A Quantitative Approach to Credit Risk Management in the Underwriting Process for the Retail Portfolio

Andreea Costea

The core of this paper encloses a mathematical approach of credit risk management, based on a scorecard model used in the bank's underwriting process. The main purpose of this paper is to present how to develop, validate and apply a rating model in practice. Using 21568 loan applications provided by one of the largest banks from Romania, a scorecard is built for the underwriting purposes. The customer data used in the modeling is based on socio-demographic characteristics. The model is developed according to a set of statistical methods for parameter estimation. A real-life example of how to use such a model in the strategic decisions of a bank is presented. The cut-off score for the acceptance of the applications is calibrated to a potential risk appetite of the main four banks in Romania. From an evaluative perspective, this paper is compatible with an exploratory approach to quantitative research methodology.

Keywords: Credit risk management, Basel III, Retail Scorecard, Cut-off calibration

JEL Classifications: G17 Financial Forecasting and Simulation

1. Introduction

The management of credit risk is found unquestionably among the most important topics of study within financial risk management both at university and banking institutions level. The financial crisis, the regulatory framework introduced by the Basel Committee, and the need for an adequate credit risk management, drive banks to develop appropriate statistical models.

The motivation for choosing this theme of study comes with the pragmatic necessity to develop a credit risk model, based on real data. This model estimates the probability of default for each customer and is used to implement a risk-based threshold for the acceptance of future customers.

Work methodology consists of a combination of qualitative research, such as legal analysis, and quantitative research, respectively statistical analysis. The study of international law in matters of banking regulations reveals the main issues to be considered by all banks in the European banking system. Statistical analysis applied to the development of scorecards represents a practical implementation of the legislation.

The personal contribution of the author consists in integrating the qualitative with the statistical approach. The qualitative approach presents a pragmatic analysis of international banking legislation. The public disclosure reports issued by four banks in Romania are included in the analysis. The statistical approach is centered on development of a credit risk model, known as a Retail Scorecard. The model is based on data provided by one of the largest banks in Romania. The main objective of the present paper is to highlight both the elements of international banking laws and how to create an effective risk model that respects the law and can be applied in practice.

2. Review of the scientific literature

Depending on the purpose for which they are used, credit risk models are divided into two categories. Therefore, we distinguish between individual loss models and portfolio loss models.

Examples of individual loss models are Merton's model (1974) and Logistic Regression.

Merton's model (1974) applies the option pricing theory developed by Black & Scholes (1973), for modeling the company's debt. According to the model, if a company enters into default at the time of payment of the debt, its assets are less than its liabilities. Capital structure of the company is supposedly composed of equity, equal to the positive part of the difference between the asset value of the company and zero coupon bond with maturity T and face value D. Therefore, the share capital of the company is modeled as an european option on the asset value, with maturity T and strike D.

Logistic regression is one of the most used tools in applied statistics and discrete data analysis. Its purpose is to identify a model that can describe the relationship between a variety of independent variables (X1, X2, ..., Xn) and a dependent variable (Y). More specific, it uses customer characteristics (age represents X1, X2 represents income, etc.) as independent variables and the dependent variable Y is dichotomous.

$$Y = \begin{cases} 1 \text{ for } \textit{non-default} \text{ customers} \\ 0 \text{ for } \textit{default} \text{ customers} \end{cases}$$

The estimated probability of default is modeled by the following equation:

$$P\,(Y=1) = L(\beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \ldots + \beta_n \cdot X_n + \varepsilon), \quad (1)$$

where β_k are logistic regression coefficients, X_k are independent variables, ε is standard deviation, and L is logit function defined by $L(x) = \frac{1}{1+e^{-x}}$. (2)

The expression (2) can be rewritten as follows:

$$\ln\left(\frac{P\,(Y=1)}{1-P\,(Y=1)}\right) = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \ldots + \beta_n \cdot X_n + \varepsilon. \quad (3)$$

At the left side we find the logodds and the logistic regression express the logodds as a linear function of customer's characteristics.

Logistic regression model is the approach used practice to estimate the rating models used in Basel III. The model demonstrates its applicability and robustness through its use by all banks have adopted, according to regulatory requirements of the European Union, the IRB approach.

The portfolio loss models are built on the idea presented by Vasicek (1987). The simplicity of his idea led to the extension of his model both in regulatory requirements (Basel III) and in the financial industry (JP Morgan, Credit Suisse, KMV).

The model presented by Vasicek (1987) provides a strong mathematical background on which individual losses are aggregated at the portfolio level. The model assumes n customers with individual losses L_i and describe the aggregate loss distribution as $L = \sum L_i$.

As such, banks have developed the Vasicek model in refined forms that are used to measure potential losses in the value of a portfolio within a specified period and for a given confidence interval (α = 0.05 and α = 0.01). The standard industry models that use this approach are J.P. Morgan's CreditMetrics, CreditPortfolioManager's KMV, Credit Suisse Financial Products' Credit Risk +.

3. Research methodology and data & modeling techniques

Banks play a key role in the financial system and the economy of each country. For the purpose of reducing the credit risk in their banking system, many countries have introduced international prudential standards issued by the Basel Committee. Many banks use internal rating models and in Romania the most widely used model is the one based on logistic regression, known as Scorecards.

3.1 Research methodology

The need for a prudent behavior is becoming increasingly important at international level and is supported by the introduction of rules in the banking sector. Banks in Romania have the obligation to issue, on a yearly basis, the report on transparency and public disclosure. These reports are intended to meet the requirements on public disclosure regarding the risk assessment and capital management in banking sector. Their purpose is to ensure an appropriate level of transparency and provide an overview of risk management within the institution which disclosed it.

The qualitative research is focused on the banking legislation and on the way in which this is accomplished by the major banks in Romania, e.g. Raiffeisen Bank, BCR, UniCredit Tiriac Bank and Banca Transilvania. The purpose of this analysis is to provide an integrated overview of the situation in the banks in Romania. Reference year for these reports is 2014.

A comparative approach is used in this analysis. Elements which have a significant impact on credit risk are taken into consideration in the comparative approach, in order to generate a qualitative and graphic analysis.

The Scorecard developed is intended to measure the risk associated with each individual customer of the bank. The estimation of the model is taking into account historical information. This Scorecard may be used in the following activities:

- approve the credit decisions, respectively accepting the performing customers and rejecting the risky customers;
- estimate the credit risk expected loss and calculating provisions requirement at the bank;

- estimate the unexpected loss in credit risk and determine both economic and regulatory capital;
- determine risk-adjusted interest rates, by reducing the rates for low-risk customers and introducing additional requirements for clients with increased risk.

The scorecard was developed based on the logistic regression. This approach is in line with the methods used in the risk management of the Romanian banks. The estimation of the Scorecard was developed using Excel and the statistical program R.

3.2 Data and modeling techniques

Data about 21,568 customers of a bank in Romania was used to develop the score card. The data contains information starting with the year 2014-2016, which means that the results are representative for the current portfolio.

The variables considered in the Scorecard have been both qualitative (residential place, marital status, etc.) and quantitative (income, etc.). For both categories of variables, customers are grouped according the values of these variables into groups or ranges of values, called attributes.

The advantages associated with the grouping approach are the following:

- understanding the customer portfolio by observing distributions on groups;
- existence of a distinct group for missing values;
- obtaining a relationship between groups in each variable and credit risk.

Within the development sample were recorded also missing values, so they were treated as separate attributes. Therefore, it excludes the possibility that the actual results would be negatively affected by an unfavorable data grouping.

The approach used in grouping data is given by the ratio between the distribution of performing customers and distribution of defaulted customers within this attribute.

For example, for the income attribute between 1017 and 1468, we will consider the following value:

$$V(\omega) = -\ln \frac{\dfrac{\#\{\,defaulted\ customers\ with\ income\ ranging\ between\ 1017\ and\ 1468\}}{\#\{defaulted\ customers\}}}{\dfrac{\#\{performing\ customers\ with\ income\ ranging\ between\ 1017\ and\ 1468\}}{\#\{performing\ customers\}}} = -0.39 \quad (4)$$

This formula is called Weight of Evidence (WOE), because it defines the discrimination power of each attribute. It can be interpreted as follows:

- a negative value of WOE indicates that within this group there is a lower percentage of the performing customers compared to the percentage of the defaulted customers distribution; also a negative value may also indicate that the ratio between performing customers and defaulted customers is smaller within this group compared with the rest of the portfolio;
- a positive value of WOE indicates that this group is safe; therefore, a higher WOE implies a safer group.

In computation of WOE it is necessary to divide the variable into groups, in order for WOE to record different values. This is necessary because a variety of WOE values indicates a high power of discrimination of the model.
After grouping the variables and calculating the WOE, the following steps are required for the final selection of variables to be included in the model:

- testing distribution and economic relevance of each variable;
- testing the discrimination power at individual level of each variable.

After the selections of variables which will be part of the model, an additional test is required, that will be performed on sample data validation. The validation sample is composed of a subset of the sample that covers all the historical information collected by the bank.

After the model is validated, each client in the portfolio has attached a credit score and a probability of default. Using this information together with the key risk metrics from the reports of the four major banks to derive a quantitative cut-off score for the underwriting process. The cut-off is set such that the risk of the clients approved by the model is in line with the risk of the current portfolio of each bank. This approach provides an idiosyncratic threshold for each bank, aligned with its risk strategy.

4. Estimation results and discussion

In this section we present the estimation results obtained through a comparative analisys of reports disclosed by four banks from Romania, but also we present a model that can be used to determine and identify the characteristics of performing customers in the credit risk management.

4.1. The comparative approach for credit risk according to disclosed reports

The legal framework provided by Regulation 575/2013 presents, in Part VIII, the main duties that financial institutions must achieve regarding the requirements on transparency and public disclosure.

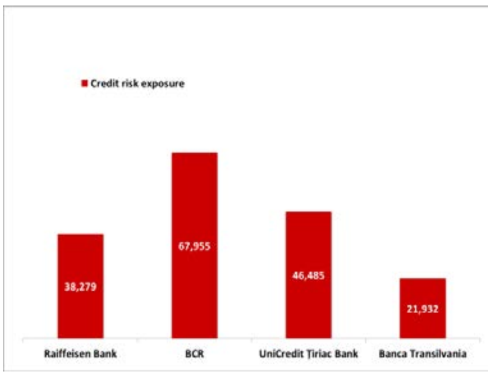
Within Part VIII, Title II, Article 435 it states that financial institutions are required to publish their objectives and policies related to risk management. Requirements related to disclosure of information on capital requirements are presented in Article 438. Information on credit risk adjustments is presented in Article 442. The institutions are required to disclose information on approaches and methods adopted for determining specific and general adjustments for credit risk.

Articles 452-453 present the issues related to qualification requirements for certain instruments or methodologies. These requirements of Regulation 575/2013 are applicable to the institutions using the IRB approach for credit risk.

With the aim of presenting the comparative approach on transparency reports disclosed by the four banks, we consider as benchmarks: credit risk exposure, risk weighted assets (RWA), adjustments for impairment and non-performing loans (NPL).

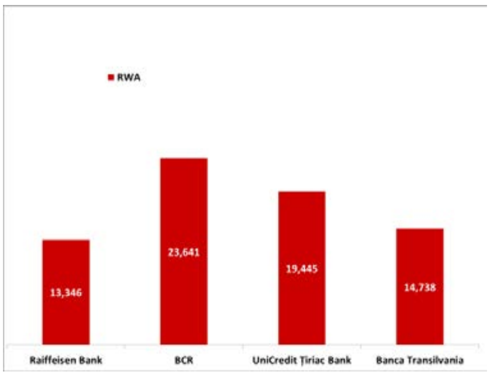
At the level of the four banks, the exposure to credit risk is analyzed and quantified in order to identify the amount of necessary provisions. BCR has the largest portfolio, with high volumes for the retail loans.

Figure 1.
Credit risk exposure



Source of data: Figure realized by author

Figure 2.
RWA



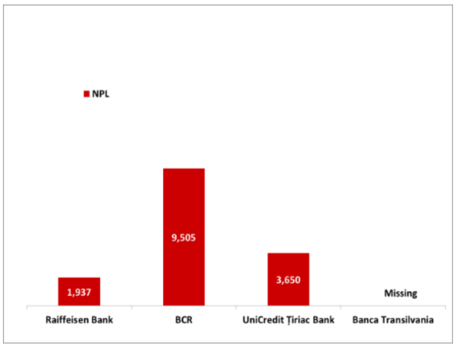
Source of data: Figure realized by author

Closely related to exposure to credit risk is the RWA indicator, which measures the bank's risk-weighted assets. The indicator is used to calculate the minimum amount of capital required in the bank to cover unexpected losses from credit risk.

In this case it is observed that the higher RWA value is recorded by BCR. The average weight used in RWA has the largest value in case of Banca Transilvania, reaching 67%. Considerably lower values are realized by UniCredit Tiriac Bank (42%), BCR (35%) and Raiffeisen Bank (35%).

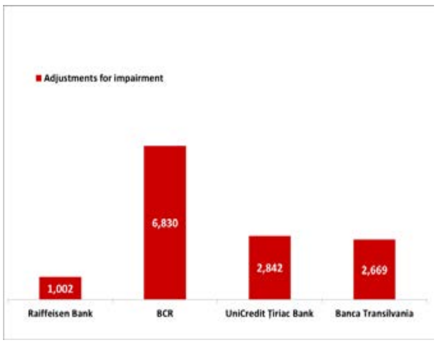
The NPL levels must be presented in a meaningful comparative analysis. The NPL level impacts the banks' recoveries through losses arising from the sale of loans in default. It can be seen that the highest level of NPL is recorded in BCR, followed by UniCredit Tiriac Bank and Raiffeisen Bank. In addition, the size of NPL is closely related to the size of exposure to credit risk and risk of each bank. Even if BCR has the largest amount of NPL, it is important to note that also exposure to credit risk records the highest values in case of this bank. The accumulation of NPL for BCR could be explained by the financial crisis, combined with a "not to sell" strategy regarding the NPL in the market. In 2015, the year prior to which transparency reports were disclosed, there has been a change in the risk strategy of BCR. The bank launched in the market substantial packages of NPL, the most important package being estimated at around 4,500 to 6,000 mio RON[1]. Thus, by selling these packages, the bank could reach a NPL level comparable with the other banks in the market.

Figure 3.
NPL



Source of data: Figure realized by author

Figure 4.
Adjustments for impairment



Source of data: Figure realized by author

[1] Hostiuc, C.,2015. BCR is preparing to put up for sale in bulk NPL of 1-1.5 bln. Euro. The transaction of the banking market is named "Neptune". Ziarul Financiar, online at [http://www.zf.ro/zf-24/bcr-se-pregateste-sa-scoata-la-vanzare-in-bloc-credite-neperformante-de-1-1-5-mld-euro-tranzactia-anului-pe-piata-bancara-poarta-numele-de-cod-neptun-14104446] visited 17.04.2016

Banks create provisions, known as adjustments for impairment to record further losses from credit risk. As previously mentioned, it is natural to find that BCR has the highest value adjustments for impairment. Subsequently, UniCredit Tiriac Bank records the value of adjustments for impairment in close correlation with the level of NPL, which is also applicable in the case of Raiffeisen Bank. In the case of Transilvania Bank the author cannot issue an opinion, because the level of NPL has not been disclosed.

The necessity to adopt a prudential behavior can be seen from the interdependence of the four elements included into comparative analysis. The unfavorable situations in credit risk have direct repercussions on the performance of the bank. These effects can me mitigated through an optimal capital adequacy and a prudent approach regarding underwriting loans. The risk strategy for Romanian banking system is aligned with internationally regulated requirements for credit risk. This implies an appropriate level of risk and expected loss provisioning for all four banks that are part of the present analysis.

4.2 Statistical modeling of credit risk

In this section we present the estimation of a model that distinguishes with high accuracy between performing customers (i.e. customers that repay the loan) and defaulted customers. These models can be used to determine and identify the characteristics of performing customers in the underwriting process. In this way the risk frameworks assures a safe growing of the bank's portfolios, bounded by the risk covered by provisions and internal capital.

4.2.1 Scorecard estimation based on data from the banking market in Romania

In order to realize the Scorecard, the sample was divided as follows: a development sample and a validation sample. The scorecard is estimated based on the development sample and the validation sample it used to verify the evolution of the discriminatory power in time. The development sample contains 15,104 observations and the validation sample 6,464 observations.

We present in the following sections the steps necessary for the development of the scorecard model:

- description of individual factors
- selection criteria for the variables of the model
- estimation of the model in R statistical program
- analysis of discriminatory power of the model based on development and validation samples

I. Description of individual factors (WOE grouping)

First of all, the computation was performed in order to find the number of performing customers and the number of customers in default. Subsequently, the aggregate distribution was estimated at the level of variable for each attribute. The aggregate distribution was divided into distribution of performers customers ("Distribution of goods") and distribution of defaulted customers ("Distribution of bads").

By knowing the values for the two distributions the WOE is calculated according to the formula:

$$\text{WOE} = -\ln \frac{\frac{\#\{default\ cusotmers\ from\ attribute\}}{\#\{default\ customers\}}}{\frac{\#\{performing\ customers\ from\ attribute\}}{\#\{performing\ customers\}}} \qquad (5)$$

Product variable

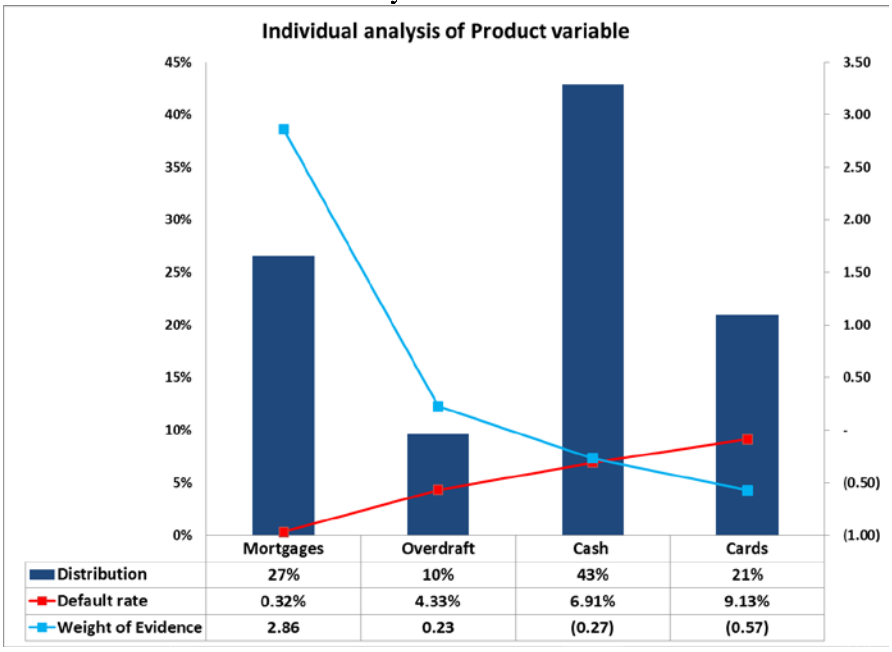
As it can be seen from the individual analysis of the Product variable, 43% of customers are choosing the Cash product. This product has a default rate, expressed in absolute terms, of 6.91%. This rate is considerably higher than Mortgages (0.32%), which is the safest product both in terms of probability of default (expressed in absolute terms) and in terms of WOE (which records a positive value and close of 3).

The WOE development by type of product complies with the economic significance because the mortgage products have a collateral allocated, thus making the product safer in terms of credit risk.

Among the products without collateral, the Overdraft is the safest because in most cases it is granted based on the salary transfer to the bank.

Chart 1

Individual analysis of Product variable



Source of data: Chart realized by author

Table 1. WOE Product shows how the groups are divided within WOE variable, the computations of the total and differentiated distribution (goods vs bads) and the computation of the indicator Information value for the Product variable.

WOE Product

Table 1

Product	Performing numbers	Default numbers	Group	WOE
PRIVATE MORTGAGE	3,949	13	Mortgages	2.86
AMERICAN MORTGAGE	47	-	Mortgages	2.86
HOUSING MORTGAGE	16	-	Mortgages	2.86
OVERDRAFT	1,455	63	Overdraft	0.23
CASH	6,473	447	Cash	(0.27)
CARDS	3,164	289	Cards	(0.57)
Total	**15,104**	**812**		

Product	Distribution	Default rate	Distribution of goods	Weight of Evidence	Distribution of bads	Information value
Mortgages	27%	0.32%	28%	2.86	2%	75%
Overdraft	10%	4.33%	10%	0.23	8%	0%
Cash	43%	6.91%	42%	(0.27)	55%	3%
Cards	21%	9.13%	20%	(0.57)	36%	9%
	100%	**5.38%**	**100%**		**100%**	**76%**

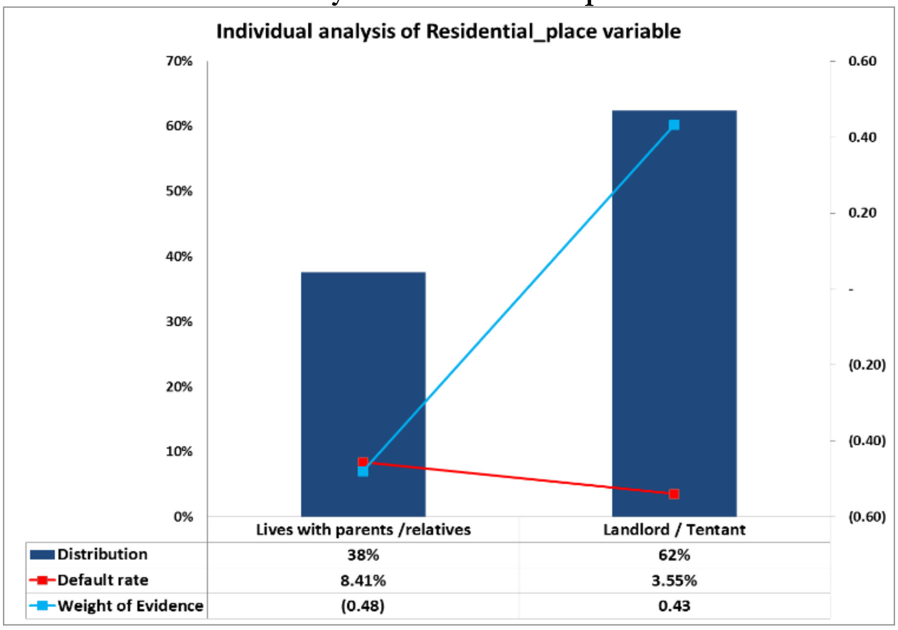
Source of data: Table realized by author

Residential Place variable

After the analysis of this variable it can be stated that the highest percentage of customers belongs to "Landlord/Tenant" attribute. This attribute recorded a default rate of 3.55%, lower than the 8.41% default rate associated with the "Lives with parents/relatives" attribute.

Chart 2

Individual analysis of Residential place variable



Source of data: Chart realized by author

Therefore, the "Landlord / Tenant" attribute has a positive value of WOE, which means this group is safer comparing with the "Lives with parents / relatives" attribute.

The table contains the groups within WOE variable, and the computation of the indicator Information value for the Residential Place variable.

WOE Residential_place

Table 2

Residential_place	Performing numbers	Default numbers	Group	WOE
Locuieste cu parintii/rudele	5,674	477	Lives with parents /relatives	(0.48)
Proprietar fara ipoteca	8,221	315	Landlord / Tentant	0.43
Proprietar cu ipoteca	925	1	Landlord / Tentant	0.43
Altele	251	19	Landlord / Tentant	0.43
Inchiriata de la stat	17	-	Landlord / Tentant	0.43
Inchiriata de la particular	11	-	Landlord / Tentant	0.43
Proprietate cooperativa	4	-	Landlord / Tentant	0.43
Total	**15,104**	**812**		

Residential_place	Distribution	Default rate	Distribution of goods	Weight of Evidence	Distribution of bads	Information value
Lives with parents /relatives	38%	8.41%	36%	(0.48)	58.74%	10.73%
Landlord / Tentant	62%	3.55%	64%	0.43	41.26%	9.69%
	100%	**5.38%**	**100%**		**100%**	**20.43%**

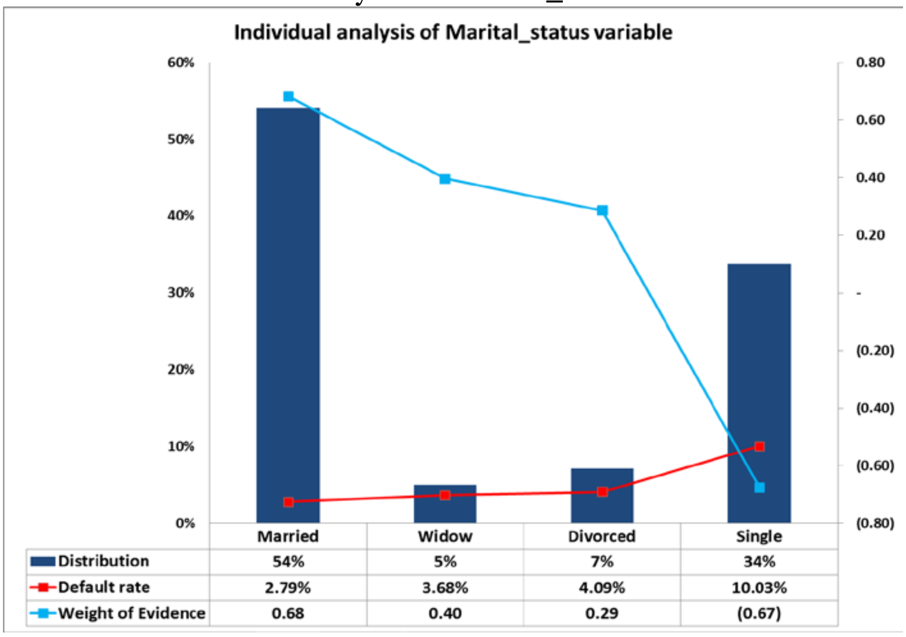
Source of data: Table realized by author

Marital Status variable

Individual analysis of the marital status variable reveals that the safest group, both in terms of the probability of default and in terms of WOE is the group "Married ". Registering the lowest probability of default and the highest value of WOE, the group "Married" is the most performing in the entire variable marital status. In contrast, the riskiest group is "Single" which has the highest probability of default by 10.03% in absolute terms and the lowest value of WOE with -0.67.

Chart 3

Individual analysis of Marital_status variable



Source of data: Chart realized by author

According to Table 3. WOE Marital_status it can be seen that was kept the order form the development sample. Moreover, the Information Value indicator record the highest value if the extreme group of the variable, as for "Married" group with 18.72% and "Single" group with 20.86%.

WOE Marital_status

Table 3

Marital_status	Performing numbers	Default numbers	Group	WOE
Single	5,103	512	Single	(0.67)
Married	8,165	228	Married	0.68
Divorced	1,075	44	Divorced	0.29
Widow	761	28	Widow	0.40
Total	**15,104**	**812**		

Marital_status	Distribution	Default rate	Distribution goods	Weight of Evidence	Distribution of bads	Information value
Married	54%	2.79%	56%	0.68	28.08%	18.72%
Widow	5%	3.68%	5%	0.40	3.45%	0.67%
Divorced	7%	4.09%	7%	0.29	5.42%	0.51%
Single	34%	10.03%	32%	(0.67)	63.05%	20.86%
	100%	**5.38%**	**100%**		**100%**	**19.39%**

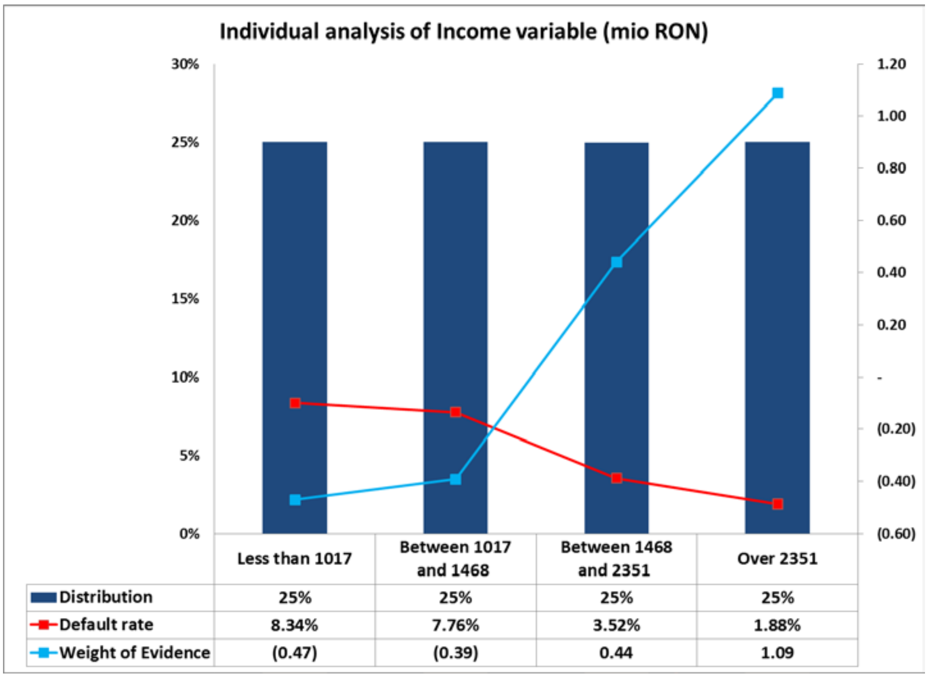
Source of data: Table realized by author

Income variable

Individual analysis of Income variable (mio RON) from Chart 4 presents the customers distribution by equal-weighted intervals (the income segments were calculated as that customers distribution remain uniform). The income intervals were identified according to income quantiles, respectively, 25%, 50% and 75% of sample development.

Chart 4

Individual analysis of Income variable (mio RON)



Source of data: Chart realized by author

The Table 4. WOE Income (mio RON) contain information related to number of performing clients from each group, their distribution and the Information value of Income variable.

WOE Income (mio RON)

Table 4

Income (mio RON)	Performing numbers	Default numbers	Group	WOE
1017	3,776	315	Less than 1017	(0.47)
1468	3,778	293	Between 1017 and 1468	(0.39)
2351	3,774	133	Between 1468 and 2351	0.44
40622	3,776	71	Over 2351	1.09
	15,104	**812**		

Income (mio RON)	Distribution	Default rate	Distribution goods	Weight of Evidence	Distribution of bads	Information value
Less than 1017	25%	8.34%	24%	(0.47)	38.79%	40.48%
Between 1017 and 1468	25%	7.76%	24%	(0.39)	36.08%	4.59%
Between 1468 and 2351	25%	3.52%	25%	0.44	16.38%	4.02%
Over 2351	25%	1.88%	26%	1.09	8.74%	18.67%
	100%	**5%**	**100%**		**100%**	**67.76%**

Source of data: Table realized by author

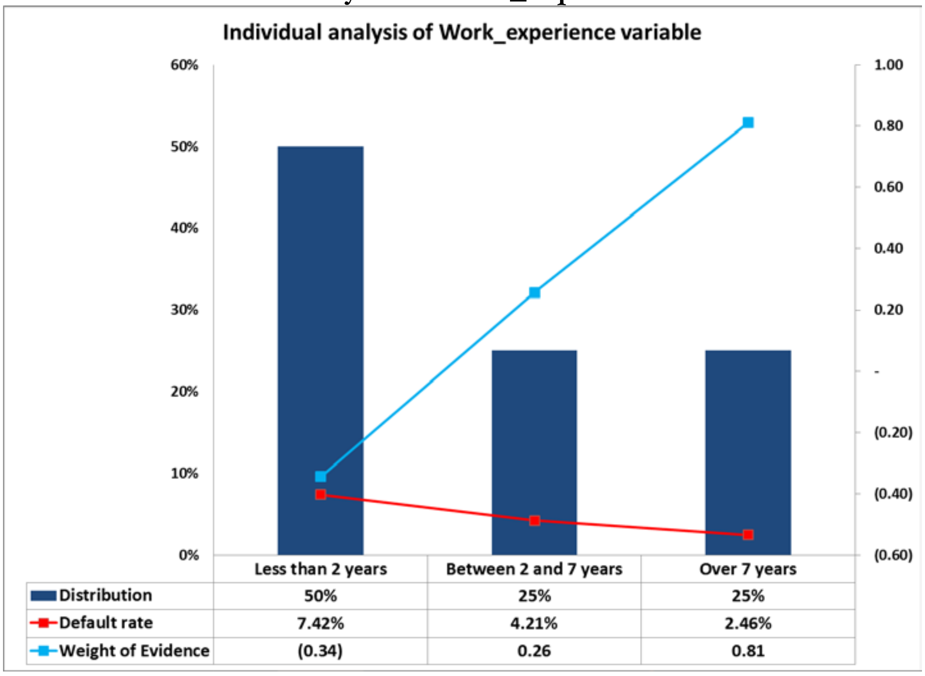
The WOE indicator has economic meaning as follows: the riskiest group (i.e. with a negative -0.47 WOE) is the group of customers with the lowest incomes, i.e. below 1017 RON. Also, the groups arranged in ascending order of income have a lower risk as measured by the increasing trend of WOE.

Work experience variable

After analyzing the Work experience variable we observe that most default customers belong to the group "Less than 2 years" professional experience.

Chart 5

Individual analysis of Work_experience variable



Source of data: Chart realized by author

Table 5. WOE Work_experience indicates the method of calculation of the WOE indicator and the value Information indicator. Both the percentage associated with default rate and value of WOE have economic significance because the upward trend of WOE justify the risk associated with each group: the work experience is great, the credit risk is low. In the case of the group "Over seven years" value of 0.81 WOE express its low risk compared to a negative value of WOE -034 related to customers from group "Less than 2 years".

WOE Work_experience

Table 5

Work_experience	Performing numbers	Default numbers	Group	WOE
2	7,552	560	Less than 2 years	(0.34)
7	3,776	159	Between 2 and 7 ye	0.26
44	3,776	93	Over 7 years	0.81
	15,104	812		

Work_experience	Distribution	Default rate	Distribution goods	Weight of Evidence	Distribution of bads	Information value
Less than 2 years	50%	7.42%	49%	(0.34)	68.97%	6.88%
Between 2 and 7 years	25%	4.21%	25%	0.26	19.58%	1.47%
Over 7 years	25%	2.46%	26%	0.81	11.45%	11.61%
	100%	5%	100%		100%	19.96%

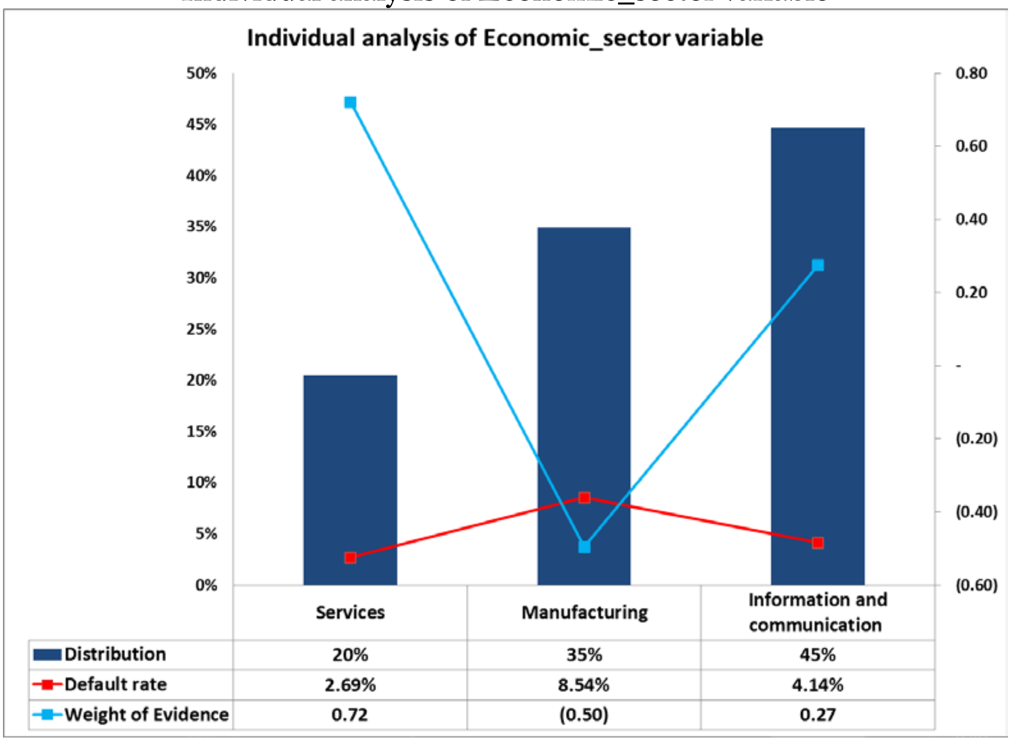
Source of data: Table realized by author

Economic sector variable

Analysis of Economic sector variable shows that the group "Services" has the lowest rate of default respectively 2.69%, while the group "Manufacturing" is at the opposite end, recording a 8.45 % percentage.

Chart 6

Individual analysis of Economic_sector variable



Source of data: Chart realized by author

Economic sector variable has economic significance because hectic development of intangible economy is reflected in the degree of stability of the activities associated with services. From this perspective, employees from service sector hold a favorable position in the labor market, with a stable income and financial security. It is therefore natural that these employees have the lowest credit risk. From a statistical view, the group "Services" is aligned with economic reality, this group having a positive indicator WOE of 0.72, being the highest value of the indicator variable of all groups associated economic sectors.

The table of Economic Sector variable provides information about each attribute of the variable. Also, distributions can be seen the performing customers and default customers distributions, which are treated separately.

WOE Economic sector

Table 6

Economic_sector	Performing numbers	Default numbers	Group	WOE
Professional, scientific and technical activities	413	3	Services	0.72
Financial and insurance activities	269	2	Services	0.72
Education	409	6	Services	0.72
Public administration and defence	1,232	40	Services	0.72
Human health and social work activities	682	25	Services	0.72
Real estate activities	83	7	Services	0.72
Mining and quarrying	194	1	Manufacturing	(0.50)
Electricity and gas	149	2	Manufacturing	(0.50)
Water supply	323	23	Manufacturing	(0.50)
Construction	382	28	Manufacturing	(0.50)
Altele	877	68	Manufacturing	(0.50)
Accommodation and food service activities	172	15	Manufacturing	(0.50)
Manufacturing	2,974	292	Manufacturing	(0.50)
Agriculture, hunting and forestry	201	21	Manufacturing	(0.50)
Information and communication	550	6	Information and communication	0.27
Transportation and storage	719	31	Information and communication	0.27
Wholesale and retail trade	1,512	95	Information and communication	0.27
Missing	3,963	147	Information and communication	0.27
Total	**15,104**	**812**		

Economic_sector	Distribution	Default rate	Distribution goods	Weight of Evidence	Distribution of bads	Information value
Services	20%	2.69%	21%	0.72	10.22%	7.79%
Manufacturing	35%	8.54%	34%	(0.50)	55.42%	10.76%
Information and communication	45%	4.14%	45%	0.27	34.36%	2.99%
	100%	**5.38%**	**100.00%**		**100%**	**21.54%**

Source of data: Table realized by author

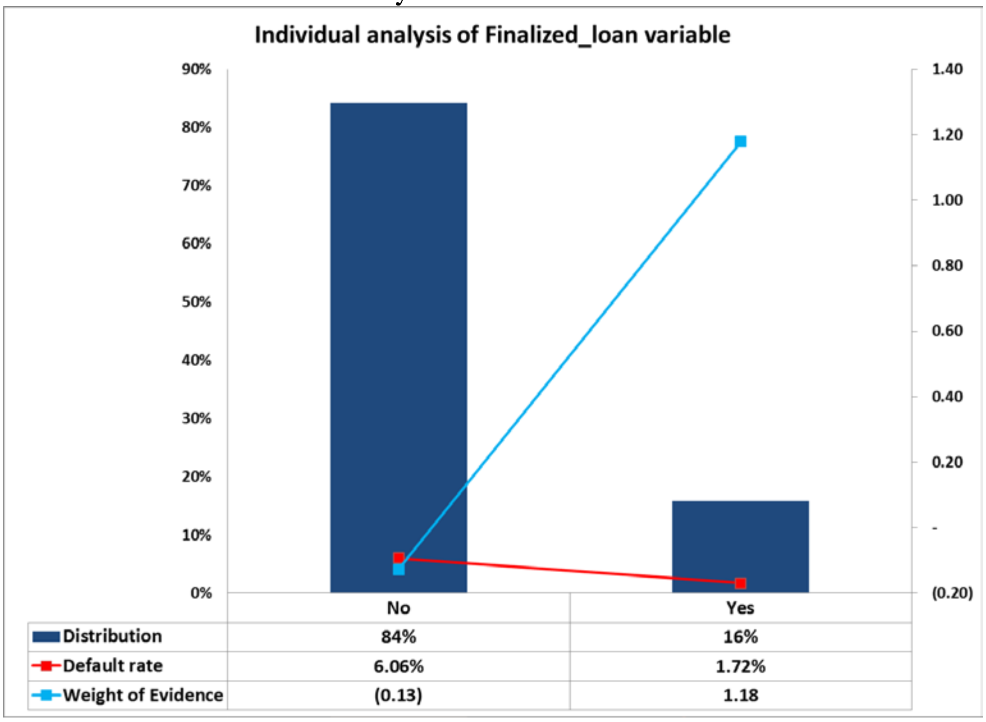
Finalized loan variable

After analyzing the finalized loan variable we observe that the development provides qualitative information about the creditworthiness of customers. Concerning the finalized loan variable, 84% of customers from the development sample had no previous loans which were completed on the application date.

Group of clients with loans repaid contains customers with banking experience. For this reason, economic expectations are that these most experienced customers to be more confident in terms of credit risk. In statistical terms, the trend variable is similar with economic expectations, as such the group of customers with completed loans having a low risk, measured by a positive value of WOE 1.18.

Chart 7

Individual analysis of Finalized loan variable



Source of data: Chart realized by author

Table 7. WOE Finalized_loan provides information about the distribution of these customers, divided into distribution of customers who have repaid loans, respectively 16%, and distribution customers who have not repaid loans respectively 94%.

WOE Finalized_loan

Table 7

Finalized_loan	Performing numbers	Default numbers	Group	WOE
0	12,718	771	No	(0.13)
1	2,386	41	Yes	1.18
Total	**15,104**	**812**		

Finalized_loan	Distribution	Default rate	Distribution goods	Weight of Evidence	Distribution of bads	Information value
No	84%	6.06%	84%	(0.13)	94.95%	1.45%
Yes	16%	1.72%	16%	1.18	5.05%	13.39%
	100%	**5.38%**	**100%**		**100%**	**14.83%**

Source of data: Chart realized by author

II. Selection criteria for Scorecard variables

The selection criteria for the variables that were included in the final Scorecard are related to the percentage of relevant information that each variable contains, as well as the indicator WOE. A positive indicator WOE indicates increased stability within that group, while a negative value indicates an increased risk associated with those customers. The most important variables show a considerable difference between the extremes of WOE groups, indicating high discrimination power of the performing customers to enter into default.

In order to calculate the dispersion of the WOE within groups, we use an indicator known in the banking industry as the Information value (IV), which is calculated as follows:

Information value = SUM (Distribution of goods − Distribution of bads) * (− LN (Distribution of bads / Distribution of goods)) (6)

or

Information value = SUM (Distribution of goods − Distribution of bads) * WOE (7)

Within Table 8. we observe the information value percentages computed for each variable from development sample:

Information value

Table 8

Attribute	Description	Information value
ID	Unique ID	
REQUEST_DATE	Date of loan request initiation	
PRODUCT	Product type	75.92%
BIRTH_DATE	Birth Date of solicitor	13.93%
DISTRICT	Solicitor's district of residence	0.20%
AREA	Area of residence (county capital, urban or rural area)	4.17%
RESIDENTIAL_PLACE	Living Status (owner, tenant, cohabitant, etc.)	20.43%
EDUCATION	Highest Education Level of solicitor	9.82%
MARITAL_STATUS	Marital Status of solicitor	19.39%
HOUSEHOLD_MEMBERS	Number of household members	26.04%
NO_OF_DEPENDENTS	Number of dependends	0.00%
INCOME	Monthly Accepted Income (RON)	67.76%
EMPLOYMENT_DATE	Date of Employment at current work place	19.96%
BUSINESS_SINCE	Date of business start of Current Employer	5.52%
LEGAL_FORM	Legal form of Current Employer	4.57%
ECONOMIC_SECTOR	Industry of Current Employer (aggregated)	21.54%
EMPLOYEE_NO	Employee number of Solicitor's Current Employer	9.85%
CLIENT_SINCE	Date of first bank relation	3.61%
DEBIT_CARD	Indicator if solicitor has debit cards in BCR at loan request time	1.72%
CURRENT_ACCOUNT	Indicator if solicitor has current accounts in BCR at loan request time	1.69%
SAVING_ACCOUNT	Indicator if solicitor has saving accounts in BCR at loan request time	0.00%
SALARY_ACCOUNT	Indicator if solicitor has salary accounts in BCR at loan request time	1.90%
FOREIGN_ACCOUNT	Indicator if solicitor has accounts in foreign currency in BCR at loan request time	0.09%
FINALIZED_LOAN	Indicator if solicitor has repayed loans in BCR at loan request time	14.83%
DEPOSIT	Indicator if solicitor has deposits in BCR at loan request time	0.48%
DEFAULT_FLAG	= 1, if client reaches 90 days past due in 1 a one year time window from loan origination = 0, otherwise	

Source of data: Table realized by author

In the Scorecard were considered only variables that exceed the 15% threshold for Information indicator value. It is considered that the variables having a percentage over 15% in Information Value incorporate relevant information from economically view and worth to be included in the model.

III. R output and final Scorecard

After the computation of Information value and a preselection data, the information related to WOE for each variable in the development sample were introduced in R to be subjected to statistical processing.

After analyzing the results values, observed p-value greater than 5% were observed for the coefficients of the variables WOE_AGE and WOE_HOUSEHOLD_MEMBERS variables. Therefore coefficients of these variables are not significantly different from 0 and a new model is estimated by eliminating these two variables.

Final results related to the output from the estimation in the R code were summarized in Table 9. It is noted that all the regression coefficients are significantly different from 0.

Output R

Table 9

Coefficients:

	Estimate	Std. Error		z value	Pr(>|z|)
(Intercept)	-2.85287	0.04575	-62.363	2.00E-16	***
WOE_PRODUCT	-0.99479	0.07434	-13.382	2.00E-16	***
WOE_RESIDENTIAL_PLACE	-0.97032	0.09238	-10.503	2.00E-16	***
WOE_MARITAL_STATUS	-0.66022	0.06541	-10.094	2.00E-16	***
WOE_INCOME	-0.36282	0.07645	-4.746	2.08E-06	***
WOE_EMPLOYMENT_DATE	-0.61182	0.09867	-6.2	5.63E-10	***
WOE_ECONOMIC_SECTOR	-0.63688	0.0885	-7.196	6.20E-13	***
WOE_FINALIZED_LOAN	-0.66988	0.12712	-5.27	1.37E-07	***

Source of data: Table realized by author

For a more intuitive understanding, the final value of PD from the regression is converted into score points by the following formula:

Points = Scalar – Factor * ln (PD/(1-PD)) (8)

Points = 217 – 72 * ln (PD/(1-PD))

The formula allocates points from 0-1000 and is used in practice by credit institutions to measure credit risk. The final score of a customer can be calculated based on the marginal contribution of each variable as follows:

Points = 217 – 72 * ($\beta_0 + \beta_1$ WOE$_1$ + ... + β_7 WOE$_7$) = $\sum_{i=1}^{7}(217/7 - 72^* \beta_0/7 - 72^* \beta_i \, WOE_i)$

Final Scorecard and allocation of points for each attribute

Table 10

Variable	Group	Score
Product	Mortgages	265
	Overdraft	77
	Cash	41
	Cards	19
Residential_place	Lives with parents /relatives	27
	Landlord / Tentant	91
Marital_status	Married	93
	Widow	79
	Divorced	74
	Single	28
Income (mio RON)	Below 1017	48
	Between 1017 and 1468	50
	Between 1468 and 2351	72
	Over 2351	89
Work_experience	Below 2 years	45
	Between 2 and 7 years	72
	Over 7 years	96
Economic_sector	Services	93
	Manufacturing	38
	Information and communicat	73
Finalized_loan	No	54
	Yes	117

Source of data: Table realized by author

For each of the seven variables we have a WOE allocated to each group, the final contribution of each variable, depending on the individual attributes being shown in Table 10. For example, a client obtains a total score of 844 points as follows:

1. 265 points for applying for a Mortgage

2. 91 points because it is Landlord / Tenant

3. 93 points for married

4. 89 points because it has come across 2,351 RON

5. 96 points for having over 7 years of experience his work

6. 93 points for working in a company from the service sector

7. 117 points because he previously had a loan which it paid.

This client's probability of default corresponding to the 844 points equals to:

PD = 1/(1+exp(-Scor+217)/72) = 0.02%

IV. Analysis of discriminatory power of the model based on development and validation samples

The role of credit score is to order customers based on individual credit risk. After estimating the model it is necessary to test its accuracy when put into practice. Therefore, the goal of this test is to see if the scores assigned to variables discriminate between performing and defaulted customers.

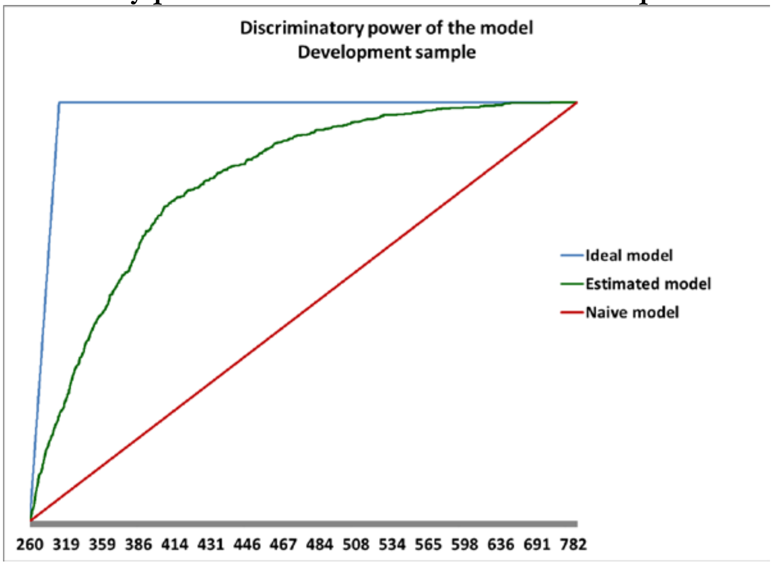
We represent graphically the discriminatory power of the model at the level of scores. Chart 8 presents the discriminatory power of the model. It can be stated that the estimated model tends towards the ideal model, which entails a high discriminatory power of the model.

In order to test the discriminatory power of a model on a given sample, the customers are scored with the model and then ordered by their score. An ideal model concentrates all defaulted customers in the riskiest scores. A naïve model distributes the defaults randomly across the scores. A model with high discriminatory power concentrates most of the defaulted customers in the scores that have the most increased risk. Both in practice and in theory, an estimated model should be as close as possible to the ideal model.

The model developed in the present article has a high discriminatory power, based on the development sample (see Chart 8.). This can be attributed to an overfitting of the model's parameters, because the sample on which we present the results in Chart 8. is the same as the sample on which the model has been developed. That is why we also test the discriminatory power of the model on a disjoint sample of clients, namely the validation sample. The results presented in Chart 9. indicate a high discriminatory power on the validation sample, underlying the robust nature of the model.

Chart 8

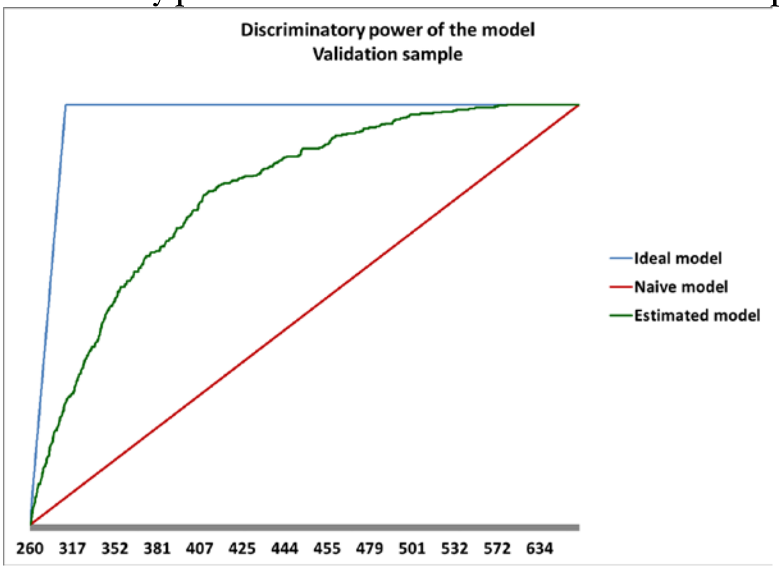
Discriminatory power of the model based on development sample



Source of data: Chart realized by author

Chart 9

Discriminatory power of the model based on validation sample



Source of data: Chart realized by author

4.3 Calibration of the application cut-off scores for the risk appetite of a bank

In this section we present the implementation of the Scorecard in the process of granting new loans. Credit decisions are a crucial instrument in risk management of the bank. The importance of correlate these decisions with analyzes and tools of risk management is essential to ensure a favorable climate for the development of the real economy. Also, the use of instruments of credit risk management in an integrated manner into management decisions of a bank is a requirement covered by Regulation 575/2013 of the European Union and Regulation 5/2013 of the National Bank of Romania. The analysis presented in this section is aligned with regulatory requirements and provides a potential tool for risk management applied within the banking institutions in Romania.

For a realistic approach we use financial information published in the public disclosure report, presented in Section 1, in order to estimate a maximum of the probability of default accepted during loan process. Providing a maximum threshold for clients admitted in portfolio leads to the stabilization of the portfolio risk and appropriateness to the bank's risk strategy. If this threshold is specified in the risk strategy of the bank, it can be referred as such in the selection cut-off algorithm, to calculate the cut-off score. For purposes of actual analysis and according to data publicly available, the maximum threshold must be approximated based on the provisions and total exposure in the Retail portfolio (where this information is available) or to the whole portfolio of the bank.

The provisions on the performing portfolio are representative for estimating the expected loss on portfolio level, therefore they are estimated as a product of the probability of default, loss given default and exposure given default. Loss given default, in a portfolio of retail, can be approximated to 45% of the exposure given default (value of 45% represents a standard regulatory threshold for loss given default). Thus, from the data on exposures and transparency provisions, published within the reports, we can get an approximation of the probability of default specific to the current portfolio of the bank. This probability is a threshold indicative of the bank's risk appetite and risk strategy and will be used as the maximum threshold for the PD accepted on loan process.

Consider the Scorecard, statistically estimated based on historical data in Section 2.2.3. The Scorecard is based on data from the banking market in Romania and can be considered as being representative of the situation of the Romanian banking market.

The results of the individual analysis are presented in Table 11. We notice that the bank with highest PD at the level of current portfolio is obtained for Banca Transilvania, with a value of 18%. Therefore, we can assume that Banca Transilvania will credit only those customers who obtain PD values lower than 18%, based on the Scorecard. Therefore all customers with a PD over 18% will be rejected for loans, i.e. all customers with a score of less than 326. Based on this information we observe that Banca Transilvania rejects a number of 10% of applications from validation sample.

Cut-off calibration

Table 11

Bank	Provision Individuals	Exposure Individuals	Expected PD Individuals	Score Cut-off	Rejection rate
BCR	6,830	67,953	15%	343	12%
Raiffeisen Bank	572	12,032	7%	403	30%
Unicredit Ţiriac Bank	190	4,879	6%	418	36%
Banca Transilvania	2,669	21,932	18%	326	10%

Source of table: Table realized by author

The highest degree of rejection is recorded by UniCredit Tiriac Bank, which does not accept customers with a score of less than 418. This rule provides a selection of customers so that the PD does not exceed the threshold of 6% recorded in the bank's current portfolio.

5. Conclusions

The models developed for credit risk management illustrates the manner in which modern techniques used in credit risk management are used to optimize capital in banks by an appropriate selection of customers applying for loans.

Technical concepts presented in this article are a short introduction in the financial modeling using quantitative methods. The results are to provide an overview of the necessary theoretical steps necessary in the development of a rating model.

Development of mathematical models for risk management has been possible due to the successful implementation of elements of probability theory in the financial environment. These tools allow to extract relevant information from existing data of the banks and therefore provides the Management Board with important resources in the decision making process.

In our example we have used the logistic regression, which is the mathematical model used most frequently in practice. The model has become a standard and it is highly regulated in the international regulations, like the Basel Internal Rating Based approach.

The second chapter includes part of the work, which was structured on two levels: o research in qualitative legislation international banking issued by the Basel Committee and a practical part where a model rating was developed for use in credit risk management. Rating model was developed based on credit application data related to a portfolio of clients in the banking market in Romania.

In the qualitative analysis the banking legislation was interpreted by reference to Regulation 575/2013, which sets the benchmark in international banking legislation. Regulation 575/2013 mentions clear transparency criteria that banks must comply, evidenced by frequent publication of information relating with annual credit risk. In this regard, we analyzed four reports published by banks representative for the Romanian banking system. This information is centralized in the case study in terms of credit risk, in a comparative analysis. The information in Sections 2.1 and 2.2 are combined in an instrument that is used in the underwriting decisions. The Scorecard asses the risk of each individual customer and can be used to bound the risk of the new production by a cut-off score. The cut-off scores are calibrated based on the risk appetite specific to each of the four banks analyzed.

The limits of the current research are represented by a limited access to confidential information. Specifically, the scorecard was estimated based on the public information, which means that a different approach to credit risk management can have a chance of success.

Possible future research directions can focus on the implementation of scorecards in other processes within banks, such as the following: estimation of provisions, estimation of economic capital, or risk adjusted interest rates.

References

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring. Journal of the Operational Research Society, 54(6), 627-635.

Bellotti, T., & Crook, J. (2009). Credit Scoring with Macroeconomic Variables Using Survival Analysis. Journal of the Operational Research Society, 60(12), 1699-1707.

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. Journal Of Political Economy, 81(3), 637-654.

Chen, T., Liao, H., & Lu, C. (2009). Bank Credit Risk and Structural Credit models: Agency and Information Asymmetry Perspectives. Journal Of Banking And Finance, 33(8), 1520-1530.

Credit Suisse Financial Products (1997). CreditRisk+: A Credit Risk Management Framework.

Crouhy, M., Galai, D., & Mark, R. (2000). A comparative analysis of current credit risk models. Journal Of Banking & Finance, 24(1-2), 59-117.

Derbali, A., & Hallara, S. (2012). The Current Models of Credit Portfolio Management: A Comparative Theoretical Analysis. Int. J. Manag. Bus. Res., 2(4), 271-292.

Elizalde, A. (2006). Credit risk models II: structural models. CEMFI Working Paper, 0606, 1-27.

Hand, D. J., & Henley, W. E (1997). Statistical Classification Methods in Consumer Credit Scoring: A Review. Journal of the Royal Statistical Society: Series A (Statistics in Society), 160(3), 523-541.

Hao, C., Alam, M., & Carling, K. (2010). Review of the literature on credit risk modeling: development of the past 10 years. Banks And Bank Systems, 5(3), 43-56.

Jacod, J., & Protter, P. (2004). Probability Essentials. Berlin, Heidelberg: Springer Berlin Heidelberg.

Merton, R. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. The Journal Of Finance, 29(2), 449-470.

Moon, T. H., & Sohn, S. Y. (2010). Technology Credit Scoring Model Considering Both SME Characteristics and Economic Conditions: The Korean Case. Journal of the Operational Research Society, 61(4), 666-675.

Pampel, F. (2000). Logistic regression. Thousand Oaks. Calif.: Sage Publications.

Papoulis, A., & Pillai, U. (1984). Probability, Random Variables, and Stochastic Processes (2nd ed.). New York: McGraw-Hill.

Rajaratnam, K., Beling, P., & Overstreet, G. (2010), Scoring Decisions in the Context of Economic Uncertainty. Journal of the Operational Research Society, 61(3), 421-429.

Rosenberg, E., & Gleit, A. (1994). Quantitative Methods in Credit Management: A Survey. Operations Research, 42(4), 589-613.

Ruxanda, G. (2009). Analiza multidimensională a datelor. Suport de curs. Academia de Studii Economice.

Thomas, L. C. (2000). A Survey of Credit and Behavioural Scoring: Forecasting Financial Risk of Lending to Consumers. International Journal of Forecasting, 16(2), 149-172.

Vasicek, O. (1987). Probability of Loss on Loan Portfolio. KMV Corporation.

Regulamentul nr. 575/2013 al Parlamentului European şi al Consiliului Uniunii Europene din 26 iunie 2013 privind cerinţele prudenţiale pentru instituţiile de credit şi societăţile de investiţii şi de modificare a Regulamentului (UE) nr. 648/2012

Banca Transilvania (2014). Raport privind cerinţele de transparenţă şi publicare a informaţiilor, online at [https://www.bancatransilvania.ro/files/ guvernanta-corporativa-bt/cerinte_de_publicare_var_finala.pdf] accessed on 18.03.2016

BCR (2014). Raport privind cerinţele de transparenţă şi publicare a informaţiilor, online at [https://www.bcr.ro/content/dam/ro/bcr/www_bcr_ ro/Investitori/Transparenta-si-publicare/Raport transparenta_si_publicare_2014.pdf] accessed on 18.03.2016

Raiffeisen Bank S.A. (2014). Raport privind cerinţele de transparenţă şi publicare a informaţiilor, online at [https://www.raiffeisen.ro/wps/wcm/connect /8f245860-dce6-48b6-b926-c8eaddf7 b83f /31.12.2014+Raport+privind +cerintele+de+transparenta+si+de+publicare.pd f?M OD =AJPERES &CAC HEID=8f245860-dce6-48b6-b926-c8eaddf7b83f] accessed on 18.03.2016

UniCredit Ţiriac Bank (2014). Raport privind cerinţele de transparenţă şi publicare a informaţiilor, online at [https://www.unicredit.ro/content/dam/cee 2020-pws-ro/ Documente PDF /Institutional-RapoarteBasel/Raport %20de%20Publicare%20a%20informatiilor%202014. pdf] accessed on 18.03.2016

Hostiuc, C., 2015. BCR se pregăteşte să scoată la vânzare în bloc credite neperformante de 1-1,5 mld. euro. Tranzacţia anului pe piaţa bancară poartă numele de cod „Neptun“. Ziarul Financiar, online at [http://www.zf.ro/zf-24/bcr-se-pregateste-sa-scoata-la-vanzare-in-bloc-credite-neperformante-de-1-1-5-mld-euro-tranzactia-anului-pe-piata-bancara-poarta-numele-de-cod-neptun-14104446] accessed on 17.04.2016