
Calculating the Probability of Returning a Loan with Binary Probability Models

Associate Professor PhD Julian VASILEV (e-mail: vasilev@ue-varna.bg)
Varna University of Economics, Bulgaria

ABSTRACT

The purpose of this article is to give a new approach in calculating the probability of returning a loan. A lot of factors affect the value of the probability. In this article by using statistical and econometric models some influencing factors are proved. The main approach is concerned with applying probit and logit models in loan management institutions. A new aspect of the credit risk analysis is given. Calculating the probability of returning a loan is a difficult task. We assume that specific data fields concerning the contract (month of signing, year of signing, given sum) and data fields concerning the borrower of the loan (month of birth, year of birth (age), gender, region, where he/she lives) may be independent variables in a binary logistics model with a dependent variable "the probability of returning a loan". It is proved that the month of signing a contract, the year of signing a contract, the gender and the age of the loan owner do not affect the probability of returning a loan. It is proved that the probability of returning a loan depends on the sum of contract, the remoteness of the loan owner and the month of birth. The probability of returning a loan increases with the increase of the given sum, decreases with the proximity of the customer, increases for people born in the beginning of the year and decreases for people born at the end of the year.

Key words: probability models, probit model, logit model, econometrics, credit risk analysis, SPSS

JEL classification: C – Mathematical and Quantitative Methods

INTRODUCTION

Giving a loan is a risky operation. Credit risk analysis is usually focused on the factors influencing the decision of a credit institution to give or not to give a loan. Risk management calculates the portfolio selection. Calculating the probability of returning a loan may be done in two aspects – full payment of a loan or partial payment. Making clusters of customers and making customer profiles are usual techniques. Financial analysis focuses on special instruments but not on the face-to-face connection between the person who gives the loan and the person who takes the loan. Expectations from both sides have to be evaluated. More over sound statistics and econometrics have

to be applied in order to find an appropriate model which allows calculating the probability of returning a loan.

Analysis of an actual economic phenomenon has to be made – giving a loan. Appropriate methods of observation have to be applied. Mathematical and statistical tools have to be used to analyse an economic phenomenon. Empirical determination of economic laws is an appropriate approach in the studied case. The art of making social research of economic phenomena consists of standard methods as well as a great deal of creative inspiration. Finding some assumptions and checking their validity for a certain case study may lead to conclusions which may be applied to other money lending institutions. The right interpretation of available data will give the researcher the possibility to accept or reject initially formulated hypotheses. So calculating the probability of returning a loan is a specific econometrics research. Calculating the probability of returning a loan has several aspects: economic, mathematical and statistical.

Economic theory is quite reserved in giving some dependencies between the probability of returning a loan and the factors affecting the decision of the credit institution. Economic theory says that some people are given loans, others – not. Some loans are returned as whole, others – partially, others – not. This is the obvious situation in economics. We have to find some explanation of these phenomena.

Monetary theory as part of economic theory has an explanation. Since the country does not emit new banknotes, loan owners cannot pay the principal and the interest as a whole sum, because new banknotes are not emitted. The explanation is a bit rough, but it is clear. It does not describe the economics as a whole system with its specific features.

The mathematical theory may build an equation with several independent variables (factors) and with one dependent variable (the resulting variable) which best describes the phenomenon. Since we have to calculate the probability of returning a loan, probit and logit models are the most appropriate ones. They are based on the binary regression. Since we may predict some but not all factors a coefficient of determination is calculated.

According to Gujarati (2012) econometrics is mainly interested in empirical verification of economic theory. Data analysis is mainly concerned with collecting, processing, and presenting data in the form of charts and tables. Mathematical statistics gives a lot of examples in the sphere of trade, but not in loan estimation. The main methodology of an econometric research consists of the following steps: (1) defining a hypothesis, (2) making a mathematical model, (3) obtaining data, (4) an estimation of the parameters of the model, (5) testing of hypothesis, (6) making a prediction and (7) adapting the model into practice.

It should be mentioned that the probability of returning a loan is a statistical phenomenon, not a deterministic one. The probability may depend on various factors such as – sum of the loan, gender of the customer, age of the customer, year of contract, month of contract, region where the customer lives. All these data are stored in the databases of credit institutions. Historical data may be used to calculate the probability of returning a loan.

The relationship between the independent (input) variables and the dependent (output) variable may be strong or weak. Statistics can never establish a causal connection. We have to make assumptions and check them.

We may use correlation analysis to check the strength of an association between two variables – for instance between the sum of the loan and the probability to be returned. In regression analysis we try to estimate the average value of one variable on the basis of the exact values of other variable. We may predict the probability of returning a loan only by the sum of contract. This is a one-factor model. We may continue with two-factor models by adding other variables to the model. We may continue with three-factor models by adding other variables to the model. In regression analysis the dependent variable is assumed to be random. It has to have a probability distribution.

THEORETICAL BACKGROUND

Probit and logit models are part of qualitative response regression models (QRRM). The dependent variable (the probability of returning a loan) has two values – “yes” or “no”. The dependent variable is binary. QRRM are used in political science where voters have to choose between two parties. QRRM are used in social analysis of households. A family may has a house or it may not have a house. A certain medicine is effective or not for a certain illness. It should be marked that the regressand Y is qualitative. QRRM are also called probability models.

The **linear probability model** (LPM) has the following equation:

$$Y_i = B_1 + B_2 * X_i + u_i \quad [1]$$

Y is the dependent variable. The values of X_i are the different independents variables. The LPM seems to be like a standard regression model. The Y variable is dichotomous. Some problems with LPM have to be marked. It is assumed that u_i values have normal distribution. But when the sample size is large the OLS (ordinary least squares) estimators are usually normally distributed. The disturbances in LPM are usually homoscedastic. The third problem is not fulfilment of $0 \leq Y_i \leq 1$. Calculating the R^2 of the

LPM is a disputable question. Some authors (John Aldrich and Forrest Nelson) think that the coefficient of determination (R^2) should not be calculated in qualitative dependent variable models.

LPM is simple but not very correct. The prediction is that with the increase of X, Y also increases or decreases. So the problem is with the assumption that there is a linear dependency between X and Y. So we may assume that with the increase of the sum of a loan, the probability to be returned increases or decreases.

The equation of the **logistic distribution function** (LDF) is the following:

$$P_i = \frac{e^z}{(1 + e^z)} \quad [2]$$

Where $z = B_1 + B_2 * X_i$

P_i is between 0 and 1. In the logistic distribution function B_1 and B_2 cannot be calculated by the OLS method.

The LDF function works with grouped data showing the relative frequency. For each value of X total number of cases is counted and the number of cases with Y value equal to 1. Or in other words, for each level of X, we compute the probability of returning a loan. The LDF describes the **probit model**. It is sometimes called a normit model.

A comparison of binary logit and probit models is done by Cakmakyapan and Goktas (2013). Both models are appropriate when the dependent variable is binary. They use Monte Carlo simulation to compare both models.

Probit models in credit institutions are used by Teker et.al. (2013). They try to compose a new rating methodology and provide credit notches to 23 countries which of 13 are developed and 10 are emerging. They use 11 variables in the probit model. The factor analysis eliminates 5 of them.

Logit regression is used in credit institutions to find the factors affecting the decision of customers to use e-banking (Annin, K. et. al, 2014). They used a sample with 241 customers. They used binary logistics regression to analyze responses. The most significant factors are the chosen bank, occupational status and monthly income.

Probit models are used in loan repayment of yam farmers in Ghana (Wongnaa and Awunyo, 2013). A random sampling technique is used to select 100 farmers. A questionnaire is made to collect data. They found the factors affecting the loan repayment performance by the adoption of a probit model. They proved that the age has a positive effect on loan repayment.

Estimating the coefficients in a binary regression may be made by their statistical significance. Some articles (Fitrianto and Cing, 2014) focus on theoretical aspects of binary logistics regression.

METHODOLOGY OF THE RESEARCH AND INTERMEDIATE RESULTS

Analyzing the structure of the dataset

We have cross-sectional data from a credit institution. The dataset consists of 699 records for contracts signed between June 2013 and March 2014. The usual period of a contract is 30 days. Customers have to return the loan within 30 days. So we have actual data on the first of August 2014. Some of the fields in the dataset concern the contract: month of contract, year of contract, sum of contract, finished (yes or no). Other fields in the dataset concern the customer: age, gender, month of birth, region (where he/she lives). We use individual data (for the entire population). A population includes each element from the set of observations. The statistics term “population” represents all possible measurements. We do not make a sample. The inferences in this paper are limited to the observed credit institution. They may be valid for other credit institutions or not. Since credit institutions work with confidential data (names of customers, their addresses, number of contracts, credit history), we do not use customer specific data. We have individual (ungrouped) data. We may easily make pivot tables to get grouped data.

Part of the initial dataset with values of individual cases

Table 1

Month of contract	Year of contract	Sum contract	Finished (0 - false; 1 - true)	Age	Month of birth	Gender (0-male; 1-female)	Region
6	2013	220	0	32	5	0	1
6	2013	50	0	70	2	0	1
6	2013	110	0	32	5	0	1
6	2013	70	0	30	12	0	6
6	2013	115	0	30	12	0	6
6	2013	100	0	32	5	0	1

All the fields have numeric values. The region of the loan owner is saved in the database as the name of the town/village, where he/she lives. We have made coding of the region by replacing the text of the town/village with a corresponding number. People who live in the town of the credit institution have region “1” and those who live at most remote places – “16”. So the region coding shows the proximity to the credit institution with a numeric value on an ordinal scale.

Using aggregated data

Before testing the probit model we need to have aggregated data. We have computed aggregated data in Excel.

Aggregated data version 1

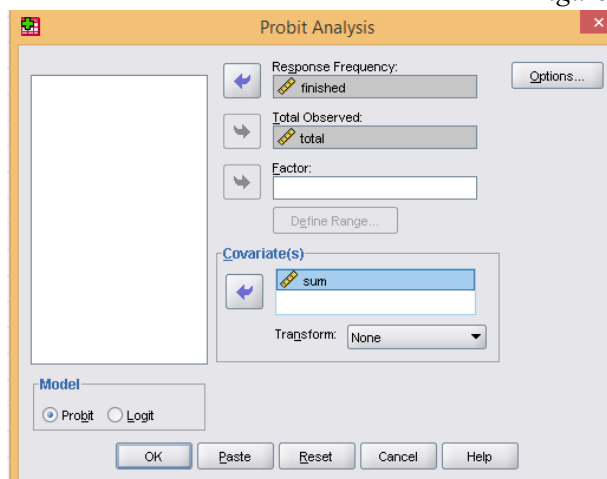
Table 2

Sum	Finished	Total
100	263	555
200	33	59
300	20	50
400	10	13
500	5	8
600	4	6
700	6	7
800	1	1

If we look at the first row, if the sum of contract is less than or equal to 100, 555 loans are given, 263 of them are returned. So the probability of returning a loan, if the sum of contract is less than or equal to 100, is 47%. The aggregated data are copied into SPSS (Hadzhiev et. al., 2009). The analysis starts with a probit model (Analyze/Regression/Probit).

Parameters of probit analysis 1 within SPSS

Figure 1



The Chi-square tests show that the model is adequate. The probit parameters (sum and intercept) are significant at Alfa = 0.05. Their significance is 0.015 and 0.021.

Parameter estimates of the probit model 1

Table 3

	Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
PROBIT ^a	sum	,001	,000	2,421	,015	,000	,002
	Intercept	-,186	,081	-2,303	,021	-,266	-,105

a. PROBIT model: PROBIT(p) = Intercept + BX

The interpretation of the probit model estimates is the following. If the sum of loan increases with 100, the probability of returning it increases with 10% (0.001 * 100). Expected responses with the probit model are calculated by SPSS.

Cell count and residuals of probit model 1

Table 4

	Number	sum	Number of Subjects	Observed Responses	Expected Responses	Residual	Probability
PROBIT	1	100,000	555	263	260,917	2,083	,470
	2	200,000	59	33	30,341	2,659	,514
	3	300,000	50	20	27,911	-7,911	,558
	4	400,000	13	10	7,819	2,181	,601
	5	500,000	8	5	5,148	-,148	,644
	6	600,000	6	4	4,103	-,103	,684
	7	700,000	7	6	5,055	,945	,722
	8	800,000	1	1	,758	,242	,758

If we look at the sum 300, we take into consideration loans with sums between 200 and 300. The empirical observation shows that 20 loans of 50 were returned – or it is 40%. But the probit model calculates that the probability of returning a loan with sum of contract between 200 and 300 is 55.8%.

It is obvious that most of the loans 555 of 699 are with sum less than or equal 100. It is appropriate to make a second probit model with loans below 100. In MS Excel we filter the dataset and we use a formula to calculate the rank.

Filtering the dataset and calculating the rang

Table 5

Sum_contract	Finished (0 - false; 1 - true)	Rank
50	0	5
70	0	7
100	0	10
65	0	7
50	0	5

The formula for calculating the rank is the following:

$=ROUND(C3/10;0)$

C3 is the sum of contract for the first loan.

With a Pivot table we compute total observed and finished.

Aggregated dataset 2 for probit model 2

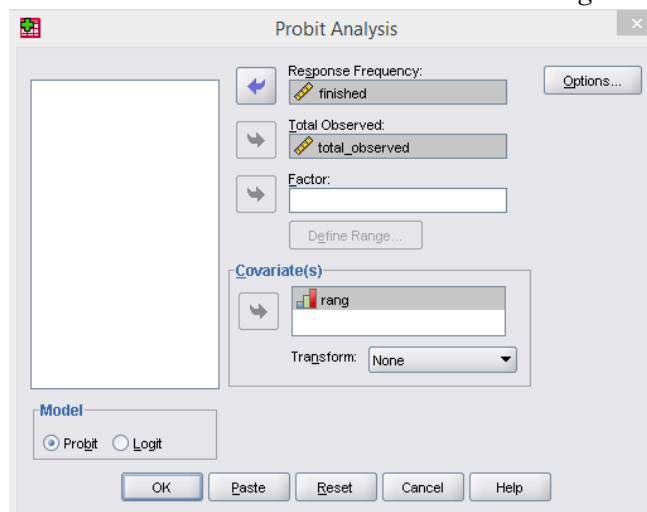
Table 6

Row labels	Count of finished (0 - false; 1 - true)	Total observed (0 - false; 1 - true)
1	19	42
2	59	124
3	37	86
4	30	57
5	56	97
6	15	40
7	8	21
8	16	28
9	5	15
10	18	45
Grand total	263	555

Now we may use again SPSS to calculate and test the probit model 2 (Analyze/Regression/Probit).

Parameters of probit analysis 2 within SPSS

Figure 2



The parameter estimates table shows the significance of the probit model parameters (rang and intercept). Both of them have significance greater than 0.05 (relatively 0.571 and 0.894). The conclusion is that the second probit model cannot be made. So we cannot expect the probability of returning a loan to increase or decrease with the increase with 10 levs of the sum of contract. Our conclusion may be confirmed by the Cell counts and residual table for the second probit model.

Cell counts and residual table for the second probit model

Table 7

	Number	rang	Number of Subjects	Observed Responses	Expected Responses	Residual	Probability
PROBIT	1	1,000	42	19	20,569	-1,569	,490
	2	2,000	124	59	60,153	-1,153	,485
	3	3,000	86	37	41,320	-4,320	,480
	4	4,000	57	30	27,123	2,877	,476
	5	5,000	97	56	45,707	10,293	,471
	6	6,000	40	15	18,663	-3,663	,467
	7	7,000	21	8	9,701	-1,701	,462
	8	8,000	28	16	12,805	3,195	,457
	9	9,000	15	5	6,791	-1,791	,453
	10	10,000	45	18	20,165	-2,165	,448

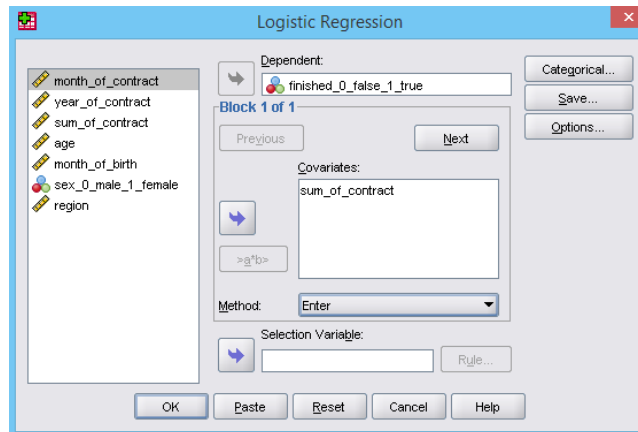
It is obvious that for contracts with sums between 1 and 100 levs, the probability of returning a loan is between 44.8% and 49%.

Using cross-sectional data – values of individual cases
Estimating one-factor models

The **binary logistics model** may be estimated in SPSS. We may use the values of individual cases instead of using aggregated data.

Parameters of the binary logistics model in SPSS

Figure 3



SPSS says „Estimation terminated at iteration number 3 because parameter estimates changed by less than 0.001”. The variables in the equation are the following: $B_0 = -0.18$ and $B_1 = 0.02$. The significance of B_0 is 0.060. The significance of B_1 is 0.021.

A binary logistics model with an independent variable “sum of contract” and including the constant in the equation is adequate. The binary logistics equation is mentioned in equation [2].

$$P = \frac{e^z}{(1 + e^z)} \quad [3]$$

Where $z = B_0 + B_1 * X = -0.18 + 0.02 * X$
 X is the sum of contract.

For a loan of 100 levs (approximately 54.12 EURO), the probability of returning it is computed in the following way.

$$Z = -0.18 + 0.02 * 100 = 1.82$$

$$P = \frac{e^{1.82}}{1 + e^{1.82}} = \frac{6.17}{7.17} = 0.86 = 86\% \quad [4]$$

For a loan of 50 levs, the probability of returning it is computed in the following way.

$$Z = -0.18 + 0.02 * 50 = 0.82$$

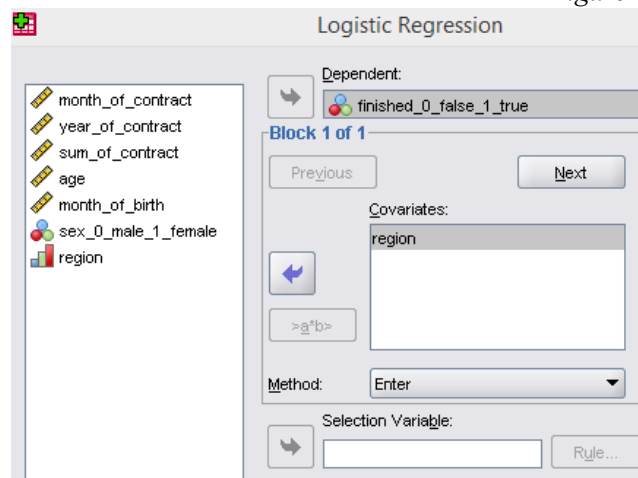
$$P = \frac{e^{0.82}}{1 + e^{0.82}} = \frac{2.27}{3.27} = 0.69 = 69\% \quad [5]$$

If we exclude the constant from the equation, the B_1 variable in the equation is not statistically significant (Sig. = 0.134).

We may make a binary logistics regression with an independent variable “region”.

Parameters of the binary logistics regression with an independent variable “region”

Figure 4



We try the regression including the constant in the equation. $B_0 = 0.103$. $B_1 = -0.62$. B_0 is not statistically significant (Sig = 0.285). B_1 is statistically significant (Sig = 0.016).

We try the regression excluding the constant in the equation. $B_1 = -0.145$. B_1 is statistically significant (Sig = 0.024).

A binary logistics model with an independent variable “region” and excluding the constant in the equation is adequate.

The equation is the following:

$$P = \frac{e^z}{(1 + e^z)} P = \frac{e^z}{(1 + e^z)} \quad [6]$$

Where $z = B_1 * X = -0.145 * X$ $z = B_1 * X = -0.145 * X$
 X is the region.

In our case study the region changes from 1 to 16, where 1 is relative to people living in the town of the credit institution, and 16 – for people living far away.

For $X = 1$ the calculation is the following.

$$Z = -0.145 * X = -0.145 * 1 = -0.145$$

$$P = \frac{e^z}{1 + e^z} = \frac{e^{-0.145}}{1 + e^{-0.145}} = 0.46 = 46\% \quad [7]$$

So for people living in the town of the credit institution the probability of returning a loan is 46%.

For $X = 8$ we calculate the probability of returning a loan.

$$Z = -0.145 * X = -0.145 * 8 = -1.16$$

$$P = \frac{e^z}{1 + e^z} = \frac{e^{-1.16}}{1 + e^{-1.16}} = 0.24 = 24\% \quad [8]$$

For people living far from the credit institution, the probability of returning a loan decreases.

A binary logistics model may be checked by entering the month of the contract as an independent variable. B_0 (Sig=0.513) and B_1 (Sig=0.659) are not statistically significant. **The month of the contract cannot be an independent variable in a binary logistics model.**

A binary logistics model may be checked by entering the year of the contract as an independent variable. B_0 (Sig=0.489) and B_1 (Sig=0.489) are not statistically significant. **The year of the contract cannot be an independent variable in a binary logistics model.**

A binary logistics model may be checked by entering the age as an independent variable. B_0 (Sig=0.122) and B_1 (Sig=0.076) are not statistically significant. **The age cannot be an independent variable in a binary logistics model.**

A binary logistics model may be checked by entering the month of birth as an independent variable. B_0 (Sig=0.031) and B_1 (Sig=0.006) are statistically significant. **The month of birth can be an independent variable in a binary logistics model.** $B_0 = 0.339$. $B_1 = -0.059$

The equation is the following:

$$P = \frac{e^z}{(1 + e^z)} \quad [9]$$

$$\text{Where } z = B_0 + B_1 * X = 0.339 - 0.059 * X$$

X is the month of birth.

In our case study the month of birth changes from 1 to 12, where 1 is relative to people born in January and 12 – for people born in December.

For $X = 1$ (people born in the first month of the year) we calculate the probability of returning a loan.

$$Z = 0.339 - 0.059 * X = 0.339 - 0.059 * 1 = 0.28$$

$$P = \frac{e^z}{1 + e^z} = \frac{e^{0.28}}{1 + e^{0.28}} = 0.57 = 57\% \quad [10]$$

57% of the people born in January are expected to return their loans.

For $X = 6$ (people born in the sixth month of the year) we calculate the probability of returning a loan.

$$Z = 0.339 - 0.059 * X = 0.339 - 0.059 * 6 = -0.015$$

$$P = \frac{e^z}{1 + e^z} = \frac{e^{-0.015}}{1 + e^{-0.015}} = 0.50 = 50\% \quad [11]$$

50% of the people born in June are expected to return their loans.

A binary logistics model may be checked by entering the gender as an independent variable. B_0 (Sig=0.097) is not significant, but B_1 (Sig=0.020) is statistically significant. **The gender cannot be an independent variable in a binary logistics model.**

Estimating a three-factor model

We may conclude that the sum of contract, the region and the month of birth influence in great extend the probability of returning a loan. We may try to make a three factor binary logistics model by including and by excluding the constant in the equation.

A three factor binary logistics model including the constant in the equation

Table 8

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	sum_of_contract	,002	,001	6,646	1	,010	1,002
	region	-,069	,026	6,936	1	,008	,934
	month_of_birth	-,056	,021	6,915	1	,009	,945
	Constant	,328	,176	3,468	1	,063	1,388

a. Variable(s) entered on step 1: sum_of_contract, region, month_of_birth.

The constant is not statistically significant (Sig=0.63). The other variables (B₁, B₂ and B₃) are statistically significant. We may not make an equation. We may try a three factor model by excluding the constant from the equation.

A three factor model by excluding the constant from the equation

Table 9

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	sum_of_contract	,002	,001	10,619	1	,001	1,002
	region	-,056	,025	4,979	1	,026	,946
	month_of_birth	-,025	,013	3,579	1	,059	,975

a. Variable(s) entered on step 1: sum_of_contract, region, month_of_birth.

We may construct an equation at Alfa = 0.06.

$$P = \frac{e^z}{(1 + e^z)} \quad [12]$$

Where $z = B_1 \cdot X_1 + B_2 \cdot X_2 + B_3 \cdot X_3 = -0.023 \cdot X_1 - 0.056 \cdot X_2 + 0.02 \cdot X_3$

X₁ is the month of birth, X₂ is the region and X₃ is the sum of contract.

A numeric example may be given. For instance a borrower, born in June (X₁ = 6), living in the town of the credit institution (X₂ = 1), wants to take a loan of 100 levs (X₃ = 100).

$$z = -0.023 * X_1 - 0.056 * X_2 + 0.02 * X_3 = -0.023 * 6 - 0.056 * 1 + 0.02 * 100 = 1.806$$

$$P = \frac{e^z}{(1 + e^z)} = \frac{e^{1.806}}{1 + e^{1.806}} = \frac{6.086}{7.086} = 0.8589 \quad [13]$$

The probability to return the loan for the mentioned borrower (born in June, living in the town of the credit institution and wants to take a loan of 100 levs) is 86%.

RESULTS

If we look at the signs of B_1 , B_2 and B_3 in the equation, we may conclude the following. With the increase of the month of birth, the probability of returning a loan decreases (-0.025 is a negative number). With the increase of the remoteness of a loan owner, the probability of returning it decreases (-0.056 is a negative number). With the increase of given sum, the probability of returning a loan increases (0.02 is a positive number).

The conducted methodology of research in this paper proves that the month of signing a contract, the year of signing a contract, the gender and the age of the loan owner do not affect the probability of returning a loan. The probability of returning a loan depends on the sum of contract, the remoteness of the loan owner and the month of birth. The probability of returning a loan increases with the increase of the given sum, decreases with the proximity of the customer, increases for people born in the beginning of the year and decreases for people born at the end of the year. People born in the beginning of the year are more willing to return their loan than those who are born at the end of the year. People who live in the town of the credit institution are more inclined to return their loans than people living far away. Loans with comparatively greater sums are more likely to be returned than loans with comparatively small amount of money.

CONCLUSIONS

Calculating the probability of returning a loan may be done by using the whole dataset from a loan institution (micro data for the entire population). Applying statistical and econometric methods the probability may be calculated. The most appropriate method concerns the qualitative response regression models. The simplest model is the linear probability model (LPM). Taking into consideration the problems with the LPM some other models may be used – such as logit and probit models. Data at individual level and grouped data are analyzed.

It is proved that the month of signing a contract, the year of signing a contract, the gender and the age of the loan owner do not affect the probability of returning a loan. It is proved that the probability of returning a loan depends on the sum of contract, the remoteness of the loan owner and the month of birth. The probability of returning a loan increases with the increase of the given sum, decreases with the proximity of the customer, increases for people born in the beginning of the year and decreases for people born at the end of the year.

The analysis conducted in this paper is quantitative but several qualitative conclusions may be brought out. People born in the beginning of the year are more willing to return their loan than those who are born at the end of the year. People who live in the town of the credit institution are more inclined to return their loans than people living far away. Loans with comparatively greater sums are more likely to be returned than loans with comparatively small amount of money.

Future research may be focused on the applying a tobit model or an ordered probit model or a nominal probit model on the same dataset. Duration models are also appropriate and applicable on a dataset containing data for given loans. Instead of calculating the probability of returning a loan with a duration model we can find the most important factors determining the duration of a loan contract. The inferences in this paper are limited to the observed credit institution. They may be valid for other credit institutions or not. Future research may be focused on the analysis of datasets of several credit institutions operating within a geographic region. The same hypotheses may be tested with the same methodology. In this case a sampling method may be applied. The conclusions will be valid for several credit institutions, not only for one – as it is in this paper.

REFERENCES

1. Annin, K. and others. (2014). Applying Logistic Regression to E-Banking Usage in Kumasi Metropolis, Ghana. *International Journal of Marketing Studies*; Vol. 6, No. 2, pp. 153-162.
2. Cakmakyapan, S. and Goktas, A. (2013). A comparison of binary logit and probit models with a simulation study. *Journal of social and economic statistics*, Vol. 2, No 1, pp. 1-17.
3. Fitrianto, A. and Cing, N. 2014. Empirical Distributions of Parameter Estimates in Binary Logistic Regression Using Bootstrap. *Int. Journal of Math. Analysis*, Vol. 8, no. 15, 721 – 726.
4. Gujarati, D. (2004). *Basic econometrics*. 4th edition Available: <http://egei.vse.cz/english/wp-content/uploads/2012/08/Basic-Econometrics.pdf> <30.10.2014>.
5. Hadzhiev, V. and others. (2009). *Statistical and econometric software*, Varna: Nauka I ikonomika (in Bulgarian).
6. Teker, D. et.al. (2013). Determination of Sovereign Rating: Factor Based Ordered Probit Models for Panel Data Analysis Modelling Framework. *International Journal of Economics and Financial Issues*. Vol. 3, No. 1, pp.122-132.
7. Wongnaa, C. and Awunyo-Vitor, D. (2013). Factors Affecting Loan Repayment Performance Among Yam Farmers in the Sene District, Ghana. *Agris on-line Papers in Economics and Informatics*, Vol. 5, N 2, p.111-122.