Iranian Economic Review 2022, 26(3): 707-720
DOI: 10.22059/ier.2022.87472

RESEARCH PAPER

Identification and Evaluation of the Factors Affecting Credit Risk Management Using the Multinomial Logistic Regression Model

Nasrin Motedayen[a], Rafik Nazarian[b,*], Marjan Damankeshideh[c], Roya Seifipour[d]

a, b, c, d. Faculty of Economics, Islamic Azad University, Tehran Central Branch, Tehran, Iran

Received: 28 November 2021, Revised: 12 March 2022, Accepted: 26 April 2022

Abstract
The aim of this study is to identify and evaluate the factors that influence credit risk employing a multinomial logistic regression approach. For this purpose, in the first phase, indicators that affect credit risk assessment of natural customers were identified using documentation and library method. Then, the final data on the indicators were collected, including 7330 files of natural customers of Mellat Bank, and multinomial logistic regression was employed in studying the indicators of credit risk assessment of the bank's natural customers in the four classes of timely receipt, overdue, deferred, and non-performing loans. The results of the estimated model show that the indicators of gender, loan value, age, installment interval, previous loan, occupation, loan repayment term, number of installments, quantity of each installment, loan extension, type of collateral, average balance, facility interest rate, type of facility, and education level have a significant impact on the credit risk of real customers.
Keywords: Natural Customers, Credit Risk, Credit Scoring, Multinomial Logistic Regression.
JEL Classification: E51, G21.

Introduction

Financial risks include uncertainties associated with any type of financing, including credit, trade, investment, and operational risks (Han & Wang, 2021). Among these risks, credit risk related to customers is of particular importance (Bellotti and Crook, 2008).

Financial risks include uncertainties associated with any type of financing, including credit, trade, investment, and operational risks (Han & Wang, 2021). Among these risks, credit risk related to customers is of particular importance (Bellotti and Crook, 2008).

Many credit scoring models have been developed by bankers and researchers to decide whether to accept a loan. The term credit scoring refers to a variety of statistical methods that attempt to provide a quantitative measure of how likely a customer is to engage in a certain behavior (e.g., nonpayment of a loan) based on their creditworthiness so that when a bank receives a credit application, it decides whether or not to grant credit to the applicant (Haralambie, 2016). Meanwhile, most of the existing credit scoring methods traditionally follow the perspective of binary classification and divide the potential risk into two classes: good and bad credit risk. In the first case, the rejection of the application of an applicant who is able to repay the loan means the loss of the credit position. In the second case, approval of the credit application of an applicant who is unable to repay the credit means financial loss. Therefore, the second scenario is the most vulnerable scenario to reduce the bank's claims. This is because the risk of lending to a bad customer is certainly worse than not lending to a

*. Corresponding author email: r_nazarian@yahoo.com

good customer (Bandyopadhyay et al., 2011). Therefore, it is very important for the bank to properly evaluate the credit in the context of lending to the customer (Moradi and Rafiei, 2019). However, binary classification is not very accurate and banks need to subdivide their customers more carefully according to their domestic policies. In this case, they will be able to make a variety of customer interaction decisions depending on which class the customer belongs to (Mohammadi and Pirmohammadiani, 2015). Therefore, in an innovative action, considering the credit policy of the bank under review (Mellat Bank), according to experts by default and repayment behavior, the present study uses a multinomial logistic regression approach to study different types of customers in four categories rather than a binary classification. Multinomial logistic regression approach is a statistical method with the ability to calculate different regressions and correspond to different quartile points, which in addition to express a more complete and comprehensive picture of the data, provide the opportunity to measure the relationship between independent variables and the desired logics of the dependent variable without the need for normal data and even in the presence of remote locations (Eyvazi, 2016). Furthermore, since the objective of credit managers is to minimize risk and maximize bank profits, it is necessary to establish a decision criterion to distinguish between those who receive credit and those who do not. This is possible through a detailed analysis of the demographic, social and economic characteristics of the applicants (Ngo et al., 2021). Accordingly, through a review of the relevant literature, this study identifies a wide range of individual, economic and social factors as independent variables for customer validation and examines the extent to which the factors can and should be validated to influence customers.

The remainder of this paper is organized as follows. Section 2 provides the theoretical foundation and literature review. In Section 3, research methodology is elaborated, and in Section 4, research models and variables are described. Section 5 presents the research findings, and finally Section 6 concludes the paper and gives some recommendations.

Theoretical Foundations and Literature Review

Credit risk associated with customers is of particular concern, and managers should provide an appropriate solution to assess and identify customer risk to enable efficient allocation of credit facilities (Bellotti & Crook, 2008). Debtors' non-payment at maturity can have a variety of reasons. Regardless of the reason, non-payment has many negative effects on the financial relations of the society, the most important of which are loss of public trust, reduction in long-term transactions, reduction in bank lending and profitability, and reduction in cooperation between individuals and the banking system (Mehrabian and Seifipour, 2016). Previous empirical studies show that high levels of non-performing loans are usually responsible for collapse (e.g. Gup and Kolari, 2005; Samad, 2012) as well as increased vulnerability of the banking system and the financial sector as a whole (Niinimaki, 2012). In other words, the deterioration of banks' asset quality not only destabilizes the banking system, but can also reduce a country's productivity and economic welfare (Nikolopoulos and Tsalas, 2017). Indeed, the economic and social well-being of countries is highly dependent on the behavior of the commercial banking sector. Banks provide the credit needed to sustain manufacturing, agricultural, commercial, and service enterprises, create jobs, and increase purchasing power, consumption, and savings. Therefore, a banking collapse affects the social fabric of a country in general and can quickly affect other financial sectors. Thus, it is crucial that lending decisions are made as prudently as possible while keeping the decision-making process efficient and effective (Dahooie et al., 2021). So, it is important to establish a decision criterion to distinguish between those who receive credit and those who do not. This is possible through a detailed analysis of the demographic, social and economic characteristics of the applicants (Nego et al., 2021).

In this regard, Marrez and Schmit (2009) examined the impact of socio-demographic characteristics such as age, gender and marital status on customer credit risk. Male customers are more likely to default than female customers, but the repayment rate is the same for male and female customers. As for the age factor, the 20–25 age group is more likely to default than the 61–70 age group. Single customers are 36% more likely to be in debt than married customers. Dehmardeh et al. (2012) confirmed that the value of facilities received from the bank, employment of the borrower's spouse, status of the returned check, repayment term of installments, marital status, current property and assets of the borrower, and current residential status of the borrower affect the credit risk of customers. Bhole and Ogden (2010) and Gomez and Santor (2003) argued that group lending provided greater protection to the financial institution than individual lending. In this context, other empirical studies showed that there were many factors that influenced the credit risk of commercial banks. In these studies, various methods have been used to validate the customers. The following is a summary of some of these studies.

Ngo et al. (2021) conducted a study to investigate the factors affecting credit risk in commercial joint stock banks' lending. The results showed that profitability and inflation have positive effects on credit risk, while bank capital, bank size, economic growth, and loan to deposit ratio have negative effects. Hatefi Majoomard et al. (2021) conducted a study to determine the factors affecting credit risk management in banks listed in Tehran Stock Exchange. They came to the conclusion that credit performance, inflation, equity ratio, as well as GDP growth to assets ratio variables have significant relationship with credit risk. Also, equity to assets ratio and net profit to assets ratio have no significant relationship with credit risk, and GDP growth rate has a significant relationship with assets and credit risk. Moqaddaseh et al. (2021) investigated the impact of demographic variables on the responsibility of bank customers. The results showed that age, income, education level, property, gender, current ownership, and the value of facilities received have a significant impact on credit risk and the distinction between two groups of good and bad customers. Gudde Jote (2018) studied the determinants of loan repayment (non-performing loans and deferred loans) and used the logistic model. The results indicated that six variables had a statistically significant impact on the probability of loan repayment. These important variables are: the education level, the type of loan, the degree of relationship and proximity of the borrower to the institutions, family size, and income from financial activities of loans and education.

Mohammadi et al. (2019) conducted a study to investigate the factors affecting the probability of default of loans to bank customers. They found that the variables of the customer's monthly income, the type of relationship between the borrower and the guarantor, the guarantor's guaranteed capital, the customer's experience and job stability, the repayment term of the loan, and the history of the customer's relationship with the bank had an inverse effect on the probability of default, and also the variable of the loan value had a direct effect on the possibility of loan default. In a study, Ebrahimi et al. (2009) investigated the key factors for banks' success in receiving long-term receivables and their ranking using a fuzzy hierarchical process approach. The result of prioritizing the main criteria of the effective indicators of the factors affecting the receipt of long-term receivables of Mehr Eghtesad Bank of Yazd Province using the fuzzy hierarchical process approach technique showed that the training and awareness of bank employees ranked first, an appropriate system of evaluation and credit scoring of customers ranked second, use of a comprehensive system for tracking claims ranked third, adequate supervision in the use of facilities ranked fourth, reduction in inflation and interest rates of the banking system ranked fifth, increase in unemployment ranked sixth, personality trait (reliability and dependability) ranked seventh, and increase in penalties for late payment of the facility ranked eighth in the collection of long-term receivables of the bank. Gutu et al. (2017) investigated the factors affecting repayment performance of female borrowers of top financial institutions in Ethiopia using quantile logistics model on 182 females including 85 overdue loans and 97 non-performing loans. According to the main objective of the study, nine independent variables were selected for analysis. Determinants are six important variables and the remaining three variables have a small effect. Age of borrowers, educational level of borrowers, appropriateness of loan for intended purpose, type of residence, use of loan for intended purpose, and number of family members are the variables that affect loan repayment performance among female borrowers.

Mohammadi et al. (2017) studied the factors affecting the formation of long-term loan in the Iranian banking system using a descriptive survey research method. The results of the exploratory factor analysis of the study showed that the most important factors affecting the growth of long-term receivables in the country's banking system are as follows: non-compliance of documents and guarantees with the bank's internal credit rules and policies, technical and special credit constraints, inadequate process of monitoring facilities and receipt of receivables, incomplete guidelines for determining interest rates and liabilities, legal requirements and restrictions, economic sanctions, business environment and credit scoring, which are considered as the underlying causes and reasons of long-term receivables. Fakoor et al. (2016) conducted a study to identify and rank the factors affecting the reduction of long-term receivables in the Ayandeh Bank based on the network analysis procedure. The results showed that the three criteria of overdue loans, granted loans, and non-performing loans were effective in reducing long-term receivables in the bank, with the criterion of granted loans having the highest priority. For the criterion of deferred loans, the sub-criterion of management weakness after the granting of credit, for the criterion of credits granted the sub-criterion of mismanagement of financial resources, and finally for the criterion of non-performing loans the sub-criterion of failure to act promptly and decisively in the case of defaulters have the highest priority.

According to the literature and theoretical foundations described, the research hypothesis is as follows:

The application of personal, social and economic conditions of natural customers has a significant impact on the bank's credit risk management.

Research Methodology

This research is applied in terms of purpose, it is causal-correlation in terms of the nature and method of inference, and is post-event / retrospective in terms of research design. The statistical population in this study is the natural customers of Mellat Bank branches over the period 2013-2017 in the six provinces of Tehran, East Azerbaijan, Bushehr, Fars, Kurdistan, and Khuzestan, including the files of about 7330 natural customers. After data collection, Excel was used for summarizing and calculating. The final analysis was then performed using a multinomial logistic regression model. The first model, estimated in the structure of the logit model, calculates the probability of non-repayment of the loan using the variables introduced, according to the following equation:

$$P_i = \frac{e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_i}}{1 + e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_i}}$$

$$\log\left(\frac{p(Y=1|X)}{p(Y=0|X)}\right) = \beta X = (b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n)$$

$$p(Y = 1|X) = p(Y = 0|X)e^{\beta X}$$

$$p(Y = 1|X) + p(Y = 0|X) = 1$$

$$p(Y = 0|X)e^{\beta X} + p(Y = 0|X) = 1$$

$$p(Y = 0|X) = \frac{1}{(e^{\beta X} + 1)}$$

$$p(Y = 1|X) = \frac{e^{\beta X}}{(e^{\beta X}+1)} = \frac{1}{(e^{-\beta X}+1)} \quad (1)$$

Where P indicates the two modes of probability of default and non-default for loan borrowers. In the estimated three-stage model, $Y = 0$ indicates no default, $Y = 1$ indicates a default up to a maximum of three months (more than 61 days equal to the previous due date), and $Y = 2$ indicates default more than three months with deferred loan. In the present study, the multinomial logistic regression and Stata software are employed to estimate the model.

Research Models and Variables

Table 1. Independent Research Variables in the Multinomial Logistic Model

Variable type	Variable symbol	Variable	How the variable is measured
Individual conditions	X1	Gender	Men=0, women=1
	X2	Age	The age of the borrowers at the time of obtaining the loan
Economic conditions	X3	Loan value	Loan value
	X4	Loan repayment term	Loan repayment term by month
	X5	Installment interval	Installment interval (months)
	X6	Number of installments	Number of installments
	X7	Quantity of each installment	The quantity of each installment per month according to the interest on the facility and the repayment term
	X8	Loan extension	Extension (extension=1, no extension=0)
	X9	Previous loan	Virtual variable; If the person has already received a loan, 1 and otherwise 0
	X10	Real estate collateral	Real estate collateral=1, non-real estate collateral=0
	X11	Average balance	The average balance of the facility applicant at the time of obtaining a loan (million Rials)
	X12	Facility interest rate	Facility interest rate depending on the type of contract
	X13	Type of facility	Type of facility (capital=1, current=0)
Social conditions	X14	Level of Education	Education level; diploma and below=0, bachelor=1, master and above=2
	X15	Job	Job; government job=1, self-employment=0

Source: Research finding.

Reviewing the literature and following certain studies[1], this paper deals with the customer credit rating and its impact on credit risk. A review of the literature has shown that most previous studies have classified customers into only two classes, good customers and bad customers, where bad customers are those who have spent more than 90 days of a credit commitment. This classification is very general and does not take into account the different intervals and types of default. Banks, on the other hand, need to classify customers more carefully according to their domestic policies so that they will be able to make a variety of decisions on how to interact with customers depending on which class the customer belongs to. Therefore, in the present paper, considering the credit policy of the bank under study (Mellat Bank), according to experts and taking into account the default and repayment behavior of customers and multinomial logistic regression approach, three classes for different types of customers have been defined. The independent research variables and how they are measured are listed in Table 1.

Descriptive Findings

Table 2 contains descriptive statistics that include the central characteristic, dispersion, and deviation from symmetry for the research variables.

Table 2. Descriptive Statistics of the Research Variables

Variable	Average	Mean	Maximum	Minimum	Standard deviation	Kurtosis	Skewness
X1	0.277	0	1	0	0.447	0.999	1.998
X2	42.544	40	95	18	11.208	1.045	4.311
X3	2,010,000,000	63,000,000	2,670,000,000,000	1,250,000	69,700,000,000	38.214	1461.403
X4	54.917	36	480	1	44.271	3.713	22.602
X5	1.036	1	2	1	0.186	4.976	25.757
X6	52.784	36	240	1	39.089	2.943	12.634
X7	7,662,265	1,636,000	1,340,000,000	-902000	65409045.000	18.770	377.513
X8	0.422	0	2	0	0.504	0.431	1.475
X9	0.229	0	1	0	0.420	1.289	2.661
X10	0.112	0	1	0	0.316	2.455	7.029
X11	11,129,034	0	907,000,000	0	58,384,943	7.536	73.091
X12	10.112	5	29	1	7.241	0.537	1.577
X13	0.250	0	1	0	0.433	1.156	2.336
X14	0.220	0	2	0	0.432	1.630	4.406
X15	0.237	0	1	0	0.425	1.235	2.526

Source: Research finding.

In Table 2, the infimum and supremum are given as the bounds of observations for each of the research variables. The difference between the two values indicates the amplitude of change, which is the main indicator of dispersion. The index is strongly influenced by distant observations; therefore, the index of standard deviation was used in the dispersion analysis.

1. These studies are: Moqaddaseh et al. (2021), Gudde Jote (2018), Ume et al. (2018), Mohammadi et al. (2009), Ebrahimi et al. (2009), Keyqobadi et al. (2012), Albadvi et al. (2014), Nawai & Shariff (2012), Safari et al. (2010), Tari et al. (2010), Marrez & Schmit (2009), and Heydarpour & Karzabhi (2009).

Inferential Findings

Unit Root Test (Study of Stationarity)

Before modeling the research, in order to avoid false regressions in the research, the significance of the variables was first examined, for which the augmented Dick-Fuller (ADF) unit test was used. On the basis of the tests performed, the question of whether the time series used had a stationary process (with zero accumulated degree) or a divergent process (with non-zero accumulated degree) was investigated. Considering that the probability value of the unit root tests is below 0.05 in all the following cases (Table 3), the statistical assumption that all the above variables have a single root is rejected. Therefore, these variables are Stationarity. Thus, the model can be estimated without concern about false regression. The results are provided in Table 3.

Table 3. Results of Multinomial Logistic Model Estimation

Variable	Class I	Class II	Class III
C	0.079 (0.03)	0.45 (0.00)	0.87 (0.00)
X15	-0.25 (0.01)	-0.16 (0.02)	-0.11 (0.04)
X14	-0.32 (0.00)	-0.45 (0.03)	-0.56 (0.03)
X13	-0.055 (0.00)	-0.042 (0.00)	-0.035 (0.02)
X12	-0.049 (0.00)	-0.038 (0.00)	-0.029 (0.02)
X11	-0.075 (0.00)	-0.084 (0.00)	-0.102 (0.04)
X10	-0.078 (0.02)	-0.059 (0.03)	-0.043 (0.01)
X9	0.110 (0.03)	0.086 (0.04)	0.073 (0.02)
X8	0.072 (0.02)	0.066 (0.01)	0.058 (0.00)
X7	0.075 (0.00)	0.061 (0.03)	0.047 (0.01)
X6	-0.041 (0.00)	-0.038 (0.00)	-0.022 (0.00)
X5	-0.14 (0.02)	-0.11 (0.00)	-0.081 (0.04)
X4	-0.035 (0.00)	-0.029 (0.03)	-0.021 (0.02)
X3	4.15 (0.02)	6.19 (0.02)	8.98 (0.02)
X2	0.068 (0.03)	0.072 (0.01)	0.084 (0.02)
X1	0.076 (0.00)	0.081 (0.03)	0.095 (0.03)

Source: Research finding.

The coefficient of variable X1 (gender) in the estimated model for the customer classes equals 0.459, 0.948, and 2.042, respectively. This coefficient indicates that as the gender of the borrower changes, the probability of default increases. It shows that when other conditions being equal, changing the gender of the borrower, the log-likelihood of default increases by 1.58%, 2.58%, and 7.70%, respectively. In other words, when the gender of the borrower changes, the probability of default is greater than one.

The coefficient of variable X2 (age) is 0.068, 0.072, and 0.084. This coefficient indicates that as the age of the borrower increases, the probability of default increases. In this case, the log-likelihood of default increases by an average of 0.85, 1.101, and 1.132 units.

The coefficient of variable X3 (loan value) equals to 4.15, 6.19, and 8.98. This coefficient indicates that, other things being equal, the log-likelihood of default increases by 0.68, 0.94, and 1.105 units on average as the loan value increases.

The coefficient of variable X4 (repayment term) equals -0.035, -0.029, and -0.021. This coefficient indicates that the probability of default decreases as the repayment term of the loan increases. It shows that, other things being equal, when the loan repayment term increases by 1%, the log-likelihood of default decreases by 0.67%, 0.52%, and 0.38%, on average, for the different customer classes, respectively.

The coefficient of variable X5 (installment payment interval) in the estimated model equals -0.14, -0.11, and -0.81. This coefficient indicates that as the interval between installments increases, the probability of default decreases. It shows that, if the other conditions being equal, when the installment payment interval increases by 1%, the log-likelihood of default decreases by 0.65%, 0.51%, and 0.24%, respectively, for the different classes of customers.

The coefficient of variable X6 (number of installments) in the estimated model equals -0.041, -0.38, and -0.22. This coefficient indicates that as the number of installments increases, the probability of default decreases. It shows that if the other conditions being equal, as the number of installments increases, the log-likelihood of default decreases by 0.53%, 0.42%, and 0.28% for the different classes of customers, respectively.

The coefficient of variable X7 (quantity of each installment) in the estimated model equals 0.75, 0.061, and 0.048. This coefficient indicates that as the quantity of each installment of the borrower increases, the probability of default decreases. It shows that if the other conditions being equal, as the quantity of each installment of the borrower increases, the log-likelihood of default decreases by 0.65%, 0.54%, and 0.48% on average. In other words, if the borrower has a longer work experience, the probability of default decreases to the probability of on-time repayment.

The coefficient of variable X8 (loan extension) in the estimated model equals 0.072, 0.066, and 0.056. This coefficient indicates that the borrower's extension increases the probability of default. It shows that, if the other conditions being equal, as the loan is extended, the log-likelihood of default increases by 0.53%, 0.47%, and 0.41% on average.

The coefficient of variable X9 (previous loan) in the estimated model equals 0.110, 0.86 and 0.073. This coefficient indicates that the probability of default is higher if the borrower has a previous loan. It shows that if the other conditions being equal, with the previous loan, the log-likelihood of default increases by 0.77%, 0.62%, and 0.58% on average.

The coefficient of variable X10 (real estate collateral) in the estimated model equals -0.078, -0.059, and -043. This coefficient indicates that the probability of default decreases if the borrower has real estate collateral. It shows that if the other conditions being equal, with the borrower's real estate collateral, the log-likelihood of default decreases by an average of 0.56%, 0.46%, and 0.38%.

The coefficient of variable X11 (average balance) in the estimated model equals -0.075, 0.084, and -0.102. This coefficient indicates that as the average balance of the borrower increases, the probability of default decreases. It shows that, other things being equal, as the average balance of the borrower increases, the log-likelihood of default decreases by 0.67%, 0.52%, and 0.46% on average.

The coefficient of variable X12 (interest rate of the facility) in the estimated model equals 0.049, 0.038 and -0.029. This coefficient indicates that the probability of default decreases when the interest rate of the borrower's facility increases. It shows that, other things being equal, when the interest rate of the borrower's facility changes, the log-likelihood of default decreases by 0.48%, 0.42%, and 0.37% on average.

The coefficient of variable X13 (type of facility) in the estimated model equals -0.055, -0.042, and -0.035. This coefficient indicates that the probability of default decreases when the type of borrower facility is capital. It shows that, other things being equal, changing the type of facility of the borrower decreases the log-likelihood of default by 0.56%, 0.50%, and 0.45% on average.

The coefficient of variable X14 (education level) in the estimated model equals -0.32, -

0.45, and -0.56. This coefficient indicates that the probability of default decreases when the borrower's education level increases. It shows that, other things being equal, with a change in the level of the borrower's education, the log-likelihood of default decreases by 0.35%, 0.31%, and 0.27% on average.

The coefficient of variable X15 (occupation) in the estimated model equals -0.25, -0.16, and -0.11. This coefficient indicates that the probability of default decreases if the borrower has an occupation. It shows that if the other conditions being equal, changing the type of occupation held by the borrower decreases the log-likelihood of default by 0.63%, 0.51%, and 0.42% on average.

For the model estimated for the customers, the correct prediction rate of the logit model was 83.95%. After calculating the sensitivity level and the detection level of the model to study the discriminative power of the two classes (good customers and bad customers in this case), a curve called ROC is used. This 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 corresponds to one minus the degree of detection and whose vertical axis corresponds to the sensitivity degree of the model. The closer this curve is to the upper left corner (1,0), the greater is the strength of the model and the distinction between the two classes. At the point (1,0) both the degree of sensitivity and the degree of detection of the model are the highest and equal to one.

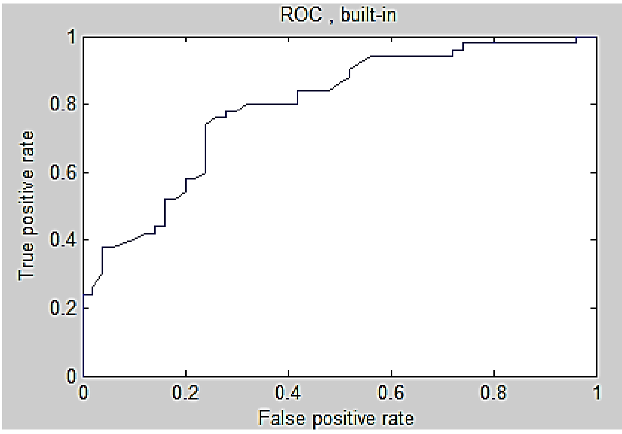

Figure 1. Drawing the ROC Curve for the Multiple Logit Model
Source: Research finding.

The ROC curve is used to evaluate the performance of the model. It is an indicator to evaluate the accuracy of the model. The area under the ROC curve is called the Area under the Curve or AUC, and indicates the level of detection performance compared to the full scoring function. The higher the detection curve, the closer it is to the full scoring function. The value under the curve is between zero and one hundred percent. As shown in Table 4, the area under the logit curve is 0.875. In the case of random prediction of customer behavior, the probability of correct prediction is 0.5 and in the case of prediction with the logit model, this probability is 0.87.

Table 4. The Result of the Calculations Related to the ROC Curve of Legal Customers

95% significance level		Prob. value	Standard deviation	Area under the curve
Upper border	Lower border			
0.935	0.798	0.000	0.024	0.875

Source: Research finding.

Only significant variables are shown in the estimated models. Due to the significance of the coefficients, the probability distribution function in terms of default for different classes is plotted below.

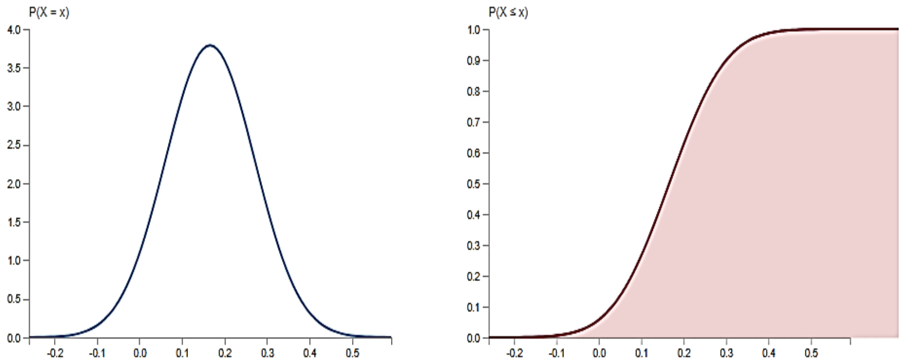

Figure 2. Probability Distribution Function
Source: Research finding.

Analysis and Interpretation of Model Coefficients

In this part of the study, the research model is estimated. The dependent variable of the model is credit from natural customers, which includes four different classes: timely receipt, overdue, deferred, and non-performing loans. Independent variables used in this study are a set of personal information (gender, age), economic (loan value, loan repayment term, installment plan, number of installments, quantity of each installment, loan extension, previous loan, real estate collateral, average balance, facility interest rate, type of facility) and social (education level, occupation) traits of customers, most of which are continuous and numerical variables. Using the data on the above variables for 7330 natural customers of Mellat Bank, the significance of the coefficients was examined at a confidence level of 95%. The results are presented in Table 5.

Table 5. Results of the Model Estimation

Variable	Symbol	Timely receipt loan		Overdue loan		Deferred loan		Non-performing loan	
		Coefficient	Possibility	Coefficient	Possibility	Coefficient	Possibility	Coefficient	Possibility
Gender	X1	0.0000108	0.994	0.0549797	0.534	-0.0000556	0.998	-0.0000145	0.000
Age	X2	-1.57E-07	0.998	-3.22E-03	0.371	3.11E-06	0.997	2.23E-07	0.100
Loan value	X3	3.75E-13	0.000	-6.88E-14	0.902	-5.52E-13	0.000	-7.50E-13	0.000
Loan repayment term	X4	2.34E-07	0.997	-2.18E-03	0.598	3.81E-03	0.000	-1.82E-07	0.240
Installment interval	X5	0.000001	1.000	0.4826938	0.134	-0.1371174	0.107	0.0000198	0.102
Number of installments	X6	-6.66E-07	0.993	-4.83E-03	0.286	-8.72E-03	0.000	3.88E-07	0.023
Quantity of each installment	X7	1.08E-11	0.342	9.56E-10	0.141	1.29E-10	0.451	-9.94E-14	0.000
Loan extension	X8	0.9999857	0.000	-0.1421921	0.073	-0.9998108	0.000	-0.9999725	0.000
Previous loan	X9	-1.10E-06	0.999	-1.24E-01	0.180	-3.99E-05	0.999	-9.97E-07	0.774
Real estate collateral	X10	0.0000198	0.994	0.3648199	0.017	0.1174386	0.003	0.0000478	0.000
Average balance	X11	-5.64E-12	0.664	-1.56E-09	0.036	-1.26E-09	0.000	7.32E-12	0.000
Facility interest rate	X12	-1.32E-06	0.991	-1.63E-02	0.011	-8.98E-06	0.996	7.77E-07	0.001
Type of facility	X13	-0.00000917	0.996	0.1352416	0.185	0.0001275	0.996	0.0000246	0.000
Level of education	X14	-1.57E-05	0.992	-6.18E-01	0.000	-2.94E-02	0.235	-9.37E-06	0.008
Job	X15	4.92E-07	1.000	-9.81E-02	0.293	-1.11E-04	0.996	-1.56E-06	0.656
_cons		1.000003	0.000	2.888492	0.000	4.31337	0.000	4.000003	0.000

Source: Research finding.

According to the results of the regression (Table 5), variables X2 (age), X5 (installment plan), X9 (previous credit), and X15 (job) have no significant effect on the reduction of the bank's long-term receivables. Similarly, variable X1 (gender) have a significant effect only on non-performing loans. The significance of variable X1 (gender) on timely receipt, overdue and deferred receivables is not confirmed at the 95% confidence level. Variable X3 (loan value) has a significant effect on timely receipt, deferred and non-performing loans, and its significance on overdue receivables is not confirmed at the 95% confidence level. Variable X4 (loan repayment term) has a significant effect only on deferred receivables, and its significance on timely receipt, overdue, and non-performing receivables is not confirmed at the 95% confidence level. Variable X6 (number of installments) has a significant effect on deferred and non-performing receivables, and its significance on timely receipt and overdue receivables is not confirmed at the 95% confidence level. Variable X7 (quantity of each installment) has a significant effect on non-performing receivables, and its significance on timely receipt, overdue and deferred receivables is not confirmed at the 95% confidence level. Variable X8 (credit extension) has a significant effect on timely receipt, deferred and non-performing receivables, and its significance on overdue receivables is not confirmed at the 95% confidence level. Variable X10 (type of collateral) has a significant impact on overdue, deferred and non-performing receivables, and its significance is not confirmed at the 95% confidence level for timely receipt receivables. Variable X11 (average balance) has a significant impact on overdue, deferred and non-performing receivables, and its significance is not confirmed at the 95% confidence level for timely receipt receivables. Variable X12 (facility interest rate) has a significant impact on overdue and non-performing receivables, and its significance is not confirmed at the 95% confidence level for timely receipt and deferred receivables. Variable X13 (type of facility) has a significant impact only on non-performing receivables, and its significance for timely receipt, overdue, and deferred receivables is not confirmed at the 95% confidence level. Variable X14 (education level) has a significant effect on overdue and non-performing loans, and its significance is not confirmed at the 95% confidence level for timely receipt and deferred loans.

Conclusion

In the present study, an attempt was made to investigate in more detail the effectiveness of the factors affecting the credit risk of natural customers of the bank using a combined method. For this purpose, firstly, using documentation and library method, the existing and available scientific literature was reviewed in order to enumerate the indicators that are effective in assessing the credit risk of natural customers. Then, considering the credit policy of the bank under study (Mellat Bank), according to the information provided by the experts and taking into account the deferment and repayment behavior of customers and the regression approach as an econometric model, four classes for different types of customers were defined to use the results of the model estimation. Since most of the studies in this area have examined customer credit in the two classes of good and bad, in this study, by extending the credit classification to the four classes of timely receipt, overdue, deferred and non-performing, the effect of this classification on the validity and effectiveness of the factors was examined. This was possible due to the availability of a large amount of correct and accurate data from all classes (the information from 7330 natural customers of Mellat Bank was used in this study). The results show that a wide range of personal information (gender, age), economic (loan value, loan repayment term, installment plan, number of installments, quantity of each installment, loan extension, previous loan, real estate collateral, average balance, facility interest rate, type of facility) and social (education level, job) traits, which have been examined in previous studies as effective indicators for assessing the credit risk of natural customers, failed in some classes (timely receipt, overdue, deferred and non-performing loans) in the calculation of customer credit. Accordingly, this study made it clear that the existing valuation parameters were not consistent with historical data. And this is the answer to why most current credit scoring methods are not accurate in assessing credit scoring. Therefore, dealing with such parameters can jeopardize the credit system. Indeed, the consequences of a credit system with inappropriate parameters in lending to the banking system jeopardize banks' credit risk management and reduce banks' ability to generate returns.

In fact, the complexity of people's behavior due to their influence on environmental conditions means that no absolute predictive models are available to explain customers' credit scoring. This has led to the fact that the various methods and techniques used in customer authentication cannot provide conclusive and very good results, so the results of these methods are relatively measured.

According to the results of this study, the following recommendations are made to improve the credit scoring system of the country's banks:

- Based on the method used in the study, one of the most important results is the initial identification of the main parameters that affect the credibility of natural customers. Depending on the final research results and the different impact of variables affecting validation in each class, it is also necessary to identify and always review other effective variables during the scientific process and according to the opinion of experts.
- Since the customers' credibility is not easy to quantify, the current models are the only risk mitigation tool. In other words, the dependence of these models on accurate, precise, and huge data is unavoidable. Therefore, access to such data will be a basic requirement for building models. Accordingly, it is necessary to improve databases and information systems with current and past financial and credit data of bank customers.
- By introducing a credit scoring system for customers based on the research and review model and improving the process of selecting banks' credit scoring indicators by combining qualitative and quantitative methods, a positive change can be achieved in improving credit risk management.

Finally, due to the different importance of variables in each class, it is essential to use and study newer parameters to discuss customer credit scoring in future research.

References

[1] Ebrahimi, O. (2019). *Investigating the Key Success Factors of Banks in Receiving Non-Current Receivables and Ranking Them with Fuzzy AHP Approach; Case Study: Mehr Eghtesad Bank, Yazd Province* (Unpublished Master Thesis, Yazd University of Science and Art, Iran).

[2] Albadvi, A., Mokhatab Rafiei, F., & Qasemi, Z. (2014). *Application of Electromagnetic Algorithm in Credit Clustering of Bank Customers* (Unpublished Master Thesis, Tarbiat Modares University, Iran).

[3] Tari, F., Qasemi, A. R., & Amir Kavasemi, Sh. (2010). Designing a Credit Model for Banking Customers and Its Role in Reducing Credit Risk of Banks; Case Study: Eghtesad Novin Bank (Unpublished Master Thesis, Allameh Tabataba'i University, Iran).

[4] Heydarpour, F., & Karzebhi, M. (2009). Designing a Model for Credit Evaluation of Legal Costumers in Banks by Using 5C Index. *Financial Knowledge of Securities Analysis, 2*(2), 135–154.

[5] Dehmardeh, N., Shahraki, J., Saifuddinpour, S., & Esfandiari, M. (2002). Loan Customers Validation Using Credit Scoring model (Case Study: Sepah Bank Branches, Zahedan). *Public Management Research, 18*(5), 135–152.

[6] Safari, S., Ebrahimi Shaqaqi, M., & Sheikh, M. J. (2010). Credit Risk Management of Legal Customers in Commercial Banks with Data Envelopment Analysis Approach (Credit Rating). *Management Research in Iran, 14*(4), 137–164.

[7] Eyvazi, E. (2016). *Difference between Exchange Rate Effects in Different Stocks of Stock Return Distribution-Quantile Regression Approach* (Unpublished Master Thesis, Allameh Tabataba'i University, Iran).

[8] Fakoor, E. (2016). *Identifying and Ranking the Factors Affecting the Reduction of Non-Current Receivables in Ayandeh Bank Based on ANP Method; Case Study: Ayandeh Bank Branches in Tehran* (Master Thesis, Islamic Azad University, Iran).

[9] Keyqobadi, A. R., Nemati, A., & Khodami, V. (2012). *Credit Rating of Business Units Receiving Financial Facilities Based on Financial Statements* (Unpublished Master Thesis, Islamic Azad University, Iran).

[10] Mehrabian, A., & Seifipour, R. (2016). Pathology of Current Receivables in the Iranian Banking System. *Financial Economics*, *10*(36), 73–86.

[11] Mohammadi, T., Shakeri, A., Eskandari, F., & Karimi, D. (2017). Factors Affecting the Formation of Non-Current Receivables in the Banking System: A Case Study. *Parliament and Strategy*, *89*(24), 269–299.

[12] Mohammadi, Sh., & Pirmohammadiani, R. (2015). Behavioral Scoring of Bank Customers Using Data Mining Approach and Hierarchical Analysis Process. *Soft Computing and Information Technology*, *4*(3), 66–80.

[13] Mohammadi, M. H. (2019). *Investigating the Factors Affecting the Probability of Default on Loans to Customers of the International Bank of Afghanistan in Herat, Afghanistan* (Unpublished Master Thesis, Ferdowsi University of Mashhad, Iran).

[14] Moqaddaseh, M. (2021). *The Effect of Literacy Level and Income on Banking Customers' Accountability* (Unpublished Master Thesis, Ayatollah Boroujerdi University, Iran).

[15] Hatefi Majoomard, A. (2021). *Identification of the Effective Factors on Credit Risk Management of Banks Listed on the Tehran Stock Exchange* (Unpublished Master Thesis, Yazd University of Science and Art, Iran).

[16] Bandyopadhyay, S., Saha, S., & Pedrycz, W. (2011). Use of a Fuzzy Granulation–Degranulation Criterion for Assessing Cluster Validity. *Fuzzy Sets and Systems*, *170*(1), 22-42.

[17] Bellotti, T., & Crook, J. (2008). Modelling and Estimating Loss Given Default for Credit Cards. Retrieved from https://www.researchgate.net/publication/215991677_Modelling_and_predicting_loss_given_default_for_credit_cards

[18] Bhole, B., & Ogden, S. (2010). Group Lending and Individual Lending with Strategic Default. *Journal of Development Economics*, *91*(2), 348-363.

[19] Dahooie, J. H., Hajiagha, S. H. R., Farazmehr, S., Zavadskas, E. K., & Antucheviciene, J. (2021). A Novel Dynamic Credit Risk Evaluation Method Using Data Envelopment Analysis with Common Weights and Combination of Multi-Attribute Decision-Making Methods. *Computers & Operations Research*, *129*, 105223.

[20] Gomez, R., & Santor, E. (2003). Do Peer Group Members Outperform Individual Borrowers? A Test of Peer Group Lending Using Canadian Micro-Credit Data. Retrieved from Bank of Canada.

[21] Gudde Jote, G. (2018). Determinants of Loan Repayment: The Case of Microfinance Institutions in Gedeo Zone, SNNPRS, Ethiopia. *Universal Journal of Accounting and Finance*, *6*(3), 108-122.

[22] Gup, B. E., & Kolari, J. W. (2005). *Commercial Banking: The Management of Risk*. New Jersey: John Wiley & Sons Incorporated.

[23] Gutu, F., Mulugeta1, W., & Birlie, B. (2017). Determinant Factors Affecting Loan Repayment Performance of Women Borrowers from Micro Finance Institutions in Southwest Ethiopia: Evidence from Four Woredas Around Gilgel Gibe Hydroelectric Power Dam. Global Journal of Management and Business Research, Retrieved from https://journalofbusiness.org/index.php/GJMBR/article/view/2229

[24] Han, Y., & Wang, T. (2021). Semi-Supervised Clustering for Financial Risk Analysis. *Neural Processing Letters*, *53*, 3561-3572.

[25] Haralambie, Maria–Monica. (2016). Corporate Qualitative and Quantitative Assessment. *The Audit Financial Journal*, *14*(140), 868-868.

[26] Marrez, H., & Schmit, M. (2009). Credit Risk Analysis in Microcredit: How does Gender Matter. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1022.6488&rep=rep1&type=pdf

[27] Moradi, Somayeh, & Rafiei, Farimah. (2019). A Dynamic Credit Risk Assessment Model with Data Mining Techniques: Evidence from Iranian Banks. *Financial Innovation*, *5*(1), 1-27.
[28] Nawai, N., & Shariff, M. N. M. (2012). Factors Affecting Repayment Performance in Microfinance Programs in Malaysia. *Procedia - Social and Behavioral Sciences*, *62*, 806-811.
[29] Ngo, T., Le, V., & Le, H. (2021). Factors Affecting Credit Risk in Lending Activities of Joint-Stock Commercial Banks in Vietnam. *Journal of Eastern European and Central Asian Research (JEECAR)*, *8*(2), 228-239.
[30] Niinimaki, J-P. (2012). Hidden Loan Losses, Moral Hazard and Financial Crises. *Journal of Financial Stability*, *8*(1), 1-14.
[31] Nikolopoulos, K. I., & Tsalas, A. I. (2017). Non-performing Loans: A Review of the Literature and the International Experience. In *Non-Performing Loans and Resolving Private Sector Insolvency* (47-68). Berlin: Springer.
[32] Samad, A. (2012). Credit Risk Determinants of Bank Failure: Evidence from US Bank Failure. *International Business Research*, *5*(9), 10-15.
[33] Ume, S. I., Ezeano, C. I., & Obiekwe, N. J. (2018). Analysis of Determinant Factors to Loan Repayment among Broiler Farmers in Enugu State, Nigeria. *International Journal of Environmental and Agriculture Research (IJOEAR)*, *9*, 2454-1850.