Designing Credit Risk Early-Warning System for Individual and Corporate Customers of the Bank Using Multiple Logit Comparison Model and Survival Function

Mirfeiz Fallah Shams
Associate Professor of Azad University, Tehran markaz branch, Tehran, Iran
Islamic Azad University of Central Tehran Branch (Modern Financial Risk Research Group)
fallahshams@gmail.com

Hossein Jahangirnia
Department of Accounting, Qom Branch, Islamic Azad University, Qom, Iran
hosein_jahangirnia@yahoo.com

Reza Gholami Jamkarani
Qom branch, Islamic Azad University, Qom, Iran
accountghom@gmail.com

Hmaidreza Kordlouie
faculty of management and accounting, islamic azad univercity, islamshahr branch
Islamic Azad University of Eslamshahr (Modern Financial Risk Research Group)
(Corresponding author)
hamidreza.kordlouie@gmail.com

Roya Derakhshani
Finance and accounting, Faculty of Humanities, Qom Islamic Azad University, Qom,
YousefAbad, Jahanara , Tehran , Iran
roya.derakhshani@gmail.com

Submit: 12/1/2020 Accept: 17/11/2020

ABSTRACT

This article aims to estimate the credit risk of individual and corporate customers of Iran's banking system. The estimation of credit risks of banks, financial institutions and insurance companies is not possible without an accurate credit scoring of the customers. Credit scoring or credit rating is a process in which the credit amount of individual and corporate customers of the financial-credit institution and banks is measured using the information provided by the customers. The process makes it possible to obtain a wider knowledge of the people's situation to repay the credit received and, or to measure the loan default probability. The statistical data of 399 individual customers and 780 corporate customers from 2011 to 2019 (7500 data) are used to design credit risk models in this study. Multiple Logit Regression, Survival function and Support Vector Machine (SVM) are used to design credit risk models. The results indicate that the selected factors have a significant impact on the customer default probability and credit risk calculation, based on personality, financial and economic characteristics. The Comparison of the results obtained from the accuracy of the forecast shows a higher explanatory power of the Support Vector Machine model and survival function than the Multiple Logit model for both groups of customers

Keywords: Credit risk, Natural and Juridical customers, Financial characteristics, Multiple Logit model, Back up Vector Machine.

1. Introduction

The nature of the banks' activities is such that, although they do not show any signs of crisis or bankruptcy, they may have invisible crisis which are hidden in various forms. This forms of crises have had the authorities of financial systems' regulatory and executive institutions consider the risk managements of financial institutions (especially banks) in a more serious and expert way – due to the consequences and costs which financial instability can create for banks, financial and credit institutions, the economy and various people and institutions (Abdoli and Fard-Hariri, 2015).

According to a report published on the website of the Central Bank of Iran, the average ratio of non-current facilities to total facilities during the years 2011 to 2019 was almost 10%, which indicates the inability of the banking system to recognize the possibility of default on bank installments. Again, according to the report of the Central Bank of Iran, the receivables of banks at the end of 2019 are 1936 thousand billion Tomans, which considering the ratio of non-current receivables equal to 10%, the volume of non-current receivables is more than 193 thousand billion Tomans since 193 thousand billion Tomans is a high figure. It is necessary to consider the optimal method for determining the probability of default before granting the facility

Although there are numerous definitions of risk, it can be argued that all these definitions are intended to express situations in which three common factors exist. Risky situations are the situations in which:

1) An act or action produces more than one result.
2) There is no awareness of the results until the results are tangible.
3) At least one potential result can have adverse consequences (Ahmadizadeh, 2006).

A risk contains qualitative and quantitative aspects. Analysts have been looking for quantitative tools to measure risk for a long period of time, and despite many advances, it is still one of the most dynamic scientific disciplines. Risk measurement is the main focus of companies, banks, investment managers, legislators and market regulators. In the last two decades, the risk factors have led to the bankruptcy of big companies and banks such as Orange County (Due to interest rate fluctuations), Barings Bank (due to the fluctuations in the Japan's stock market), Natwest and UBS (Due to the mistakes in transaction option pricing), and Procter and Gumble (due to a very high risk acceptance) (Hanifi,2002).

At the level of financial and credit institutions and among the various risks, credit risk is of great importance and requires affective management. A credit system can employ quantitative and qualitative methods to assess credit risk. If the credit institution is unable to obtain required information about the potential capabilities of the facility applicant, it must use a qualitative method to assess credit list. In other words, it has to specify a list of special factors about the facility applicant such as his/her credit in the past – which is usually determined by credit rating agencies and cannot be found in Iran, the applicant's wealth, his/her interest fluctuations and whether or not he/she should include collateral in the facility contract.

Credit institutions and banks need a credit rating system of their customers for two main reasons. The function of customer credit rating system for banks and credit institutions is that using this system and based on the existing withheld rates, they can reduce their credit portfolio risk to the least possible amount and select the most reliable and least risky customers from the facility applicants. In credit institution where it is possible to determine the facility rate based on risk and credit rate of the customers, the credit rating system can be helpful in designing their credit portfolio based on the principle of diversity.

In this study, using the Multiple Logit regression and Backup Vector Machine methods, we examine and evaluate the credit risk of natural and juridical customers, to minimize the risk of granting facilities to them.

This study consists of five sections. In the following comes the second part, in which the theoretical and empirical literature of the research will be reviewed. In the third section, research methodology and the models used will be explained. The fourth section is dedicated to examine the estimated model and finally, the results and conclusions and suggestions will be presented.

2. Research Literature

2.1. The definition of risk

Risk is defined as the probability of changes in the predicted benefits of a decision, event or situation in the future (Rahnama Management Dictionary). Probability indicates that variable's changes are not certain. If there was a sufficient certainty in the changes, certain changes would be covered in the framework of the predicted benefits, while the impossibility of prediction caused by the probability of changes, has made it a risk in benefits. Changes refers to any decrease or increase in benefits. It means that not only the undesirable changes, but also desirable changes in this sense are also covered in the risk framework. Decision, event or state refers to the intentionality or unintentionality of the conditions in which risk prevails.

In general, the comprehensive definition of risk in banking is as follows: Risk is the potential loss that either arises directly from the income and capital losses or indirectly arrives from the restrictions that reduces the ability of bank in achieving the financial and business goals. These restrictions would reduce the bank's ability in business administrations or achieving the benefits resulted from the various risk situations.

2.2. The types of bank risks

Todays, it is evident that efficient banking system is one of the crucial factors affecting the development of economies of the countries. Taking any step to improve and upgrade banking system will enhance the quality of cash flow, investment and resource allocation, and will lead the potential scattered and hidden facilities to be used for development and public welfare of the country. In this process, the banks face numerous risks, which can be classified into 4 categories (Abdolo and Fard-Hariri, 1394):

1) Credit risks: credit risks are the most important risks that banks face and they arise due to the possibility of non-timely repayment of facilities and their interests.
2) Operational risks: operational risks are among the other risks that bank faces. These risks arise from fraud, negligence, technological problems, failure in performing tasks on time, etc. And put the bank at the danger of bad reputation and legal issues. These risks are caused by internal and external activities (people, systems, processes and condition changes). These risks include the losses caused by fraud, misuse, inappropriate implementation of laws, changing the rules and not complying with the rules in a timely manner, delays in banking processes, physical and environmental damages. Unlike credit and market risks, operational risks are not a profitable source and must be directly recognized and controlled.
3) Liquidity risk: liquidity risk arises due to the time interval between obligation maturities and investment maturities and put the bank at the danger of lack of liquidity in fulfilling its short-time obligations. A bank must always be preparing to meet its short-term obligations to maintain depositors' trust.
4) Investment risks: are the market risks arose from the fluctuations and falling prices of the market-influenced investments such as stock prices of companies and other assets, and exposes the bank to the risk of lack of capital and loss of public trust.

With the globalization of the economy and the intensification of competitions between banks, todays, the profit margins of traditional banking activities have decreased which increased the risk in banks. Given the problems faced by banking institutions, and regarding its important role in achieving competitive advantage, risk management should be highly considered in these situations and other financial and credit institutions.

Credit risk is the probability of loss due to non-repayment of the principal and interest of loans and credit facilities by customers received them, or the possibility that the value of some of the bank's assets, especially granted facilities would decrease or devalued. Regarding that banks' capital is smaller relative to the total value of their assets, even if a small percentage of the loans are uncollectible, the bank will run the risk of bankruptcy. credit risk of the banking industry is either caused by customers or the bank itself when unable to pay debts (Fallah-shams and Tehrani, 2005). Default risk is considered as one of the most important components of credit risk, and occurs when the borrower does not fulfill its obligations to the lender on the maturity date due to inability of unwillingness. This is one of the oldest and most important risks that affect financial institutions, because the default of a small number of customers can inflict irreparable damages on an organization (Khani, 2007).

3.2. A review on the previous studies

Altman (1968) measured the credit risk of corporate bonds using the multivariate scoring model, known as

Z-score model (discriminant analysis). Altman Z-score model is a discriminant analysis model that employing significant financial ratio values to distinguish companies with financial distress (bankrupted) from the ones without such problems. Given that in most cases the non-payment of loans leads the companies to encounter financial distress, it is possible to predict credit risk using this model.

Bolton et al. (2009) evaluated the credit risk concept. They indicated that financial crisis has led to a renewed focus on the conflict of interest between institutions in their credit ratings (CRAs). They evaluated the Credit Regulatory Organization's conflict model of credit risk to attract more business, the credit card issuers' engagement to the most favorable rating (exporter purchase) and the effectiveness of a number of regulatory solutions suggested by CRAs in their research. They found that Credit Rating Agencies (CRAs) are more likely to be overvalued and rating in the institutes' nominal values. Their model predicts that CRAs are more likely to perceive credit risk during a recession than a boom period. Thus, their study shows that Credit Rating Agencies place lower value on customers during a recession period, while overestimate their credit value during boom times, which leads to an increase in borrower financing costs.

Hilscher and Wilson (2012) examined whether credit risk and credit rating criteria are sufficient for managing banks' risk. This study evaluated the information about credit rating of companies from a positive and normative perspective. If we consider the credit rating as an information index of credit risk, these indicators should determine what is about a risk-averse investor in a systematic risk condition. Also, they have to determine if the rating of incorrect actions is taken from the default probability or not. They examined the factors affecting this issue, using a simple model based on publicly available financial information. However, the information ranking provided by these two, is related to the measurement of ways of the risk exposure and diversity sharing in probability as defaults. They concluded that due to the multidimensional nature of the credit risk, it is not possible to measure all credit risk and validation information. As a result, rating may be misinterpreted.

West (2014) examined the credit scoring of customers based on the neural network method in the banks of United Kingdom. This study attempts to estimate the credit risk of natural customers of UK banks. The 5C system is used to select the independent variables affecting credit risk. There are five main criteria used in 5C system to evaluate customers' credit: character, capacity, capital, collateral and condition. In this study, 404 customers who took loans from the bank between 2010 and 2013 were selected by random sampling method. Using logistic regression, credit risk was estimated and its affecting factors were identified and weighted, finally, a model was designed for credit ranking of customers. The 9 variables chosen from 5C system included: education, age, amount of loan, amount of installments, repayment time, housing status, collateral type, interest rate, occupation status. Variables education and interest rate did not have any significant relationship with the credit risk of natural customers in this study. Finally, placing the variables affecting the credit risk of each of the customers in the final model, the probability of default for granted facilities and credit customers were rated.

Liao (2015) examined the foreign exchange credit risk management of bank customers using data analysis approach. Presenting a model resulted from data analysis of 319 various banks' foreign exchange customers' information between 2012 and 2014, it was concluded that foreign exchange credit risk management became possible with the classification of customers. The customers' classification based on the banks' share showed that 18% of creditworthy customers are high risk, 8 % of are medium risk and 75% are low risk. Also, 4% average customers are high risk and 96% of them are low risk. All un-creditworthy customers are in the high risk category. It was also found that the customer's type affects the sum of balance of original past due and the original balance sum of deferred account is affected by the balance sum of the original pas due.

Feng (2016) investigated the credit risk management in Chinese banks using parametric and non-parametric methods. The information of 187 facility files of banks' natural costumers were evaluated in this study. Data analysis was performed using logit and CART methods and the results analyzed based on the hypothesis that there is a significant relationship between individual characteristics of costumers and the risk of granting credit to them. Factors affecting the risk of granting credit to customers were also identified.

Azab-Mazar and Rouyintan (2005) investigated about the factors affect credit risk of bank customers in the case study of Agricultural Bank (Bank-e-Keshavarzi). The qualitative and financial information of a random sample of 200 companies that received credit facilities from the branches of Agricultural Bank in Tehran Province in 1999-2004 was analyzed. In their study, they identified 36 qualitative and financial variables and using logit analysis, 17 variables with significant impact on credit risk and the distinction between too groups of creditworthy and un-creditworthy customers were selected. The final model was fitted using these variables. The results showed that logit model has a strong capability in estimating the factors affecting credit risk.

Mousavi and Gholipour (2009) ranked the credit rating criteria of banks' customers using Delphi approach. For this purpose, they collected and analyzed the required information from experts and person in charges of selected banks' facility sections by questionnaire. The results, confirmed the economic and financial theories related to the factors affecting credit risk. All the factors affecting the credit risk of banks' juridical customers did not weight equally, some variables were more important in evaluating the credit risk of customers.

Issazadeh and Oryani (2010) conducted a research on the ranking of juridical customers of Agricultural Bank based on credit risk, using data coverage analysis method. Employing cluster sampling method, they evaluated 286 companies receiving facilities from Agricultural Bank branches located in the east and west of Tehran Province. After extracting inappropriate data, 75 companies that received facilities using a 24-month installment contact with a maturity of the end of May 2005, were analyzed. Finally, using data covering analysis method, technical efficiencies were calculated and the companies were ranked. The results show that 15 companies (20% of all companies surveyed) are on the border of efficiency and were considered fully efficient. Also, the average technical efficiency was 78%, which means that in total, the companies' usage of inputs and production factors was 22% more than required, so they had low profitability.

Mirzaei. et al. (2011) investigated the factors affecting the credit risk of juridical individuals of the banks (Case study of Bank Melli Iran branches, Tehran). Using logistic regression model, a random sample of 455 customers (323 creditworthy and 132 un-creditworthy customers) of juridical companies that received credit facilities from Bank Melli Iran branches in Tehran in 2008 were examined. In the fitted model, the significance of the efficient and total regression was examined using Wald and LR statistics respectively (significance level of 95%). The results showed that regarding statistical indicators, these functions are significant considering their coefficients and discriminant power, with high level of validity.

Mirghafouri and Aashouri (2015) evaluated the credit risk of bank customers. They noted that the present banks use different credit rating methods to evaluate the financial performances of companies applied for credit facilities in order to manage credit risk. They employed one parametric (logistic regression) and one non-parametric (division tree and regression) method to create a credit rating model. They analyzed 282 small and medium enterprises' data, that borrowed from one of Tejarat Bank branches in Tehran Province. 13 financial ratios were used as indicators to determine the financial status of selected companies. Employing these two methods, effective ratios and the accuracy of the methods in customer classification were determined. The outcomes showed that non-parametric methods have competitive accuracy with parametric methods.

Mohammadian et al. (2016) investigated the credit risk of juridical customers using the back-up vector machine model and hybrid model of genetic algorithm in Tejarat Bank. For this purpose, they used the financial variables of 282 companies received facilities from Tejarat Bank between 2008 and 2011. To optimize the inputs of the back-up vector machine, genetic algorithm is used. The remarkable ability of genetic algorithm in selecting the optimal points assures the used that the suggested optimal points are more proper points. The findings showed that GA-AVM hybrid model performs better than SVM model in identifying creditworthy and un-creditworthy customers and predicting customers' credit risk.

Eskandari and Rouhi (2016) investigated the credit risk management of bank customers using the revised decision vector machine by genetic algorithm form a data analysis approach. They presented a model using data analysis to predict the customers' receipt rate index. This index is considered as a newer method to measure customer risk instead of measuring the default probability or customers' demands. Due to the low accuracy of prediction methods, they are widely examined and modeled in various researches. They used modeling by decision vector regression with optimization by genetic algorithm as the method of selecting model parameters.

3. Research methodology

The purpose of this paper is to provide a comparative bank credit risk model using Multiple Logit regression, Survival function and Support Vector Machine. In order to be able to predict costumers' credit risk and design a early warning system for both individual and corporate customers using Multiple Logit models, survival function and Support Vector Machine,we used 80% of the sample to fit the model and the remaining 20% of the sample to test the model.

In the estimated models of this study, customers are divided into four groups. The first group are the customers who have fulfilled all their obligations and have not deferred their installments, the second group consists of the customers who delayed up to 10 days, the third group indicates a repayment delay of up to three months, maximum (more than 61 days regarding the previous maturity), finally, the forth group represents the customers with a repayment delay of more than three months equal to the overdue receivables. The variables used in juridical and natural customers' section are introduced in the following.

Table 1. Explanatory variables used in the model

Symbol	Explanatory variables in individual customer	Symbol	Explanatory variables in corporate customer section
X1	Gender	X1	Company size
X2	Age	X2	Bounced check
X3	Customer credit history (bounced check)	X3	Inflation rate
X4	6-month average of account balance	X4	Economic Growth Rate
X5	Credit balance	X5	Sanction
X6	Employment status	X6	Exchange Rate
X7	Work experience	X7	Net profit margin
X8	Marital status	X8	Operating profit margin
X9	Housing status	X9	Return on asset
X10	Job type (employed-unemployed)	X10	Return on capital
X11	Installment amount	X11	Return on working capital
X12	Repayment period	X12	Current ratio
X13	College education	X13	Instant ratio
X14	Inflation late	X14	Cash adequacy ratio
X15	Economic Growth rate	X15	Cash working ratio
X16	Exchange rate	X16	Debt collection period
		X17	Debt ratio
		X18	Current debt to special value ratio
		X19	Loan financial burden ratio
		X20	Financial costs to net profit
		X21	Financial costs to operating profit

4. Model estimation

4.1 Multiple Logit Models

Unlike binary models, multiple logit models divide banks and credit institutions' customers into more than two categories. For instance, the customers can be divided based on their credit status into creditworthy, relatively creditworthy, relatively un-creditworthy, un-creditworthy and extremely un-creditworthy customers. The table below provides the operational definition of each group.

The probability that a customer falls in one of 5 categories

The operational definition of bank customers' categories

Dependent variable	Credit status	Number of deferred installments
Y=1	creditworthy	0
Y=2	relatively creditworthy	Up to 2 installments
Y=3	relatively un-creditworthy	3 to 4 installments
Y=4	un-creditworthy	Up to 18 installments
Y=5	extremely un-creditworthy	More than 18 installments

$$P_1 = F\left(\sum_{k_1}^{K_1} \beta_{k_1} x_{k_1}\right)$$

$$P_2 = \left[1 - F\left(\sum_{k_1}^{K_1} \beta_{k_1} x_{k_1}\right)\right] F\left(\sum_{k_2}^{K_2} \beta_{k_2} x_{k_2}\right)$$

$$P_3 = \left[1 - F\left(\sum_{k_1}^{K_1} \beta_{k_1} x_{k_1}\right)\right]\left[1 - F\left(\sum_{k_2}^{K_2} \beta_{k_2} x_{k_2}\right)\right] F\left(\sum_{k_3}^{K_3} \beta_{k_3} x_{k_3}\right)$$

$$P_4 = \left[1 - F\left(\sum_{k_1}^{K_1} \beta_{k_1} x_{k_1}\right)\right]\left[1 - F\left(\sum_{k_2}^{K_2} \beta_{k_2} x_{k_2}\right)\right]\left[1 - F\left(\sum_{k_3}^{K_3} \beta_{k_3} x_{k_3}\right)\right] F\left(\sum_{k_4}^{K_4} \beta_{k_4} x_{k_4}\right)$$

$$P_5 = \left[1 - F\left(\sum_{k_1}^{K_1} \beta_{k_1} x_{k_1}\right)\right]\left[1 - F\left(\sum_{k_2}^{K_2} \beta_{k_2} x_{k_2}\right)\right]\left[1 - F\left(\sum_{k_3}^{K_3} \beta_{k_3} x_{k_3}\right)\right]\left[1 - F\left(\sum_{k_4}^{K_4} \beta_{k_4} x_{k_4}\right)\right]$$

4.2. Estimation of Multiple Logit model for individual and corporate customers

The credit risk model is estimated for 399 samples of individual customers, using data about the mentioned variables. In the fitted models, the coefficient and total regression significances are examined using F statistics (F is used to evaluate the significance of the regression model) at 95% significance level, as well as the absence of regression collinearity and heteroscedasticity of variance. The outcomes of model estimation are shown in the following Table 2.

Table 2 shows that the coefficients of all variables (at 90% significant level) have a significant difference from zero. For example, the customer credit history coefficient (bounced check) indicates that the probability of default in repayment of obligation increases with the increase in the bounced check. Also, in the case of other variables such as exchange rates and inflation, it can be said that since the growth of the exchange rate and inflation will make the conditions for fulfilling the pledge difficult, the probability of non-repayment of debts will increase.

Table 2: The results of logit model estimation for individual customers

Forth Group	Third group	Second group	First group	Variable
(0.00) -0.045	(0.00) 0.87	(0.00) 0.45	(0.03) 0.079	C
(0.01) 0.103	(0.03) 0.095	(0.03) 0.081	(0.00) 0.076	X1
(0.01) 0.092	(0.02) 0.084	(0.01) 0.072	(0.03) 0.068	X2
(0.03) 9.34	(0.02) 8.98	(0.02) 6.19	(0.02) 4.15	X3
(0.01) -0.017	(0.02) -0.021	(0.03) -0.029	(0.00) -0.035	X4
(0.02) -0.065	(0.04) -0.081	(0.00) -0.11	(0.02) -0.14	X5
(0.02) -0.014	(0.00) -0.022	(0.00) -0.038	(0.00) -0.041	X6
(0.00) -0.028	(0.01) -0.047	(0.03) -0.061	(0.00) -0.075	X7
(0.00) -0.038	(0.00) -0.058	(0.01) -0.066	(0.02) -0.072	X8
(0.02) -0.059	(0.02) -0.073	(0.04) -0.086	(0.03) -0.110	X9
(0.00) -0.024	(0.01) -0.043	(0.03) -0.059	(0.02) -0.078	X10
(0.02) 0.124	(0.04) 0.102	(0.00) 0.084	(0.00) 0.075	X11
(0.01) -0.019	(0.02) -0.029	(0.00) -0.038	(0.00) -0.049	X12
(0.04) -0.024	(0.02) -0.035	(0.00) -0.042	(0.00) -0.055	X13
(0.02) 0.63	(0.03) 0.56	(0.03) 0.45	(0.00) 0.32	X14
(0.01) -0.045	(0.04) -0.11	(0.02) -0.16	(0.01) -0.25	X15
(0.02) 0.056	(0.01) 0.042	(0.02) 0.028	(0.00) 0.015	X16

Table 3: The results of the logit model estimation for corporate customers

Forth Group		Third group		Second group		First group	
Coefficient	Variable	Coefficient	Variable	Coefficient	Variable	Coefficient	Variable
0.598	C	0.876	C	1.124	C	0.845	C
-0.213	X1	-0.249	X1	-0.368	X1	-0.459	X1
0.429	X2	0.342	X2	0.245	X2	0.142	X2
2.461	X3	2.461	X3	1.124	X3	0.798	X3
-0.640	X4	-0.640	X4	-0.854	X4	-1.151	X4
0.405	X5	0.328	X5	0.286	X5	0.254	X5
0.452	X6	0.318	X6	0.246	X6	0.189	X6
-1.007	X7	-1.007	X7	-1.354	X7	-2.102	X7
-0.319	X8	-0.319	X8	-0.745	X8	-0.987	X8
-0.315	X9	-0.389	X9	-0.425	X9	-0.542	X9
-1.136	X12	-1.136	X12	-1.598	X12	-1.899	X12
-1.258	X13	-1.258	X13	-1.451	X13	-1.795	X13
-0.218	X14	-0.325	X14	-0.486	X14	-0.598	X14
-0.058	X16	-0.058	X16	-0.075	X16	-0.098	X16
0.396	X17	0.353	X17	0.258	X17	0.254	X17
0.456	X19	0.329	X19	0.248	X19	0.198	X19
0.569	X20	0.412	X20	0.399	X20	0.315	X20
0.860	X21	0.751	X21	0.683	X21	0.548	X21

The coefficient of variable X1 (company size) for customers' groups in the estimated model is 0.459,0.368,0.249 and 0.213, respectively. The coefficient indicates that the probability of default decreases with an increase in the size (assets) of the borrowing company. Assuming that the other conditions are stable, this coefficient shows that with increasing the size of the borrowing company, the logarithm of chance decreases by 0.459,0.368,0.249 and 0.213 for each group, in favor of delayed repayment. In others words, with an increase in the borrower company's assets, the probability of untimely repayment to the probability of timely repayment is less than 1. Meanwhile, for the fourth group of the companies defaulted 6 months of installments, with an increase in the company's assets, the default probability is 0.808. also, the change of variable 9X (return on assets) for the third group, whose number of differed installment is in the third place of ranking, is -0.389, and indicates that with the increase in the return on assets of the company, the default probability decreases. The coefficient shows that considering that the conditions are stable, if the return on assets of the company increases by 1%, the logarithm of chance decreases by 0.67% in the favor of non-timely repayment. In other words, if the return on assets increases, the probability of repayment default to the probability of timely repayment, is 0.67 (less than 1). The coefficient for the second group is -0.425, which indicates that if the return on assets increases for the companies placed as second default level, the probability of default in repayment to the probability of timely repayment is 0. 65 (less than 1). The coefficient is equal to -0.542 for the first group, which indicates that if the return on assets increases for this group, the probability of default in repayment to the probability of timely repayment is 0.58 (less than 1). The calculations show that as the number of defaulted installment increases, they lead to an increase in positive indexes affect the company's repayments, or the probability of default is decreased for the companies.

In Table 4, namely the classification table, the predicted value for the dependent variable Y (in the fitted equation) at the threshold of 0.5, in terms of their placement regarding threshold, are classified compared to the observed real values.

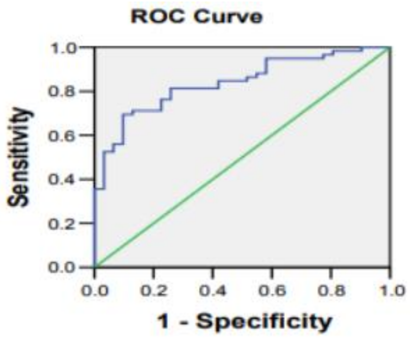
Table 4: The predictive power of Multiple Logit model

The prediction of Legal Customers' Logit Model					The prediction of Individual Customers' Logit Model					Customer type
total	deferred	Bad debts	Past due date	Trustworthy costumers	total	deferred	Bad debts	Past due date	Trustworthy costumers	
6950 92.66%	163 2.17%	123 1.64%	150 2%	6514 86.85%	270 67.66%	2 0.50%	7 1.75%	6 1.50%	255 63.90%	Trustworthy costumers
204 2.72%	29 0.38%	37 0.49%	86 1.14%	52 0.69%	44 11.02%	8 2.00%	16 4.01%	15 3.75%	5 1.25%	Past due date
220 2.93%	67 0.89%	89 1.18%	19 0.25%	45 0.6%	73 18.29%	5 1.25%	57 14.28%	8 2.00%	3 0.75%	Bad debts
126 1.68%	19 0.25%	65 0.86%	27 0.36%	15 0.2%	12 3.00%	8 2.00%	2 0.50%	2 0.50%	0 0%	deferred
7500 100%	278 3.70%	314 4.18%	282 3.76%	6626 88.34%	399 100%	23 5.76%	82 20.55%	31 7.76%	263 65.91%	total

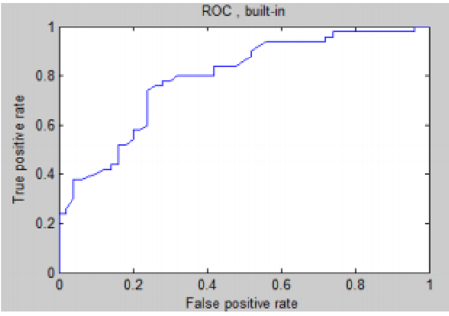
The correct prediction ration of the estimated Logit model is equal to 83.95% and 89.44% for the individual and corporate customers, respectively. As indicated in Table 5, the area under individual customers' logit cure is 0.839 and the corporate customers' logit is 0.875. As the customers' behavior is predicted randomly, the probability of correct prediction is 0.83 and 0.87. In the following, a curved called ROC is used to examine the distinguishing power of two groups (creditworthy and un-creditworthy customers in this research). The curve is drawn from the point (0,0) in the lower left corner to the point (1,1) in the upper right corner in the coordinate plane whose horizontal axis is -1 degree of detection and vertical axis is the degree of model sensitivity. The closer the curve is to the upper left corner (1,0), the greater the strength of the model and the distinguishing power between the two groups is. At (0,1) point, the sensitivity and detection degree of the model are both at their maximum value (1).

So, the individual and corporate customers' models are shown in Graph 1 and 2, respectively.

As can be seen in Figures (1) and (2), the slope of both ROC charts is more than 45 degrees, which indicates its high reliability in predicting and distinguishing power of the two groups.



Graph 1: Drawing the ROC curve for individual customers' logit model



Graph 2: Drawing the ROC curve for corporate customers' logit model

According to the models above, a summary of the results of customers' ROC curve's calculation can be seen in Table 5.

Table 5: The result of calculation of customers' ROC curve

Model	The under the curve area	Standard deviation	Prob value	95% significance level	
				Lower limit	Upper limit
Multiple Logit for individual customers	0.875	0.024	0.000	0.798	0.935
Multiple Logit for legal customers	0.896	0.016	0.000	0.798	0.935

4.3. Survival function

Survival analysis examines the time until an event occurs. First, Kaplan and Mir (1985) developed an estimator to estimate the survival function for censored and uncensored data. Then, Cox (1972) introduced hazard function to make a connection between the personal characteristics of the individuals and the time until the occurrence of desired even (default, in this research). The random variable T is the time until default or survival time for a customer, in the present research. In other words, this variable indicates the time duration that takes for each customer to default position. Usually, in the data related to loan or credit, the customer repayment information is displayed on a monthly basis. According to the common definitions in the banking industry, if the customer delays in installment repayments for 3 month or more, continuously, he/she is considered as a defaulted person.

This definition is also confirmed by Basel Committee. If the time to customer default is indicated by random variable T, the probability of customer default before the time ($(t.\,X)$) t will be as follows:

$$PB(t.\,X) = \Pr\{T < t\}$$

It should be noted that the above mentioned possibility may also relate to some personal characteristics of customers indicated by X.

$PB(t.\,X)$ for the probability distribution function of the random variable T and (PB' ($t.$ X) of the corresponding density function can also be defined. Accordingly, the probability of default in $(t,\ t + \delta t]$ period will be as follows:

$$\Pr\{t < T < t + \delta t.\,X\} = PB\,'\,(t.\,X)\delta t$$

The survival probability function in which the default occurs after time ($(t.\,X)$), is as follows:

$$PG\,(t.\,x) = 1 - (t.\,X) = \Pr\{T \geq t\} = \int PB\,'\,(u.\,X)du$$

Hazard rate function is another basic concepts of survival analysis and shows that if the customer does not have any default in (t,0] period, how much the default rate in $[t.\ t+\delta t)$ will be. This conditional probability function is defined as follows:

$h(t.\ X) = \Pr\{t < T \leq t + \delta t; T > t; X\}$

The survival function is estimated using Cox proportional hazard method, the results of which are shown in Table (6):

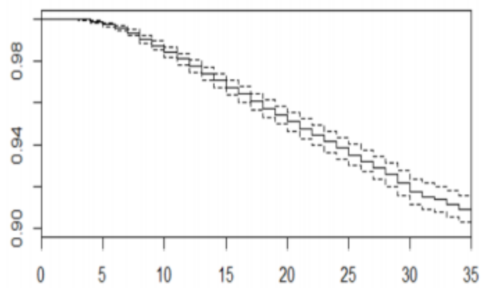
Table 6: Cox model estimation for individual customers' hazard rate

Forth group	Third group	Second group	First group	Variables
(0.03) -0.68	(0.04) -1.53	(0.02) -1.98	(0.00) -2.45	C
(0.00) 3.19	(0.02) 2.14	(0.03) 1.58	(0.01) 1.38	X1
(0.03) 0.71	(0.01) 0.62	(0.01) 0.53	(0.00) 0.36	X2
(0.03) 1.57	(0.03) 1.15	(0.02) 0.95	(0.03) 0.84	X3
(0.04) -0.47	(0.01) -0.63	(0.04) -0.75	(0.02) -0.92	X4
(0.02) -0.33	(0.02) -0.41	(0.03) -0.59	(0.02) -0.72	X5
(0.03) -0.22	(0.01) -0.35	(0.03) -0.42	(0.03) -0.55	X6
(0.01) -0.07	(0.00) -0.12	(0.03) -0.17	(0.01) -0.21	X7
(0.02) -0.39	(0.03) -0.52	(0.02) -0.66	(0.02) -0.78	X8
(0.01) -0.02	(0.02) -0.05	(0.01) -0.09	(0.01) -0.12	X9
(0.01) -0.19	(0.04) -0.26	(0.04) -0.34	(0.02) -0.48	X10
(0.02) 1.29	(0.03) 1.06	(0.02) 0.92	(0.02) 0.81	X11
(0.02) -0.35	(0.02) -0.41	(0.04) -0.52	(0.02) -0.67	X12
(0.02) -0.18	(0.03) -0.27	(0.01) -0.35	(0.02) -0.46	X13
(0.02) 0.45	(0.02) 0.38	(0.02) 0.29	(0.03) 0.20	X14
(0.00) -0.05	(0.02) -0.09	(0.01) -0.12	(0.03) -0.18	X15
(0.03) 0.41	(0.03) 0.32	(0.03) 0.25	(0.00) 0.22	X16

Table 7: Cox model estimation for corporate customers' hazard rate

Forth group	Third group	Second group	First group	Variables
(0.03) -0.56	(0.04) -1.12	(0.01) -1.86	(0.00) -2.99	C
(0.00) 3.99	(0.02) 3.16	(0.00) 2.12	(0.00) 1.89	X1
(0.03) 0.74	(0.01) 0.60	(0.04) 0.48	(0.00) 0.32	X2
(0.05) 1.39	(0.02) 1.05	(0.03) 0.86	(0.00) 0.79	X3
(0.00) -0.42	(0.04) -0.63	(0.00) -0.75	(0.01) -0.89	X4
(0.00) -0.33	(0.00) -0.41	(0.01) -0.59	(0.01) -0.75	X5
(0.03) -0.25	(0.01) -0.35	(0.02) -0.42	(0.01) -0.62	X6
(0.02) -0.07	(0.01) -0.12	(0.00) -0.17	(0.02) -0.28	X7
(0.01) -0.39	(0.00) -0.52	(0.04) -0.66	(0.03) -0.77	X8
(0.00) -0.04	(0.02) -0.08	(0.04) -0.12	(0.00) -0.19	X9
(0.00) -0.19	(0.03) -0.24	(0.00) -0.34	(0.04) -0.51	X12
(0.03) 1.37	(0.04) 1.19	(0.00) 0.99	(0.04) 0.92	X13
(0.02) -0.35	(0.00) -0.41	(0.02) -0.52	(0.00) -0.62	X14
(0.01) -0.18	(0.04) -0.27	(0.03) -0.39	(0.00) -0.51	X16
(0.01) 0.52	(0.05) 0.45	(0.04) 0.31	(0.02) 0.27	X17
(0.00) -0.08	(0.04) -0.14	(0.04) -0.19	(0.00) -0.24	X19
(0.01) 0.54	(0.00) 0.37	(0.00) 0.28	(0.01) 0.19	X20
(0.02) 0.45	(0.01) 0.39	(0.02) 0.32	(0.01) 0.25	X21

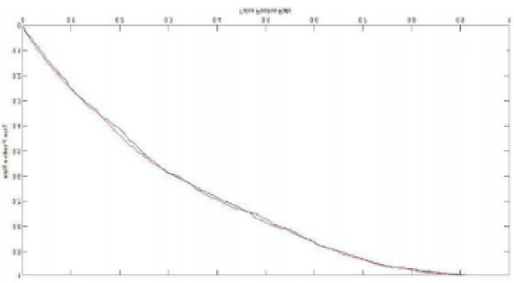
The best set of predictor variables can be identified using multivariate stepwise regression. Backward removal and forward selection are The most common regression used in analyses. It is obvious that the results are the same in all methods. In the present research, the both methods are used for multivariate analysis. Due to the similarity of their results however, only the results obtained from the backward removal method using Stata software are reported in the table above. In this technique, first, all variables are entered into the model, then, based on the significant test, the variable with the least amounts of predictive power based on the probability level (Prob) are removed. The variables with the highest level of significance (above the significance level of 0.05, typically) are removed from the model. Then, the model is re-estimated with the remaining variables and this process continues until all variables are significant. It should be noted that in order to perform multivariate analysis, all variables ae entered into the model to examine the effectiveness of each factors of survival and risk rate of the customers. This process is conducted when the other variables are controlled. Based on this, the variables load amount, number of installments, gender, age, marital status, job type, etc. are included in the Cox model. The graph below shows the adjusted survival function.



Graph 3: Survival probability chart

As shown in the figure above, first, the survival function takes the value of 1. The reason is the definition we have for default, which is that default occurs when a person has not paid his/her installments for 3 months, so there is no default in the first 3 months. The reason is approximately 10% of the data have defaulted, the survival function tends to be 0.9.

Due to the discrete times, the survival function is stepped. Graph 4 represents the ROC for the survival probability function:



Graph 4: ROC graph of the survival probability function

4.4. SVM model structure

SVM formulation is based on the concepts of structural risk minimization, which has advantages over experimental risk minimization. Traditional neural network models are based on empirical risk minimization. SVM, minimizes the maximum limit of the expected risk. This feature enhanced the generalizability of SVM and is the target of statistical learning theories. If we consider S_h as the hypothesis space with dimensions VC, h and as:

$$S1 \subset S2 \subset \cdots S\infty$$

Structural minimization of risk is minimizing the following relationship:

$$\min R_{epm}[f] + \sqrt{\frac{h \ln\left(\frac{2l}{h} + 1\right) - \ln\left(\frac{\delta}{4}\right)}{l}}$$

The VC dimension actually represents the capacity of a function set. SVM is an algorithm that finds specific types of linear models that achieve maximum hyper plane.

Maximizing the margin of the hyper plane maximizes the separation between the categories. Back up vectors are the closes training points to the maximum margin of the hyper plane. These are the only vectors (points) used to define the boundary between categories (Shin, 2005). Table 8 represents the prediction of results of SVM models in models ran for eight times. In the present research, to run the models, MATLAB and the functions of this software are used.

Table 8: Prediction results of SVM model for individual customers

total			experimental			Educational			
total	Un-trustworthy	trustworthy	total	Un-trustworthy	trustworthy	total	Un-trustworthy	trustworthy	
399	80	319	202	39	163	197	41	156	Run1
0.618			0.612			0.613			
399	53	346	198	23	175	201	30	171	Run2
0.637			0.635			0.613			
399	28	371	188	12	176	211	16	195	Run3
0.732			0.711			0.695			
399	33	366	193	15	178	206	18	188	Run4
0.712			0.704			0.732			
399	59	340	203	19	184	196	40	156	Run5
0.685			0.698			0.712			
399	93	306	203	35	178	196	58	138	Run6
0.598			0.587			0.614			
399	68	331	223	24	199	176	44	132	Run7
0.743			0.709			0.743			
399	31	368	193	13	185	206	19	187	Run8
0.754			0.695			0.710			

According to the table above, the prediction accuracy of SVM model is about 63% for un-creditworthy and 73% for creditworthy customers.

Table 9: Prediction results of SVM model for corporate customers

total			experimental			Educational			
trustworthy		total	Un-trustworthy	trustworthy		total	Un-trustworthy	trustworthy	
7500	1250	6250	03	650	3350	3500	600	2900	Run1
0.648			0.665			0.625			
7500	1358	6142	4000	720	3280	3500	638	2862	Run2
0.678			0.698			0.649			
7500	1658	5842	4200	751	3449	3300	907	2393	Run3
0.736			0.725			0.702			
7500	1245	6255	4100	652	3448	3400	593	2807	Run4
0.689			0.678			0.648			
7500	1890	5610	4400	895	3505	3100	995	2105	Run5
0.798			0.769			0.754			
7500	1896	5604	4500	864	3636	3000	1032	1968	Run6
0.690			0.756			0.741			
7500	1654	5846	4250	850	3400	3250	804	2446	Run7
0.785			0.735			0.721			
7500	1795	5705	4350	759	3591	3150	1036	2114	Run8
0.662			0.674			0.698			

According to the table above, the prediction accuracy of SVM model is about 74% for un-creditworthy and 75% for creditworthy customers.

Table 10: Measuring the SVM model predictive power

The prediction of Legal Customers' Logit Model					The prediction of Individual Customers' Logit Model					Customer type
total	deferred	Bad debts	Past due date	Trustworthy costumers	total	deferred	Bad debts	Past due date	Trustworthy costumers	
6678 89.04%	117 1.56%	121 1.61%	115 1.53%	6325 84.33%	298 74.68%	2 0.75%	5 1.25%	4 1.00%	286 71.67%	Trustworthy costumers
223 2.97%	52 0.69%	29 0.38%	101 1.34%	41 0.54%	38 9.52%	5 1.25%	10 2.50%	18 4.51%	5 1.25%	Past due date
274 3.65%	67 0.89%	152 2.02%	15 0.20%	40 0.53%	55 13.78%	1 0.25%	49 12.28%	2 0.50%	3 0.75%	Bad debts
325 4.33%	187 2.49%	87 1.16%	32 0.42%	19 0.225	8 2.00%	3 0.75%	2 0.50%	2 0.50%	1 0.25%	deferred
7500 100%	423 5.64%	389 5.18%	263 3.50%	6425 85.66%	399 100%	23 3.00%	82 16.54%	31 6.51%	263 73.93%	total

Evaluating the table above, it is concluded that SVM model is able to predict 75% of the experimental data, correctly.

4.4 Comparison of the models' predictive power can be seen in Table 11, for two models in 5 categories: creditworthy, past due, doubtful, deferred and total.

Table 11: Comparative evaluation of predictive power of the models

The prediction of Legal Customers' Logit Model					The prediction of Individual Customers' Logit Model					Model type
total	deferred	Bad debts	Past due date	Trustworthy costumers	total	deferred	Bad debts	Past due date	Trustworthy costumers	
7500 100%	278 3.7%	314 4.18%	282 3.76%	6626 88.34%	399 100%	23 5.76%	82 20.55%	31 7.76%	263 65.91%	Logit model
7500 100%	423 5.64%	389 5.18%	263 3.50%	6425 85.66%	399 100%	23 3.00%	82 16.54%	31 6.51%	263 73.93%	SVM Model

5. Conclusion and suggestions

One of the important issues regarding bank lending is the probability of non-repayment of loans by borrowers. Estimating that a company may bankrupt in the future is of a great significance for lenders and creditors. So, finding a model that provides the best classification for companies has always been crucial. Therefore, the need to identify the factors causing non-repayment of loans, in necessary to reduce and control credit risk and improve the credit process. This study aimed to estimate the credit risk of individual and corporate customers of the Iran banking system. 399 individuals and 780corporate customers from 2011 to 2019 (7500 data) customers are used in this study. For this purpose, Multiple Logit regression and Support Vector Machine were used. The results showed that the considered components based on personality, financial and economic characteristics had significant impacts on the customers' default probability and credit risk calculation. The outcomes of the statistical tests indicated that SVM model had a higher accuracy in predicting customer credit risk. As showed by prediction and modelling results, it can be said that these models have the ability necessary to minimize credit risk. Prediction results are acceptable in all proposed models according to the good fit criteria. However, as a result of the research of Mohammadian et al. (2016), the GA-SVM hybrid model has shown better performance in identifying customers than the SVM. Also, by observing the results of Mirghofori and Assyrian study, it can be understood that non-parametric methods have competitive accuracy with parametric methods. Regarding the outcomes, we suggest that a database containing financial, economic, personal and managerial data of customers be stablished and updated exclusively in these institution and banks. Also, the software system based on these models should be prepared and launched for the use of experts in this institute. Hybrid methods should be used in risk prediction to increase the forecast accuracy.

References

1) Eskandari, Meysam Jafari and Rouhi, Milad (2016), Credit Risk Management of Bank Customers Using the Revised Decision Vector Machine by Genetic Algorithm with Data Analysis Approach, Quarterly Journal of Asset Management and Financing, No. 1, Pp: 12-38.
2) Tehrani, Reza, Fallah Shams, Mir Feyz (2005), Risk Credit Model Designing and Explaining in the Iran Banking System, Journal of Social Sciences and Humanities, Shiraz University, No. 43, Pp: 45-60.
3) Abdoh Tabrizi, Hossein, Raadivar, Meysam (2009), Measuring and Managing of Market Risk, Value at Risk Approach, Pishbord Publication, No. 5, Pp: 34-52.
4) Sham Al-Dini, Akbar (2010), Principles of Credit Risk Management, Saderat Bank, No. 20, Pp: 84-87.
5) Khodaverdi, Omid (2009), Credit Risk Scoring of the Insured Customers Using Smart Methods (Case Study: An Export Credit Institution), Master Thesis, University of Tehran.
6) Khalili-Iraqi, Maryam (2005), Credit Risk Management Using Decision Making Models, Quarterly Journal of Economic Research, No. 16, Pp: 183-212.
7) Abdoli, Ghahraman , Fard-Hariri, Alireza (2015), Modeling the Credit Risk Assessment of Juridical Customers of Refah Bank, Quarterly Journal of Applied Theories of Economics, No. 1, Pp: 1-24.
8) Arab Mazar, Abbas, Rouyintan, Pooneh (2006), Factors Affecting Credit Risk of the Bank Customers, Case Study of Agricultural Bank, Quarterly Journal of Economic Researches, No. 6, Pp: 45-80.
9) Issazadeh, Saeid, Oryani, Bahareh (2010), Juridical Customers' of the Banks' Rating based on Credit Risk Using Data Envelopment Analysis: A Case Study of Agricultural Bank Branches, Quarterly Journal of Economic Researches and Policies, No. 18 (55), Pp: 59-86.
10) Salahi, Mohammad (2011), Review and Prioritization of the Factors Affect Bank Customers' Credit Using AHP Method, Case Study of Sina Bank, School of Management, Financial Management, University of Tehran.
11) Fard-Hariri, Alireza (2008), Risk Modeling and Credit Ranking of Bank Systems' Customers, Master Thesis, Faculty of Economics, University of Tehran.
12) Mousavi, Seyyed Reza, Gholipour, Elnaz (2009), Credit Criteria Ranking of Bank Customers using Delphi Approach, The First International Conference on Bank Service Marketing.
13) Mirzaei, Hossein, Nazarian, Raafik and Bagheri, Rana (2011), Investigating the Factors Affect Juridical Customers of the Banks' Credit Risk (Case Study: Melli Bank Branches, Tehran Province), Quarterly Journal of Economic Research, 16th Year, Pp: 67-98.
14) Mirghafouri, Seyyed Habibollah, Ashouri, Zohreh (2015), Credit Risk Assessment of Bank Customers, Quarterly Journal of Business Management Research, Volume 7, No. 13, Pp: 147-166.