JEL Classification: O470

УДК 336.77: (336.717-061)

STEP-BY-STEP METHOD OF CREDIT SCORING MODEL DEVELOPMENT FOR RISK DYNAMICS ESTIMATION AND MANAGEMENT

PYSANETS K.[1]
[1]Kyiv National University of Technologies and Design

Keywords:
Credit scoring, scoring model, credit risk, risk dynamics, credit risk management.

ABSTRACT

Purpose. Create scoring parametric model based on the concept of survival for application stage of credit cycle in the Ukrainian consumer loans market which allows to take into account factors' influence change on default risk in dynamics.

Methodology. Scientific and special methods were applied during the research: the historical, system and structure methods, methods of logical and economic analysis, economic-mathematical modeling methods as well as number of probability theory and mathematical statistics methods.

Results. Scoring model for assessment of credit risk dynamics was built based on proposed step-by-step method. Significant factors which determine default risk during whole loan term were defined. Credit scoring modeling methods were improved by consideration of factors influence change on credit risk measure in dynamics.

Scientific novelty. Step-by-step method is proposed for credit scoring model development which allows to assess credit risk dynamics considering factors influence change on credit risk over time.

Introduction. Credit scoring is an effective tool for credit risk assessment in modern risk management. The purpose of implementation process of credit scoring is to improve the mechanism of credit risk management through effective borrowers' differentiation. Modern methodology of scoring is well developed both theoretically and practically. Conceptual logic of credit scoring development and its use are well presented in fundamental works of Anderson R. [1] and Siddiqi N. [2].

Financial institutions build credit scoring to assess borrowers' creditworthiness in various stages of the credit cycle, based on different information and methods of scoring function construction. Thus, according to the stages of the credit cycle, we can distinguish application, collection and behavioral scoring. Based on the information used in the process of scoring creation, two types of scoring can be defined: internal credit scoring and credit bureau scoring. The principal difference between the two approaches is that the lender uses only own information that is available for him, but the credit bureau has complete information about the borrower's credit history, gathered from number of banks and other lenders from the whole market.

Methods of credit scoring models development can be classified into parametric and nonparametric. Parametric methods include generic Bayesian parametric method, linear probabilistic approach, logistic regression etc. Nonparametric methods include expert method, a generic Bayesian nonparametric method, neural networks, nearest neighbors' approach, genetic algorithms etc. Basic statistical methods for credit scoring development are well represented in the literature, particularly by Anderson R. [1]. Expert scoring development methods in Ukraine were developed by Volyk N. [3] and Kaminsky A. [4]. The application of neural network technology, in particular for collection scoring were investigated by Kaminsky A. [5]. Malyugin V. in his article [6] compares different algorithms of application scoring: linear discriminant analysis algorithm, the logit models algorithm of binary choice, decision tree algorithm. In the research [7] Delamaire L. et al. describe genetic algorithms, decision tree method, cluster analysis approach, neural networks method for scoring that identifies fraud. Number of works was analyzed in the monograph of Thomas L. [8] and in the research of Man R. [9].

All the presented works consider the dependent variable in models as binary type that takes the value «Good» if the result is positive event and «Bad» otherwise. For example, for application scoring of biennium loans variable is set to «Bad» if borrower's delinquency has been lasted for 90 days or more. The borrower is considered «Good» if the delay in payments was absent. Building a collection scoring, variable «Good» characterizes the borrower who pays the debt to the lender, and the variable «Bad» describes borrowers who do not pay. The principal similarity among all these definitions is that borrowers with «Bad» and «Good» statuses are considered "in average" for the whole loan term, i.e. without reference to the time when they became «Bad» or «Good». Even in 1992 Narain B. [10] had proposed to include time parameter to credit scoring model that allows to evaluate not only the fact of the event (loan default), but the time of its occurrence, thereby opening new opportunities for credit risk management.

Today, there is a large number of scientific works, devoted to research of credit scoring with a time dependent parameters and the theoretical basis of it discloses the survival concept. This concept was originated in biology, but today its main ideas are used in many fields of knowledge, including economy. The essence of this concept is introduced for consideration and study of survival function $S(t) = P(T > t)$, where T - time of the organism death (in biology) or time to debt occurrence of the borrower (economy). This approach is actively developing in credit risk management. Credit scoring models are based on the concept of survival are presented in [10] - [15]. For example, Baesens B. et al. [11] describe neural network technology use for assessment of the borrower default at the date of the loan application.

Bellotti T. and Crook J. proposed credit scoring model for estimation of dynamic of borrowers' probability of default [13] on the application stage, estimating parameters by maximum likelihood method.

The main idea of the use of an approach based on the concept of survival for credit scoring is that it allows building more adequate models of lending activities. Those borrowers who took credit and did not return it from the first payment bring much greater losses than those borrowers who stopped paying on the loan after a series of payments. In the second case, borrowers can even return the entire principal of the loan before the time of default. Thus, differentiation of borrowers' time to default is the actual task which can be solved by building scoring models for credit activities.

It is worth to mention that the majority of works are devoted to modeling of credit risk dynamics are based on the concept of survival and use addition of time parameter to the "classical" model scoring.

Setting objectives. In this paper we present a model of credit scoring in which all scoring weights depend on the time parameter, and suggest implementation of such a model to differentiate debtors for application stage of the credit cycle. Thus, the objective of this work is to create scoring parametric model based on the concept of survival and its application to develop improved application scoring model for the Ukrainian market. The primary interests are the model coefficients dependence on time parameter and characteristics that lead to differentiation of debtors.

Research results. The idea of building models of scoring systems to assess the dynamics of credit risk is innovative for the Ukrainian consumer lending market despite the fact that it was described in scientific works in the 90's of the XX century [11]. This is due to the fact that the massive lending in Ukraine began in the middle of the first decade of the XXI century, and domestic financial institutions until relevant period did not have a sufficient amount of information and material resources for such researches. The stimulus to find ways of the effectiveness improvement of credit risk management, including the use of models of risk dynamics assessment is caused by two main factors: the growing competition and the financial crisis of 2008-2009.

Let's consider these factors in more detail. The first factor is the competition in the segment of consumer lending. It requires financial institutions to conduct continuous improvement of credit risk management mechanisms that today mainly based on models of "risk-profitability". The model with lower losses on risk assessment error offers better loan conditions, such as lower interest rate. Another example is inclusion of default time parameter to the models to evaluate the dynamics of borrowers' risk and profitability. For example, consider borrowers, which despite of the fact of default time, cover interest and major part of loan principal.

Flexible credit risk management, which takes into account the time by default, allow for financial institution to keep client by offering a shorter loan term, or fewer loan amount. Another example is to evaluate borrowers who repay early.

The second factor was the financial crisis of 2008-2009, which allowed to show advantages of credit scoring systems for credit risk dynamics estimation over generic credit risk assessment which consider risk state for fixed time horizon. Models with time parameters allow to include macroeconomic parameters and provide stress test with the expected time of the crisis occurrence, as shown in the publications of Bellotti T. and Crook J. [13].

Consequently, the inclusion of time parameter into scoring systems models improves the credit risk management quality and it is a promising area of research for the Ukrainian market of consumer lending.

Explained by the first factor of need in scoring systems models improvement for borrowers' risk assessment we propose to build a modified model of Cox D. [14], the essence of which is to define the most important risk factors and assess their impact on the dependent variable taking into account differences in the degree of their influence over time, from loan issue to its repayment.

Let's consider Cox model with logic basic function in the next form:

$$h(x_i, t) = \ln(\frac{P(i,t)}{1 - P(i,t)}) = \beta_0(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_n x_{in}, \quad (1)$$

$h(x_i, t)$ – borrower's i risk function at time t, presented as logarithm of event (default) probability to event absence (non-default) probability ratio; $P(i,t)$ – probability that borrower i will take default status at time moment t; x_{ij} – characteristic j of borrower i, where $i = \overline{1,k}$, $j = \overline{1,n}$; $\beta_0(t)$ – basic risk function (without influence of factors x_{ij}), which depends on time parameter t; β_j – model coefficient, that describes significance of the characteristic j.

The model (1) comprises risk of borrowers as proportional in time, expressed by coefficient $\beta_0(t)$. In addition, set of characteristics x_{ij} is defined by their influence significance of default rate for the fixed period of time, which is basically limited by loan term (for example, 2 years). Because of this fact, some variables, which significantly influence the level of credit risk during certain time periods of loan term, can be missed. Another drawback of the model is equal level of borrowers' characteristic influence on default level over time (coefficients β_j do not depend on time parameter in model (1)).

Let's consider a model to assess the level of default risk of the borrower that takes into account the dynamics of influence power of risk factors during time interval of evaluation:

$$h(x_i, t) = \ln\left(\frac{P(i,t)}{1 - P(i,t)}\right) = \beta_0(t) + \beta_1(t)x_{i1} + \beta_2(t)x_{i2} + \ldots + \beta_n(t)x_{in}. \quad (2)$$

To estimate the parameters of the model we propose to use "step-by-step" method. In the first step we determine the time periods and for each factor calculate dynamics of "information value" indicator [2, pp. 81-82]. In the second step based on correlation and regression analysis, «p-value» indicator and the business logic, in each period we determine significant independent variables and estimate logistic regression model parameters. In case of the quarterly data and the time horizon of two years, it is required to build 8 models (we will name them “models from the second step” further). In the case of monthly data, we should build 21 models (three months need for default happening). In the third step we build models to assess the significance of each factor dependence β_j on the time parameter t (we will call them “models from the third step” further). These can be, for example, models of linear or nonlinear regression. In the fourth step we combine obtained models from the second and third step into a basic model of the form (2).

We built the model of credit risk assessment of borrowers and conducted estimation of its parameters for application stage of the credit cycle using data from more than 16,000 borrowers in Ukraine for consumer loans, issued in 2012 with full credit history for two-year period.

The result of the first step - an assessment of the information value – gave 46 (among 62 initially selected) factors with strong, average or weak predictive power for at least one period of selected time horizon of two years. In the second step correlation matrix was built and dependent variables were discarded according to the principle: among two correlated factors we excluded the one that has less information value. Then, for each month the model of logistic regression was built (21 models), which excluded factors with «p-value» less than 0.05 [16, p. 266-237]. Thirteen independent factors included in 21 models and their characteristics are given in Table 1.

Table 1

Factors of step-by-step model for borrowers' credit risk assessment

#	Factor name	Information value, t=4	Information value, t=24	Max correlation coefficient with other factors, modulus	p-value, t=4	p-value, t=24
1	Applicant age	0,55	0,14	0,433	0	0
2	Job experience at the last place of work	0,44	0,10	0,433	0	0
3	Loan term	0,11	0,10	0,138	0	0
4	Number of requests to Credit bureau for last month	0,22	0,09	0,165	0	0
5	Marital status	0,37	0,07	0,192	0	0
6	Registration region	0,16	0,06	0,159	0	0
7	Broker presence in loan obtaining process	0,01	0,05	0,165	0	0,001
8	Residence term at the last place	0,08	0,04	0,297	0,002	0,001
9	Third party person alliance	0,05	0,04	0,087	0	0
10	Loan purpose	0,13	0,03	0,163	0	0
11	Number of active loans	0,01	0,03	0,104	0,031	0
12	Gender	0,04	0,02	0,212	0	0
13	Availability of telephone number	0,10	0,02	0,156	0	0

Source: Calculated by the author

Therefore, estimating models parameters from the second step, we have received 21 values for each coefficient $\beta_0, \beta_1, \beta_2, ..., \beta_{13}$ for model (2). These coefficients show the dynamic of impact strength of relevant characteristics on the level of borrowers' default risk. Two types of factors behavior over time can be defined, with increasing or downward dynamics. Two parameters demonstrate this below in Fig. 1.

For all estimates of coefficients with increasing (decreasing) dynamics we detect significant growth (fall) from the fourth to the sixth month from the date of loan issue, followed by a period of moderate growth (decline). Since 18th month, value of coefficient estimate stabilizes and tends to constant. The ration of scattering coefficient of factor to its mean value can range from 20% to 90% (Fig. 1).

To simulate the curves of received form, we propose to use the logistic trend that will be applied in the next step.

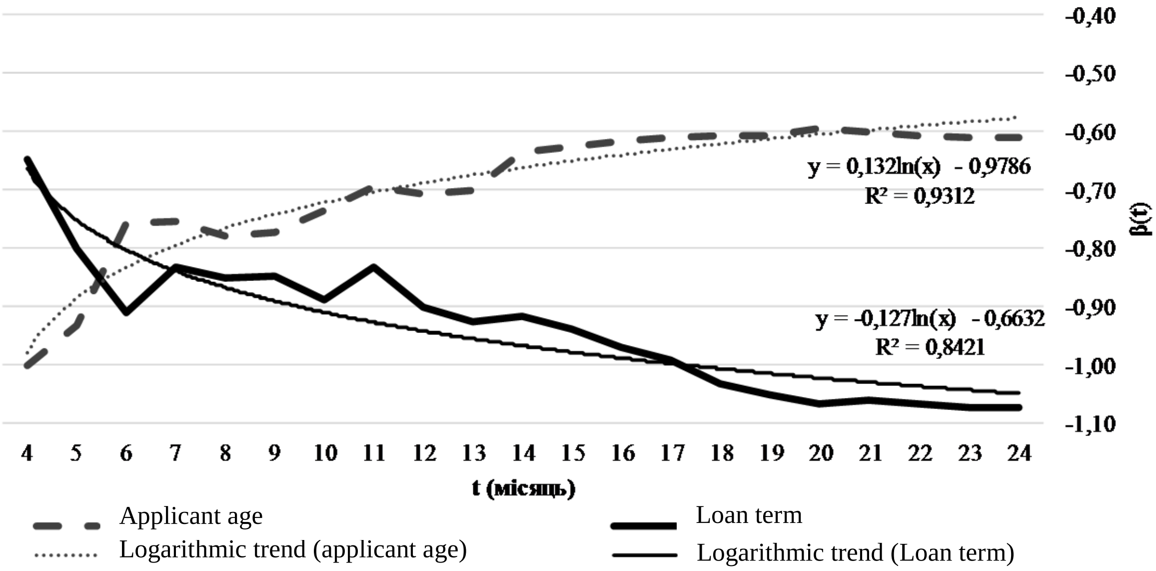

Figure 1. Types of coefficients dynamics from models of the second step
Source: Calculated by the author

In the third step we assessed 14 models to get a set of functions, each of which describes the ratio of the respective factor depending on the time parameter. Thirteen of these models are designed to factors listed in the Table 1, and one more - for constant. To estimate coefficients from the third step we used least squares method. Depending equation for the coefficient j is:

$$\beta_j(t) = \alpha_{0j} + \ln t \cdot \alpha_{1j}, \tag{3}$$

where $\beta_j(t)$ – the coefficient of the model from the second step for characteristic $j = \overline{1,n}$; t – time parameter; α_{0j} and α_{1j} – coefficient of the model from the third step.

Unlike linear trend and exponential regression, this assessment approach provides a smaller error for the early and late periods. Coefficients of all characteristics near t, except loan term and number of active loans have growing trend. This means that their impact on the default risk decreases with time (see table 2).

Implementing the fourth step, models from the second and the third steps are combined into basic model of the next form:

$$\begin{aligned} h^*(x_i, t) = {} & (0{,}57 + 0{,}66 \cdot \ln t) + (-1{,}2 + 0{,}2 \cdot \ln t) \cdot x_{i1} + (-0{,}74 + 0{,}1 \cdot \ln t) \cdot x_{i2} + \\ & + (-0{,}43 - 0{,}2 \cdot \ln t) \cdot x_{i3} + \ldots + (-0{,}48 + 0{,}06 \cdot \ln t) \cdot x_{i13}, \end{aligned} \tag{4}$$

where $h^*(x_i, t)$ – estimated risk function, that depends on time parameter t and on vector of parameters x_i for borrower i.

Table 2

Factors and coefficients estimated in step-by-step model of borrowers' default risk dynamics estimation

j	Factor j name	Estimate α_{0j}^{*}	Estimate α_{1j}^{*}
1	Applicant age	-1,20	0,20
2	Job experience at the last place of work	-0,74	0,10
3	Loan term	-0,43	-0,20
4	Number of requests to Credit bureau for last month	-1,76	0,18
5	Marital status	-0,71	0,04
6	Registration region	-0,94	0,09
7	Broker presence in loan obtaining process	-0,29	0,04
8	Residence term at the last place	-0,42	0,05
9	Third party person alliance	-1,15	0,15
10	Loan purpose	-0,94	0,08
11	Number of active loans	0,05	-0,19
12	Gender	-0,56	0,08
13	Availability of telephone number	-0,48	0,06
14	Constant	0,57	0,66

Source: Calculated by the author

Conclusions. We have built the model (4) for evaluating borrowers' credit risk dynamics and estimated its coefficients for consumer credit segment of the Ukrainian market which allows to manage credit risk more efficiently. The higher value of the risk function, the higher probability of borrowers' default. The main difference of proposed approach is improvement of risk parameters influence compared to the Cox model. Information value indicator allowed to define significant risk factors considering the dynamics of their impact on the risk of default, and step-by-step method - to estimate the influence of factors on the dependent variable in each time moment with monthly periodicity. There were determined 13 most significant variables (Table 2) which can be useful for risk managers to estimate default risk on application stage of credit cycle in the Ukrainian consumer segment of credit market.

References

1. Anderson, R. A. (2007). *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management.* Oxford University Press: UK.
2. Siddiqi, N. (2006). *Credit risk scorecards: developing and implementing intelligent credit scoring*. Hoboken.: John Wiley & Sons.
3. Volyk, N. H. (2008). Skorynh yak ekspertnyi metod otsiniuvannia kredytnoho ryzyku komertsiinoho banku pry spozhyvchomu kredytuvanni [Scoring as expert method of credit risk evaluation of commercial banks in consumer lending]. *Visnyk Zaporiz'koho natsional'noho universytetu. [Bulletin of Zaporizhsky National University]. Ser.: Ekonomichni nauky, 1,* 40-44.

4. Kaminskiy, A. B. (2006). Ekspertna model' kredytnoho skorynhu pozychal'nyka banku [Experts credit scoring model for borrowers]. *Bankivska sprava, 1,* 75-81 [in Ukrainian].
5. Kaminskiy, A. B. & Sikach, V. O. (2011). Neyromerezhevi tekhnolohiyi v upravlinni portfelem prostrochenoyi zaborhovanosti [Neural network technologies in management of arrears]. *Mizhvid. nauk. Zb. Modelyuvannya ta informatsiyni systemy v ekonomitsi [Modeling and information systems in economics], 84,* 5-19 [in Ukrainian].
6. Maliuhyn, V. Y. & Hryn', N. V. (2010). Ob effektyvnosty statystycheskykh alhorytmov kredytnoho skorynha [The effectiveness of statistical credit scoring algorithms]. *Bankovskyy Vestnyk [Bank bulletin], 4,* 39-46 [in Ukrainian].
7. Delamaire, L., Abdou, H. & Pointon, J. (2009). Credit card fraud and detection techniques: a review. *Banks and Bank Systems, 4(2),*57-68.
8. Thomas, L. C., Edelman, D. B. & Crook, J. N. (2002). *Credit scoring and its applications: SIAM monographs on mathematical modeling and computation*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
9. Man, R. *Survival analysis in credit scoring: A framework for PD estimation*. Retrieved from http://essay.utwente.nl/65049/1/Thesis% 20Ramon%20Man.pdf.
10. Narain, B. (1992). Survival analysis and the credit granting decision. Thomas, LC, Crook, JN, & Edelman, DB (eds). *Credit Scoring and Credit Control. Oxford University Press,*109-121.
11. Baesens, B., Van Gestel, T., Stepanova, M., Van den Poel, D. & Vanthienen, J. (2005). Neural network survival analysis for personal loan data. *Journal of the Operational Research Society, 56,* 1089-1098.
12. Stepanova, M. & Lyn, T. (2002). Survival Analysis Methods for Personal Loan Data. *Operations Research*, *50(2),* 277-289.
13. Crook, J., Bellotti, T. (2011). Forecasting and stress testing credit cards default using dynamic models. *Working paper, 11-35, Credit Research Centre, University of Edinburgh Business School.*
14. Cox, D. & Oakes, D. (1984). *Analysis of Survival Data: monograph.* London: Chapman and Hall.
15. Pysanets, K. (2013). *Survival concept for credit scoring models // Ynnovatsyonnie y ynformatsyonnie tekhnolohii v razvytyi natsyonal'noi ekonomyky: teoryia y praktyka: Monohrafyia [Innovative and information technologies in the development of national economy: theory and practice: Monograph]*. T. S. Klebanova (Ed.). Moscow: Nauchnye tekhnolohyi.
16. Hrin, V. H. (2005). *Ekonometrychnyy analiz: pidruchnyk [Econometric analysis: textbook]*. (A. Oliynyk & R. Tkachuk Trans.). Kiev: «Osnovy» [in Ukrainian].