Practice of Data Envelopment Analysis (DEA) in Discriminate Analysis (DA) and it’s usage in ranking of credit customer of Bank Mellat

Gholam Reza Bayati
ISLAMIC AZAD UNIVERSITY - Central Tehran Branch
Bayati114@yahoo.com

Mr. Farhad Hosseinzadeh Lotfi (Ph.D)
ISLAMIC AZAD UNIVERSITY - Central Tehran Branch
Hosseinzadeh.lotfi@iauctb.ac.ir

Abstract

Failure to pay the granted loans in a timely manner is seen as one of the most important challenges faced regarding facilities granting in the banking system of the country which results in outstanding, delayed, and doubtful claims accounts.

One of the suitable solutions to face such challenge is to separate clients with good records from those having bad records prior to grant any loan. Measuring the credit risks of the clients before granting the relevant loan may be considered as a certain solution to reduce the credit risk. The purpose of this study is to present a certain model, which chooses effective features of the client in measuring the credit risk by taking benefit from the elites’ judgment and applies data coverage analysis in discriminate analysis (DA) to identify clients with good records from those with bad records. Such presented model was tested by using dataset of bank’s clients and through considering 20 features from 500 borrowers. Through utilization of Delphi method, eight effective features of the clients in predicting the credit risk were identified, and further, based on the number of the outstanding installments, the clients were classified into two groups of having good records (G1) and bad records (G2). Later on, through placing the data in DA model, $\alpha_{1,...,8}$, $\beta_{1,...,8}$ and $d$ weights relevant to the super-planes separating the two foregoing groups were obtained. Based on such weights, it is possible to determine that the new data shall be placed in which group. Considering the high precision of prediction of the model in classification of the clients into two groups with good and bad records prior to grant loan, such model may be used as a decision-assisting system to help those charge of facilities in the banks. Therefore, not only loans will not be paid to those clients with bad records, but also certain measures (inter alia, obtaining sufficient pledge) may be taken to grant loans to clients with good and/or bad records. Hence, in addition to take a practical step towards reducing the bank’s claims, the level of satisfaction of the clients with good records may also be increased.

Keywords: Data Envelopment Analysis (DEA), Discriminate Analysis (DA), Ranking of credit customer of Bank.
Introduction

Studying economics at the international level indicates the fact that the always there is a near relationship between countries in investment and level of economic progress. This means that countries with efficient model in allocation of capital in different economic sectors, often have higher economic development and social welfare. Mobilization and allocation of resources for investment in economic activities is done through financial markets which The credit market is part of the market. Most of the banks, raise funds and allocate them to the different economic sectors. But it should be noted that on the one hand, the financial resources are supplying the needs of banks and on the other hand, banks must optimize allocation of their limited financial resources to produce goods and services. This means that the firm is in efficient level, because in terms of economic theory, performance is the results of optimizing production and allocation of resources, thus, the bank is seeking to grant loans to companies that, while having a low risk, have Returns proportionate with the profit of facilities. This can be fulfilled only when banks be able to identify their credit customers both natural and legal and be able to categorize them based on the ability and willingness of the full and timely repayment of obligations using financial and non-financial appropriate criteria. Because in such a system, the facilities granted to applicants who had a lower credit risk and the possibility of repayment of debts due is higher. Because this can be used as a source of next financial funds, in this case, it has an important role in increasing investment, growth and economic development of the country. (Safari et al., 2010). In this study, discriminate analysis of customers applying for credit facilities of Mellatbank branches using data envelopment analysis is examined.

Statement of the problem

Failure to timely repayment of loans is one of the most important challenges in the banking system is granting facilities, resulting in non-performing loans, past due and doubtful payments. One of the solutions to meet these challenges, is the separation of good customers from bad ones accounts before granting a loan. Assessing the credit risk of customers before lending can be considered as a way to reduce credit risk. (Kazemi, 2013)

In the past, decisions were made about lending to customers of traditional banks and based on personal judgment about the risk of default. However, the increasing demand for banking facilities by firms and households, on the one hand and increasing competition from large commercial and the efforts of banks and financial institutions in the country to reduce the risk of non-repayment of loans, on the other hand resulted in the application of new methods including statistical methods in this field. Today, to predict a possible failure to repay loans and to classify their customers, banks benefit from credit ratings of their customers. Saving time, cost savings, removing personal judgments and increasing accuracy in assessing applicants' types of facilities are among the advantages of it. In this connection, various statistical methods such as discriminant analysis and DEA have been used.

Discriminant analysis was used to find a rule to distinguish between two or more of multivariate. The most important application of discriminate analysis is rating. Deciding whether an observation with P variables belongs to which one of the many rival community is called rating. For example, suppose a client with different characteristics such as age, sex, educational level, qualification such as being Well-known customer or not, the turnover on the account, the amount of credit in the community, workplace and residence, introducer, jobs and time of engagement in it, place of business, business licenses, being famous in the city or area or neighborhood, being native or non-native, culture, religion and adhering to it, past financial records and facilities, indicators of character and so on entered the bank to get facilities. By statistical methods how can we decide if he is good or bad
customer? In the other words, is he Creditworthy or not and whether or not we lend he our facility? Discriminant analysis and classification methods are used to answer this question. Numerous applications for discriminant analysis and classification can be imagined. In fact, in many cases, identify the type of client can be based on visible signs based on Discriminant analysis and classification methods. According to the above description, this study seeks to answer two questions: 1. Are the indicators of financial and non-financial measures affecting credit risk? 2. Does the discriminant analysis methods and data envelopment analysis has the necessary efficiency for credit ratings of natural and legal customers of banks? The importance of this issue has caused the current research to rank the risk of Mellatbank customers under management of region 2 branches in Tehran who are applying for credit facilities with regard to financial and non-financial aspects and using discriminant analysis and data envelopment analysis. In this study by using data envelopment analysis in discriminant analysis, a model is designed which can be achieved by the use of customer data in the past and present time, predict good or bad customers (creditworthy or not) in the future.

**Background Information**

Accreditation means the assessment of repaying the loan and financial facilities by applicants and probability of non-repayment of funds received. The first article of DEA was published in 1978 as evaluating the efficiency of decision making units. DEA methodology is a nonparametric method. The procedure is used for defining the boundaries of performance and calculates the relative efficiency of each observation based on the deviation from the most efficient observations. One reason is that the power and authority of DEA without determining the weights of the input and output data and even with low data can also measure relative performance. Maher and Senin 1997 used neural networks and logistic regression in order to increase the prediction accuracy of the customer's credit rating. Adel Azar and colleagues in 2006 in a study designed to evaluate the performance of the branches of insurance using DEA's technique. Foros(2008) to find a new way for predicting the credit risk of Polish companies used the DEA method. Tsulas (2009) in a study as a usefulness and effectiveness modeling of bank branches using DEA model presented a framework for assessing the overall performance of the branch in terms of efficiency and effectiveness.


For the first time a model designed to measure the degree of credit risk in 1909 by John Murray on bonds. In 1966 to determine the company's bankruptcy Logistic regression analysis was employed by Beaver. Later this model was used to measure credit risk of issued corporate bonds. One of the studies about the measurement of credit risk of corporate bonds was conducted in 1968 by Altman using multivariate model scoring and it was known as the Z-score model. Altman's Z-score model is indiscriminant analysis model, using values of important financial ratios attempts to distinguish bankrupt companies and non-bankrupt companies. Saunders and Alan used the model to predict the credit risk of borrowers and concluded that this model has high power to predict the credit risk. In the late 1970s linear and situational probability models have been proposed. And using mathematical planning models for predicting bankruptcy occurred in the 1980 and 1990. The main purpose of these
models was to remove the assumptions and limitations of previous techniques, to improve the reliability and accuracy of the classification. In the early 1990s, Decision Support Systems were used in combination with multi-system decision to solve the financial problems of classification (Safari et al., 2010).

**Importance of the issue**

Nowadays, many changes have been made in knowledge management, but despite the importance of credit risk in banking activities, it seems that a coherent and organized movement has not performed to create models of credit risk in the country. The country's financial markets on the one hand, lack of credit risk indices and rating agencies is sensed clearly. On the other hand, in the field of lending to customers, it has not been seen an integrated process and regularly evaluation of the credit risk and their ranking and the credit ceiling based on the risk. And now experts of credit committees try to determine them (Safari et al., 2010) which sometimes associated with errors. The issue arises because, for example, when a customer to go to the bank and asks for facilities, with customers classified information in the past and the information we received from them at the present time, we do not know whether it he/she is good or bad customer facilities applicant? Is creditworthy or not? Can afford to pay his/her mortgage or not? In case of delay in repayment of loans, can he/she afford delay payment of a fine or not? And dozens of other questions that the bank branch raises with the challenge whether or not agree with the customer's request? The assessment of credit customers is essential for banks so that it did not prevent the Bank's prosperity, its progress and survival. To survive, banks like any other organization need to increase efficiency. They try to make optimal use of available resources. Current status and strategies (decision) Optimization, provides good service to always ensure customer satisfaction and to enjoy sufficient and reliable profitability. Bank customers as recipients of funds if not properly verified banks and as a result the community will face serious problems. Credit rating contains ways which helps financial institutions to make decision on acceptance or rejecting requests for facilities. This decision is one of the most important credit processes. Each customer's credit rating, in fact, is a number that reflects the applicant's credit facilities at a certain point of time and reliable method for assessing the credit history of clients.

**Research purposes**

Determine and establish the parameters and variables that affect the ranking (grading) Credit Customers of bank. The model is designed to distinguish between two or more customers and to classify them.

**Theoretical Framework**

The process of evaluating credit customers of Mellat bank branches is as follows:
1) First, through questionnaires distributed among the population, we determine importance of evaluation criteria by professionals and The Delphi method.
2) then we analyze the results obtained with the help of mathematical modeling DEA and discriminant analysis and using GAMS software.
The research model

For a rule to distinguish between two or more multivariate societies from one another and ranking
them in terms of being efficient or not, there are several methods. Among them Discriminant analysis
combined with DEA are of the most important ways in which the use of mathematical techniques,
avoiding mental style and taste, and the high accuracy of the conclusions, had a special position in
recent decades.
Discriminant analysis concerns with finding a rule to distinguish between two or more of multivariate
communities. The application of discriminant analysis is for classification. Deciding on \( P \) observed
belong to which of the two or more rival communities is called Classification.
data covering analysis is a linear programming and a useful tool in the management of assessing
performance. And several inputs for production and multiple output is used.
data covering analysis, compares every single decision maker with the best single decision and units
in which equal input is used to reach equal output and this Determines efficient and inefficient units.

Research questions

How can one determine a function for distinguishing between the two sets?
How to predict a new feature vector in which the collection will be put?
How can grant a degree of membership for new members?

Research Methodology

Type of research is development of practical knowledge in a particular field in terms of the aim. This
research is intended to apply the results of its findings to resolve the problem of finding a rule to
distinguish between two or more of multivariate (consumer credit). The research is descriptive and
non-laboratory research. And the method used in this research is survey.

Research community

Given that this research is done using mathematical modeling, there's no need to sampling and the
entire population will be considered. Research community of bank credit customers under the
management of branches of Mellat bank in region 2 in Tehran. Since research using mathematical
models to analyze audit and DEA are examined, The target population should be homogeneous, We
checked the coordinates of clients that have this feature. Since obtaining indicators, The following
equation for input and output must be met:
The number of DMUs ≤ 10 (Number of entries + number of output) * 3 Or decision-making units ≥
(total index) * 3
Therefore, the more the population, because of more indicators, the more research would be desirable.

Place and time territory

Credit Customers of Mellat Bank in branches under the management of the bank's branches in Region
2 of Tehran from 2011 to 2014

Subject domain
Determining evaluation indicators of Credit customer credit of Mellatbank branches under the management of branches of Region 2, Tehran, by Delphi method to help analyze the results of discriminant analysis model, data covering analysis and Vector specifications of each customer and thus its performance in the credits.

Methods and tools for data collection

For information library resources including books, articles, Internet sites and graduate student theses related to the subject were used. It has been done by directly visiting bank branches and staff and using the information contained in it is organization. The study of documents have been made available.

Materials Analysis

By selecting indicators as inputs and outputs, credit Customer performance of bank branches is calculated to the discriminate analysis and data envelopment analysis the amount of output to input. In this activity, it should use one of the existing software used for data envelopment analysis Which the GAMS software used.

Definitions of terms and variables

Validation: evaluating the affordability of the applicants to repay loans and financial facilities and the risk of non-repayment of funds received by them.
Credit risk: is the potential that a bank borrower or by his record fails in implementing its obligations In front of the bank within the period specified.
Performance: In terms of economic theory, efficiency is the result optimizing the production of goods and services and the allocation of resources.
Discriminant analysis and classification: Discriminant analysis is concerned with finding a rule to distinguish between two or more of multivariate. The most important application of discriminant analysis is ranking.
The decision on a P observation of variable belongs to which of two or more competing society, is called Classification. For example, suppose a client with different characteristics such as age, sex, educational level, qualification such as being Well-known customer or not, the turnover on the account, the amount of credit in the community, workplace and residence, introducer, jobs and time of engagement in it, place of business, business licenses, being famous in the city or area or neighborhood, being native or non-native, culture, religion and adhering to it, past financial records and facilities, indicators of character and so on entered the bank to get facilities. By statistical methods how can we decide if he is good or bad customer? In the other words, is he Creditworthy or not and whether or not we lend he our facility? Discriminant analysis and classification methods are used to answer this question. Numerous applications for discriminant analysis and classification can be imagined. In fact, in many cases, identify the type of client can be based on visible signs based on Discriminant analysis and classification methods.
DEA method: DEA is meant a mathematical programming model to evaluate the efficiency of decision making units which has several inputs and outputs. In DEA Model with input view, we are seeking to gain technical inefficiency that should reduce input and Output remains unchanged, and the unit's efficiency at the border.
In The output point of view, we are looking towards the fact that we should raise output without having to change the input to achieve the efficiency frontier.

Classification: deciding whether A P observation belongs to which of two or more competing society, is called Classification.

**data analysis**

DEA and discriminant analysis techniques is the point used in this study. These techniques and the extensive use of it has been introduced in the second section and has in this chapter. After data collection, DA model is formulated using GAMS software. Based on the results of resolving model, we start analyzing the information described above.

**Definition 3.1.**

Let S1 and S2 are set to non-empty in E_n, we call Hyperplane H = \{x: p^t x = \alpha\} a hyper plane S1 and S2 separator, If p^t x \geq \alpha for each x \in S_1 and p^t x \leq \alpha for each x \in S_2.

**Discriminant analysis**

Discriminant analysis is a technique or method to separate the different groups, Observations and assign new observations to the previously defined categories. The point of trying to find differences is to separate the numerical values of the groups as much as possible. To classify these groups is the emphasis on obtaining a rule that enable us to observe the new specific groups. The main goal of these techniques is to obtain a membership base in anticipation of a new view. In Discriminant analysis, this rule can be obtained from Parametric and non-parametric methods. This classification can be done by one of the statistical methods or Programming. In statistical methods of Discriminant analysis it is assumed that the data are multivariate normal distribution. When their distribution is ignored, statistical methods are not used. (Barkhordari Ahmadi, 2005)

**A description of the DEA**

Chaornesand Rorez and Rhodes in their paper, defined this technique as follows: Data envelopment analysis is a mathematical programming model for the observed data which is the new method for estimating the weight of experimental or provides a function of the efficiency frontier which is the basis of modern economics. (Rezai et al., 2008)

**Collective model**

While the output of the input of the axis while maintaining a given level acts is appropriate and possible to reduce the amount of data and vice versa, the data-driven model output increase received In proportional. Collective model is a model that simultaneously reduces the data to consideration.(Tehran and Amin, 2011)

Collective model was also introduced by Charnes, Cooper, Golani, Seafor, Stutes. The general form of the linear programming model are based on model (2-19):

\[
\begin{align*}
\text{Max} & \quad \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \\
\text{S.t.} & \quad \sum_{i=1}^{m} \lambda_ix_{ij} + s_i^- = x_{ip}, \quad i=1,...,m \\
& \quad \sum_{j=1}^{n} \lambda_jy_{rj} - s_r^+ = y_{rp}, \quad r=1,...,s \\
& \quad \sum_{j=1}^{n} \lambda_j = 1 
\end{align*}
\]
\[ \lambda_j, s_i^-, s_r^+ \geq 0 \quad i=1,...,m, \quad r=1,...,s \quad j=1,...,n \]

In this model, \( s_r^+ \) covariates related to The limitation of \( r \)-th output
And \( s_i^- \) auxiliary variable \( i \) is related to restrictions on input states.
The review Unit is efficient when the value of all covariates in the optimal solution is zero.
The \( s_i^- = 0 \) and \( s_r^+ = 0 \).
When a non-efficient unit which The auxiliary variables are non-zero. Covariates reflect the resources or the lack of efficiency in input and outputs corresponding to their limits. (Rezai et al., 2008)
Collective model to move up or down along the input axis and also to move the left or right along the output axis is valid. The optimal solution of the problem is not changed by The transmission input and output. (Mansouri, 2005)

In DEA a same set of factors is measured for each unit that any agent is as an input or as an output. Efficient and inefficient units can be classified into two groups. Group membership is unknown at the beginning and the aim is to determine this point that the units belong to which group. Using the performance obtained from the DEA and the threshold value is equal to a separation of the two groups. Each unit has a performance value less than 1 is classified as inefficient and units with the performance value more than 1 are classified as efficient. Using the optimal set of weights, for each unit a corresponding hyperplane is determined and tries to classify the unit. In DEA corresponds to each unit, a linear programming problem is solved and a hyperplane is created. After solving the problem of linear programming, hyperplanes constitute a linear piece of the border. (Barkhordari Ahmadi, 2004)

Discriminant analysis (Hossein Lotfi Zadeh, 2005) (Freed, N 1981) is a method for separating a set of observations which is classified in two groups so that for all observations a same set of factors are measured and a set of weights searched in a way that observation is the best for isolation. The resulting set of weights for new observations used to predict group membership. First, it is assumed that members are specified groups then a set of weights and a threshold value is defined for a hyperplane which separates the two groups apart. This separator hyper plane is used for prediction of The new observations membership. Suppose \( n \) observations each with \( k \) parameters (indexes, agent) are
independent and $z_{ij}$ are factors of $i$ ($i = 1, \ldots, k$) from j-th observation ($j = 1, \ldots, n$). Let $n_1$ observed in the first group (G1) and $n_2$ observed in the second group (G2) is that $n_1 + n_2 = n$ $V$ G1 $U$ G2 = G.

The target is to find weight sets $\alpha_i$ and a clear threshold value $d$ for a hyperplane so that the hyperplane tries to divide it into two groups as the following: (Saghaei, 2008)

$$\sum_{j=1}^{k} \alpha_i z_{ij} \geq d, \quad j \in J_2 = \{ j / z_j = (x_j, y_j) \in G_2 \}. \quad (3-4)$$

**DEA - Discriminant analysis**

Before explaining the process of DEA-DA, about the DEA and DA and what features it offered. Figure (3-1) and (3-2) respectively show the DEA and DA. Figure (3-1) has two axes $y_1 / x$ and $y_2 / x$ and all observations are shown. DEA algorithm efficient frontier (a-b-c-d) is determined, then the performance-per-view (DMU) calculate by comparison with the efficient frontier. DEA efficient frontier is piecewise linear set of lines that part of the convex hull is the result of all DMUs and in Figure 2 is shown.

In addition, in figure (3), the relationship between Linear and observations has been shown. Separating line is determined by minimizing the sum of deviations (MSD). The line located between the two groups of observations.
Suppose there are \( n \) observations and each observation has \( k \) independent factor to determine its performance, which are expressed by \( z_{ij} \). In DA, all observations are sorted into two groups of \( G_1 \) and \( G_2 \) with respectively \( n_1 \) and \( n_2 \) observations (members). So the \( n_1 + n_2 = n \).

Given these assumptions, DA may be assuming a linear separating hyperplane separates the two groups or that some of the observations in the group \( G_2 \) are categorized in group \( G_1 \) and some observations in the group \( G_1 \) is classified in group \( G_2 \). A number of observations that cannot be clearly classified, are located in a set named overlap set, representing the incorrect classification of DA. In DA, separating hyperplane is expressed by a linear function. DEA technique can clear the screen by piecewise linear form. Sueyoshi used this assumption and presented a new technique. The new technique is coalescence of mass integration of DEA model and DA with the criterion of obtaining minimum of the sum of the deviations. This technique is known as DEA / DA which is a two-step calculation process.

**Effective institutional index on bank credit customers**

In this study, Due to the multitude of indicators and criteria that can be used to assess bank credit customers, measures taken to identify indicators by referring to References including books and articles and the opinions and views of experts in bank. Many studies have been conducted to evaluate Credit Customers and each is appropriate to the situation and conditions in the target bank and a society that bank credit customer reviews in which it is located, some indicators to assess the customer's bank creditors have been considered. Due to the nature of the model used in this study, this research must set a number of parameters that affect the customer's credit rating and collect information about them.

Library studies and interviews with experts Bank and considering the importance of indicators in assessing bank credit customers, according to banking experts, indicators that were less important or for them the information was not available, were removed. And finally 8 indicators that were most effective were chosen which are as follows: Education, occupation, age, amount of collateral, loan amount, Loan interest rate (%), the monthly installment amount and the penalty rate (percentage).

**Grouping data**

In this study, using data envelopment analysis -Discriminant analysis is used. As mentioned in previous sections, in discriminant analysis, Unlike DEA, groups are specified in a way that predict the new observations belong to which group. In other words, Discriminant analysis finds new sampling method to predict membership have been set in groups. In Discriminant analysis, it is assumed that all observations (\( G \)) are divided into two groups. The first group (\( G_1 \)) and second one (\( G_2 \)) which respectively have \( n_1 \) and \( n_2 \) members. Based on the above explanation, in the absence of overlaps, clear screen can be split between the two groups to form \( p^T z = d \) in which \( p \) is normal vector and \( d \) is a fixed amount. Accordingly, in order to collect data, a data set of 500 were selected randomly from real clients of Mellat Bank who have been granted facilities during the period 2011 to 2013 and information on education indicators, occupation, age, bail, loan amount, loan interest rate (percent), the amount of monthly payments and the penalty rate (percent) were classified. The data were divided based on each client's mortgage arrears, creditworthy groups (\( G_1 \)), this means that customers who
have not instalment arrears,(Table 1) and not creditworthy group (G2), those who had installments arrears (Table 2), respectively.

Accordingly, the number of customers who were in the group G1 was 287 and the number of people who were placed in Group G2 was 213.

The research model

In this study, using data envelopment analysis-Discriminant analysis we rate bank credit customer of Mellat bank in region 2 of Tehran In 2011 and 2013 as described above. In the third chapter of this study, this technique was introduced. The model used in this study was a new technique coalesced the collective model with standard DA-DEA and minimized the total deviation. This technique is known as DEA / DA which is a two-step calculation process.

First stage:
Categories of data sets in G1 and G2 or an overlap G1 ∩ G2.
Second stage:
Categories reuse common data G1 ∩ G2 in each of G1 and G2.
In the absence of overlapping, algorithm can be one-step process to separate the two groups, when there is overlap between the two groups. A multi-step algorithm is required.

The general form of this method is as follows:
1. first stage (to determine the classification and overlap):

\[
\begin{align*}
\text{Min} & \quad \sum_{j \in J_1} s_{1j} + \sum_{j \in J_2} s_{2j} \\
\text{s.t.} & \quad \sum_{i=1}^{k} \alpha_i z_{ij} + s_{1j}^+ - s_{1j}^- = d, \quad j \in J_1, \\
& \quad \sum_{i=1}^{k} \beta_i z_{ij} + s_{2j}^+ - s_{2j}^- = d - \eta, \quad j \in J_2, \\
& \quad \sum_{i=1}^{k} \alpha_i = 1 - 2u, \\
& \quad \sum_{i=1}^{k} \beta_i = 1 - 2v, \\
& \quad s_{1j}^+ , s_{1j}^- \geq 0, \quad j \in J_1, s_{2j}^+ , s_{2j}^- \geq 0, \quad j \in J_2, \\
& \quad \alpha_i \geq 0, \beta_i \geq 0, \quad i = 1, \ldots, k.
\end{align*}
\]

That \(s_{1j}^+\) and \(s_{1j}^-\) in \(J_1\) are positive and negative deviation linear function of \(\sum_{i=1}^{k} \alpha_i z_{ij}\), the \(d\) value of the group \(G_1\). Positive deviation \(s_{1j}^+ > 0, j \in J_1\) shows an example of incorrect classification of the \(j\)-th observation in \(G_1\) which is included in the model (8-3) to reduce the incorrect classification.

The negative deviation \((s_{1j}^- > 0, j \in J_1)\) shows an example of correct classification which gives details about deviations in \(G_1\). The same can be said for \(G_2\) also. \(s_{2j}^+ \) and \(s_{2j}^-\) \(j \in J_2\) are positive and negative deviation of linear function \(\sum_{i=1}^{k} \beta_i z_{ij}\) the value of \(d-\eta\) are in the group \(G_2\) which a positive number \(\eta\) has been added to avoid the obvious answer in (1-4). In this case a negative deviation \((s_{2j}^- > 0, j \in J_2)\) show an example of an incorrect classification of group. While positive deviation \((s_{2j}^+ > 0, j \in J_2)\) show correct classification.

The statues of \(m\)-th observation of the most recent sample, the value of which is expressed by \(z_{im}\), can be defined as follows:

\begin{align*}
& (a) \quad \text{if } \sum_{i=1}^{k} \alpha_i^+ z_{im} \leq d^- < \sum_{i=1}^{k} \alpha_i^- z_{im} \quad \rightarrow \quad \sum_{i=1}^{k} \alpha_i^+ z_{im} \leq d^- < \sum_{i=1}^{k} \beta_i^- z_{im} \text{then, } \\
& \quad z_{im} \in G_1 \cap G_2 \\
& (b) \quad \text{if } \sum_{i=1}^{k} \alpha_i^+ z_{im} \geq d^+ \geq \sum_{i=1}^{k} \beta_i^+ z_{im} \geq d^+ \quad \text{then } z_{im} \in G_1 . \\
& (c) \quad \sum_{i=1}^{k} \alpha_i^- z_{im} < d^- \quad \sum_{i=1}^{k} \beta_i^- z_{im} < d^- \quad \text{then } z_{im} \in G_2 . \\
& (d)
\end{align*}
The second phase (to achieve overlap)

There overlap (\(G_1 \cap G_2\)) determines the new observation belongs to both groups. If such a situation occurs, DEA / DA must predict that whether new observation belongs to \(G_1\) or \(G_2\).

Model (7-3) has been proposed to address overlap determined by the first stage of DEA / DA:

\[
\begin{align*}
\text{Min} & \quad \varphi = \sum_{j \in J_1} s_{ij}^+ + \sum_{j \in J_2} s_{ij}^- \\
\text{s.t} & \quad \sum_{i=1}^{k} Y_i z_{ij} + s_{ij}^+ - s_{ij}^- = d, \quad j \in J_1, \\
& \quad \sum_{i=1}^{k} Y_i z_{ij} + s_{ij}^+ - s_{ij}^- = \eta, \quad j \in J_2, \\
& \quad \sum_{j=1}^{k} Y_i = 1, \\
& \quad s_{ij}^+, s_{ij}^- \geq 0, \quad j \in J_1, s_{ij}^-, s_{ij}^+ \geq 0, \quad j \in J_2, \\
& \quad \gamma_i \geq 0, \quad i=1,\ldots,k. 
\end{align*}
\]  

(9-3)

After Optimized \(y^*\) obtained from Model (9-3), all observations which are in \(G_1 \cap G_2\), are classified by comparison of \(\sum_{i=1}^{k} Y_i z_{ij}\) with time threshold \(d^*\) In each group \(G_1\) or \(G_2\) as follows:

- a) \(\forall j \in J_1 \cap J_2, \sum_{i=1}^{k} Y_i z_{ij} > d^* \Rightarrow z_{ij} \in G_1\)
- b) \(\forall j \in J_1 \cap J_2, \sum_{i=1}^{k} Y_i z_{ij} < d^* \Rightarrow z_{ij} \in G_2\)

Two conditions for classification of a number of observations belong to the overlap \((G_1 \cap G_2)\) are used.

Analysis Tool Data

Required calculations done by the software GAMS.

Research variables

In this study, the indicators chosen for rating of the bank’s clients are shown in the attached tables (1) and (2). Variables shown \((I_1, \ldots, I_8)\), are defined in each table is defined as follows:

- \(I_1\): customer education
- \(I_2\): job
- \(I_3\): the age of the customer
- \(I_4\): credit rate (percent)
- \(I_5\): penalty rate (percent)
- \(I_6\): Security Deposit
- \(I_7\): loan amount
- \(I_8\): the amount of monthly installment

Analysis of the results

As mentioned, the data collected in the two groups \(G_1\) and \(G_2\) in order to perform calculations and obtaining outcome in the form of Discriminant analysis model of GAMS software. Weights values \(\alpha_1,\ldots,8\) and \(\beta_1,\ldots,8\) and the threshold value \(d\) are shown in tables (1) and (2).
Table 2. calculated weights of $G_2$

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.38640133</td>
<td>0</td>
<td>0</td>
<td>0.61359867</td>
<td>0</td>
</tr>
</tbody>
</table>

The threshold value is obtained as follows:

$$d^* = 15.95$$

However, to achieve weight values $\alpha_1, \ldots, \alpha_8$ and $\beta_1, \ldots, \beta_8$ and the threshold value $d$, you can determine the $m$-th status of last sampling which its value is expressed by $z_{im}$ as follows:

a) if $\sum_{i=1}^{k} \beta_i^* z_{im} \leq d^* < \sum_{i=1}^{k} \alpha_i^* z_{im}$ then $z_{im} \in G_1 \cap G_2$

b) if $\sum_{i=1}^{k} \alpha_i^* z_{im} \geq d^*$ then $z_{im} \in G_1$.

c) if $\sum_{i=1}^{k} \alpha_i^* z_{im} < d^*$ then $z_{im} \in G_2$

Example 1:
Consider a real customer example with indicators $I_1, \ldots, I_8$ (Table 3):

Table 3. indicators of customer A

| Customer name | 10 | 40 | 35 | 13 | 19 | 177 | 142 | 19 |

Then we have:

$$\sum_{i=1}^{8} \alpha_i^* z_{im}$$

As you see, numerical values $\sum_{i=1}^{8} \alpha_i^* z_{im}$ and $\sum_{i=1}^{8} \beta_i^* z_{im}$ both are larger than the threshold value $d$, so the second case (Case B) is established and $z_{im}$ belongs to $G_1$. In other words, it is predicted that if the customer A has received bank facilities, he/she will be considered creditworthy customer.

Example 2:
Consider Customer B with real customer indicators $I_1, \ldots, I_8$ (Table 4):

Table 4. indicators of customer B

| Customer name | 10 | 35 | 30 | 12 | 19 | 3 | 30 | 10 |

Then we have:

$$\sum_{i=1}^{8} \alpha_i^* z_{im}$$
As you see, Numerical values $\sum_{i=1}^{8} \alpha_i z_{ia}$ and $\sum_{i=1}^{8} \beta_i z_{ia}$ both are lower than the threshold value $d$, so the third case (Case B) is established and $z_{ia}$ belongs to $G_2$. In other words, it is predicted that if the customer B has received bank facilities, he/she will be considered not creditworthy Customer.

**Example 3**
Consider Customer C with real customer indicators I$_1$, ..., I$_8$ (Table 5):

<table>
<thead>
<tr>
<th>Customer name</th>
<th>I$_1$</th>
<th>I$_2$</th>
<th>I$_3$</th>
<th>I$_4$</th>
<th>I$_5$</th>
<th>I$_6$</th>
<th>I$_7$</th>
<th>I$_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>10</td>
<td>37</td>
<td>27</td>
<td>4</td>
<td>12</td>
<td>27</td>
<td>20</td>
<td>6.06</td>
</tr>
</tbody>
</table>

Then we have:

$$\sum_{i=1}^{8} \alpha_i z_{ia} = 6.06$$

$$\sum_{i=1}^{8} \beta_i z_{ia} = 7.27$$

Just as you can see, $\sum_{i=1}^{8} \alpha_i z_{ia} < d < \sum_{i=1}^{8} \beta_i z_{ia}$, so the first case (Case A) is established. In other words, there is overlap. The overlap $(G_1 \cap G_2)$ determines the new observation belongs to both groups. If such a situation occurs, DEA / DA must predict whether the new observation belongs to $G_1$ or $G_2$.

In order to determine whether the customer is creditworthy or bad account, Model of overlap As previously noted, We solve. After running the model and gaining $\gamma_1^*$, we have:

$$d = 4.39$$

So when $\sum_{i=1}^{8} \gamma_i^* z_{ic} = -4.306 > d^* = 4.39$

Therefore, It is expected that the client C Bad customer's account in the group.

**Example 4**
Consider customers D with real indicators of I$_1$, ..., I$_8$ in table 6:

<table>
<thead>
<tr>
<th>Customer name</th>
<th>I$_1$</th>
<th>I$_2$</th>
<th>I$_3$</th>
<th>I$_4$</th>
<th>I$_5$</th>
<th>I$_6$</th>
<th>I$_7$</th>
<th>I$_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>16</td>
<td>40</td>
<td>38</td>
<td>13</td>
<td>19</td>
<td>68</td>
<td>18</td>
<td>24</td>
</tr>
</tbody>
</table>

So we have:
Just as you can see, $\alpha_1$, $\beta_1, \gamma_1$, so the first case (Case A) is established. In other words, there is an overlap. Existence of overlap ($G_1 \cap G_2$) determines that the new observation belongs to both groups. If such a situation occurs, DEA / DA must predict that whether new observation belongs to $G_1$ or $G_2$. In order to determine whether the customer is creditworthy or non-creditworthy, we solve

Model of the overlap mentioned above: After running the model and gaining $\gamma_i^*$, we have:

$$
\sum_{i=1}^{8} \beta_i z_{ic} = (0 * 16) + (0.38640133 \times 40) + (0 * 38) + (0 * 13) + (0 * 19) + (0.61 \times 68) \\
+ (0 * 18) + (0.18 \times 24) = 57.18 > \frac{d^*}{2}
$$

Just as you can see, $\sum_{i=1}^{8} \alpha_i^* z_{ia} < \sum_{i=1}^{8} \beta_i^* z_{ia}$, so the first case (Case A) is established. In other words, there is an overlap. Existence of overlap ($G_1 \cap G_2$) determines that the new observation belongs to both groups. If such a situation occurs, DEA / DA must predict that whether new observation belongs to $G_1$ or $G_2$. In order to determine whether the customer is creditworthy or non-creditworthy, we solve

Model of the overlap mentioned above: After running the model and gaining $\gamma_i^*$, we have:

$$
\sum_{i=1}^{8} \gamma_i^* z_{ic} = (0.34 \times 16) + (0.043 \times 40) + (0.2 \times 38) + (0.23 \times 13) + (-0.19 \times 19) + (0.042 \times 68) \\
+ (-1.17 \times 18) + (1.499 \times 24) = 32.128 > \frac{d^*}{2} = 4.39
$$

Since $\sum_{i=1}^{8} \gamma_i^* z_{ic} = -4.306 > \frac{d^*}{2} = 4.39$, therefore, it is predicted that customer D is creditworthy.

Conclusion:

In this study, we used discriminant analysis to check the customer's credit rating until through its results be able to predict new credit customers' rating. There are two types of modeling point of view. The first idea tries to minimize the number of errors in the two series. The second view tries to minimize the errors in modeling hyperplane. The second approach was used in this study. And it is assumed that two is the number of classes of customers. The results show the index review, indicators of job type, penalty rates between good customers, the standards of education, the amount of collateral and the amount of installment loans between non-creditworthy clients, and those who have not paid installments and loan, and more weight is important. In general it can be said that the findings suggest that Banks when lending to their customers should be careful on job characteristics, penalty rates, education, the amount of collateral and the amount of installment loans. This model can be used as a decision support system to help authorities of bank loans. In this way, Not only of lending to non-creditworthy customers prevented, but also measures can be taken (including getting adequate collateral) for lending to non-creditworthy customers or creditworthy ones. So in addition to taking practical steps, to reduce the volume of bank's claims can increase the level of customer satisfaction.

Level of customer satisfaction also increased creditworthy.

Suggestions:

1. The use of the pattern obtained in the investigation for credit rating of credit customers
2. Use of information obtained and analyzing them to establish plans and use in relevant decision-making
3. The rating corporate natural credit clients of banks with this method and with regard to financial indicators
4. Developing customer rating models into three groups:
References


Saghaei, Mahdi. " Discriminate Analysis (DA) And Data Envelopment Analysis (DEA) For Interval Data", ISLAMIC AZAD UNIVERSITY, Ghaemshahr Branch, Autumn 2008

