International Journal of Finance and Banking Research
2022; 8(1): 39-47
http://www.sciencepublishinggroup.com/j/ijfbr
doi: 10.11648/j.ijfbr.20220801.16
ISSN: 2472-226X (Print); ISSN: 2472-2278 (Online)

Science Publishing Group

The Paradox of Credit Scoring Model Deterioration

Mykyta Voloshyn

Department of Risk-management, Microfinance Organization "ForzaCredit", Kyiv, Ukraine

Email address:
mivoloshyn@gmail.com

To cite this article:
Mykyta Voloshyn. The Paradox of Credit Scoring Model Deterioration. *International Journal of Finance and Banking Research*. Vol. 8, No. 1, 2022, pp. 39-47. doi: 10.11648/j.ijfbr.20220801.16

Received: January 8, 2022; **Accepted**: January 27, 2022; **Published**: February 9, 2022

Abstract: Scoring models are widely renowned and used in financial organizations in a variety of fields, but most importantly – to predict and control credit risk. This article addresses practical problems of scorecards usage after its implementation. With time its predictive power tends to deteriorate, but not always it is applicable to completely rebuild the model fast enough due to lack of time/human/financial resources. Then cut-off, which was set when the scorecard was initially implemented, should be corrected to achieve optimal (i.e. cash-flow maximizing) performance. The literature on ways to maintain the existing model over time and manage the cut-offs is extremely scarce. The article is built on simple yet fundamental analytical explanations of scorecard performance dynamics, derived from practical experience. Results are backed by a numerical example, which shows the efficiency of different managerial decisions regarding cut-off setting in the paradox zone. The main conclusions are the following. In the most common case, the optimal reaction on model deterioration would be to counterintuitively narrow down the reject zone via cut-offs, which results in higher sales amount and even more increased risk ratios, but maximizes cash-flow in given conditions. This is the core of the scoring model deterioration paradox. It arises from the fact that when the scorecard deteriorates the high-risk segments of clients are actually becoming less risky and hence more profitable. This affects cut-offs, which must be applied to reject the riskiest loss-making segment of loan applications.

Keywords: Credit Scoring, Scoring Model, Scorecard Lifecycle, Model Risk, Cut-offs

1. Introduction

Scoring models are widely renowned and successfully used in financial organizations in a variety of fields. Most importantly – to predict and control credit risk, which is one of the key profitability drivers.

Literature on the topic of scorecard development and usage is elaborated in detail and represented by Lewis [8], Siddiqi [12], Anderson [1], Thomas [13], and other authors.

In the most common case, the scoring model takes a number of parameters derived from the client`s loan application, financial behavior, and external data sources (for example, number of closed loans according to credit history or level of client`s monthly income) and transforms it mathematically to the probability of default (PD) or credit score.

Deterioration of scorecard predictive power (discriminative ability) is a recognized problem that can diminish the efficiency of this risk-management tool. The article by Tikhonov, Masyutin, Anpilogov describes destructive consequences of such model dynamics [14]. Kelly states that once a scoring system is installed, its performance is likely to deteriorate until such a time that replacement is deemed necessary [5]. In practice, the lending organization seeks maximal performance of scoring models. But development and implementation of a new scorecard can take both time, human and financial resources. In such case, the period arises during which the organization has to operate the deteriorated model.

Lewis states that most organizations set minimum score levels at which they are willing to accept applicants, which is referred to as a "cut-off," and can represent a threshold risk, profit, or some other level, depending on the organization's objectives in using the scorecard [8].

This article will be focused on questions of accurate choice of cut-off level. In particular, setting cutoff when credit scorecard loses its predictive power.

In literature about cut-off setting most attention is paid to setting it for the new model just after the first implementation. Works on the topic of ways to maintain the existing model over time and manage the cut-offs are extremely scarce.

Siddiqi examines the situation when there appears a disparity between expected and predicted performance of scorecard. In such case three actions can be taken to resolve the problem: change cut-offs, change policy rules, and rebuild/refresh the scorecard [12].

The work of Jung, Thomas, and Mee Chi So is the closest to the topic of the current article [4]. The authors emphasize that lenders do adjust their cut-offs but normally in the light of policy decisions or because they believe the scorecard is being used on a new population. But they should adjust the cut-offs regularly to deal with the dynamics of the scorecard irrespective of any changes in policy or populations. In the article, the theory of replacement and maintenance of industrial equipment and complex systems is proposed to be applied for scorecards. The mathematical model is built to answer the questions of when to rebuild the model, when and how to readjust cut-offs to achieve maximal profit. The current article will consider the problem from a slightly different point of view, from which there occur some new unobvious conclusions.

2. The Paradox

As an illustrative example consider the credit default scoring. It aims to predict the probability of default (PD), i.e. the probability of credit becoming "Bad". Every company can have its own definition of "Bad" or "Default" loan. For instance, in microcredit organizations such definition of default is often used: the first payment is fully or partially overdue on 30 days or more, and at the same time the total amount of payments is lower than the disbursement amount. So, the scorecard helps to control actual BadRate (the ratio between number of "Bad" and all loans) by anticipating it with PD and cutting off the riskiest operations (i.e. rejecting the riskiest loan applications).

The ideal scoring model distinguishes "Bad" and "Good" clients absolutely accurately, so BadRate is equal to PD (prediction of BadRate).

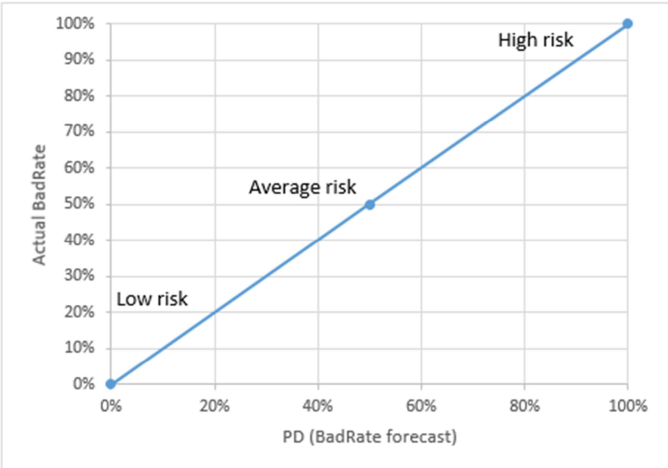

Figure 1. *Ideal model.*

Without loss of generality, this plot can be built between BadRate and credit score. Then dependency between these two parameters will be inverted since higher credit score depicts lower BadRate.

The predictive power of the scorecard is usually measured by Gini coefficient. For example, by Thomas [13] and Tikhonov, Masyutin, Anpilogov [14]. If such model would classify all clients only into two categories (PD=0% and 100%) and classification appear to be totally correct (BadRate=PD), then Gini coefficient would be equal to 100%. For this explanation, let the distribution of clients across the PD scale be continuously uniform. Hence, Gini will become less than 100% but will still be extremely high.

The next question is how to determine the cut-off PD which will serve as a boundary between approved and rejected applications. Anderson states that there are three approaches to the choice of cut-off [1].

1) To fix the aim value of BadRate and reach it through correcting cut-off and hence ApproveRate (ratio between number of approved applications to all processed loan applications). When the new better scorecard is introduced this means reaching the same BadRate and increasing ApproveRate.
2) To fix the aim value of ApproveRate. When the new better scorecard is introduced this means reaching the same ApproveRate and decreasing BadRate.
3) Ideally, lenders should try to maximize profit. Lewis was the first to highlight the obvious approach of setting the cut-off to the lowest score with a contribution greater than or equal to zero, which implies accepting any account that provides a profit [8].

Jung, Thomas, and So also state that profitability of the portfolio should be the dominant objective of a lender [4].

So, the most sophisticated (and obvious) approach implies maximization of profit or net cash-flow. Cut-offs derived from this strategy will be optimal. Under such a strategy scorecard becomes a tool that helps to process the incoming flow of applications (reject a part of it) in such a way, that maximizes net cash-flow. That makes perfect sense if the final high-level purpose of an organization is earning profit (accountant or cash). Note, the organization might need to control not only the ability to generate profit in a long-term perspective, but also limit the average time of achieving the break-even.

Cash-flow depends on multiple factors: interest rate, loan term (including tendency of clients to early repayment), penalties (including tendency to overdue), loan amount, etc. But the most important factor is BadRate.

Determining the dependency between Cash-flow and BadRate helps to derive the maximal acceptable BadRate which leads to generation of zero cash-flow. This task can be performed via mathematical modeling of all contributing factors or via statistical analysis. We prefer the second due to its simplicity. The example of methodology is the following. The data sample is formed which consists of a sufficient number of matured loans (in rows) and net cash-flow, default flag (0 for "Good" and 1 for "Bad"), PD, or another risk indicator – in columns. Net cash-flow can be calculated as the sum of all payments made by the client on a concrete loan minus disbursement amount minus costs that occur after a credit decision. A risk indicator (PD) is used to cluster the sample on segments with different BadRate. Then for each cluster average net cash-flow is calculated and BadRate-CashFlow chart is plotted. Usually, this dependency is described by a concave downward curve. According to Anderson, the profit increases initially as risk improves, but reaches a peak and then decreases, eventually becoming a loss. In most situations, the graph turns down at low-risk levels but does not go negative [1].

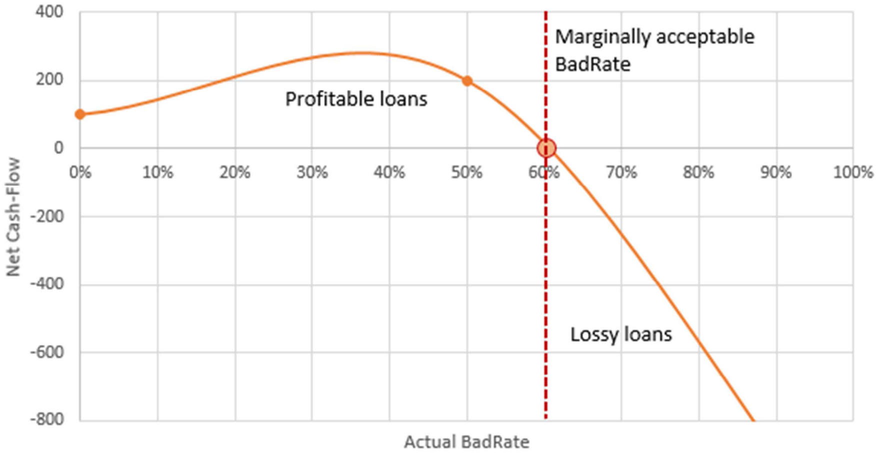

Figure 2. *Finding maximal acceptable BadRate.*

Segments with the lowest risk are usually not very profitable because they tend to lower loan amounts and terms. For segments with high risk credit losses start to overweight the incomes and hence the whole segment eventually becomes lossy.

For the given example the break-even point is BadRate=60%. Lower BadRate corresponds to profitable segments of loans and higher – to lossy.

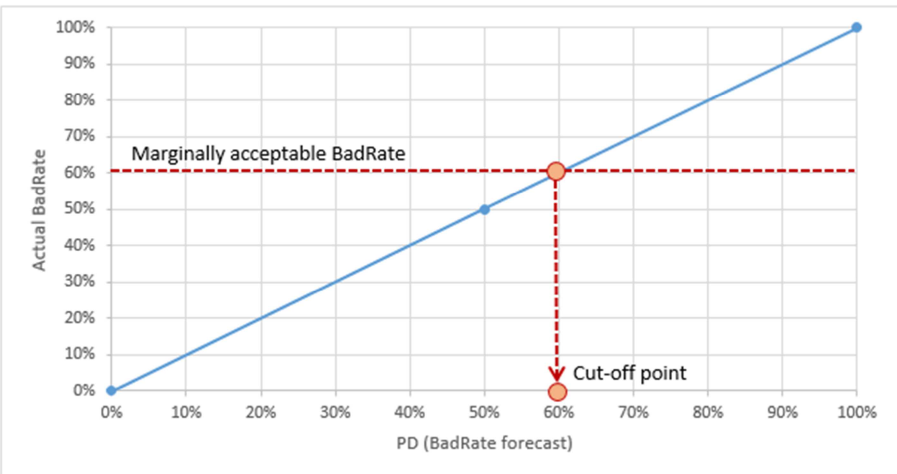

Figure 3. *Cut-off for cash-flow maximization for ideal model.*

After applying this marginally acceptable BadRate to the ideal model, the optimal cut-off level of PD=60% is chosen. All loss-making applications are rejected.

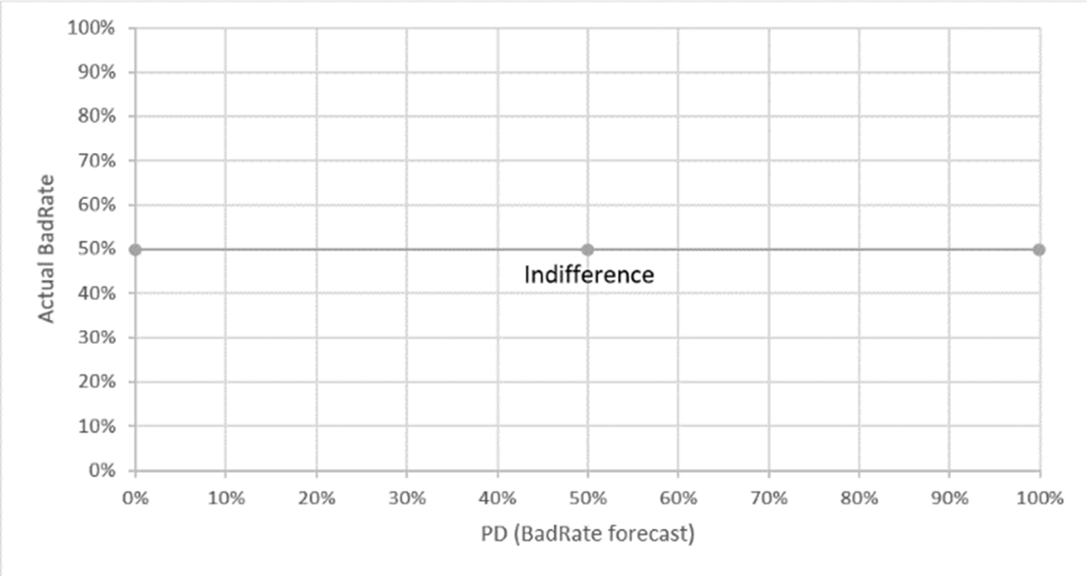

Figure 4. *A model with zero Gini (random model).*

If the ideally powerful model is also ideally stable, it continues to work and efficiently maximize the net cash-flow. But what happens if its predictive power deteriorates? The PD-BadRate chart flattens. In the extreme case, Gini drops to zero which means that scorecard is not distinguishing between "Bad" and "Good" clients at all. Across all PD segments, only one BadRate will be observed – the average BadRate characteristic for the incoming flow of clients (applications).

According to previous calculations, the marginally acceptable BadRate is 60% while the average BadRate is 50%. Hence, the strategy that maximizes net cash-flow would be approving all applications. Of course, the cash-flow earned will be significantly lower in comparison to utilization of the ideal model. But since the organization does not have it anymore, the best decision would be 100% ApproveRate.

If marginally acceptable BadRate would be lower than average BadRate, the organization cannot reach profits without a powerful scorecard and will need to cease any disbursements. But in practice, this situation is rare since scorecard is processing applications that passed policy rules designed to cut-off most of the loss-generating segments.

Next, assume that the ideal model steadily deteriorates over time through the average model to zero-Gini model.

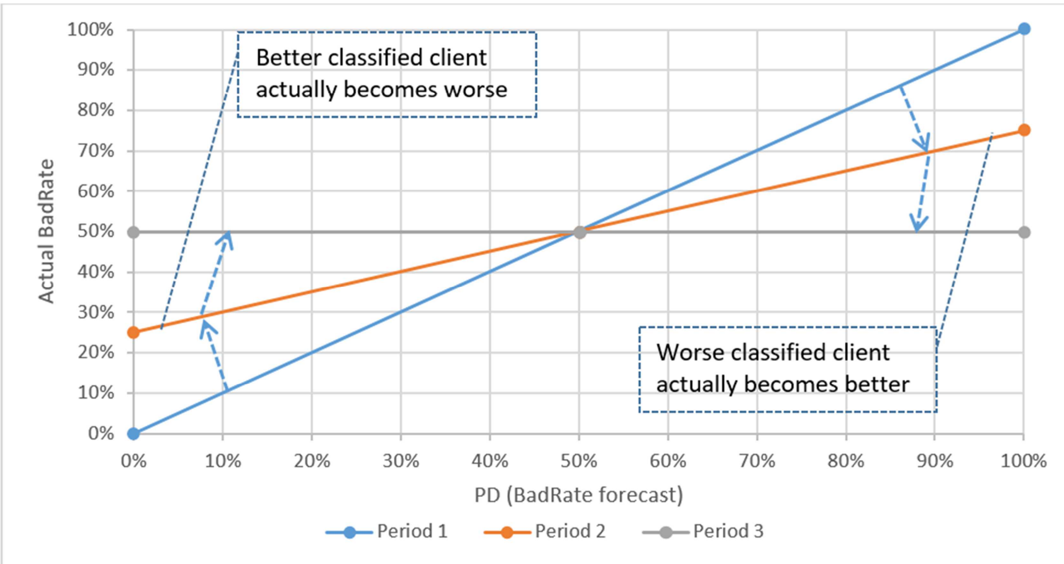

Figure 5. *Deterioration of model.*

The lines rotate around the average BadRate which equals 50% in this example. The gradual flattening of PD-BadRate dependency is observed. As a result, high-risk clients which were rejected by the ideal model actually become better from both risk and cash-generating perspective. The loss-making segment is shrinking. According to the cash-maximizing strategy, in the second and third periods, cut-offs must be adapted to changes in the predictive power of the model.

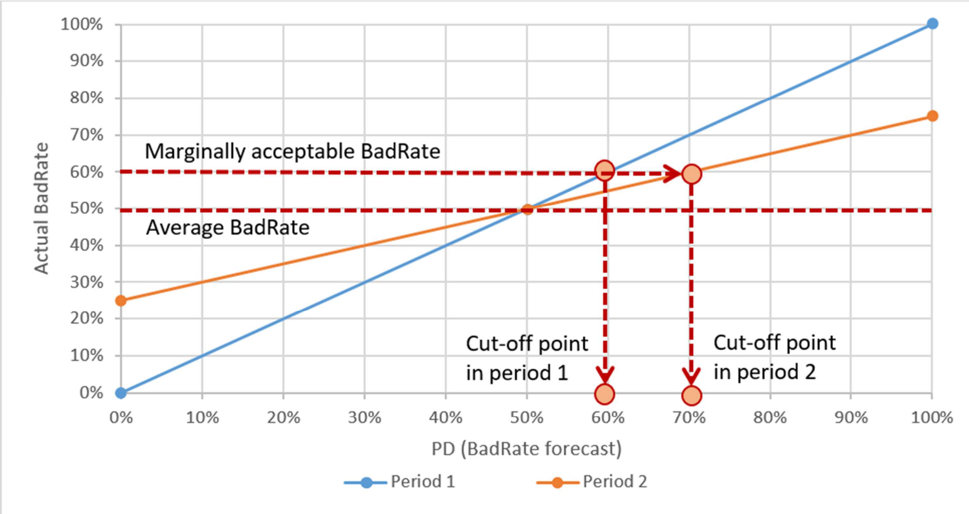

Figure 6. *Cash- flow maximizing cutoffs for ideal and average models (average BadRate is lower than marginally acceptable).*

Due to deterioration of predictive power of model and decrease of BadRate in high-risk segments the optimal cut-off which corresponds to acceptable BadRate migrates from 60% for the ideal model to 70% for the average one.

This is the core of the paradox. When predictive power of model deteriorates and hence average BadRate of disbursed loans increases, the optimal cash-maximizing decision would be increasing of cut-off PD, increasing disbursement amount, and even more increasing BadRate. In other words, the lower is the ability to predict the risk, the higher will be the ApproveRate and sales amount.

This conclusion is valid when marginally acceptable BadRate is higher than average BadRate. Which in most practical cases is true. If marginally acceptable BadRate would be lower than average, then results will be the opposite.

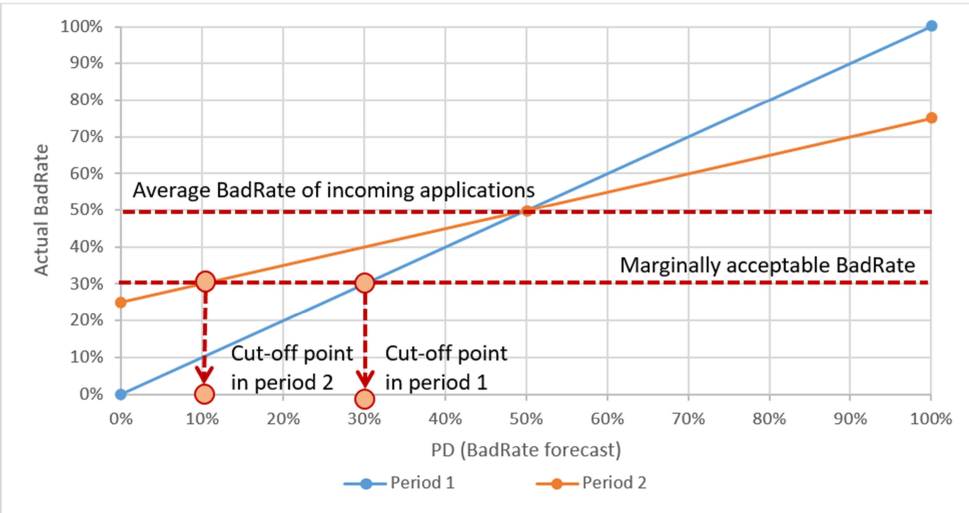

Figure 7. *Cash-flow maximizing cutoffs for ideal and average models (average BadRate is higher than marginally acceptable).*

Here average BadRate is still 50% and marginally acceptable BadRate is lowered to 30%. In such case, the optimal cut-off for the ideal model is 30% and for the average model, it decreases to 10%. Such dynamic is aligned with common sense and no paradox occur here.

As a sidenote. When overtime PD-BadRate chart stays the same and only distribution of clients across PD scale changes, the average BadRate also changes, as well as Gini, ApproveRate, and, subsequently, cash-flow. But optimal cut-off will remain constant. Cut-off will need to be corrected only if it`s derived not from cash-maximization strategy, but from the strategy of retaining certain fixed levels of BadRate or ApproveRate.

3. Practical Consequences

The considered hypothetical example is only for explanatory purposes. In practice, figures are not so exaggerated. Neither ideal models nor worthless zero-Gini models are used in banks and other lending organizations.

Also, the distribution of clients across PD scale is not uniform but close to lognormal, exponential, or other. The distribution of clients and predictive power of scorecard can change simultaneously and affect the results. The predictive power of scorecard can change asymmetrically. For example, deteriorate for low-risk clients only. But such nuances do not diminish the practical importance of obtained conclusions.

Once the new scorecard is implemented, its predictive power usually does not tend to increase. At least it is not intended. Apart from short-term volatility, in perspective, it can only gradually deteriorate (slower or faster). In an ideal situation when any slight deterioration or opportunity to enhancement is detected, the model must be rebuilt and implemented. But in practice, there is a window between Gini of the newly built model and marginal Gini when the scorecard needs to be replaced. According to the life-cycle of the scoring model, it must be rebuilt only if such action is profitable. Since development and implementation have their own cost, the expected benefit of the new scorecard must exceed it. More information on the life-cycle of scoring models can be found in Jung, Thomas, and So [4].

Also, development and implementation of a new scorecard demand a considerable amount of time, which can reach up to a month or more, depending on the technological and organizational level of the company. This time might include hovering while freshly-disbursed loans will get matured, according to Siddiqi [12]. During this period between the decision to rebuild the model and its actual implementation, the old model will be inevitably utilized. To maximize cash-flow, cut-offs must be corrected in the described way.

The first step is to assess the BadRate of the whole incoming flow of applications without applying any cut-offs. It can be done through extrapolation of BadRates onto the given distribution of applications across the PD scale. If the average BadRate is lower than marginally acceptable, then the organization is in a paradox zone.

The second step is implementation of the new scorecard. A better model gives off the lower ApproveRate since it is capable of isolating the wider segment of loss-making clients. So the implementation of the new model will harm sales amount. It is counterintuitive to say that the better the organization “understands” the risk profile of incoming clients, the less it can disburse. But the more it will earn.

Additionally, if the goal of the organization is not maximal efficiency of disbursements, but fast expansion and seizing a substantial market share, maybe it doesn’t need a powerful scorecard that limits sales. For example, it can buy a minimal amount of external data from credit bureaus or other companies, having a positive effect on decreasing costs and maintaining a high level of sales.

The third step is monitoring and updating cut-off when the predictive power of the existing scorecard deteriorates. Due to that, even if cut-off remains the same, BadRate will increase. Despite that, to maximize cash-flow, cut-off must be moved up towards even higher BadRate. Such a decision is also highly counterintuitive. In practice, it can even be the case when a weak model cannot distinguish any lossy segment and nothing is left except approving all applications. It maximizes sales amount and cash-flow in given conditions. It is the best way to handle the worst situation in which cash-flow is the lowest. But it could be much lower if the risk manager will try to hold the increase of BadRate by applying stricter and stricter cut-off. Marginally it can lead to cessation of any disbursements.

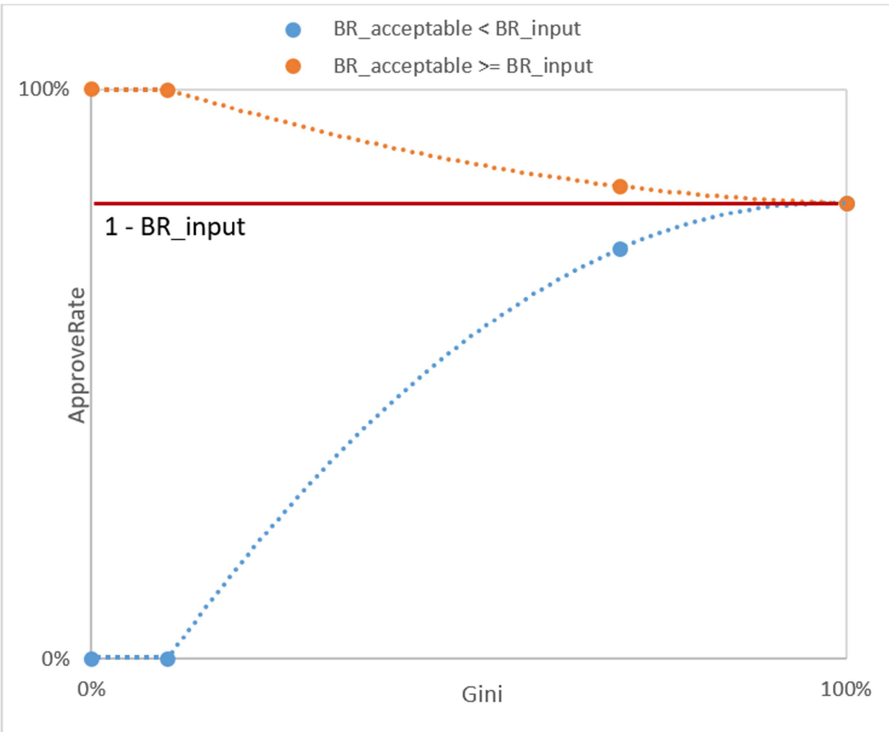

Figure 8. *Dependency between ApproveRate and Gini of the scorecard under cash-flow maximizing strategy.*

Of course, if the scorecard deteriorates as well as cash-generating ability, the management can decide to wait until renewed scorecard will be up and running. That will lead to a more efficient allocation of funds in the future. But it must be a conscious decision to the detriment of the strategy of cash-flow maximization.

The fourth step is to implement a new model when cost-benefit analysis shows that it will be an efficient action. Then predictive power of the utilized model grows again, sales amount drops, cash-flow increases.

As a consequence, when the scorecard deteriorates, cut-off (and hence ApproveRate) must be changed to fit the cash-flow maximizing strategy. For the ideal model, ApproveRate will be equal to 1 minus BadRate of the incoming flow of applications since such model is capable of distinguishing and cutting off all “Bad” loans. When the scorecard starts to deteriorate, ApproveRate will grow (when marginally acceptable BadRate is higher than BadRate of the incoming flow of applications) or decline (when marginally acceptable BadRate is lower than BadRate of incoming flow of applications). Eventually, it will lead to approval of all loan applications or rejecting them all.

As a sidenote. According to Thomas, increasing cut-off can lead to an increase of predictive power and Gini coefficient and vice versa [13]. It occurs because of widening of PD range, on which scorecard is allowed to perform. This dependency does not lead to any further changes of cut-off but helps to understand the consequences of such managerial decision.

These conclusions are quite universal. They are valid not only for credit default scorecards but also for scorecards with other targets (profit-scoring, collection scoring, retention scoring, anti-fraud, and any other), for all types of clients (corporate, SME, retail), for all types of products (payday loans, installments, credit cards, mortgage and other), for all types of scorecards which allow determining probability of forecasted event or score (linear regression, logistic regression, decision trees, random forest, gradient boosting, neural networks and other).

4. Numerical Example (Case Study)

The methodology of the numerical study is the following. First, the simulation of the incoming flow of applications is conducted. Three abstract time periods are considered, though which the hypothetical scorecard deteriorates. In each period 1000 loan applications are simulated. For each time period, the distribution of applications across the PD scale is the following.

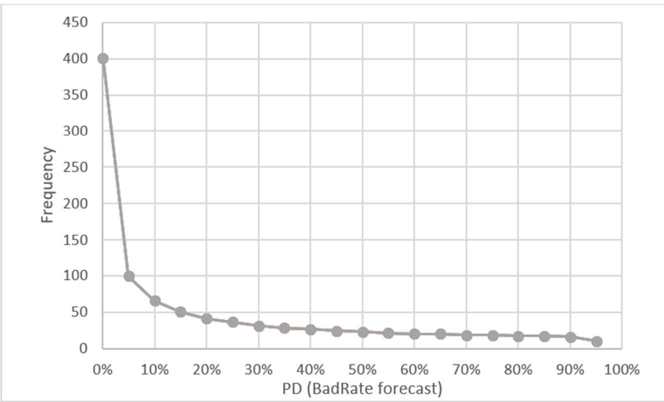

Figure 9. *Distribution of applications across PD scale.*

Then the outcome is simulated in the “IsBad” field, where 0 stands for “Good” and 1 for “Bad” loan.

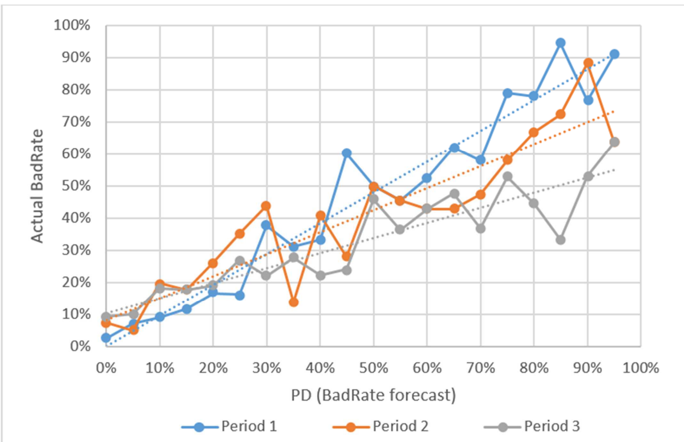

Figure 10. *PD-BadRate charts for three periods.*

Gini of the scorecard deteriorates in three periods and amounts to 77%, 58%, and 45% respectively.

Then cash-flow is introduced into calculations. For illustrative purposes, +700 conventional units for “Good” loans and -1000 for “Bad” are utilized.

As a result, the table with three columns (Period, PD, IsBad, CashFlow) and 3000 rows is formed (thousand for each time period) as input data. Then results are grouped across PD ranges and periods.

Table 1. *Cash-flow for each PD subsegment of applications.*

PD	Period 1	Period 2	Period 3
[0%; 5%)	263 700	231 400	217 800
[5%; 10%)	58 100	61 500	53 000
[10%; 15%)	36 000	24 100	25 800
[15%; 20%)	25 500	20 400	20 400
[20%; 25%)	17 500	10 700	15 800
[25%; 30%)	15 700	3 800	8 900
[30%; 35%)	2 000	-1 400	10 500
[35%; 40%)	5 000	13 500	6 700
[40%; 45%)	3 600	200	8 700
[45%; 50%)	-8 000	5 600	7 300
[50%; 55%)	-3 600	-3 600	-1 900
[55%; 60%)	-1 600	-1 600	1 800
[60%; 65%)	-4 000	-600	-600
[65%; 70%)	-7 400	-600	-2 300
[70%; 75%)	-5 400	-2 000	1 400
[75%; 80%)	-12 200	-5 400	-3 700
[80%; 85%)	-11 200	-7 800	-1 000
[85%; 90%)	-16 300	-9 500	2 400
[90%; 95%)	-10 200	-13 600	-3 400
[95%; 100%)	-9 300	-4 200	-4 200

In the first period, the optimal cut-off is applied (45%). In the second and third periods the risk manager has three options:

1) to recalculate cut-offs in each period based on cash-flow maximizing strategy;
2) to keep cut-off on the constant level of 45%;
3) to lower cut-off to keep BadRate on a constant level (9%).

The resulting cash-flow under these three strategies will be the following.

Table 2. *Cash-flow for each strategy.*

#	Strategy	Period 1	Period 2	Period 3
1	CashFlow → max	427 100	369 800	374 900
2	Cut-off → const	427 100	364 200	367 600
3	BadRate → const	427 100	337 400	217 800

So, the cash-flow maximizing strategy is capable of generating extra 5 600 conventional units in period 2 (+2%) and 7 300 in period 3 (+2%) in comparison to constant cut-off strategy and extra 32 400 in period 2 (+9%) and 157 100 in period 3 (+42%) in comparison to constant BadRate strategy. These results are reached by applying different cut-offs.

Table 3. *Cut-offs for each strategy.*

#	Strategy	Period 1	Period 2	Period 3
1	CashFlow → max	45%	50%	50%
2	Cut-off → const	45%	45%	45%
3	BadRate → const	45%	20%	5%

5. Conclusions

In the article, it is shown that an intuitive and logical decision to tighten the cut-offs down when the scorecard deteriorates can lead to a decrease in the cash-generating ability of a lending organization. First, the organization should assess the BadRate of the incoming flow of loan applications (after applying policy rules) on credit scorecard. If this BadRate is lower than marginally acceptable (i.e. the firm can earn profit without applying any scorecard), then the organization is in a paradox zone. In most practical cases it will be true. Otherwise, it would be quite a harsh business environment, in which the overall profitability and survival of a firm highly depend on the predictive power of scorecards which tend to deteriorate.

The paradox of credit scorecard deterioration can be stated as follows. When the predictive power of the model deteriorates and hence the average BadRate of disbursed loans increases, the optimal cash-maximizing and profit-maximizing decision would be increasing of cut-off PD, increasing disbursement amount, and even more increasing BadRate. At least, the risk manager should not try to compensate for the deterioration of risk ratios by decreasing cut-offs. That will have a negative impact on cash-flow.

The paradox arises from the fact that when the scorecard deteriorates the high-risk segments of clients are actually becoming less risky and hence more profitable. This is important since cut-off is applied for the riskiest segment to reject the loss-making loan applications.

The marginal cases are examined. An absolutely ideal scoring model with Gini=1 will lead to a maximal level of rejects which is equal to BadRate. Since the model will know all Bads and cut them off efficiently. A random model with Gini=0 cannot separate any lossy segment of clients so it cannot justify any rejects of applications. Real-life models will be somewhere in the middle.

The numerical example of scorecard deterioration over three time periods is considered. The consequences of applying different strategies are shown. Fixating the BadRate with lower cut-off results in the lowest cash-flow and rejection of profitable segments of clients.

References

[1] Anderson, Raymond. 2007. The credit scoring toolkit, theory, and practice for retail credit risk. Management and decision automation, Oxford University Press, Oxford.

[2] Board of Governors of the Federal Reserve System, 2011. Supervisory Guidance on Model Risk Management: SR Letter 11-7 Attachment.

[3] Jung K. M., Thomas L. C., Mee Chi So. 2013. Time varying or static cut-offs for credit scorecards. Journal of the Operational Research Society, Palgrave Macmillan; The OR Society, vol. 64 (9), pages 1299-1306, September.

[4] Jung K. M., Thomas L. C., Mee Chi So. 2015. When to rebuild or when to adjust scorecards [Link].

[5] Kelly, Mark Gerard. 1998. Tackling change and uncertainty in credit scoring. PhD thesis The Open University.

[6] Kritzinger, Nico. 2017. Improving credit risk measurement and management: A new application of statistical techniques. Thesis submitted for the degree Philosophiae Doctor in Risk Analysis at the Potchefstroom Campus of the North-West University. October.

[7] Kritzinger, N. & Van Vuuren, G. W., 2018. An optimized credit scorecard to enhance cut-off score determination. South African Journal of Economic and Management Sciences 21 (1), a1571. https://doi.org/10.4102/sajems.v21i1.1571.

[8] Lewis E. M. 1992. An Introduction to Credit Scoring, 2nd Ed. Athena Press: San Rafael, CA.

[9] López-Ratón M., Rodríguez-Álvarez M., Cadarso-Suárez C., Gude-Sampedro F., 2014. OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. Journal of Statistical Software. October 2014, Volume 61, Issue 8.

[10] Madzova Violeta, Ramadini Nehat. 2013. Can credit scoring models prevent default payments in the banking industry in the period of financial crisis? International Journal of Business and Technology. Volume 2, issue 1, article 5. November.

[11] McNab H., Wynn A., 2000. Principles and Practice of Consumer Credit Risk Management. CIB Publishing, Canterbury.

[12] Siddiqi, Naeem. 2006. Credit risk scorecards, developing and implementing intelligent credit scoring, John Wiley & Sons, Inc. Hoboken, NJ.

[13] Thomas, L. C., 2009. Consumer Credit Models: Pricing, Profit, and Portfolios Oxford University Press. Oxford, UK.

[14] Tikhonov Roman, Masyutin Aleksey, Anpilogov Vadim. 2019. Relationship between Credit risk management and financial performance: empirical evidence from microfinance banks in Kenya. Research Journal of Finance and Accounting. Vol. 7, No. 6.

[15] Verstraeten G., D. Van den Poel. 2003. Quantifying credit-scoring performance. Transactions on Information and Communications Technologies. Volume 29, WIT Press.