UDC 336.7

Nikita Lukashevich
PhD (Economics), Associate Professor,
Business and Commerce Department, St. Petersburg State Polytechnical University, Russia
29 Polytechnicheskaya Str., St. Petersburg, 195251, Russia
lukashevich@kafedrapik.ru

A PRACTICAL APPROACH TO VALIDATION OF CREDIT SCORING MODELS

Abstract. *Introduction.* The implementation of the Third Basel Accord raises many technical and methodological issues regarding the development and validation of credit risk models and makes these issues much more important. Bank regulators will pay more and more attention to testing model validation processes in order to examine the predictive accuracy of banks' credit scoring models. Lenders therefore need to develop and apply the approaches for operational monitoring of predictive accuracy and modifying the cutoff value as one of the most important parameters of credit scoring models. The author poses the receiver operating characteristic (ROC) curve technique that can be successfully used to validate credit scoring models. *The purpose of the research* is testing the application of ROC curve technique for estimating the validity and predictive accuracy of credit scoring models. *Results.* The main parameters of credit scoring models validation were summarized. The possible criteria for determining the acceptable cutoff value for credit scoring models are presented. The approbation of the ROC curve technique is given by comparing two logistic models developed on the factual statistical data. *Conclusions.* The ROC curve technique can be applied successfully to estimate validity and compare credit risk models. The research gives recommendations of how to apply the proposed technique in the validation process. The area for the further research can be the consideration of the ROC curve technique in terms of the economic benefits and losses from the true and false classified credit applications.
Keywords: credit risk; credit scoring; logistic regression; validity; ROC curve.
JEL Classification: C52, E42, E58, G21

Н. С. Лукашевич
кандидат экономических наук, доцент, Санкт-Петербургский государственный политехнический университет, Россия
О ПРАКТИЧЕСКОМ ПОДХОДЕ К ВАЛИДАЦИИ КРЕДИТ-СКОРИНГОВЫХ МОДЕЛЕЙ
Аннотация. Внедрение Базельского соглашения вызывает много важных технических и методологических вопросов, касающихся разработки и проверки достоверности кредит-скоринговых моделей. Кредиторы должны вырабатывать и применять подходы для оперативного мониторинга точности прогноза и изменения порогового значения как одного из ключевых параметров. Целью исследования является апробация ROC-анализа для оценки достоверности моделей кредитного скоринга. Автором обобщены основные параметры кредит-скоринговых моделей и представлены критерии определения приемлемого порогового значения. Апробация ROC-анализа осуществлена путем сравнения двух логистических моделей, разработанных на основе статистических данных. Исследование позволило выработать рекомендации о применении подхода на основе ROC-анализа в процессе валидации. Областью дальнейших исследований может стать применение ROC-анализа с точки зрения выгоды или потерь от правильно либо ложно классифицированных кредитных заявок.
Ключевые слова: кредитный риск, кредитный скоринг, логистическая регрессия, валидация, ROC-кривая.

М. С. Лукашевич
кандидат економічних наук, доцент, Санкт-Петербурзький державний політехнічний університет, Росія
ПРО ПРАКТИЧНИЙ ПІДХІД ДО ВАЛІДАЦІЇ КРЕДИТ-СКОРИНГОВИХ МОДЕЛЕЙ
Анотація. Упровадження Базельської угоди викликає багато важливих технічних і методологічних питань, які стосуються розробки та перевірки достовірності кредит-скорингових моделей. Кредитори повинні виробляти та застосовувати підходи для оперативного моніторингу точності прогнозу і зміни порогового значення як одного із ключових параметрів. Метою дослідження є апробація ROC-аналізу для оцінки достовірності моделей кредитного скорингу. Автором узагальнено основні параметри кредит-скорингових моделей та представлено критерії визначення прийнятного порогового значення. Апробацію ROC-аналізу проведено шляхом порівняння двох логістичних моделей, розроблених на основі статистичних даних. Дослідження дозволило виробити рекомендації щодо застосування підходу на основі ROC-аналізу в процесі валідації. Сферою подальших досліджень може стати використання ROC-аналізу з точки зору вигоди чи втрат від правильно або помилково класифікованих кредитних заявок.
Ключові слова: кредитний ризик, кредитний скоринг, логістична регресія, валідація, ROC-аналіз.

Introduction. In modern conditions, the problem of credit risk management is becoming increasingly important. The requirements for the reliability of the banking system, imposed by the various regulatory bodies, credit terms and the number of credit operations, success of which directly depend on the economic situation of the borrowers, are constantly growing. In accordance with the Basel Capital Accord, known as Basel III, it is recommended for the estimation of credit quality to use an approach based on the internal banking ratings and according to which it is required to develop the mathematical models to estimate the probability of default. The analyst can use the abbreviated, structural and credit scoring models that have the greatest practical interest and allow estimating the borrowers' credit rating. The implementation of Basel III raises many technical questions regarding the development and validation of credit risk models. It also makes the validation of credit risk models much more important. Bank regulators will pay more and more attention to model validation processes in order to examine the accuracy of banks' credit scoring models.

The purpose of the research is testing the application of ROC for estimating the validity and predictive accuracy of credit scoring models. As the information base for research the impersonal sample of the individual borrowers was captured. Based on the sample, using logistic regression as the traditional statistical tool to estimate the probability of default, a credit scoring model was designed for testing ROC curve technique.

Brief Review of Literature. There are so many papers used intelligent and statistical techniques in credit scoring since the 1930s. D. J. Hand and W. E. Henley (Hand & Henley, 1997) reviewed several statistical classification models in consumer credit scoring. H. A. Abdou and J. Pointon (Abdou & Pointon, 2011) reviewed articles based on credit scoring applications in various areas especially in finance and banking based on statistical techniques. Their study also include some of data mining techniques, and comparison of different techniques accuracy for different datasets, they conclude that there is no overall best statistical technique in building scoring models. L. C. Thomas (Thomas, 2000) describes a variety of approaches to development of credit scoring models, among which the statistical and neural network methods are traditionally used in practice and implemented in most modern banking software products. All recommendations of how to choose an approach are detailed in the articles [1; 2; 3]. The practical credit scoring models, developed on the basis of the statistical, neural networks or fuzzy sets methods, and the comprehensive interpretation of the peculiarities of their application for the purpose of credit risk analysis are presented by Y. Yingxu (Yingxu, 2007), B. W. Yap, S. H. Ong and N. H. M. Husain (Yap, Ong & Husain, 2011), C.-F Tsai and J.-W. Wu (Tsai & Wu, 2008). There are so many validation and test methods such as accuracy rate, type I and II errors, areas under ROC curve that are mostly used in the research. D. J. Hand and R. J. Till (Hand & Till, 2001) examined the important problem of variables selection in the scorecard using logistic regression. The authors presented approach to variables selection depending on the calculated values of the area under the ROC curve. A. Ben-David and E. Frank (Ben-David & Frank, 2009) dealt with the problem of comparing the expert and statistical credit scoring models in terms of classification accuracy and validity. T. Seidenfeld (Seidenfeld, 1985) touched the problem of scoring rules calibration for the first time.

Results. Regardless of the used approach, an important prerequisite for the effective implementation of credit scoring models is the reasonable choice of their parameters, required for decision making on crediting, as well as the estimation of the predictive accuracy of the models, that defines the classification accuracy of the borrowers. To resolve this problem it is possible to use ROC curve analysis [10]. In signal detection theory, a receiver operating characteristic (ROC), or simply ROC curve, is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold is varied. ROC curve analysis is widely used in various fields such as the theory of signal detection, the diagnostic tests in medicine [11], a comparison of models and algorithms in the theory of management decisions [12; 13]. The ROC curve allows constructing the dependence of the number of correctly classified positive examples on the number of incorrectly classified negative examples [14].

$$TPR = \frac{TP}{TP+FN},$$

$$FPR = \frac{FP}{TN+FP},$$

$$TNR = \frac{TN}{TN+FP},$$

$$FNR = \frac{FN}{FN+TP},$$

where TP (true positives) – the true classified positive outcomes (true positive outcomes); TN (true negatives) – the true classified negative outcomes (true negative outcomes); FN (false negatives) – the positive outcomes classified as the negative one (false negative outcomes); FP (false positives) – the negative outcomes classified as the positive one (false positive outcomes).

Parameter TPR determines the sensitivity of the model. The model, possessing the high sensitivity, provides the greater probability of the correct recognition for the positive outcomes. Parameter TNR determines the specificity of the model. The model with high specificity provides a greater probability of the correct recognition for the negative outcomes. Briefly summarized, the model with a high specificity corresponds to a conservative credit policy (a high level of rejected credit applications) and the model with a high sensitivity – a risky credit policy (a high level of approved credit applications). In the first case, the losses from credit risk are minimized, and in the second one the loss of economic benefit is minimized. The last important parameter of credit scoring models is the threshold (limit) value C (cutoff point). This value is essential in order to apply the model in practice and classify the new outcomes. Choosing the threshold value, the analyst can control the probability of the correct recognition of the positive and negative outcomes. When reducing the threshold value, the probability of the erroneous recognition of the positive outcomes (false positive outcomes) increases and conversely, when maximizing, the probability of the incorrect recognition of the negative outcomes increases (false negative outcomes).

The ROC curve represents a set of coordinates, specified by TPR and (1 TNR) at different values of C. For the perfect classifier the graph for the ROC curve passes through the upper left corner, where the share of the false positive outcomes is equal to zero. Therefore, the closer the curve to the upper left corner, the higher the predictive capability of the model. The diagonal line (the socalled line of nodiscrimination or random guess) corresponds to the «bad» classifier. Parameter AUC is calculated as the area under the ROC curve using, for example, trapezoid rule [7] and takes values in the interval [0; 1]. The high value for AUC is evidence of the high quality of the model in terms of its predictive capability.

The key problem in the ROC curve analysis is to determine the acceptable threshold value on the basis of the formalized ROC curve. The possible criteria for determining the acceptable threshold value among k possible values are presented below:

1. Ensuring the minimum allowable value of the model sensitivity TPR_{min} (criterion K_1):

   $TPR_k \geq TPR_{min}$.

2. Ensuring the minimum allowable value of the model specificity TNR_{min} (criterion K_2):

   $TNR_k \geq TNR_{min}$.

3. Ensuring the maximum value of total sensitivity and specificity of the model (criterion K_3):

   $max\ \{(TNR_k + TPR_k)\}$.

4. Ensuring a balance between sensitivity and specificity of the model (criterion K_4):

   $min\ \{|TPR_k - TNR_k|\}$.

5. Ensuring the maximum value of Youden's index (criterion K_5) [15]:

   $max\ \{(TPR_k+TNR_k-1)\}$.

6. Ensuring the maximum value of reliability index (criterion K_6)

   $max\ \{(\frac{TN_k+TP_k}{TN_k+TP_k+FN_k+FP_k})\}$.

7. Ensuring the minimum sum of losses from classification errors (criterion K_7):

   $min\ \{(S_{FP}\cdot FP_k + S_{FN}\cdot FN_k)\}$,

where S_{FP} – cost of the false positive outcome;
S_{FN} – cost of the false negative outcomes.

The greatest practical interest provides the last criterion. On the one hand, it allows linking classification errors with economic indicators, but, on the other hand, the determination of the false outcomes cost is a difficult problem, requiring special research, that significantly limits the application of this criterion in practice. The analyst can roughly calculate the cost of classification errors for each false outcome on the basis of data on overdue debt and credit conditions.

Two credit scoring models based on logistic regression were defined during the statistical processing. Due to correlation between predictors, the parameters of the model may be inaccurate, resulting in a significant number of the false outcomes.

The matrix of pair correlation coefficients is formed and presented in Table 1. The conclusion about partial multicollinearity can be made. In this case, it is formally possible to obtain estimates of the model parameters and their exact values, but they will not be stable and will affect the predictive accuracy of the models. Considering that the research objective is testing the application of the ROC curve analysis in banking practice rather than getting the adequate practical credit scoring models, parameters of the models were found.

Tab. 1: **The matrix of pair correlation coefficients (significant coefficients highlighted)**

Q_i	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	Q_9	Q_{10}
Q_1	1,000	0,146	**0,314**	-0,085	-0,189	0,017	-0,050	0,182	0,071	-0,053
Q_2	0,146	1,000	-0,231	0,204	-0,143	0,230	-0,154	0,026	0,021	0,083
Q_3	**0,314**	-0,231	1,000	**-0,280**	0,147	-0,117	-0,189	0,199	-0,015	-0,080
Q_4	-0,085	0,204	**-0,280**	1,000	**-0,259**	0,092	0,136	-0,115	-0,150	0,054
Q_5	-0,189	-0,143	0,147	**-0,259**	1,000	0,039	-0,001	0,163	-0,021	-0,084
Q_6	0,017	0,230	-0,117	0,092	0,039	1,000	-0,111	0,025	0,007	-0,075
Q_7	-0,050	-0,154	-0,189	0,136	-0,001	-0,111	1,000	0,081	-0,073	-0,230
Q_8	0,182	0,026	0,199	-0,115	0,163	0,025	0,081	1,000	**0,253**	-0,179
Q_9	0,071	0,021	-0,015	-0,150	-0,021	0,007	-0,073	**0,253**	1,000	**-0,283**
Q_{10}	-0,052	0,082	-0,080	0,053	-0,084	-0,075	-0,230	-0,179	**-0,283**	1,000

Source: Author's development

To build the first model

$$Z_1=-0,17Q_1-0,04Q_2+1,9Q_3+0,5Q_4+0,3Q_5+$$
$$+0,58Q_6+1,7Q_7+4,8Q_8+0,9Q_9+0,21Q_{10}-7,2$$

based on logistic regression the method of step-by-step inclusion with Wald test is used.

For the second model

$$Z_2= 1,79\ Q_3 + 1,53\ Q_7 + 4,9\ Q_8 - 6,89$$

the same settings for logistic regression are used, but with the forced inclusion of all factors.

The models include the following factors: Z – default («yes» or «no»), Q_1 – gender, Q_2 – age, Q_3 – marital status, Q_4 – record of service, Q_5 – type of employer, Q_6 – credit history, Q_7 – savings Q_8 – the ratio of income to expenses, Q_9 – income variation, Q_{10} – security for credit.

On the basis of the formalized logistic regression models the main parameters and criteria $(K_1$-$K_6)$ were calculated to conduct the ROC curve analysis. The results of calculations only for the first model Z_1 are presented in Table. 2. The calculated parameters allowed making the ROC curves for both models, presented in Figure 1, and define the rational threshold value C.

Despite the various parameters and methods of logistic regression construction, the predictive accuracy of both models is the same because of the similar values of AUC, obtained by summing the figures in the corresponding row in Table 2. This fact can be explained by the sufficient correlation between factors. The curves are closer to the diagonal line of random guess that confirms the fact of the correspondence between both models and «bad» classifier.

The rational threshold value was found using criteria K_3, K_5 and K_6 and equal to 0.60 for all criteria (see the underlined figures in Table 2).

The balance between sensitivity and specificity for the model Z_1 is achieved at the threshold value 0.35, as shown in Figure 2.

Conclusions. Summing up, we can say that the ROC curve analysis can be applied to solve the following tasks in credit risk management: 1) estimation and comparison of the predictive accuracy, sensitivity and specificity of credit scoring models; 2) determination of the rational threshold value for credit scoring models; 3) parameters of credit scoring models assessed by the ROC curve analysis may be used as the indicators showing the need for adjusting the model (classifier). The lower sensitivity of the model, increase in the number of the false positive outcomes are some examples of such indicators.

Tab. 2: **The ROC curve analysis results**

Model	**Parameters**	**The threshold value, *C***																				
		0	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	1
Z_1	TP	35	35	35	35	35	32	31	28	25	24	23	22	22	18	15	14	12	9	8	5	0
	TN	0	0	2	5	9	11	17	19	23	24	28	29	30	30	30	31	33	34	35	35	35
	FN	0	0	0	0	0	2	4	7	9	11	12	13	13	17	20	22	23	26	27	30	35
	FP	35	35	33	30	26	25	18	16	13	8	7	6	5	5	5	3	2	1	0	0	0
	TPR	1,00	1,00	0,95	0,88	0,80	0,74	0,65	0,60	0,52	0,50	0,45	0,43	0,42	0,38	0,33	0,31	0,27	0,21	0,19	0,13	0,00
	FPR	1,00	1,00	0,94	0,86	0,74	0,69	0,51	0,46	0,36	0,25	0,20	0,17	0,14	0,14	0,14	0,09	0,06	0,03	0,00	0,00	0,00
	TNR	0,00	0,00	0,06	0,14	0,26	0,31	0,49	0,54	0,64	0,75	0,80	0,83	0,86	0,86	0,86	0,91	0,94	0,97	1,00	1,00	1,00
	FNR	0,00	0,00	0,00	0,00	0,00	0,06	0,11	0,20	0,26	0,31	0,34	0,37	0,37	0,49	0,57	0,61	0,66	0,74	0,77	0,86	1,00
	K_1	1,00	1,00	0,95	0,88	**0,80**	0,74	0,65	0,60	0,52	0,50	0,45	0,43	0,42	0,38	0,33	0,31	0,27	0,21	0,19	0,13	0,00
	K_2	0,00	0,00	0,06	0,14	0,26	0,31	0,49	0,54	0,64	0,75	**0,80**	0,83	0,86	0,86	0,86	0,91	0,94	0,97	1,00	1,00	1,00
	K_3	1,00	1,00	1,00	1,02	1,05	1,05	1,13	1,14	1,16	1,25	1,25	1,26	**1,28**	1,23	1,19	1,22	1,21	1,18	1,19	1,13	1,00
	K_4	1,00	1,00	0,89	0,73	0,54	0,44	0,16	**0,05**	0,12	0,25	0,35	0,40	0,43	0,48	0,52	0,60	0,68	0,76	0,81	0,88	1,00
	K_5	0,00	0,00	0,00	0,02	0,05	0,05	0,13	0,14	0,16	0,25	0,25	0,26	**0,28**	0,23	0,19	0,22	0,21	0,18	0,19	0,13	0,00
	K_6	0,50	0,50	0,53	0,57	0,63	0,61	0,69	0,67	0,69	0,72	0,73	0,73	**0,74**	0,69	0,64	0,64	0,64	0,61	0,61	0,57	0,50
	AUC	0,00	0,06	0,08	0,09	0,04	0,12	0,03	0,05	0,05	0,02	0,01	0,01	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00

Source: Author's development

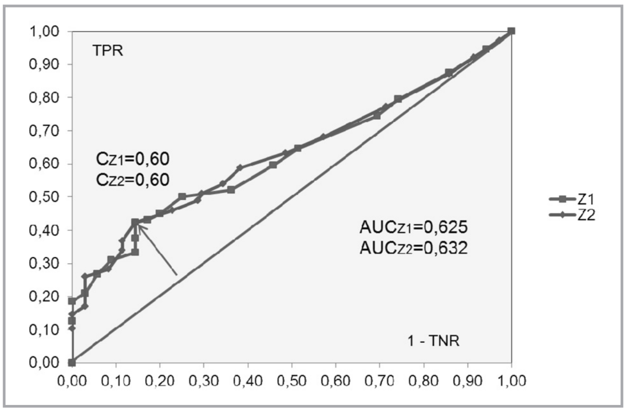

Fig. 1: **The constructed ROC curves for both models**
Source: Author's development

Thus, the research shows the possibility of application of the ROC curve analysis in solving practical problems of credit risk and predictive capability estimation. The area for the further research can be, first of all, consideration of the ROC curve analysis in terms of the economic indicators, for example, the economic benefits and losses from the true and false classified credit applications. This fact takes into account the results of the bank financial activity. Secondly, it is very important to discuss the influence of the adjustable model parameters on AUC that will provide sufficient ground for recommendations how to configure classifiers with the best predictive capability. Finally, the priority task for the future research is to develop approach of the ROC curve analysis application for the situation of more than two classes of the borrowers.

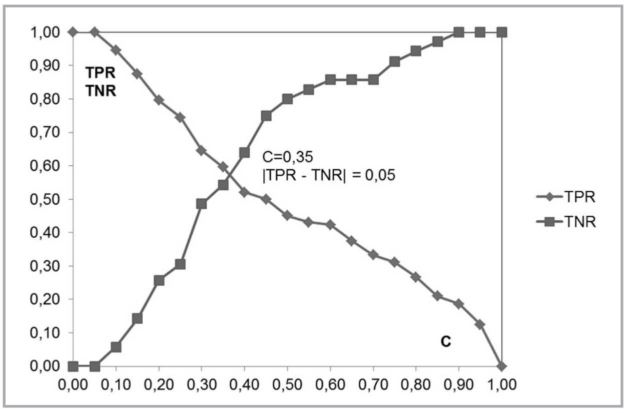

Fig. 2: **The balance between sensitivity and specificity for the model Z_1**
Source: Author's development

References

1. 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.
2. Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: a review of the literature. *Intelligent Systems in Accounting, Finance and Management, 18*(2), 59-88.
3. 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.
4. Yingxu, Y. (2007). Adaptive credit scoring with kernel learning methods. *European Journal of Operational Research, 183*(3), 1521-1536.
5. Yap, B. W., Ong, S. H., & Husain, N. H. M. (2011). Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems with Applications, 38*(10), 13274-13283.
6. Tsai, C.-F., & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications, 34*(4), 2639-2649.
7. Hand, D. J., & Till, R. J. (2001). A simple generalization of the area under the ROC curve to multiple class classification problems. *Machine Learning, 45*(2), 171-186.
8. Ben-David, A., & Frank, E. (2009). Accuracy of machine learning models versus «hand crafted» expert systems – A credit scoring case study. *Expert Systems with Applications, 36*(3, Part 1), 5264-5271.
9. Seidenfeld, T. (1985). Calibration, coherence, and scoring rules. *Philosophy of Science, 52,* 274-294.
10. Egan, J. P. (1975). *Signal detection theory and ROC analysis.* New York: Acad. Press.
11. Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240,* 1285-1292.
12. Drummond, C., & Holte, R. (2004). What ROC curves can't do (and cost curves can). *Proceedings of ROC Analysis in Artificial Intelligence* (pp. 19-26). Valencia, Spain.
13. Stein, R. M. (2005). The relationship between default prediction and lending profits: Integrating roc analysis and loan pricing. *Journal of Banking and Finance, 29,* 1213-1236.
14. Fawcett, T. (2004). *ROC Graphs: Notes and Practical Considerations for Researchers.* Kluwer Academic Publishers, Netherlands. Retrieved from http://home.comcast.net/tom.fawcett/public_html/papers/ROC101.pdf
15. Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3,* 32-35.

Received 24.03.2014