
Credit Risk Assessment Utilizing Data Reduction Technique for Individual Loaning in Financial Institutes (Case Study: Tejarat Bank, Rasht, Iran)

Abbas Mahmoudabadi^{1,*}, Matin Mehrshad¹, Mohammad Reza Aminnaseri²

¹Department of Industrial Engineering, Mehr Astan University, Astaneh-Ashrafieh, Guilan, Iran

²Department of Industrial Engineering, Tarbiat Modares University, Tehran, Iran

Email address:

mahmoudabadi@mehrastan.ac.ir (A. Mahmoudabadi)

*Corresponding author

To cite this article:

Abbas Mahmoudabadi, Matin Mehrshad, Mohammad Reza Aminnaseri. Credit Risk Assessment Utilizing Data Reduction Technique for Individual Loaning in Financial Institutes (Case Study: Tejarat Bank, Rasht, Iran). *American Journal of Theoretical and Applied Business*. Vol. 3, No. 1, 2017, pp. 11-17. doi: 10.11648/j.ajtab.20170301.12

Received: March 16, 2017; **Accepted:** April 25, 2017; **Published:** May 30, 2017

Abstract: Because of the nature of the financial and economic activities and they are practically accompanied with a degree of risk., banks are usually dealing with many risks, including operational, marketing, interest rate, etc. Since, credit risk has significant effects on financial banks activities in terms of loaning profits, the risk of repayment individual loans has been investigated in this research work. Two well-known regression models of Probit and Logistic have been developed based on nine extracted factors which have been investigated during the offering of loans according to the possibility of late or non-repayment. In order to minimize inter-correlation and extracting high-independency factors, the statistical technique of Principal Component Analysis (PCA), categorized as a data reduction technique, has been utilized and three factors out of nine have been omitted. One of Tejarat bank branches in the Iranian Northern Province of Guilan has been selected as case study to gather experimental data for assessing the credit risk of individual bank investors. The results of model validation revealed that the implementation of PCA method can improve the accuracy of models' outputs and Probit regression model has better results rather than Logit one.

Keywords: Non-repayment Loaning, Credit Risk Evaluation, Individual Investors, Principal Component Analysis, Probit and Logit Regression Models

1. Introduction

Risk is a part of human and organization life, so financial managers are mainly dealing with the degree of risk to make their decisions. Since, risk cannot be completely removed, scientific attitude for risk is leading to evaluate and manage it. Since, financial banks are a great pillar of the economy all around the world, they are dealing with many kinds of risks of operational, marketing, interest rate, credit and liquidity and etc because of the nature of their financial loaning [1]. One of the most important activities in banking system is loan payment and credit risk stems from this fact that one of the sides of the loaning contract cannot or will not perform its obligations. Non-repayment risk, caused by individual investor's, is still considered as a major cause of banks

failure claimed to be still rising and growing according to published researches [2]. Therefore, what is important for banks is to grant loans to individuals or organizations those meet the standards of getting loans and are able to repay their loans on time. So, it is necessary to assess the credit risk of payments when individual investors receive loans from banks.

Following the above mentioned, the main aim of this research work is to assess the risk of non-repayment loans by individual investors considering the effective factors which have been sieved using a data reduction technique of principal component analysis (PCA). This paper is organized into seven sections. After introduction, scientific background is discussed followed by developing model for assessing risk two regression models of Logit and Probit. Case study and

data are following discussed in the fourth section together with computational results and conclusion in the last section.

2. Scientific Background

2.1. Literature Review

In the field of assessing credit risk of borrowers, many studies have been conducted and results published. For example, a decision tree model was offered for bank credit risk evaluation using data-extracting techniques to identify affective factors on credit risk of customers and divided them into four groups including past due, suspected of receipt, the outstanding and ongoing and revealed that the possibility of assigning the repayment loans to each group are different customer to customer [3]. The model suggests that new criteria in analyzing the loan application have different influences on repayment loans and evaluating results showed that the proposed model can reduce non-performing loans to less than 5% and as well as banks can be classified to well performing grade [3]. Using three factors in credit scoring including re-payment period (in month), loan amount and customer's age, Dong et al. [4] utilized logistic regression with random coefficients on credit scoring of individual customers. They divided the above mentioned factors into four intervals and allocated ranking points from 1 to 4 for any interval. In terms of prediction accuracy, the performance of their proposed method was much better than logistic regression comparing to constant coefficients.

Probit analysis technique was also used to estimate the probability of credit risk [5]. Influencing factors including personal and financial attributes have been investigated for estimating credit risk for Swedish credit customers on both rejected and approved applicants. Applying a fuzzy analytical hierarchy process and computing Value at Risk (VAR for short) showed that the efficient selection of loan applicants can reduce credit risk by 80% [5]. By utilizing Meta-heuristic approach in the field of assessing financial risk, Che et al. [6] stated that while there is a strong competition between Taiwan financial institutions and banks, high proportion of outstanding deposits can be observed over the loaning systems particularly on small and medium companies in Taiwan. Many warning systems, designed for commercial loans risk, are usually composed of artificial neural network and the concept of traditional warning. Baon et al [7] used the artificial neural network (ANN) to propose a warning system for risk assessment of banking loans. Long and short term repayment abilities, performance and profit abilities and eventually degree of required warning have been used as evaluation criteria for developing the ANN model. The outputs of systems are more understandable and practical rather than ones in normal warning systems and have provided effective decision-making tools in bank loans for the companies who are dealing with loans risk assessment. Different views on assessing credit risk were pointed out for evaluating the credit risk of companies and international firms (corporate customers), banks and the other institutions

lending to these companies [8]. Company's international history, international relations and the market in countries have been identified as affecting factors to assess the credit risk of companies while developing regression model shows that the credit risk of international companies can arise from experience of companies and current business activities [8]. Data envelopment analysis (DEA) is another technique which helps to offer appropriate approaches for credit scoring [9]. In contrast to the regression analysis and neural networks models which need additional data to calculate credit point, these kinds of models only use historical records the credit information. Collecting financial data on 1061 foreign companies from the Korea Credit Guarantee and using financial ratios authors [9] combined the overall performance of a company and calculated the company credit rating.

As observed, many studies have been made for accessing credit risk, but loaning by individual investors who are more popular in developing countries should be carefully investigated for late or non repayment loans while many fields of data may be unavailable. Therefore, the main aim in this research work is to study the risk of non-repayment loans of individual investors as well as utilizing an efficient statistical technique of principal component analysis to omit factors which have inter-correlated with the other influencing factors for gathering as less necessary as data should be gathered for carrying out these kinds of studies.

2.2. Principal Component Analysis

Principal component analysis (abbreviated as PCA) is a statistical analysis technique used to determine the most influencing variables. Using the PCA, the number of trivial data is deleted to reduce the inter-correlation as well as improve the independency of variables. The main stage of PCA is Kaiser-Meyer-Olkin (KMO) test in which the KMO index checks if factors are efficiently related to the original variables or not [10]. Calculating the correlation matrix is the starting point. It is assumed that the variables are more or less correlated, but the correlation between two variables can be influenced by the others. So, the partial correlation is used in order to measure the relation between two variables by removing the effects of the remaining variables. The KMO index compares correlations between variables and those of the partial correlations. If KMO is less than 0.5, data would not be appropriate to utilize factor analyze means that there is no significant correlation, so they are suitable for statistical analysis because of less or non-correlated variables. Applying this method, the combinations of P primary variable X_1, X_2, \dots, X_p for maximum P independent component, defined as PC_1, PC_2, \dots, PC_p , are created. Each component can be determined with a presented sequence by series of equations as (1) [10].

$$\begin{aligned} PC_1 &= w_{11}x_1 + w_{12}x_2 + \dots + w_{1p}x_p \\ PC_2 &= w_{21}x_1 + w_{22}x_2 + \dots + w_{2p}x_p \\ PC_p &= w_{p1}x_1 + w_{p2}x_2 + \dots + w_{pp}x_p \end{aligned} \quad (1)$$

Where PC_i represents new generated variables known as factors, W_{ij} is weighing coefficients of primary variable and X_i is the i th primary variable. W_{ij} is estimated according to the variation of variables, so that the first component is considered as the maximum variance and the second component predicts not-intended maximum variance which has not been defined by the first component. In addition, two constraints should be applied for obtaining independency presented as equations (2) and (3).

$$w_{i1}^2 + w_{i2}^2 + \dots + w_{ip}^2 = 1, i = 1, \dots, p \quad (2)$$

$$w_{i1} \times w_{j1} + w_{i2} \times w_{j2} + \dots + w_{ip} \times w_{jp} = 0, i \neq j \quad (3)$$

For analyzing PCA, the following steps are performed:

- (1) Standardization of input variables: At the first stage, input data should be standardized in terms of scale in a way that has zero mean and deviation benchmark one. Z (Normal Standard) matrix, the matrix includes standardized values of the parameters can be obtained from equation (4) where \bar{x}_j is the average data and s_j is the related standard deviation of samples [11].

$$Z_{ij} = \frac{(x_{ij} - \bar{x}_j)}{s_j} \text{ for } i = 1, 2, \dots, n, j = 1, 2, \dots, q \quad (4)$$

- (2) Calculating the KMO factor: KMO is used in the range of zero to one. The index is derived from the equation (5), where r_{ij} and a_{ij} are respectively the correlation coefficient and partial correlation coefficient between variables i and j [12].

$$KMO = \frac{\sum \sum r_{ij}^2}{\sum \sum r_{ij}^2 + \sum \sum a_{ij}^2} \quad (5)$$

- (3) Calculating the correlation matrix (variance) for the primary variables: R matrix shows the correlation between each of the basic variables [11]. Z is standard variables of the main variables calculated by equation (6).

$$R = \frac{1}{n} Z' Z \quad (6)$$

- (4) Calculating λ eigenvalues and corresponding eigenvectors of the correlation matrix: By using equations (7) and (8) the eigenvalues and eigenvectors of any particular value are calculated [11]. Special vectors obtained, especially as coefficients for all basic variables in the respective components are formed. Solving the equation (7), where I is the identity matrix, eigenvalues (λ_n) are calculated. The variance of each principal component is calculated by equation (8).

$$\det(R - \lambda I) = 0 \quad (7)$$

$$\det(R - \lambda I) = V_h \quad (8)$$

- (5) Benchmarking the number of operations: Special value standard, variance percentage standard and cut

test standard [11] are the most important parameters for extracting the number factors that are used in this process.

- (6) Implementation of proper rotation on the components coefficient matrix: At this stage, the variables that have high coefficients in extracted main components are selected as important variables to enter to the modeling [13] and the rest will be omitted.
- (7) Eventually, after Implementation of proper rotation, the main variables are those at least one of their coefficients uses to form the relevant factor, has relatively high amount and others should be deleted.

2.3. Probit Regression Model

The Probit regression method firstly proposed in 1930 [14] is applied in two types of two-binomial (two states) and ratings variables. Probit binomial regression is used for quantitative and qualitative measures to explain and predict the dependent variable to measure two states (binary) based on a set of independent variables. Probit regression procedure is used when dependent variable is measured by a rating scale. The general form of the dependent variable in two states Probit model is defined as a function such equation (9), which can be identified as zero or one. In equation (10), X is a vector of independent variables and β is the vector of parameters that must be estimated using available data targeting to least square errors of observed and estimated amounts of dependent variable [15].

$$y_i = \begin{cases} 1 & \text{First status} \\ 0 & \text{Second status} \end{cases} \quad (9)$$

$$y_i = x_i \beta + \epsilon_i \quad (10)$$

Regarding to figure 1, it is assumed that dependent variable with binomial distribution in terms of its observations can be defined as normal distribution in which a threshold or boundary (t) has made it to the binomial distribution [14]. It is assumed that dependent variable or underlying response has a continuous distribution and follows a normal distribution function. Then Probit link function according to equation (11) is used to calculate the cumulative normal distribution function for modeling and explaining the probability of success ($P(Y = 1)$). In equation (11), symbol Φ represents the cumulative normal distribution function available in all published statistical books [15].

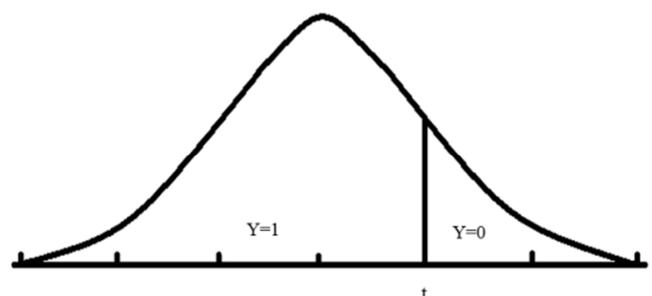


Figure 1. Continuous distribution function.

$$\text{Probit}[p(y=1)] = \phi[p(y=1)] = \phi(x\beta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x\beta)^2} \quad (11)$$

2.4. Binomial Logistic Regression (Logit)

Binomial logistic regression (Logit) is used for modeling the probability of occurrence or non-occurrence of a given situation on a number of independent variables. Since, dependent variable is a binary variable (accept or reject; one or zero), there is the possibility of using linear regression. In this case while dependent variable is a binary variable and independent variables are nominal, ordinal, interval or a combination of them, there is the possibility of using Logit regression to estimate the probability of success for dependent variable Y. In logit regression, dependent variable (Y) follows un-normal and binomial distribution. Independent variables (X_i) can also be a combination of qualitative and quantitative variables. Link function in the Logit regression is the natural logarithm (ln) chance ratio. The chance ratio indicates the likelihood of success versus failure [14] formulated as equation (12). As a result, the binomial logistic regression equation is written as the equation (13).

$$\text{Chance ratio} = \frac{\text{Probability of success}}{\text{Probability of failure}} = \frac{P(Y=1)}{1-P(Y=1)} \quad (12)$$

$$\text{Logit}[P(Y=1)] = \ln \left[\frac{P(Y=1)}{1-P(Y=1)} \right] = a + b_1X_1 + b_2X_2 + \dots \quad (13)$$

With a change in the equation (13), the odds of success can be replaced by its logarithm. For this purpose, the exponential forms of that, $\exp(X) = e^x$, is used as $\exp[\ln(X)] = X$ and Logit regression equation can be rewritten by equation (15) [14].

$$\exp \left[\log \left(\frac{P(Y=1)}{1-P(Y=1)} \right) \right] = \frac{P(Y=1)}{1-P(Y=1)} = e^a e^{b_1x_1} e^{b_2x_2} \dots \quad (14)$$

$$P(Y=1) = \frac{e^a e^{b_1x_1} e^{b_2x_2} \dots}{1 + e^a e^{b_1x_1} e^{b_2x_2} \dots} \quad (15)$$

3. Developing Risk Assessment Models

Since, dependent variable used for conducting the present study, is a binary variable (defined as un-creditworthy or creditworthy customer), two methods of Probit and Logit regression models have been used to calculate the probability of credit risk for individual customers. Principal Component Analysis (PCA) has also been used to insure existing input variable independency. In addition, the impact of its utilization on Probit and Logit models are also investigated. Results of the estimating for probability of creditworthy states have been eventually compared before and after utilizing PCA to check if PCA is useful for improving the accuracies of the proposed models.

3.1. Assessment Criteria for Credit Risk

- (1) Literature reviewing shows that there are 32 effective factors for evaluating the customers' credit of banks.

Nine factors among them have been selected after an interview with managers in the branch of Tejarat Bank Rasht, Guilan, Iran. Data corresponding to nine selected criteria of creditworthy and non-creditworthy customers have been collected over March 21, 2014 to March 20, 2015 (1393 in lunar calendar). The population is a sample size of 45 consisting detailed information on 20 samples on non-creditworthy and 25 on creditworthy customers.

- (2) Dependent variable: The repayment status is considered as dependent variable. It is a discrete variable known as response variable determines the customer repayment ability status. Customers who have not defaulted in their repayments are creditworthy, adopted the value (0) and non-creditworthy customers who have defaulted on their repayments adopted by (1). In this case, the result of the proposed model is defined as the probability of defaulting repayment loans. It means that loaning banks would be aware of non-creditworthy customers.
- (3) Independent variables: Dependent variables have been selected as a set of effective criteria on repayment of loaning investment described in the previous section. They are as follows:

Gender: Gender is a discrete variable includes male (0) and female (1).

Age: Age is a continuous variable reviewed in years.

Job status: Borrowers related to the employment status are divided into two discrete categories based on employment status of having income defined as (0) and no-income defined as (1).

Amount of loan: The amount of loan had been received by borrower. This is a continuous variable.

Repayment period: Repayment period is a time period (in months) that the principal and interest loan to be repaid.

Type of collateral: In the loaning process, collaterals are recommended by customers who guarantee the repayment procedure. Collaterals are in different types but in this research are divided into two types of property collateral (0) and non-property (1)

Term of continuous transactions: The period of time (in years) that the customer has an account in the bank with financial transactions.

Interest rate: Bank interest rate is defined by banks including a wide range from 4% to 29% range.

The number of supporters (guarantors): Following the bank instruction, guarantor (one or more) is required to guarantee the repayment process to the bank if borrower fails to repay that. Otherwise, the client collateral must be property.

One dependent and nine independent variables (credit status) as well as their symbols are shown in table 1. The first one in dependent variable notated by CS and the others are dependent.

Table 1. Dependent and Independent Variables and Their Symbols.

NO	Variables	Symbol
1	Credit status	CS
2	Gender	G
3	Age	A
4	Job status	JS
5	Amount of loan	AL
6	Repayment period	TL
7	Type of collateral	TC
8	Term relationship with bank	TR
9	Interest rate	IR
10	Number of guarantors	Gu

3.2. Binomial Probit Regression Model

Probit binomial regression methods (two states) for all variables have been performed using the well-known statistical software of SPSS. As shown in table 2, outputs indicated that the estimated Probit model has acceptable performance for predicting the dependent variable which is defined as the probability of non-creditworthy. Acceptable performance happens when the likelihood ratio in a significant level is less than 0.05 for confidence level of 95 percent. The observed significant surface is near to zero and less than 0.05. Parameters, statistical coefficients and their corresponding significant levels for both regressions models of Probit and Logit have also been tabulated in table 2. The first column represents notations of parameters (independent parameters). They are defined as one constant factor, three continuous variables and six discrete variables. Probit and logit regression models for each level of qualitative variables except the reference level (the last level is selected by default) are estimated as statistical coefficient and corresponding significant level. As shown in table 2, results of the Probit binomial regression model revealed that four independent variables including employment status (JS), age (A), term relationship with the bank (TR) and the interest rate (IR) have significant effects on the probability of non-creditworthy because their calculated significant levels are less than 0.05. Using experimental data, the proposed Probit model for assessing the real customer's credit risk of branches of Rasht Tejarat bank is formulated by equation (16). In other words, the probability of non-creditworthy accounts and failure to repay loans is followed by equation (16) in which the rate of each variable is defined as a nonlinear formulation.

$$[p(y=1)] = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(3.546 JS - 0.583A + 1.156 TR + 98.090IR)^2} \quad (16)$$

3.3. Logit Regression Model

The implementation of Logit regression method using experimental data indicates that the one is also capable to estimate the acceptable performance in predicting the dependent variable (significant level less than 0.05 for 95% confidence levels). Statistical coefficients and their corresponding significant level have been tabulated in table 2. Results of the Logit binomial regression model showed that three variables of employment status (JS) with

coefficient 5.924, age (A) with coefficient -0.965 and term relationship with the bank (TR) with coefficient 1.905, have significant levels less than 0.05 as shown in the fourth column of table 2. Following the above mentioned, it is concluded that the above independent variables have significant effects on the probability of non-creditworthy. The proposed Logit model for assessing credit risk is formulated as equation (17)

$$[P(Y=1)] = \frac{e^{5.924 JS - 0.965 A + 1.905 TR}}{1 + e^{5.924 JS - 0.965 A + 1.905 TR}} \quad (17)$$

Table 2. Coefficients and significance level for Probit and Logit methods before PCA.

Model	Probit Regression		Logit Regression	
	Coefficients (β)	Significant level	Coefficients (β)	Significant level
TL	33.39	0.201	54.99	0.055
TC	2.501	0.140	4.093	0.153
TR	Na	Na	Na	Na
IR	3.546	[0.033]	5.924	[0.047]
Gu	Na	Na	Na	Na
Symbol	-24.324	0.053	-40.046	0.70
TL	Na	Na	Na	Na
TC	-0.58	[0.02]	-0.97	[0.03]
TR	0.001	0.209	0.26	0.2
IR	-0.018	0.461	0.510	-0.03
Gu	1.156	[0.033]	1.91	[0.047]
Symbol	98.09	[0.043]	161.55	0.057
TL	-24.69	0.055	-40.54	0.074

Two aforesaid models have been utilized using 20 samples of non-creditworthy customers and results have been summarized in table 3. Results revealed that the Probit two-state model by accuracy of 0.3 is valid. In other words, 6 out of 20 acceptable probability (more than 0.50), put in the parenthesis, has been offered for non-creditworthy customers.

4. Performing the Principal Component Analysis

Since, all parameters in decision-making models are not necessarily defined as the same dimensions, normalization or data standardization methods are commonly used for accessing the uniform modes of independent variables.

Table 3. The probability of non-creditworthy customers before applying PCA.

No.	Probit	Logit	No.	Probit	Logit
1	0.0274	≅0	11	(0.9857)	≅0
2	(1)	≅0	12	0	≅0
3	0.1729	≅0	13	0.0150	≅0
4	(0.9354)	≅0	14	0.0001	≅0
5	0	≅0	15	(0.9531)	≅0
6	0.0065	≅0	16	(0.6780)	≅0
7	0	≅0	17	0	≅0
8	(1)	≅0	18	0	≅0
9	0	≅0	19	0	≅0
10	2.0E-13	≅0	20	0.0015	≅0

Assuming that the distribution function is normal, the data standardization method is used by equation (18), where X_{old} is initial value of parameter X_{new} is standard value of parameter. The above equation changes the values of input and output parameters to the normal standard range with average of zero and standard deviation of one.

$$X_{new} = \frac{X_{old} - \bar{X}}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}} \quad (18)$$

Following the utilization of the data standardizing process, calculating the KMO factor for standardized variables revealed that the value of the coefficient is equal to 0.497. Since it is less than 0.5, main component analysis cannot be performed on the whole variables. Therefore, the variable that is correlated with other variables must be eliminated and emitted. In order to identify and eliminate the mostly correlated variable, covariance matrix is calculated and the total amount of each row, total correlation of each variable with all other variables, have been investigated. The variable by much correlated with other variables has been removed and the KMO factor is recalculated. After elimination process, it has been concluded that the type of collateral (TC) of the input data has the biggest correlation with other variables. It has been removed and KMO has been recalculated as 0.523. In this case, results confirmed that the necessary correlation between input variables is enough to perform principal component analysis method.

Eigenvectors have been calculated using equations (7) and (8), the rotation process on the coefficient matrix has also been done as well. Three leading factors have been extracted denoted as (LF₁, LF₂, LF₃) shown in table 4. Since, the main variables are those at least one of their coefficients uses to form the relevant factor, have relatively the highest correlation. This criterion is selected as it is equal to 0.7. As shown in Table 4, two factors of gender (G) and age (A) don't have the estimated coefficients more than 0.7. Therefore they are known as the least significant influencing variables in non-repayment and should be deleted. Other variables have coefficients more than 0.7 shown is table 4. Utilizing Probit and Logit regression models remains with six independent variables and the credit status dependent variable continues. It is necessary to notice that symbol Z identifies the standard form of corresponding variable. For instance, Z(G) identifies the standard form of input variable Gender.

Table 4. Results of the rotated matrix (Eigenvectors).

Variable	(LF ₁)	(LF ₂)	(LF ₃)
Z(G)	-0.268	0.201	-0.560
Z(A)	0.516	-0.543	0.205
Z(JS)	[-0.760]	0.291	-0.035
Z(AL)	0.012	0.074	[0.891]
Z(TL)	-0.200	[-0.742]	0.283
Z(TR)	[0.714]	0.243	0.201
Z(IR)	[0.782]	0.226	0.051
Z(Gu)	0.035	[0.835]	0.163

After removing correlated variables using PCA method, Probit and Logit models have been developed considering six remaining variables of (JS, AL, TL, TR, IR, Gu). The models' outputs revealed that both models have acceptable performance for predicting the dependent variable while the significance level is near to zero. Results for both models revealed that (as shown in table 5), loan repayment period (TL), the interest rate (IR) and guarantor (Gu) have more significant impact on the probability of repayment and have significant level less than 0.05 for 95% confidence levels for Probit model shown in brackets.

By extracting and examining their impact coefficients and significance level, the models of Probit and Logit regressions have been developed as equations (19) and (20), respectively. Validation results of the above proposed models defined by equations (19) and (20) according to 20 samples of non-creditworthy customers are tabulated and shown in table 6. Calculated the probability of risk on non-creditworthy investor put in the parenthesis for more than 0.5 states that the Probit model is validated by an accuracy factor of 0.7 (14 out of 20) rather than Logit one by 0.6 (12 out of 20) credit coefficient has a better performance.

$$[p(Y=1)] = \frac{1}{\sqrt{2\pi}} e^{\left[\frac{1}{2}(-0.017 TL + 9.919 IR - 2.049 Gu)^2\right]} \quad (19)$$

$$[P(Y=1)] = \frac{e^{-0.029 TL} e^{16.311 IR} e^{-3.362 Gu}}{1 + e^{-0.029 TL} e^{16.311 IR} e^{-3.362 Gu}} \quad (20)$$

Table 5. Coefficients and significance level after PCA.

Model	Probit Regression		Logit Regression	
	Coefficients (β)	Significant level	Coefficients (β)	Significant level
Intercept	1.371	0.239	2.229	0.271
JS=0	0.173	0.823	0.317	0.815
JS=1	NA	NA	NA	NA
AL	NA	0.182	-0.001	0.230
TL	-0.017	[0.014]	-0.029	[0.023]
TR	0.056	0.303	0.097	0.351
IR	9.919	[0.017]	16.311	[0.021]
G	-2.049	[0.016]	-3.362	[0.023]

Table 6. Results for the probability of Non-creditworthy Customers after applying PCA.

No.	Logit regression	Probit regression	No.	Logit regression	Probit regression
1	(0.6625)	(0.6620)	11	(0.9932)	(0.9826)
2	(0.8561)	(0.8436)	12	0.1351	0.1379
3	(0.6501)	(0.6393)	13	(0.5507)	(0.5510)
4	(0.5351)	(0.5117)	14	(0.7050)	(0.6815)
5	0.0437	0.0508	15	(0.5853)	(0.5803)
6	(0.5115)	0.4727	16	(0.6625)	(0.6620)
7	(0.5114)	0.4909	17	0.0001	0.0018
8	(0.6656)	(0.6693)	18	(0.8366)	(0.8303)
9	0.0001	0.0020	19	(0.9912)	(0.9800)
10	0.0008	0.0055	20	0.2012	0.1865

5. Conclusion

Risk assessment is one of the main concerns in financial institutes. So, this research work focuses on risk assessment for loaning individual investors. Some affective attributes have been considered to develop both Probit and Logit models for evaluating risk of bank loaning for 45 samples of related data and their situation including credit and non-creditworthy customers. In order to check internal correlation, the well known statistical technique of principal component analysis (PCA) has been utilized and results have been compared before and after the utilization of PCA. Results revealed that utilizing PCA improves the prediction models' accuracies over estimating the probability of being non-creditworthy situation for customers.

Experimental data have been gathered for Tejarat Bank branch in Rasht, the capital city of the Iranian northern province of Guilan. Utilizing two proposed models revealed that the credit risk probability of non-creditworthy for individual customers is mainly affected by three variables of loaning, interest rate and the number of guarantor. In this case, results of running both Probit and Logit models applied on 20 samples of non-creditworthy customers show the accuracy of 0.7 for recognizing non-creditworthy customers. This research work has been examined bank loaning risk assessment for individual customers using experimental data, so developed models can be extended using more effective and detailed data. Due to the existing of a wide range between the amounts of loaning received by private customers, it is recommended to divide customers into two categories including small and large sizes of loaning. Eventually, as application, it is also suggested that two-state Probit model may be used to achieve better result rather than Logit model.

References

- [1] Ebrahim, H. Kazem, Y. Nader, M. Reza, KH., Effects of Risk Parameters (Credit, Operational, Liquidity and Market Risk) on Banking System Efficiency (Studying 15 Top Banks in Iran), *Iranian Economic Review*, (17). 2013.
- [2] Soheib, I, The Determinants of Credit Risk in Commercial Banks of Pakistan, *Journal of Poverty, Investment and Development*, (25). 2422-846X, 2016.
- [3] Mandala N., Badra C., Rian F., Assessing Credit Risk: an Application of Data Mining in a Rural Bank, *Procedia Economics and Finance*, (4). 406- 412, 2012.
- [4] Gang, D., Kin, K., and Jerome, Y., Credit scorecard based on logistic regression with random coefficients, *Procedia Computer Science*, (1), 2463–2468, 2012.
- [5] Jacobson T., Kasper R., Bank lending policy, credit scoring and value-at-risk, *Journal of Banking & Finance*, (27). 615–633, 2003.
- [6] Che, Z. H., Wang, H. S., and Chih-ling, Ch., A fuzzy AHP and DEA approach for making bank loan decisions for small and medium enterprises in Taiwan, *Expert Systems with Applications*, (37), 7189–7199, 2010.
- [7] Baon, Y., Ling, X., Hai, J., and Jing, X., An early warning system for loan risk assessment using artificial neural networks, *knowledge-Based Systems*, (14), 303-306, 2001.
- [8] Kent, E., Sara, J., Jessica, L., and Angelika, L., Modeling firm specific internationalization risk: An application to banks' risk assessment in lending to firms that do international business, *International Business Review*, (23). 1074- 1085, 2014.
- [9] Jae, H., and Young, L., A practical approach to credit scoring, *Expert Systems with Applications*, (35). 1762– 1770, 2008.
- [10] Johnson and, R., Wichern, D., *Applied multivariate statistical analysis*, 3rd Edition, Prentice-Hall Inc., Englewood Cliffs, SA, 590, 1982.
- [11] Caliendo, C., and Parisi, A., Principal component analysis applied to crash data on multilane roads, *Third international SIIV Congress, Bari, Italy, 20-22 September, ANCONA SIIV Vol. 1, Page 1-7, 2005.*
- [12] Hutcheson, G., and Nick, S., *The multivariate social scientist: Introductory statistics using generalized linear models*, Thousand Oaks, CA: Sage Publications, 1999.
- [13] Singha, K., Malika, A., Mohana, D., and Sinhab, S., Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) a case study, *Water Research*, Vol. 38, pp. 3980–3992, 2004.
- [14] Hsiao, CH., *the Econometric Of Panel Data*, *Handbook of Theory and Applications*, Vol. 28, pp. 223-241, 1993.
- [15] Cattell, R., The scree test for the number of the factor, *Multivariate Behavioral Research*, Vol. 1, April, pp 245-276, 1996.
- [16] Horowitz, J., and Savin, N., Binary Response Models: logits, probits and semi parametrics, *Journal of Economic Perspectives*, Vol. 15, NO 4, pp. 43-56, 2001.