to perform his debt obligations towards the bank in a given period. Accordingly, the

Basel Committee explains a default event on a debt obligation in the two following ways:

- It is unlikely that the obligor will be able to repay its debt to the bank without giving up any pledged collateral;
- The obligor is more than 90 days past due on a material credit obligation.

The recent financial crises have clearly shown that the effects of financial marks across the world are not limited to geographical boundaries and the poor performance of a business in a country of the world may destructively affect the financial markets in other countries that are located far away from that business. Therefore, the inability of a financial institute to collect its debt obligations can seriously threat the health and the life of the institution and, consequently, put pressure not only on a country's economy but also on other global markets. Therefore, to reduce the credit risk and the non-repayment of obligations, banks employ different methods to evaluate potential credit borrowers before offering credits to them (Jorge & Shuri, 1988).

Generally, Credit Scoring is the set of decision models and their underlying techniques that aid lenders in the granting of consumer credit. These techniques decide who will get credit, how much credit they should get, and what operational strategies will enhance the profitability of the borrower to the lenders (Thomas & Edelman, 2002).

A creditor can make revenues when they successfully predict the creditworthiness and default risk of applicants depending on the default predictor factors. Credit scoring is a proper technique that connects these factors to the probability of default (Lieli & White, 2010).

However, before offering credit to companies, their financial position needed to be examined as offering loan is very risky. On the basis of the financial position of their applicants requesting credit, banks assign credit scoring and on the basis of credit scores the bank decides whether to offer the credit to these applicants and also decides the credit limits (Samreen & Zaidi, 2012).

Kiss France (2003) has divided techniques used to measure credit risks into two major categories:

1) Parametric credit scoring models: Linear Probability Model, Logit Model, Probit Model, and Discriminant analysis models.

2) Nonparametric credit scoring models: Mathematical Planning, Classification trees (recursive partitioning algorithms), Nearest Neighbors Models, Analytical Hierarchy Process, Expert Systems, Artificial Neural Networks, and Genetic Algorithms.

Based what was mentioned, since the borrowers' default due to their unwillingness or inability to repay their debts is regarded as a serious challenge for banks and financial institutes, the present study attempts to introduce a credit scoring model used to assess the creditworthiness of banking customers using logic regression and discriminant analysis. Besides, the study aims to find a suitable model for the assessment of the creditworthiness of banking customers by comparing the two models.

2. Review of Literature

Nemours studies have been performed to predict the financial health of companies and their credit positions. Univariate statistical techniques were among the first methods used to predict the companies' insolvency. One of the oldest financial ratios that were used in 1870 to evaluate the financial position of companies was the Current Ratio. Besides, John Mori introduced a model for the first time in 1909 to evaluate and rank the credit risk of bonds.

Charles Mervin (1942) examined financial rations for a six month period for bankrupt and non-bankrupt companies. He introduced three ratios of working capital to total assets, equity to total liabilities, and current ratio as useful variables for predicting the financial positions of companies, among which the ratios of working capital to total assets was determined as the best predictors of companies bankruptcy.

William Beaver (1966) used Univariate Logistic Regression Model to predict financial distress. He mostly employed those ratios that were associated with cash flows. He defined financial distress as the companies' failure to perform their financial obligations. In addition, Beaver (1966) chose 30 financial ratios that he regarded them as the best indicators of a company's financial health and then divided them into six groups: ratios related to cash flow, the ratio of liabilities to total assets, ratio of liquid assets to total assets, ratio of liquid assets to current liabilities, and the ratios of turnover, and the ratios of net profit. This model was used later on to measure the credit risk of bonds issued by companies. Edward Altman used different combinations of financial ratios to predict companies' financial distress. In his study under the title of "The assessment of the success of the U.S. companies and production units using discriminant analysis", Altman developed a model using five important financial ratios known as *Z-Score* which is still used as an indicator to assess the companies' financial position.

Altman (1968) believed that his model could be used to discriminant bankrupt companies from non-bankrupt companies and also to evaluate commercial loans offer, internal control processes, and investment options.

Using 14 financial ratios employed in Beaver's studies and discriminant analysis technique used by Altman, Deakin (1972) developed a model to predict companies' insolvency. In his study under "A discriminant analysis of predictors of Business failure", Deakin (1972) pointed out that Beaver's technique has a stronger predictive power while Altman's approach results in better insights (Deakin, 1972).

Ohlson (1980) in a study under the title of "Financial ratios and the prediction of possible bankruptcy" employed nine financial ratios as the independent variables and data sets obtained from 105 bankrupt companies and 202 non-bankrupt companies developed a model to predict the health of companies. Like beaver, Ohlson used logit model to develop his model (Ohlson, 1980)

McKee and Greenstein (2000) have criticized the previous studies on bankruptcy. In their study, Mackay and Thin measured and compared the predictive specificity of an easily implemented two-variable bankruptcy model originally developed using recursive partitioning on an equally proportioned data set of 202 firms. They tested the predictive specificity of this model, as well as previously developed logit and neural network models, using a realistically proportioned set of 14,212 firms' financial data covering the period 1981–1990 (Mckee & Greenstein, 2000).

Shah and Murtaza (2000) introduced a model using a neural network for bankruptcy prediction. In their studies on "A neural network based clustering procedure for bankruptcy Prediction" they used the data sets of 60 bankrupt companies and 54 non-bankrupt companies for the time period 1992 to 1994.

Saunders and Allen (2002) used Altman's model for credit risk prediction in companies that received credits from banks. According to their study, if a company received a z score of lower than the critical level the company's position is regarded as unhealthy (Saunders & Allen, 2009).

Emel et al., (2003) used data envelopment analysis (DEA) and proposed a credit scoring approach. They used current financial data of 82 industrial/productive companies that comprised the credit portfolio of one of the largest Turkish banks. To do so, they examined 42 financial ratios and chose 6 important ratios to be used for credit rankings of companies (Emel et al., 2003).

Shin, Lee, and Kim (2005) employed support vector machines and developed a bankruptcy prediction model. They compared the function of their model to that of artificial neural networks. The results indicated that support vector machines have a better performance concerning both generalizability and the overall model precision than

Table 1. Models used to predict financial distress and bankruptcy

Models used to predict financial distress	Researcher(s)	Year
Univariate models	Fitzpatrick	1931
	Ransmer and Foster	1931
	Merwin	1942
	Walter	1957
	Braver	1966
Multivariate Discriminant Analysis (MDA)	Altman	1968
	Deakin	1972
	Edmister	1972
	Blum	1974
	Moyer	1977
	Altman, Halderman, and Narayanan	1977
	Altman	1983
	Booth	1983

	Fulmer, Moon, Gavin, and Erwin	1984
	Casey and Bartczak	1985
	Lawrence and Bear	1986
	Aziz, Emanuel, and Lawson	1988
	Altman	1993
	Altman	2000
	Grice and Ingram	2001
Logit and Probit Analysis	Martin	1977
	Ohlson	1980
	Rose and Giroux	1984
	Zavgren	1985
	Gentry, Newbold, and Whiteford	1985
	Lau	1987
	Platt and Platt	1990
	Koh	1991
	Lynn and Wertheim	1993
	Johnson and Melicher	1994
	Barniv, Hathorn, Megrez, and Kline	1999
	Lennox	1999
Recursive partitioning algorithms (RPA)	Barniv, Mehrez, and Kline	
	Marais, Patell, and Wolfson	1984
	Frydman, Altman, and Kao	1985
	Tam	1991
	McKee and Greenstein	2000
Artificial Neural Networks (ANN)	Odom and Sharda	1990
	Sachenberger, Cinar, and Lash	1992
	Coates and Fant	1991-1992
	Tam and Kiang	1992
	Coates and Fant	1993
	Nittayagasetwat	1994
	Serrano-Cinca	1996
	Lee, Han, and Kwon	1996
	Jo, Han, and Lee	1997
	Serrano-Cinca	1997
	Luther	1998
	Zhang, Hu, Patuwo, and Indro	1999
	Yang, Platt, and Platt	1999
	Shah and Murteza	2000

Source: Raeei and Fallahpour(2004)

artificial neural networks. To do so, they used 10 financial ratios for the time period 1996 to 1999 (Kyung & Lee, 2005).

Table 1 summarizes the studies conducted on the companies' financial positions and the model used for the prediction of financial distress and bankruptcy.

3. Statistical Models

3.1 Discriminant Analysis

Discriminant analysis dates back to the 1930s and works done by British statistician *Charle Pearson* and others in the field of groups' intervals and racial similarity coefficients. However, the method was first developed by Fischer in 1936 based on the methodology used in multivariate linear regression (matrix algebra) to solve linear problems.

Discriminant analysis is employed to categorize respondents based values (codes) of a nominal dependent variable with two or more aspects (Pardoe et al., 2007). In fact, for cases in which nominal dependent variables and independent variables are quantitative, discriminant analysis is used to predict the changes made in the dependent variables (group membership) based by the use of dependent variables. In this method, dependent variables also called predictive variables are combined to form a new variables and a *discriminant score* is assigned to each respondent (cases). This new variable called *discriminant function* is calculated in a way that the respondents are classified in different classes of the dependent (criterion) variable based on the values/scores they obtain.

The discriminant function is, in practice, seeking to find a linear combination of the independent standardized variables that creates the maximum variance among groups/classes which intra-group variance reaches the minimum level. To do so, the discriminant function is created based on groups' centroids (means) so that there is the minimum overlap among groups.

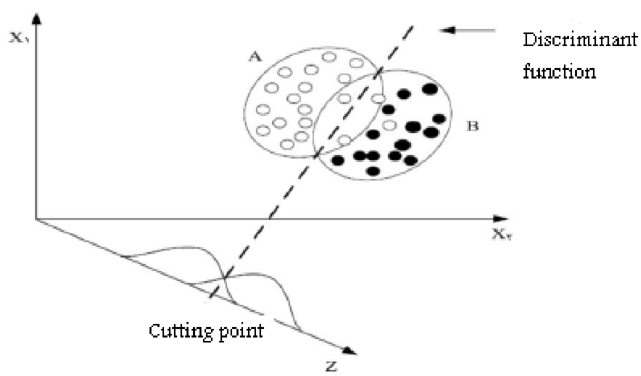

Figure 1. Discriminant analysis

To explain what was mentioned above in mathematical terms, X1, X2, ..., Xn are supposed to be independent variables and Z is a multi-level (categorical) dependent variable. Discriminant analysis aims at determining a linear function as follows:

$$Zi = \beta\alpha X_0 + \beta 1X1 + \beta 2X2 + \ldots + \beta nXn \tag{1}$$

So that the following likelihood would be at a maximum level:

$$P(Z = z|(X1, X2, \ldots, Xn) = (x1, x2, \ldots, xn) \tag{2}$$

As can be seen in Figure 1, when the dependent variable consists of two levels (groups), it is aimed to attribute the new observations (X1, X2, …, Xn) to one of these two groups based on Z (the discriminant function). To do so, β coefficient representing the share of each variable in the scoring function is chosen in a way that the resulting Z score from the above functions, to discriminant optimally between the groups. Besides, the value of Zi is calculated in a way that the intervals between means (centroids) are at the maximum level in the two groups i.e., Z_A and A_B. These requirements will be satisfied if the β function is determined as follows:

$$\beta = \sum^{-1}(XA - XB) \tag{3}$$

where:

X_A is the mean vector of n variables in the cases under study in Group A;

X_B is the mean vector of n variables in the cases under study in Group B;

∑: The variance-covariance matrix for *n* independent variables.

In discriminant analysis, the efficiency of the discriminant function in creating significant differences between groups is examined using Wilks` Lambda Test whose level of significance is determined through chi-square test (Kinnear & Colin).

As shown by the previous research, one of the applications of discriminant analysis is to predict companies' financial abilities in performing their debt obligations and repaying the loans they received. In such cases, discriminant analysis is looking for the linear function of financial and non-financial variables that creates the maximum variance between groups associated with economic enterprises through defined variables, if in-group variance were at the minimum level as possible.

There are two ways to be assured of the function of the model under study. The first way is to determine the classification specificity or, in other words, determining the ability of the model to make a distinction between healthy companies and unhealthy companies (with regard to their financial performance) in the sample under study. Another way is to determine the predicting power of the model for different samples. The classification specificity can be determined by calculating the misclassification of unhealthy companies (error Type 1) and the misclassification of healthy companies (error Type 2). The overall specificity is the sum of these two estimates. As the results of the linear discriminant function indicate, the cost of Error Type I (loan losses) is equal to the cost of Error Type II (the opportunity cost of outstanding loans). However, the cost Error Type I is in reality, greater than that of Error Type II. Therefore, it is possible to review the obtained results according to Byes decision making rules and create a cost function based on the two first and second errors.

There are a number of presuppositions when performing discriminant analysis, the most important of which are:

- Multivariate Normality: Data values are from a normal distribution;
- Equality of variance-covariance within group;
- Low multicollinearity of the variables

However, the efficiency of the model has been confirmed even when the above presuppositions are not observed totally. Especially, given that normal assumptions are usually not "fatal". The resultant significance tests may still be reliable (Darroch & Mosimann, 1985).

3.2 Logistic Regression

Logistic regression was introduced in the late 1960s and the early 1970s as a substitute for linear regression test and the discriminant function analysis. (Peng Joanne, Harry et al., 2002). When the dependent variable is nominal and independent variables are ordinal or interval, normal regression models and discriminant analysis underestimate the actual values.

Logistic regression is one of the functional techniques used to analyze classified data. For instance, if the result of an experiment is defined in terms of winning/losing, then the response variable is no longer continuous rather it is in a classified form. A type of logistic regression is binary logistic regression in which there are two variable response groups/classifications. In this model, it is sufficient to know whether an event has happened or not. Then, a discrete dependent variable such as 1 or 0 is used to show the event under study.

Logistic regression is similar to linear regression and the only difference is that coefficients are not determined in the same way by the two methods. In other words, logistic regression instead of minimizing the error squares (as done by linear regression analysis) maximizes the likelihood of the occurrence of an event. Besides, in linear regression analysis, F and T statistics are used to test the model fitness and the significance of effects of each variable in the model while in logistic regression analysis chi-square and Wald tests are employed to do so. Wald Test is determined by the use of the following equation:

$$Wald\ (Xi) = \left(\frac{\beta i}{S.E}\right)^2 \tag{4}$$

Where: βi is the variable coefficient I and S.E stands for the standard error.

In fact, the Wald Test will examine the null hypothesis stating that the values of all βs are equal to zero. In other words, the effect of all independent variable is equal to zero. If the null hypothesis is rejected the minimum value of one of the variables needs not to be zero. One of the advantages of Logit model over the other used models is that to perform the logistic regression it is not necessary to assume the equality of variances of the two groups and the multivariate normal distribution for independent variables.

In logistic regression, a concept called odd ratio (Pi/1-Pi i.e., the ratio of the likelihood of the occurrence of an event to the probability of nonoccurrence of the event) is used and the odd ratio logarithm known as Logit model is determined by Eq. 3.2.2 as follows.

$$Logit(y) = Ln(\frac{pi}{1-pi}) = \beta 0 + \beta 1Xi + \cdots + \beta nXn \quad (5)$$

Where, Pi is the probability of the outcome or event in the presence of Variable X, β_0 is the intercept of Y, β shows regression coefficient, and X stands for independent variables.

It is worth mentioning that while the range of odds ratio varies between 0 and 1, the range of the Logit of odds ratio varies from -∞ to +∞.

4. Research Methodology

The aim of the present study was to examine the application of discriminant analysis and logistic regression and to compare their efficiency in the predication and the classification of companies that fail to perform their financial obligations and companies that repay their banking credits and associated costs in time. To do so, the present study employed the following methodology:

4.1 Population and Sampling

The population under study included all legal clients in one of the branches of commercial banks in Hormozgan Province. After investigating the clients' records, a list of solvent and insolvent client companies that received their loans in 2009 was provided. Accordingly, a total number of 54 companies that had no delay in repaying their loans within a three-year period from 2009 to 2011 and a total of 46 companies with an over ninety-day delay in repaying their loans were chosen as the sample under study in the two groups of solvent and insolvent clients. Another requirement for the inclusion of companies in the sample was the accessibility of their financial data and the access to their financial accounts audited by certified auditors.

4.2 The Scope of the Study

The time span to choose the sample companies was 2009 i.e., those clients that received banking credits from one of the branches of commercial banks in 2009 were included in the sample under study. Besides, the time period to examine the financial data of the selected companies was three years before receiving loans and three years after their first default. Generally, companies' financial data, their healthiness or unhealthiness, and the occurrence of the first default lasting for more than 90 days were investigated in a three-year period from 2009 to 2011.

4.3 Research Variables

4.3.1 Dependent Variable

The quality of the loans repayment was chosen as the dependent variable that has a discrete nature in the present study and its value is either 1 or 2. Accordingly, the value of 1 was assigned to solvent customers that have repaid their loans in due time. On the other hand, insolvent customers whose default, according to Basel Committee, has lasted more than 90 days received a value of 2.

4.3.2 Independent Variables

Independent (predicting) variables were selected based on the financial data derived from the audited financial accounts of the companies under investigation. In order to eliminate the effects of the companies' size differences and specific features, financial ratios were employed instead of directing using values of financial accounts. A list of the most important financial ratios (as shown in Table 2) was provided based on the previous studies and consulting the experts in the field. Of these, 18 ratios were selected considering the two following criteria:

First, the ratios means were required to be statistically different for the solvent and insolvent companies and, second, the selected variable should have had a lower co-linearity and correlation than those of other variables.

Table 2. Independent variables and methods of estimation

Ratios	Parameter	Methods of estimation
The ratio of net profit to sales	X_1	After-tax income to total income
The ratio of gross profit to sales	X_2	Pre-tax income to total income
The ratio of gross profit to sale or loss	X_3	The ratio of gross income to total income
Ratio of return on assets **	X_4	The ratio of after-tax income to total assets
Ratio of percentage of return on assets	X_5	The ratio of after-tax income to total assets
The ratio of return on equities **	X_6	The ratio of after-tax income to total equity
The ratio of fixed assets	X_7	The ratio of after-tax income to the net fixed assets
The ratio of loan usefulness index	X_8	The ratio of return on assets to return on equities
Current ratio	X_9	The ratio of total current assets to total current liabilities
Immediate ratio	X_{10}	The ratio of the difference between current assets and inventories to total current liabilities
Liquidity ratio	X_{11}	The ratio of total cash and bank holdings to total current liabilities
Current assets ratio	X_{12}	The ratio of total current assets to total assets
Cash adequacy ratio	X_{13}	The ration of operating cash to total long-term debt, fixed asset purchases and other costs, interests, dividends payable
Cash flow ratio	X_{14}	The ratio of operational cash to total current liabilities
Working capital ratio	X_{15}	The ratio of net working capital to total assets
Inventory turnover ratio	X_{16}	The ratio of inventory to cost of goods sold multiplied by 365
Collection period ratio	X_{17}	The ratio of total accounts and trade instruments receivables to total revenues multiplied by 365
Working capital turnover ratio	X_{18}	The ratio of inventories to difference between current assets and current liabilities
Current capital turnover ratio	X_{19}	The ratio of total revenues to difference between current assets and current liabilities
Fixed capital turnover ratio	X_{20}	The ratio of total revenues to fixed assets
Total assets turnover ratio	X_{21}	The ratio of total revenues to total assets
The ratio of debt to equity	X_{22}	The ratio of total liabilities to total equities
The ratio of long-term liabilities to equity	X_{23}	The ratio of long-term liabilities to total equities
Ownership ratio	X_{24}	The ratio of total equities to total assets
Gross profit ratio	X_{25}	The ratio of after-tax income to net fixed assets
The ratio of return on net working capital	X_{26}	The ratio of after -tax income to the difference between current assets and current liabilities
Debt ratio	X_{27}	The ratio of total liabilities to total assets
Debt coverage ratio	X_{28}	The ratio of net fixed assets to total long-term liabilities
The ratio of fixed assets to equity	X_{29}	The ratio of net fixed assets to total equities
Interest coverage ratio	X_{30}	The ratio of before-tax and interest profit to paid interest expenses
The ratio of current debt to equity	X_{31}	The current ratio of total liabilities to total equities

After performing test for equality of means, two ratios were marked as * in the table as they had equal means for the two groups of companies and 11 ratios were marked with ** after running the colinearity test and were excluded from the list because of their high correlation.

5. Findings of the Study

5.1 Findings of Discriminant Analysis

After analyzing the data by discriminant analysis test through SPSS Software Package, the following findings were obtained:

The first section presents the results of Box Test to measure the equality of covariance matrices. In the first test output, logarithms of the determinants related to each group are presented in Table 3. Given that logarithms of the determinants for each group are equal to in-group logarithms of the determinants for that group, it is clear that the value of covariance matrix (1,118.060) for insolvent companies is greater than the value of covariance matrix (110.566) for solvent companies.

Table 3. Logarithms of the determinants related to each group

The Situation of company in repay	Rank	Log Determinant
healthy	18	99.981
unhealthy	18	110.566
Pooled within-groups	18	118.060

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

The results of Box Test shown in Table 4 indicate that since the level of significance is zero (Sig = 0), the assumption of equality of covariance matrices is not confirmed which may be due to the great size of in-group sample and the nonobservance of the assumption of multivariate normality. However, it has been demonstrated that the discriminant function is not limited by non-fulfillment of the above assumption.

Table 4. Results of box test

Box's M		1295.385
F	Approx.	6.090
	df1	171
	df2	27966.251
	Sig.	.000

Tests null hypothesis of equal population covariance matrices.

As was mentioned earlier, one of the most important results of discriminant analysis is Canonical Discriminant Function. As shown in Table 5, the discriminant analysis test has been able to identify a canonical discriminant function in which the great value of eigenvalues as a relative index of discrimination power of a discriminant function shows the stronger explanatory power of the resulting function. On the other hand, the canonical correlation is equal to 0.769 and since this value of this coefficient is close to 1 it is clear that there is a strong correlation between discrimination scores and the groups under study. Besides, the function is able to differentiate between groups.

Table 5. Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.444[a]	100.0	100.0	.769

a. First one canonical discriminant function was used in the analysis.

Table 6 shows the results of Wilks' Lambda Test, which is a ratio of the variances of the total discriminant scores not explained by the inter-group differences. As a result, unlike the canonical correlation coefficient, if the smaller values of Wilks' Lambda approaching to zero indicate that the inter-group means are different from each other. Accordingly, the value of Wilks' Lambda equal to 0.409 indicates a relatively reasonable difference between solvent and insolvent customers. In other words, the resulting function explains 59.1% (1–0.409) of dependent variable changes (the credit position of banking customers). In order to interpret the results of Wilks' Lambda Test more accurately it is possible to use the converted chi-square test whose value is 79.548 and since the significance level is 0, there is a significant difference between the inter-groups means ($P < 0.01$).

Table 6. Results of Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.409	79.548	18	.000

Accordingly, the discriminant function as a combination of the standardized independent variables showing the greatest inter-groups means can be written in the form of the following equation (Eq. 5.1.1) using canonical discriminant function coefficients:

$$Z\ score = -1.883-0.006X2+0.013X5-0.001X8+0.296X9+0.020X10+0.228X11-0.925X14-0.030X16-0.001X17-0.020X18+0.001X20+0.822X21-0.001X27-0.006X28+0.002X29 \quad (6)$$

The customers' credit scores can be calculated by inserting the predicting variables (financial ratios related to the customers) in Equation 3. Table 7 presents the customers' credit mean scores separately for solvent and insolvent customers which are 1.098 and -1.289 for solvent and insolvent customers, respectively. Since the number of solvent banking customers were 54 and insolvent banking customers were 46 in this study, the weighted mean of the above credit scores can be used as a measure to determine financial healthiness and unhealthiness of new customers as follows:

If new customer's credit scores > 0.00002, then the customer will be regarded as solvent (healthy). On the other hand, if new customer's credit scores ≤ 0.00002, then the customer will be regarded as insolvent (unhealthy).

Finally, the most important results of discriminant analysis are shown in Table 8 as follows. In this table, the rows show the observed groups of the dependent variable and the columns present the corresponding predicted groups. As shown in the table, 49 companies comprising 9.07% of the customers were classified correctly as solvent customers and only 5 companies, 9.3% of solvent customers were mistakenly put in the insolvent customers' group. This is not the case for insolvent customers so that in this classification 42 companies accounting for 91.3% of unhealthy companies were put acceptably in the solvent customers' group and only 4 companies accounting for 8.7% of unhealthy companies were wrongly classified as solvent customers. Although the identification of 90.7% of unhealthy companies proves the high reliability of this discriminant function in precisely predicting solvent customers, the specificity of the function is lower than its precision in classification of unhealthy companies. However, this may be regarded as an advantage since the consequences of wrong classification of unhealthy customers as healthy customers (Error Type I) are more severe than rejecting solvent customers and putting them in the unhealthy customers' group. Nevertheless, as the results of the classification table called also the confusion table indicate, the hit ratio showing the total precision of the discriminant function is equal to 91%, suggesting the high reliability of the function for the classification and discrimination of groups.

In another sense and based on the results of cross-validated classification shown in this table, 87% of healthy companies were correctly put in the solvent customers' group. This was also the case for unhealthy companies; 87% of which were correctly classified in the insolvent customers' group. This indicates that the hit ratio is equal to 87%.

Table 7. Functions at group centroids

Situation of company in repay	Function
	1
Healthy	1.098
Unhealthy	-1.289

Un standardized canonical discriminant functions evaluated at group means

Table 8. Classification results based on discriminant analysis test

		The Situation of company in repay	Predicted Group Membership		Total
			healthy	unhealthy	
Original	Count	Healthy	49	5	54
		Unhealthy	4	42	46
	%	Healthy	90.7	9.3	100.0
		Unhealthy	8.7	91.3	100.0
Cross-validated	Count	Healthy	47	7	54
		Unhealthy	6	40	46
	%	Healthy	87.0	13.0	100.0
		Unhealthy	13.0	87.0	100.0

a. 91.0% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 87.0% of cross-validated grouped cases correctly classified.

Table 9. Omnibus test of model coefficients

		Chi-square	df	Sig.
	Step	64.296	1	.000
Step1	Block	64.296	1	.000
	Model	64.296	1	.000
	Step	34.451	1	.000
Step2	Block	98.747	2	.000
	Model	98.747	2	.000
	Step	10.845	1	.001
Step3	Block	109.592	3	.000
	Model	109.592	3	.000
	Step	4.156	1	.041
Step4	Block	113.748	4	.000
	Model	113.748	4	.000
	Step	7.214	1	.007
Step5	Block	120.962	5	.000
	Model	120.962	5	.000
	Step	7.724	1	.005
Step6	Block	128.686	6	.000
	Model	128.686	6	.000
	Step	-2.100	1	.147
Step7[a]	Block	126.586	5	.000
	Model	126.586	5	.000

a. A negative Chi-squares value indicates that the Chi-squares value has decreased from the previous step.

5.2 Findings of Logistic Regression Test

This section presents the most important results obtained through logistic regression test to evaluate the efficiency of this technique in distinguishing healthy companies from unhealthy ones.

The first output is related the results of Omnibus Test shown in Table 9, as follows. Generally, the value of chi-square test shows whether independent variables affect the dependent variables or not. According to the results obtained in the seventh stage, the value of chi-square test is 123.586 with an acceptable level of fitness which is significant at the error level less than 0.01.

Table 10 presents the results of log-likelihood test and pseudo R-Square test. Pseudo coefficient including Cox and Snell R-square and Nagelkerke R-Square contains approximates of R^2 determination coefficient in linear regression which are not applied in logistic regression. Given the acceptable levels of related values of pseudo determination coefficient in the seventh stage, it can be said that the independent variables manipulated in the present study have a strong explanatory power in explaining variance and changes of the dependent variable. In other words, a value of 0.959 for Nagelkerke R-square in Stage 7 indicates that 95.9% of the dependent variable variations are explained by independent variables in logistic regression.

Given the variables used in the model and the results of Wald Test, odd ratios, and non-standardized regression coefficients; the output of Logit model can be shown as Eq. 5.2.1 as follows: 、

$$Ln\left(\frac{p}{1-p}\right) = 20.938 \pm 0.156\,X\,5 - 7.801X11 - 0.668X16 - 5.941X1 - 0.259X24 \qquad (7)$$

Finally, the most notable results of logistic regression are presented in Table 11 known as the classification table. The results of the seventh stage in the table indicate that the model correctly predicted 53 healthy companies out of a total of 54 companies and 44 unhealthy companies out of a number of 46 companies. As shown in the table, the model's sensitivity in predicting healthy companies is 98.1% and 97% in determining unhealthy companies.

5.3 Comparison of Models' Discriminator Power

A comparison of the results shown in the tables 8 and 11 suggests that the sensitivity of logistic regression test in the classification of borrower companies was 97% while discriminant analysis test predicted the healthiness or unhealthiness of companies successfully in 91% of the cases under study.

In addition, Receiver-Operating Characteristic (ROC) Curve and discriminator accuracy statistics including sensitivity and specificity at 95% of confidence interval was plotted for the two models. The area under ROC Curve as the most important index to measure the discriminator power of several test is 0.969 for logistic regression with 95% confidence interval (.929, 1). Also, the area under the curve is significantly different from 0.5 since p-value is .000 meaning that the logistic regression classifies the group significantly better than by chance. Besides, the area under the curve for discriminant analysis is .910 with 95% confidence interval (0.845, 0.975). Also, the area under the curve is significantly different from 0.5 since p-value is .000 meaning that the logistic regression classifies the group significantly better than by chance.

Table 10. Results of log-likelihood and pseudo R-Square

Step	-2Loglikelihood	Cox & Snell R-Square	Nagelkerke R-Square
1	73.692[a]	.474	.634
2	39.242[b]	.627	.838
3	28.397[c]	.666	.890
4	24.241[c]	.679	.908
5	17.027[d]	.702	.938
6	9.303[e]	.724	.967
7	11.403[f]	.718	.959

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

b. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

c. Estimation terminated at iteration number 10 because parameter estimates changed by less than .001.

d. Estimation terminated at iteration number 11 because parameter estimates changed by less than .001.

e. Estimation terminated at iteration number 15 because parameter estimates changed by less than .001.

f. Estimation terminated at iteration number 13 because parameter estimates changed by less than .001.

Table 11. Results of classification obtained from logistic regression

Observed			Predicted		
			The Situation of company in repay		Percentage Correct
			healthy	unhealthy	
Step1	The Situation of company in repay	healthy	46	8	85.2
		unhealthy	8	38	82.6
	Overall Percentage				84.0
Step2	The Situation of company in repay	healthy	48	6	88.9
		unhealthy	6	40	87.0
	Overall Percentage				88.0
Step3	The Situation of company in repay	healthy	51	3	94.4
		unhealthy	2	44	95.7
	Overall Percentage				95.0
Step4	The Situation of company in repay	healthy	50	4	92.6
		unhealthy	3	43	93.5
	Overall Percentage				93.0
Step5	The Situation of company in repay	healthy	53	1	98.1
		unhealthy	3	43	93.5
	Overall Percentage				96.0
Step6	The Situation of company in repay	healthy	53	1	98.1
		unhealthy	1	45	97.8
	Overall Percentage				98.0
Step7	The Situation of company in repay	healthy	53	1	98.1
		unhealthy	2	44	95.7
	Overall Percentage				97.0

a. The cut value is .500

Table 12. Area under curve for sensitivity and specificity

Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Predicted group	.969	.020	.000	.929	1.000
Predicted Group for Analysis1	.910	.033	.000	.845	.975

Although the area under the ROC Curve shows the strong power of the two-fitted model, this is logistic regression that, by comparison, covers a larger area under the curve and, therefore, has a higher discriminator power, which can be seen in Figure 2, visually. The curve was plotted in a coordinate sheet whose horizontal axis shows the model specificity and its vertical axis presents the model sensitivity. Besides, the curve was drawn from Point (0, 0) from the low left-side corner towards Point (1, 1) in the high right-side corner. The more the curve is closer to the low left-side corner to Point (1, 0) the model has a stronger discriminator power to distinguish between the two groups of healthy and unhealthy companies. As a result, the two model's capability can be seen in the following curve, illustrating the stronger discriminator power of logistic regression model than discriminant analysis model.

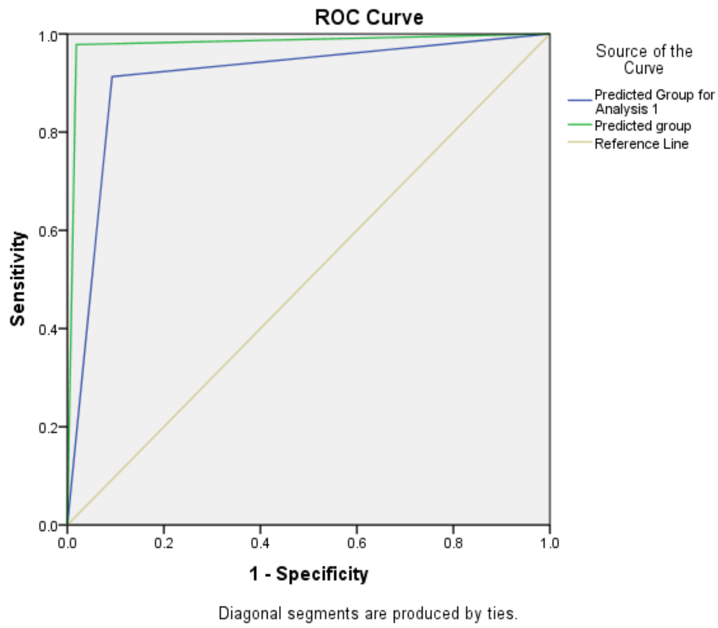

Figure 2. ROC curve

6. Conclusions

Obviously a country's economic growth is dependent on its capability and the development of financial markets within the countries. Meanwhile, banks and financial institutions play an undeniable role in the provision and distribution of financial resources, especially in those countries that the capital market capacity cannot meet financial needs of the production and industrial sectors. Since the increasing economic development is associated with the well-being of all economic sectors, the protection of banks and financial institutions against the challenges they are facing is of high significance. One of the most important of these threats and challenges is credit risk which has been an appealing area of interest over the years for the scholars in the financial field.

Similarly, the present study has attempted to identify the important financial ratios that are able to explain credit risk, using discriminant analysis and logit regression tests to develop a model to predict solvent and insolvent banking customers. The results of analyses performed by discriminant analysis suggested that the model's sensitivity in the recognition and classification of solvent and insolvent customers was equal to 91% while sensitivity of logit regression in doing so was 97%.

It was also noted that both methods are very powerful to predict and classify banking customers, although logistic regression model, by comparison, was more reliable and powerful than discriminant analysis model; a claim that is confirmed by the facts presented in the area under the Receiver-Operating Characteristic (ROC) Curve.

Undoubtedly the application of the results of the present study along with other approaches can increase the discriminator power needed for credit classification of banking customers and contribute largely to banks and financial institutions in getting rid of risks associated with credit defaults. Although more research is still needed to explore the issue in greater depths, it is suggested that researchers in their future studies examine the accuracy of the results obtained by this model in different research populations and in various time spans. Researchers are also recommended that employ other statistical and econometric techniques in future research to predict banking customers' credit scores and perform a comparative analysis of techniques used.

References

Altman, E. I. (1968). Financial ratios discriminant analysis, and the Prediction of corporate bankruptcy. *Journal of Finance, 23*(4), 589–609. http://dx.doi.org/10.1111/j.1540-6261.1968.tb00843.x

Bank for International Settlements Press & Communications. (2006). *Basel committee on banking supervision. International Convergence of Capital Measurement and Capital Standards.* Basel, Switzerland: Author.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research (Empirical Research in Accounting, Selected Studies), 4,* 71–111. http://dx.doi.org/10.2307/2490171

Chao-Ying, J. P., & Tak-Shing, H. S. (2002). Logisitc Regression Analysis and Reporting: A Primer. *Understanding Statistics*, *1*(2), 75–137.

Darroch, J. N., & Mosimann, J. E. (1985). Canonical and Principal Component of Shape. *Biometrika, 75*(1), 241–252. http://dx.doi.org/10.1093/biomet/72.2.241

Deakin, E. B. (1972). A discriminant analysis of predictors of Business failure. *Journal of Accounting Research, 101,* 167–179. http://dx.doi.org/10.2307/2490225

Emel, A. B. et al., (2003). A credit scoring approach for the commercial banking sectors. *Journal of Socio-Economic Planning Sciences, 37,* 103–123. http://dx.doi.org/10.1016/S0038-0121(02)00044-7

Glantz, M. (2003). *Managing Bank Risk*. USA: Academic Press.

Horrigan, J. O. (1968). A short history of financial ratio analysis. *Accounting Review*, *4*, 284–294.

Jorge. C. D., & Shuri, J (1988). Resource Allocation. Tehran, *Papiroos Publications*.

Kinnear, P. R., & Gray, C. D. (1999). *SPSS for Windows Made Simple* (3rd ed.). Hove: Psychology Press.

Kiss, F. (2003). Credit Scoring Processes from a Knowledge Management Perspective. *Periodica Polytechnica*, *11*(1), 95–110.

Lieli, R. P., & White, H. (2010). The Construction of Empirical Credit Scoring Models Based on Maximization Principles. *Journal of Econometrics*, *157*(1), 110–119. http://dx.doi.org/10.1016/j.jeconom.2009.10.028

Mckee, T. E., & Greenstein, M. (2000). Predicting bankruptcy using recursive partitioning and a realistically proportioned data set. *Journal of forecasting, 19*(3), 219–230. http://dx.doi.org/10.1002/(SICI)1099-131X(200004)19:3<219::AID-FOR752>3.0.CO;2-J

Ohlson, J. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*(1), 109–131. http://dx.doi.org/10.2307/2490395

Pardoe, I. et al. (2007). Graphical Tools for Quadratic Discriminant Analysis. *Technometrics, 49*(2), 172–183.

Samreen, A., & Zaidi, F. B. (2012). Design and Development of Credit Scoring Model for the Commercial banks of Pakistan: Forecasting Creditworthiness of Individual Borrowers. *International Journal of Business and Social Science, 3*(17), 2219–1933.

Saunders, A., & Allen, L. (2002). *Credit Risk Management* (3rd ed). New York: John Wiley & Sons, Ltd.

Shah, J. R., & Murtaza, M. B. (2000). A neural network based clustering procedure for bankruptcy Prediction. *American Business Review, 18*(2), 80–86.

Shin, K. S., Lee T. S., & Kim, H. J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications, 28*, 127–135. http://dx.doi.org/10.1016/j.eswa.2004.08.009

Sinky, Jr., & Joseph, F. (1992). *Commercial bank financial management* (4th ed). Toronto: Macmillan.

Thomas, L. C., Edelman, D. B., & Crook, J. N. (2002). *Credit Scoring and its Applications*. USA and Canada: SIAM. http://dx.doi.org/10.1137/1.9780898718317