Developing a Rating Model on a Statistical Basis

▪

Vasile Dedu
Academy of Economic Studies, Bucharest
Tudor Alexandru Ganea
Titu Maiorescu University, Bucharest

Abstract. *We consider that, starting from 2007, in order to deal with the competition, the banks from Romania will have to be prepared to take and effectively manage higher risks, both on their own behalf, and on the behalf of their clients, since the transition to the calculation methodology set up by the new Capital Accord (Basel II) is bound to determine the artificial decrease of the solvency indicator.*

The very conception of this article has been triggered by two significant phenomenons. First, the banks from Romania have become increasingly interested in developing and enhancing methods and procedures of risk assessment. Second, the Basel Committee on Banking Supervision, followed by the European Commission, has imposed a series of standards referring to the estimation of some crucial indicators on a banking level, under the title of „Basel II": PD (Probability of default), LGD (Loss given default) și EAD (Exposure at default).

In this respect, in 2006, the Romanian government enacted the Decree no. 99 (sanctioned and modified by the Law no. 227/04.07.2007), together with a series of regulations. The decree contains new banking regulatory provisions applicable to credit companies starting with the 1st of January, 2007, the date of Romania's adherence to the European Union.

Key words: credit risk; internal rating models; stress-testing.

▪

JEL Codes: D14, D53, F36, G21, G32.
REL Codes: 10B, 11C.

In conformity with the new regulations in force since the 1st of January, 2007, the rating systems are defined as „the complex of models, processes, control systems, data collection systems and IT systems which allow credit risk assessment, exposure assignment to different rating classes or risk groups, as well as estimation quantifying concerning the non-reimbursement situations and losses for a specific type of exposure”[(1)].

Throughout the years, the specialized works (Ganguin, Bilardello, 2005, p. 17) have focused on analyzing two components of credit risk:

- the risk of non-reimbursement, which is measured by assessing the capacity and willingness of the debtor to reimburse its debts within the established terms;
- the perspectives of retrieval, evaluating in terms of percentage the possibility to retrieve the money in case the debtor stops paying its dues.

Under the new regulations referring to credit institutions, a special focus is placed on the credit risk, as its assessment substantially influences the capital requirements that the respective institutions must meet.

Thus, when calculating the minimum capital requirements to cover the credit risk, the credit institutions may choose between the standardized approach or, with the explicit approval of the National Bank of Romania, the Internal Rating-Based Approach[(2)], in order to determine the risk weight of bank exposures.

In the standardized approach, the quality of credit is determined by means of assessments made by external credit assessment institutions or export credit agencies, recognized as eligible by the National Bank of Romania on the basis of specific criteria stipulated in the NBR’s regulations[(3)].

The methodology of determining the risk weight of bank exposures by using the internal rating-based approach, as well as the minimum conditions under which the use of this method can be approved are specified in the NBR’s regulations[(4)].

The present article is practically following the above-mentioned regulations, while putting forward a credit risk assessment mechanism using internal rating models.

The credit risk scoring, as many other ways of evaluating credit applications, represents a useful tool in determining the risk level for the credit institution’s clients. Its aim is to evaluate applications on a statistical basis, not an individual one, so as to be able to classify clients into solvents and insolvents.

The scoring stands for „an external diagnosis tool consisting in measuring and interpreting the risk to which the investor, the creditor and the company as a system in the future activity are exposed” (Anghel, 2002, p. 36).

In the past, the financial institutions would buy the scorecards from companies specialized in their production. These were provided with information from the financial institutions and used expensive mathematical models and technologies to produce the respective scorecards.

Gradually, as the IT technologies evolved, allowing the collection, storing and analysis of prodigious amounts of information with minimum costs, the financial institutions found it more profitable to produce on their own the models necessary in the classification of credit applications (Siddiqi, 2006, p. 2).

The first step that we propose in developing a credit scoring model would be setting up priorities and objectives. Among the organizational objectives envisaged by the authors of this thesis, we can include:

- Reducing losses, frauds and the risk of non-reimbursement;
- Increasing the rate of credit application approval and of market share, where the presence of a reduced risk creates an opportunity of expansion;
- Increasing the profit;
- Reducing expenses by implementing an automated credit application processing system.

In its simplest form, a scorecard consists of a group of features, statistically determined and aggregated in order to separate clients into the two above-mentioned categories: solvents and insolvents.

In the same respect, the presence of more and more substantial amounts of money on the financial market, inclusively on the Romanian one, as well as the necessity to place them as credits impose a more nuanced classification of debtors. Not all the clients considered as insolvents will fall into this category; also, not all the clients initially credited as solvent by a certain model will completely reimburse the contracted amounts of money.

An example of scorecard applied to individuals is presented in Table 1:

Table 1

1.1. OCCUPATION (CONTINUITY)	
Uninterrupted activity with the present employer for more than 5 years	4
Uninterrupted activity with the present employer for more than 3 years	3
Uninterrupted activity with the present employer for more than 2 years	2
Uninterrupted activity with the present employer for less than 2 years, but more than 3 years with the former employer	3
Uninterrupted activity for less than 2 years, both with the present, and with the former employer	0
1.2. MARITAL STATUS	
Married, widow/widower	1
Other	0
1.3. AGE (PROPENSION TO RISK)	
Under 25 years old	0
Between 26 and 35 years old	1
Between 36 and 57 years old	2
Between 57 and 70 years old	0
1.4. LIVING CONDITIONS	
Personal property (apartment/house)	2
Living with the parents/work residence	1
Rent	0
1.5. PROFESSION	
Higher education	5
Highschool education	4
Sales representative/Specialized worker	3
Retired	2
Worker	1
Seasonal employee (waiter, cook, sailor)	0
1.6. NUMBER OF PERSONS UNDER SUPPORT	
0 – 1 person	3
2 persons	2
3 or more persons	0

2.1. AVAILABLE NET INCOME	
VND < 3.33 times the credit instalment	0
VND = 3.33 – 3.99 times the credit instalment	2
VND > 4 times the credit instalment	3
2.2. DOWNPAYMENT	
Downpayment < 25%	0
Downpayment 25-29%	1
Downpayment 30-35%	2
Downpayment > 35%	4

The features used to analyze the individual can be selected from the information placed at the disposal of the credit institution. We can use demographic features (age, years of service or present residence), history of relations with the bank (the lenght of the relation, the range of used products, the debt service), properties, etc.

To each attribute (age is a feature, the 26-35 year interval is an attribute) we can assign a certain number of points after a statistical analysis and taking into account various factors, such as the discriminative force of the analyzed features and of the relations between them. The total score obtained by an applicant is the sum of the points assigned to the used features.

We must notice the fact that in Romania we are facing a major problem as to the collection and analysis of the above information. There are no databases storing positive facts concerning the life and activity of persons and companies. We are only interested in the delays in repaying our debts in due time, a highly unsatisfactory situation.

However, we must appreciate the preoccupation shown by the Credit Office in the direction of offering positive information on the banks'clients in the future, as well as of developing a scoring system which creates the risk profile of an individual debtor.

The next table is an example of a management report issued while developing a scorecard.

Table 2

Interval	Clients	Cumulated clients	Solvent clients	Cumulated Solvent clients	Insolvent clients	Cumulated Insolvent clients	Marginal badrate (%)	Cumulative badrate (%)	Approval rate (%)
24-26	120	120	119	119	1	1	0.833	0.83	1.40
22-24	140	260	139	258	1	2	0.714	0.77	3.03
20-22	230	490	229	487	1	3	0.43	0.61	5.70
18-20	400	890	396	883	4	7	1	0.79	10.36
16-18	450	1340	445	1328	5	12	1.11	0.90	15.60
14-16	*700*	*2040*	*690*	*2018*	*10*	*22*	*1.428*	*1.08*	*23.75*
12-14	1020	3060	1005	3023	15	37	1.47	1.21	35.62
10-12	1130	4190	1108	4131	22	59	1.946	1.41	48.78
8-10	1400	5590	1372	5503	28	87	2	1.56	65.08
6-8	1100	6690	1074	6577	26	113	2.363	1.69	77.88
4-6	940	7630	908	7485	32	145	3.404	1.90	88.82
2-4	640	8270	614	8099	26	171	4.062	2.07	96.27
0-2	320	8590	304	8403	16	187	5	2.18	100.00

- The row in bolded characters can be interpreted as it follows:
  - for a score between 14 and 16 points, the expected rate of non-reimbursement is 1.428% × ($\frac{10}{700}\times 100$), which means that 1.428% of the credit aplications with a score between 14 and 16 points are expected to be insolvent;
  - also, 1.08% × ($\frac{22}{2040}\times 100$) of all the applications scoring over 14 points are supposed to be insolvent;
  - the approval rate for a score of over 14 points shows that 23.75% × ($\frac{2040}{8590}\times 100$) of the credit applications are placed above the minimum threshold of 14 points.

Following the presented factors, a bank can decide to reject all the credit applications scoring a certain number of points, lower than 14, for example, or to give credits to such clients for a higher interest rate, in order to cover the assumed risks.

The risk information resulting from a scorecard, combined with other factors, such as the approval rate and the profit/loss potential corresponding to every risk category, can be used to develop new strategies to maximize revenues and minimize losses.

Based on this sort of analysis, the bank could decide that the applications scoring between 12 and 14 points receive a positive result, however at a higher interest rate.

Thus, in case of high risk clients, we propose the development of the following strategies:

- rejecting applications if the risk level is too high;
- approving a lower credit limit in case of a credit card or credit line;
- imposing a higher interest rate;
- introducing a higher evaluation for the sollicited risk insurance;
- placing the applicant on a surpervision list for potential fraudulent activity.

On the contrary, the clients scoring a superior number of points will enjoy preferrential lower rates, will receive bonus products (credit cards), will get higher credit lines, etc.

In this article, we have proposed ourselves to present an initiation, development and enhancement mechanism of an evaluation model of the clients of a commercial bank. Under the present development circumstances of the Romanian economy, the creation within every bank of specialized departments in designing and supervising the rating systems is becoming more and more necessary.

We consider that the process must include the following stages:

1. Identifying data and designing parameters;
2. Creating and using the database;
3. Developing the model;
4. Implementing the rating model;
5. Calibrating the rating model;
6. Validating the rating model.

Data identifying and parameter designing suppose an extended period of time and consistent human resources. This stage is meant to justify the necessity of producing a new and improved crediting model, as well as the necessity of identifying the most representative parameters for the segment to which the credit product is addressed.

In cases of exposure towards companies, institutions, central administration or banks, the credit institutions must collect and keep(5):

a) the complete history or ratings assigned to acknowledged debtors and guarantors;

b) the dates when the ratings have been assigned;

c) the methodology and main dates when the ratings have been assigned;

d) the identity of the person in charge with rating assignment;

e) the identity of debtors in the situation of non-reimbursement, as well as the exposures in the situation of non-reimbursement;

f) the date and circumstances of the occurence of situations of non-reimbursement and

g) information referring the probability of non-reimbursement and the non-reimbursement rate associated to each rating class; information referring to tha rating migration.

For the retail exposures, the credit institutions must collect and keep(6):

a) data used in the process of assigning exposures on rating classes and risk groups;

b) information referring to the estimation of non-reimbursement probabilities (PD), non-reimbursement-related losses (LGD) and conversion factors associated to rating classes or risk groups;

c) the identity of debtors in the situation of non-reimbursement, as well as the exposures in the situation of non-reimbursement;

d) in the case of exposures in the situation of non-reimbursement, the data referring the rating classes or risk groups to which the exposure was assigned prior to the occurence of the non-reimbursement sitation, as well as the actual value of the non-reimbursement – related losses and of the conversion factor and

e) information referring to the rate of losses for the eligible renewable retail exposures.

The NBR’s Regulation no. 15/20/2006 defines the situation of non-reimbursement as being the moment in which the following events take place, either simultaneously or consecutively(7):

a) the credit institution considers that, without taking measures to execute the guarantee, if it exists, it is improbable that the debtor should fully reimburse the credit financial obligations towards the credit institution;

b) the debtor has been late for over 90 days with the reimbursement of any significant obligation resulting from credits at the credit institution.

In certain situations, using multiple scorecards for a portfolio might produce a more adequate risk identification than using a single scorecard for all the clients.

This situation occurs when the analyzed population is composed of several sub-populations, presenting different features. The identification process of these sub-populations is known as „segmentation”.

We have emphasized two ways to determine the segmentation:

- based on expertise, the validation being realized through analytical methods;
- generating individual segments by means of statistical methods.

Whatever the method chosen, each selected segment must be large enough to correctly develop the desired model.

The Basel II Accord embraces a pragmatic approach, by defining segments as homogenous groups from the risk perspective.

We consider that a way to confirm the necessity of a segmentation is to take into account the non-reimbursement rates for different sub-populations. The method consists in analyzing these rates for different attributes of the selected featuress, thus being able to identify different segments based on these performances.

Table 3

Age	Rate of non-reimbursement (%)
Attributes	
Over 40 years old	2.1
Between 30 and 40 years old	4.3
Under 30 years old	8.8

By following the parameters defined in the prior stage, we can initiate the creation of the database needed to develop the scorecard. This database contains a set of features (predictive elements) and a target variable for each case, all this information being subsequently used to properly develop the model.

This stage is mentioned in the NBR's regulations accompanying the Emergency Government Decree no 99/2006, therefore the credit institutions must be able to provide a detailed history of losses, in function of the factors considered to be decisive in the evolution od the risk parameters[(8)].

Once the database necessary to develop the model is built, we must draw up a list of features and a target variable. The model will have as main objective establishing and quantifying the relation between the features and the client performance.

We have based ourselves on a series of features chosen according to their discriminative force, each having several expression attributes, as well as on their logistic regression. We have take into account two major aspects: the necessity of a strong statistical basis and a realistic economic approach.

The feature selection reduces the complexity of the multivaried analysis of the totality of these elements.

The first step refers to the development of a comprehensive catalogue of indicators, based on the quantitative information collected during the prior stages. This catalogue must include economic and financial indicators by means of which we must be able to evaluate the debtor's situation in terms of assets, finances and revenues.

In this way, a large list of indicators is formed, some of them similar, fact which allows us to select the ones relevant to the next stages.

The analysis of the relations between characteristics must be performed before using the regression. By studying the existent connections, we can eliminate some of the features, thus avoiding using some carrying the same information, which can affect the model relevancy.

Once forming the list of indicators, eliminating those which do not observe the work hypothesis, verifying the relations between the remaining ones, reducing once again the ones mutually influencing, we calculate the IV (Information Value) for each feature.

The initial analysis of each feature involves two main aspects:

First of all, we must evaluate the power of discrimination of each feature as a measure of performance.

The strongest features are then grouped, so as to form in the end a group of information, preferably independent, which can be used in the regression stage.

In order to establish the predictive power of each attribute, we have chosen using the WOE indicator (weight of evidence), while selecting the IV (Information Value) indicator for determining the power of discrimination (Siddiqi, 2006, p. 81).

Table 4

Value of solvency (%)	No. of companies	Company distribution (%)	Solvent clients	Solvent client distribution (SCD) (%)	Insolvent clients	Insolvent client distribution (ICD) (%)	Bad rate (%)	WOE	IV
>100	1500	6.98	1480	7.66	20	0.92	1.33	2.117134	0.142585
85.01 – 100	4500	20.93	4350	22.50	150	6.91	3.33	1.180365	0.184036
79.01 – 85	8000	37.21	7100	36.73	900	41.47	11.25	-0.12148	0.005763
50.01 – 70	5500	25.58	4800	24.83	700	32.26	12.73	-0.26164	0.01943
<50	2000	9.30	1600	8.28	400	18.43	20.00	-0.80064	0.081312
TOTAL	**21500**	**100**	**19330**	**100**	**2170**	**100**	**10.09**		**0.433126**

We consider that Table 4 is representative for the way in which the analysis of each attribute or feature should be made. The example presents the feature “value of solvency” used to analyze companies in order to determine their patrimonial value. The columns “Company distribution”, “Solvent client distribution (SCD)” and “Insolvent client distribution (DCI)” refer to the distribution of the total number of companies, solvent or insolvent clients reported to each attribute. For instance, 20.93% × $(\frac{4500}{21500}\times 100)$ of all companies, 22.5% × $(\frac{4350}{19330}\times 100)$ of the solvent clients and 6.91% × $(\frac{150}{2170}\times 100)$ of the insolvent ones have a value of solvency ranging between 85.01% and 100%.

While accepting this feature in the final scoring, it is extremely important to take into account the logical distribution, that is to start from a positive value of the WOE indicator and then gradually decrease it.

The WOE indicator measures the power of each attribute in separating the solvent clients from the insolvent ones (the probability with which a client belonging to a certain attribute is considered solvent or insolvent).

An effectve method to calculate this probability, which we used in the Table 4, is the following:

$$WOE = [\ln(\frac{DCS}{DCI})]$$

For example, for the interval 85.01% – 100%, the value of this indicator is $\ln(\frac{0.225}{0.0691}) = 1.180365$.

The IV (Information Value) indicator, which expresses the power of determination of each feature, derives from the Information Theory (Kulback, 1959, p. 205) and is measured using the following formula:

$$\sum_{i=1}^{n}(DCS_i - DCI_i) \times \ln(\frac{DCS_i}{DCI_i})$$

Based on this methodology, the interpretation of results is the following:

- a value lower than 0.02 shows an irrelevant feature, which cannot predict anything;
- a value between 0.02 and 0.1 shows a weak feature;
- a value between 0.1 and 0.3 indicates a feature with a medium discrimination power;
- IV > 0.3 expresses a strong feature, with a higher power of prediction.

We must specify that, up to this point, we have presented different types of individual analysis of each feature, but it is essential to take into consideration the later analysis of the relations between them.

The logical trend of the WOE indicator for all the attibutes of a feature

The statistical power, measured by means of the WOE and IV indicators, is not the only element we must take into account in our analysis. The power of the attributes must follow a logical and operational trend.

It is obvious that we could have distributed the companies for the value of solvency into a different group of attributes, which would have led to a higher power of discrimination.

After drawing up the list of indicators, we have eliminated the ones that did not fit into the work hypothesis, we have checked up the relations between the remaining ones, excluding those showing a mutual influence, and we have calculated the IV for each feature.

In order to maintain a limited number of indicators, we have chosen only the ones showing an IV value over a predetermined minimal threshold.

After calculating the IV (Information Value) for each feature, the following step is to assign a certain number of points to the calculated measurement.

We have started from assigning 10 points and than went on decreasing to 0. We have used the distribution of each feature.

Thus, we go back to the value of solvency:

Table 5

Value of solvency (%)	Expected weight (%)	Number of points	Expected percentage × number of points
>100	11	10	1.1
85.01 – 100	15	8	1.2
79.01 – 85	32	5	1.6
50.01 – 70	21	2	0.42
<50	21	0	0
			4.32

Supposing we have chosen 5 features, we shall draw up the following table

Table 6

Feature	IV	IV distribution (%)
Feature 1	0.3121	26.70
Feature 2	0.2467	21.10
Feature 3	0.2112	18.07
Feature 4	0.2008	17.18
Feature 5	0.1982	16.95
	1.169	

By means of the column „IV distribution”, we shall obtain a rating model.

Therefore, let’s sum up:

- we have selected a series of features with a higher power of discrimination for the representative group;
- we have double-checked, so that there are no relevant relations between the indicators;
- we have assigned points to each and every attribute of the selected features;
- we have calculated quotas by means of which we shall be able to quantify the results of the indicators for each applicant, while using an automated system.

We have envisaged the rating implementation stage as a distinct phase, following the development of the rating-based model.

We have proposed to develop a better undestanding of the economic considerations applied in the process of using a newly-conceived model, of the reports to be drawn up during the process, as well as of the strategies set up and put into practice at the level of the respective creditor.

The objective of the calibration resides in assigning a certain degree of probability referring to the incapacity of payment to every general score obtained by a certain client. The probability of non-reimbursment can be classified in its turn in over 20 categories of rating.

Establishing the probabilities of non-reimbursement for every rating represents one of the fundamental conditions related to the Internal Rating-Based Approach regulated by the Basel II Accord and imposed by the European guidelines.

In order to observe these regulations, the rating scale used during the process must comprise at least seven classes for the solvent clients ant one for the insolvent ones, excepting the retail segment.

We have emphasized the fact that the validation of internal models by the authorized supervisor represents an essential condition if a credit institution intends to use another approach than the standardized one in determining the capital requirements.

The effective use of stress-testing methodologies is extended nowadays not ony to the commercial banks, but also to the regulatory authorities and to the Central Banks. At present, all these institutions are combining stress-testing with their own macro-economic models.

Stress-testing is implemented at country level in order to assess the strenght of the financial system to unfavorable economic evolutions. This type of analysis is set up in conformity with the Financial System Assessment Programme (FSAP) deployed by the International Monetary Fund (IMF). According to the Central-European Bank, the FSAP is implemented in the following countries: Ireland (2000), Finland (2001), Luxembourg (2002), Germany (2003), Austria and Netherlands (2004), Belgium, Greece, Italy, Portugal, Spain and again Ireland (2006).

In our opinion, stress-testing should be extended to a larger geographical area.

We appreciate this direction to be important in the context of the increasingly powerful integration of the EU members, both on the economic and the financial level. We can only guess that such testing is essential in order to assess as accurately as possible the effects of changes the Basel II has brought to the EU credit institutions since January 2007. We mainly have in view the banks adopting the Internal Rating-Based approach.

Notes

(1) According to Regulation of the National Bank of Romania no. 15/20/14.12.2006 concerning the treatment of credit risk by the credit institutions and investment companies following the Internal Rating-Based Approach, Chapter V, Section 1, Art. 113.

(2) See Emergency Government Decree no. 99/6.12.2006 concerning credit institutions and capital adequacy, Chapter III, Section III, Art. 127.

(3) In accordance with Regulation of the National Bank of Romania no. 14/19/14.12.2006 concerning the treatment of credit risk by the credit institutions and investment companies following the standard approach

(4) In accordance with Regulation of the National Bank of Romania no. 15/20/14.12.2006 concerning the treatment of credit risk by the credit institutions and investment companies following the Internal Rating-Based Approach.

(5) In accordance with Regulation of the National Bank of Romania no. 15/20/14.12.2006 concerning the treatment of credit risk by the credit institutions and investment companies following the Internal Rating-Based Approach, Capitolul V, Secțiunea 1, 1.7.1, Art. 153.

(6) See Regulation of the National Bank of Romania no. 15/20/14.12.2006 concerning the treatment of credit risk by the credit institutions and investment companies following the Internal Rating-Based Approach, Capitolul V, Secțiunea 1, 1.7.2, Art. 155.

(7) See Regulation of the National Bank of Romania no. 15/20/14.12.2006 concerning the treatment of credit risk by the credit institutions and investment companies following the Internal Rating-Based Approach, Capitolul V, Secțiunea 2, 2.1., Art. 160.

(8) See Regulation of the National Bank of Romania no. 15/20/14.12.2006 concerning the treatment of credit risk by the credit institutions and investment companies following the Internal Rating-Based Approach, Capitolul V, Secțiunea 2, 2.2, Art. 165.

References

Anghel, I. (2002). *Falimentul. Radiografie și predicție*, Editura Economică, București

Chorafas, D. (2006). *Economic capital allocation with Basel II*, Oxford

Chorafas, D. (2007). *Stress testing for risk control under Basel II*, Elsevier Finance, Oxford

Ganguin B., Bilardello J. (2005). *Fundamentals of corporate credit analysis*, McGraw Hill, New York

Ganguin, B., Bilardello, J. (2005). *Fundamentals of corporate credit analysis*, McGraw Hill, New York

Kulback, S. (1959). *Information Theory and Statistics*, Hoboken, John Wiley & Sons

Loffler, G., Posch, P. (2007). *Credit risk modeling using Excel and VBA*, John Wiley & Sons

Naeem, Siddiqi (2006). *Credit Rik Scorecards. Developing and Implementing Inteligent Credit Scoring*, John Wiley & Sons, Inc

Oesterreichische Nationalbank, Financial Market Authority, "Guidelines on Credit Risk Management, Rating Models and Validation", Vienna, 2004

Ordonanța de Urgență a Guvernului nr. 99 din 6.12.2006 privind instituțiile de credit și adecvarea capitalului

Legea nr. 227 din 04.07.2007 pentru aprobarea Ordonanței de urgență a Guvernului nr. 99/2006 privind instituțiile de credit și adecvarea capitalului

Regulament nr. 13/18/14.12.2006 privind determinarea cerințelor minime de capital pentru instituțiile de credit și firmele de investiții

Regulament nr. 14/19/14.12.2006 privind tratamentul riscului de credit pentru instituțiile de credit și firmele de investiții potrivit abordării standard

Regulament nr. 15/20/14.12.2006 privind tratamentul riscului de credit pentru instituțiile de credit și firmele de investiții potrivit abordării bazate pe modele interne de rating