Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 242 (2024) 348–355

Procedia Computer Science

www.elsevier.com/locate/procedia

11th International Conference on Information Technology and Quantitative Management (ITQM 2024)

Parametric methods for precision calibration of scoring models

Mikhail Pomazanov[a], Andrey Berezhnoy[b*]

[a] *School of Finance, Faculty of Economic Sciences, National Research University "Higher School of Economics", Moscow, Russia*
[b]*Moscow State University named after Lomonosov, faculty of Mechanics and Mathematics*

Abstract

Once a scoring model has been developed for use in assessing a borrower's credit risk, under the internal ratings-based (IRB) approach, it must be calibrated to a real-world measure of default frequency. The conservativeness of the calibration is tightly controlled, if it is not violated in the allowed number of digits of the rating scale, only then the model is authorized for use. If violated, the calibration probability of default must be raised, placing additional strain on the bank's capital and reserves. In the presented paper, two new methods to improve the calibration accuracy are proposed. The methods have been tested in practice and provide significantly positive results in certain segments of the loan portfolio.

Keywords: calibration; ROC curve; scoring; Gini index; discrimination power; probability of default

1. Introduction

Supposed there is a scoring model (for example, constructs by methods [9]) for determining the rating score s such that:

- the central tendency (CT) of the default rate was statistically calculated,
- the Gini coefficient (AR) and also second-order accuracy metrics as LAR and RAR [9] have been measured on default and non-default samples.

There is required to construct a calibration function $PD(s,\vec{A}) = p_A(s)$ with corresponding to CT and accuracy metrics, where $\vec{A}$ is the required vector of parameters. There are many approaches to solving the problem of calibration function selection, the best known of which are described in the sources [12],[13]. There are several approaches in the literature that are quite difficult to implement in practice [1],[5]. We propose two new approaches for selecting a calibration function that have proven themselves in practice.

* Corresponding author. E-mail address: beradz@yandex.com

10.1016/j.procs.2024.08.175

Let there are n measurements of the score x_i for $i = 1, \dots, n$ representative to the full calibration sample.

PD, first and second-order accuracy metrics are calculated:

$$D_k(\vec{A}) = \sum_{i=1}^{k} p_A(x_i), \widehat{PD}(\vec{A}) = \frac{D_n(\vec{A})}{n}, \widehat{AR}(\vec{A}) = \frac{2}{\left(n - D_n(\vec{A})\right) \cdot D_n(\vec{A})} \cdot \sum_{k=1}^{n} D_k(\vec{A}) \cdot \left(1 - p_A(x_k)\right) - 1,$$

$$L\widehat{AR}(\vec{A}) = \frac{2}{n - D_n(\vec{A})} \cdot \sum_{k=1}^{n} \frac{1 - p_A(x_k)}{\left(k - D_k(\vec{A})\right) \cdot D_k(\vec{A})} \sum_{s=1}^{k} D_s(\vec{A}) \cdot \left(1 - p_A(x_s)\right) - 1, \quad (1)$$

$$\widehat{RAR}(\vec{A}) = 1 - \frac{2}{D_n(\vec{A})} \cdot \sum_{k=1}^{n} \left(D_n(\vec{A}) - D_k(\vec{A})\right) \cdot \left(1 - p_A(x_k)\right) \cdot \sum_{s=1}^{k} \frac{p_A(x_s)}{\left(n - D_n(\vec{A}) - \left(s - D_s(\vec{A})\right) + \frac{1}{2}\left(1 - p_A(x_s)\right)\right) \cdot \left(\left(D_n(\vec{A}) - D_s(\vec{A})\right) + \frac{1}{2} p_A(x_s)\right)}.$$

A logistic calibration is determined using the following function:

$$\frac{1}{1 + e^{s \cdot A + B}} =: PD(s, A, B),$$

where $\vec{A} = A, B$ are parameters of the calibration selected based on the equalities:

$$\widehat{PD}(A, B) = CT, \quad \widehat{AR}(A, B) = AR,$$

where CT and AR are the central tendency and the Gini index respectively which the model is configured, $\widehat{PD}(A, B), \widehat{AR}(A, B)$ the formula (1).

In practice, to find the calibration parameters (A and B), it is enough to minimize the value of the following functional:

$$F(A, B) = \frac{\left(\widehat{PD}(A, B) - CT\right)^2}{\sigma_{PD}^2} + \frac{\left(\widehat{AR}(A, B) - AR\right)^2}{\sigma_{AR}^2} \to \min_{A,B} F,$$

where σ_{PD} and σ_{AR} [4] are statistical measurement errors (standard deviations) of the probability of default and the Gini index respectively,

$$\sigma_{PD} := \sqrt{\frac{CT \cdot (1 - CT)}{n}}, \sigma_{AR} := \sqrt{\frac{1 - AR^2 + (n \cdot CT - 1) \cdot \frac{(1 - AR)^2 \cdot (1 + AR)}{3 - AR} + (n \cdot (1 - CT) - 1) \cdot \frac{(1 + AR)^2 \cdot (1 - AR)}{3 + AR}}{n^2 \cdot CT \cdot (1 - CT)}} \quad (2)$$

A_0 and B_0 are selected as the initial approximation [7] of the minimization problem $F(A, B)$:

$$\hat{a} = AR\sqrt{\pi} \cdot exp\left(\frac{AR^2 \pi}{12} \cdot \left(1 + 6 \cdot PD \cdot exp\left(-\frac{AR^2 \pi}{2}\right)\right)\right), \qquad \hat{b} = -\ln PD + \frac{\hat{a}^2}{2} - PD \cdot exp(\hat{a}^2),$$

$$A_0 = \frac{\hat{a}}{ds}, \qquad B_0 = \hat{b} - \hat{a}\frac{\langle s \rangle}{ds}, \quad (3)$$

where $\langle s \rangle$ and ds are mean and standard deviation of scores s of the sample $x_i, i = 1, \dots, n$. The algorithm for minimizing the functional $F(A, B)$ stops working when the value $\boldsymbol{F(A, B) < 1}$ is reached, which guarantees that the estimating error of the arguments is less than the statistical "one sigma".

2. Binomial test analytics of logistics calibration

To check the rating scale (calibration), a mandatory binomial test [2] is carried out on the entire sample to validate the calibration. The binomial test is applied for each element of the researched section and validates the hypothesis H_0 about the assessment of the explained variable on this element of the researched section H_0: the target variable on the considered element of the researched section is underestimated. Alternative hypothesis H_1 for a one-sided test validates the alternative statement – at the given confidence level is impossible to state that the variable is underestimated.

The null hypothesis is rejected at the given significance level α, if the number of default samples k on the considered element of the researched section is less than or equal to the critical value k^*, which is defined below:

$$k_j^* := min\left\{k \left| \sum_{i=k}^{N_j} \binom{N_j}{i} \cdot PD_j^i \cdot \left(1 - PD_j\right)^{N_j - i} \le 1 - \alpha \right.\right\},$$

where N_j is number of samples in j-th rank of the rating scale. The critical value k^* can be approximated using the central limit theorem: $k_j^* := N_j \cdot PD_j + \Phi^{-1}(\alpha)\sqrt{N_j \cdot PD_j \cdot (1 - PD_j)}$, where Φ^{-1} is standard normal distribution.

The critical value of the number of defaulted borrowers (k^*) in each rank of the rating scale is calculated for the significance level α (95% and 99%). The obtained values (k^*) for the significance level α (95% and 99%) are compared with the number of defaulted borrowers (k) for each rank of the rating scale. Moreover, if $k \le k^*$ ($\alpha =$ 95%) then the rank is in "green zone", if $k > k^*$ ($\alpha = 99\%$) then the rank is in "red zone" (hypothesis H_0 about underestimation is correct at the significance level 99%), otherwise the rank is in "yellow zone".

Practical criteria for conservatism that determine the resulting zone are given in Table 1.

Table 1. Criteria for estimating the rating scale

	Red zone	Yellow zone	Green zone
Deviation of forecast PD from DR by ranks of the rating scale for corporate and retail borrowers	3 and more ranks are in "red zone" or 5 and more ranks are in "yellow zone" or "red zone"	Other cases	2 and less ranks are in "yellow zone" and other ranks are in "green zone"

To find for calibration parameters A and B that satisfy the binomial test, it makes sense to resort to visual analysis on the $[PD{:}AR]$ plane, since the transformation from A and B to PD and AR is one-to-one, but at the same time the calculated values of PD and AR (1) have a clear interpretation. To determine the correct PD level to satisfy the binomial test for conservativeness, the entire $[PD{:}AR]$ plane can be colored based on the test color at the appropriate PD and AR values. This will allow the analyst to visually determine the appropriate PD and AR values based on the requirements being considered (satisfying binomial or other test, as well as the set of calibration requirements). It is possible to determine the minimum PD value at a given AR level for which a requirement is satisfied, such as satisfying a binomial test for conservativeness with "green zone" or "yellow". Figure 1 provides examples of color-coded planes for the binomial test of conservatism for consumer and mortgage loans segments lending models for logistic calibration:

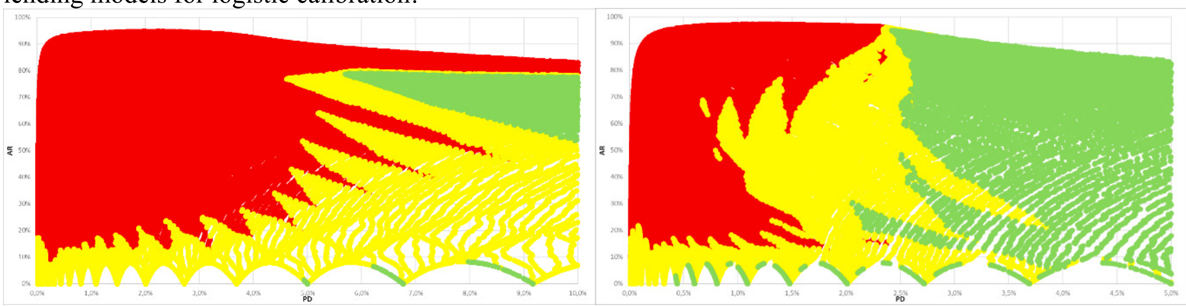

Figure 1. Coloring the $[PD{:}AR]$ plane by the color of the binomial test for conservatism for the scoring model in logistic calibration (left: for consumer loans segment, right: mortgage loans segment).

In practice, it is not necessary to color the entire plane, but rather some area of interest is sufficient, which can be realized if we consider the parallelogram $\hat{a} = 0, \dots, 6, \hat{b} = 0, \dots, 10$ in terms of formula (3), the coloring intensity or steps along the axes should be chosen based on the computing power and time available for calculation, for example, a uniform step equal to one hundredth/thousandth of a segment along each axis.

Practical examples explaining necessary to move to precision calibration: logistic calibration, statistical test is binomial test for conservatism, number of ranks of rating scale is "20+1". The results are shown in Table 2.

Table 2. Comparison of values of probability of default (Min PD) and target CT acceptable for passing the binomial test.

Model	Number of samples	AR	CT	Min PD with "green zone"	Min PD with "yellow zone"
Consumer loans segment	23 231 154	80,4%	4,76%	6,14%	5,79%
Mortgage loans segment	2 425 852	89,5%	0,48%	3,46%	2,95%

It can be seen that logistic calibration gives overestimated probability of default relative to CT in the calibration model

that satisfies the test passing criteria required for validation (comparative analysis). This will lead to increased capital/reserve requirements in the IRB.
Clearly, higher accuracy calibration models need to be explored in these segments.

3. Cubic logistic calibration

A cubic logistic calibration is a refinement of the usual logistic calibration by increasing the order of the polynomial in the argument of the logistic function. This calibration is applied using the function:

$$\frac{1}{1+e^{a_0+a_1\cdot s+a_2\cdot s^2+a_3\cdot s^3}} =: PD(s,\vec{A}),$$

where $\vec{A} := (a_0, a_1, a_2, a_3)$ are parameters of the calibration selected based on the following inequalities:

$$\left|\widehat{PD}(\vec{A}) - CT\right| < \sigma_{PD},\quad \left|\widehat{AR}(\vec{A}) - AR\right| < \sigma_{AR}, \left|\widehat{LAR}(\vec{A}) - LAR\right| < \sigma_{AR},\quad \left|\widehat{RAR}(\vec{A}) - RAR\right| < \sigma_{AR},$$

where CT is the central tendency for calibration of model, AR, LAR and RAR are the Gini index, left and right integral Gini indices (second-order accuracy metrics) respectively. σ_{PD} and σ_{AR} are estimated errors (standard deviations) of the probability of default and the Gini index respectively, calculated using formula (2), $\widehat{PD}(\vec{A}), \widehat{AR}(\vec{A}), \widehat{LAR}(\vec{A}), \widehat{RAR}(\vec{A})$ are defined in formula (1).
To find the calibration parameters $(\vec{A})$, it is enough to minimize the value of the following functional:

$$F(\vec{A}) = \frac{\left(\widehat{PD}(\vec{A}) - CT\right)^2}{\sigma_{PD}^2} + \frac{\left(\widehat{AR}(\vec{A}) - AR\right)^2}{\sigma_{AR}^2} + \frac{\left(\widehat{LAR}(\vec{A}) - LAR\right)^2}{\sigma_{AR}^2} + \frac{\left(\widehat{RAR}(\vec{A}) - RAR\right)^2}{\sigma_{AR}^2} \to \min_{\vec{A}} F.$$

There should select $\vec{A} := (B_0, A_0, 0,0)$ (A_0 and B_0 from formula (3)) as an initial approximation. There should stop the optimization process, just like for first-order logistic calibration, when $\mathrm{F}(\vec{\mathrm{A}}) < 1$ is reached.

After obtaining a solution, it must be checked to satisfy the obvious monotonicity requirement PD: $p_A(x)' \leq 0$, whence follows the condition for $f(\vec{A}, x) = a_1 + 2a_2 \cdot x + 3a_3 \cdot x^2 \geq 0$ which is equivalent to the conditions: $a_3 \geq 0, {a_2}^2 \leq 3a_1a_3$.

The main failure of the cubic logistic calibration, as well as first order (3), is a significant increase PD (up to one) for low rating points and a decrease (down to zero) for high ones. Not all scoring models have these discrimination properties as determined by their ROC curve. This leads, for example, to high expectations of a high ranking for high scores, which negatively affects the results of the binomial test of conservatism. The three-segment hyperbolic calibration model does not have this failure.

4. Three-segment hyperbolic calibration

First of all, accuracy metrics AR, LAR and RAR (1) are calculated for the scoring model. After, its trapezoidal approximation is constructed and parameters a and b are calculated for this trapezoid (Appendix 1). Then, you need to split the model's ROC curve into three segments: 1.$(0,0) - (a, R_a)$, 2. $(a, R_a) - (1-b, R_b)$, 3.$(1-b, R_b) - (1,1)$, where ($R_a := ROC(a)$ is the corresponding value of the ordinate axis for the value a of the abscissa axis, such that point (a, R_a) belongs to the ROC curve, $R_b := ROC(1-b)$ and similarly point $(1-b, R_b)$ belongs to the ROC curve) and for each segment there is necessary to calculate the Gini indices (the segment is scaled by the square $([0,1]:[0,1])$.

Figure 2 presents the geometric interpretation of the segments, the areas relative to which rectangles are presented to obtain the corresponding Gini indices for the segments (the Gini index will be equal to the ratio of the corresponding blue area to the half-area of the corresponding red rectangle).

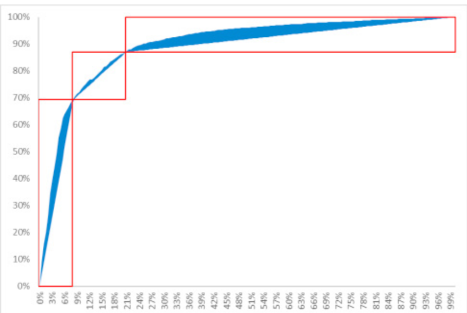

Figure 2. An example of a three-segment partition of an ROC curve. The blue background shows the areas of local ROC curves that determine the local Gini indices.

S_n and $\hat{S}_d$ are scores of non-default and default samples respectively for $n = 1, \ldots, G$ and $d = 1, \ldots, B$.

1. There is calculated trapezoidal triangulation: $\begin{cases} a, b, I := 1 - a - b \\ R_a, R_b \end{cases}$,
2. There are calculated local Gini indices of the segments: $AR_a := 2 \cdot AUC_a - 1$, $AR_I := 2 \cdot AUC_I - 1$, $AR_b := 2 \cdot AUC_b - 1$,
   where $AUC_a := \frac{1}{a \cdot G \cdot R_a \cdot B} \cdot \sum_{n=1}^{a \cdot G} \sum_{d=1}^{R_a \cdot B} \delta_{S_n}(\hat{S}_d)$, $AUC_I := \frac{1}{I \cdot G \cdot (R_b - R_a) \cdot B} \cdot \sum_{n=a \cdot G+1}^{(1-b) \cdot G} \sum_{d=R_a \cdot B+1}^{R_b \cdot B} \delta_{S_n}(\hat{S}_d)$,
   $AUC_b := \frac{1}{b \cdot G \cdot (1-R_b) \cdot B} \cdot \sum_{n=(1-b) \cdot G+1}^{G} \sum_{d=R_b \cdot B+1}^{B} \delta_{S_n}(\hat{S}_d)$, $\delta_u(w) = \begin{cases} 1, \text{if } u > w \\ \frac{1}{2}, \text{if } u = w \\ 0, \text{if } u < w \end{cases}$.
3. Need to calculate segmental probabilities of default for the given calibration $PD =: p$:

   $P_a := \frac{R_a \cdot p}{a \cdot (1-p) + R_a \cdot p}$, $P_I := \frac{(R_b - R_a) \cdot p}{(1-a-b) \cdot (1-p) + (R_b - R_a) \cdot p}$, $P_b := \frac{(1-R_b) \cdot p}{b \cdot (1-p) + (1-R_b) \cdot p}$.
4. There is calculated parameter β_I as a solution to the equation: $AR_I = 2(1 + \beta_I)\left(1 - \beta_I \cdot \ln\left(1 + \frac{1}{\beta_I}\right)\right) - 1$, with initial approximation $\tilde{\beta}_I := 1$.

After this, need to find approximate solutions to the following minimization problems:
For the left side: $\beta_a := argmin_{\beta>0} \left| AR_a - d_a(\beta) \cdot \left(2(1+\beta)\left(1 - \beta \cdot \ln\left(1 + \frac{1}{\beta}\right)\right) - 1\right)\right|$,
where $d_a(\beta) := \frac{\beta}{1-P_a} \cdot \left(P_a \cdot \frac{1 + \beta_I / P_I}{1+\beta_I} - 1\right)$, with initial approximation $\tilde{\beta}_a := 0.1$.
And for the right side: $\beta_b := argmin_{\beta>0} \left| AR_b - d_b(\beta) \cdot \left(2(1+\beta)\left(1 - \beta \cdot \ln\left(1 + \frac{1}{\beta}\right)\right) - 1\right)\right|$, where $d_b(\beta) := \beta \cdot \left(\frac{\beta_I \cdot (P_I - P_b) - P_b \cdot (1-P_I)}{P_b \cdot (1-P_I) \cdot (1+\beta_I)}\right)$, with initial approximation $\tilde{\beta}_b := 0.1$.

5. Probability of default is calculated for the quantile $x(s)$ of the score s by formulas from [8]:

$$PD(x) := \begin{cases} PD_R\left(\frac{x}{\hat{a}}, \beta_a, d_a, P_a\right) \; if \; x \in [0, \hat{a}] \\ PD_C\left(\frac{x-\hat{a}}{1-\hat{a}-\hat{b}}, \beta_I, P_I\right) \; if \; x \in (\hat{a}, 1-\hat{b}] \\ PD_L\left(\frac{x-1+\hat{b}}{\hat{b}}, \beta_b, d_b, P_b\right) \; if \; x \in (1-\hat{b}, 1] \end{cases},$$

where $\hat{a} := a \cdot (1-p) + R_a \cdot p$, $\hat{b} := b \cdot (1-p) + (1-R_b) \cdot p$,
$PD_C(y, \beta, D) := \frac{1}{2}\left(1 - \frac{y+\beta-D-2\beta D}{\sqrt{(y-\beta-D)^2 + 4\beta(1-D)y}}\right)$, $PD_R(y, \beta, d, D) := \frac{D}{2(1-d+dD)}\left(1 - D \cdot \frac{y+\beta(2d(1-D)-1)+d(1-D)-1}{\sqrt{DR(y,\beta,d,D)}}\right)$,
$DR(y, \beta, d, D) := \left(y(D + 2(1-D)(1-d)) - D(1+\beta-d(1-D))\right)^2 + 4y(1-D)(1-d+dD)((1+\beta-d)D - y(1-d))$,
$PD_L(y, \beta, d, D) := \frac{1}{2(1-dD)}\left(1 + D - 2dD - (1-D) \cdot \frac{y+\beta-dD(1+2\beta)}{\sqrt{DL(y,\beta,d,D)}}\right)$, $DL(y, \beta, d, D) := (y-\beta-dD)^2 + 4y\beta(1-dD)$.
The method of balanced approximation of the quantile $x(s)$ of the score s is presented in Appendix 2.

5. Comparison of calibration methods in consumer and mortgage banking segments

There is presents a comparison of metrics of three different calibration methods for the consumer loans segment in Table 3 (there are calibrated at one level of central tendency $CT = 4{,}76\%$):

Table 3. Comparison of metrics of three different calibration methods for the consumer loans segment.

Metrics	PD	AR	LAR	RAR
Model (no calibration)	4,76%	80,4%	64,1%	64,5%
Estimation error	0,001%	0,14%	1,39%	1,42%
Logistics calibration (LC)	4,76%	80,4%	62,7%	66,5%
Cubic logistics calibration (CLC)	4,76%	80,4%	64,4%	64,8%
Three-segment hyperbolic calibration (TSHC)	4,76%	80,0%	63,7%	63,3%

Figure 3 shows the ROC curves corresponding to the calibrations and a graph of PD depending on the score.

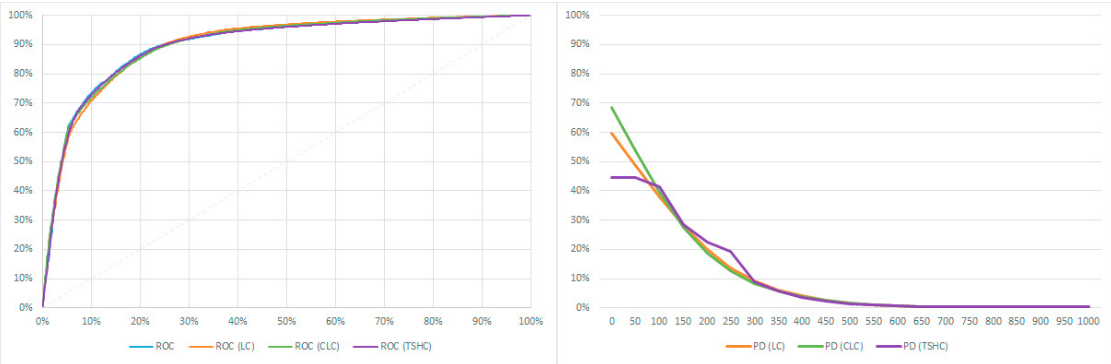

Figure 3. ROC curves and PD graph corresponding to the calibrations of the consumer loans segment model depending on the score.

There is presents a comparison of metrics of three different calibration methods for the mortgage loans segment in Table 4 (there are calibrated at one level of central tendency $CT = 0{,}48\%$):

Table 4. Comparison of metrics of three different calibration methods for the mortgage loans segment.

Metrics	PD	AR	LAR	RAR
Model (no calibration)	0,48%	89,5%	78,0%	76,4%
Estimation error	0,004%	0,33%	6,61%	5,85%
Logistics calibration (LC)	0,48%	89,5%	80,3%	74,7%
Cubic logistics calibration (CLC)	0,48%	89,5%	79,2%	75,6%
Three-segment hyperbolic calibration (TSHC)	0,48%	89,1%	77,6%	75,4%

Figure 4. shows the ROC curves corresponding to the calibrations and a graph of PD depending on the score.

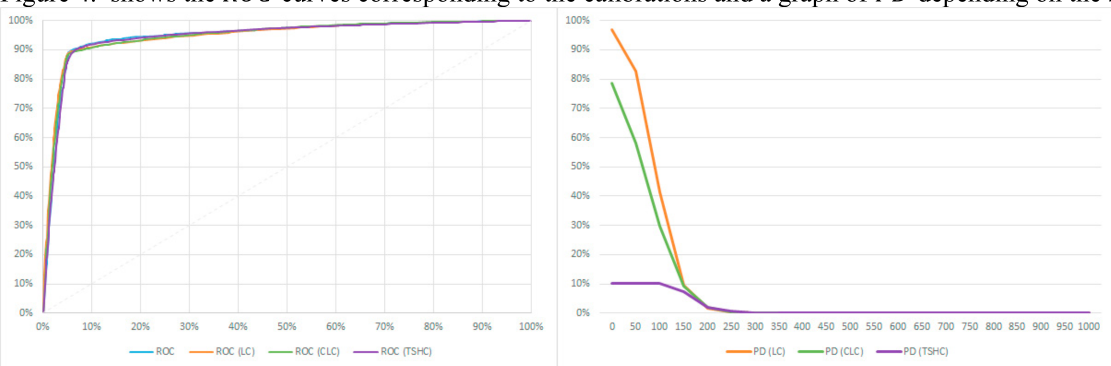

Figure 4. ROC curves and PD graph corresponding to the calibrations of the mortgage loans segment model depending on the score.

The binomial test for conservatism based on the "20+1" ranks rating scale gives the lowest acceptable values of the calibration PD at which is satisfied the "green zone". The results are presented in Table 5 and Table 6.

Table 5. The lowest calibration PD values for "green zone" of binomial test without quantile approximation $x(s)$ of score s (used all samples).

Calibration method	consumer loans segment	mortgage loans segment
Logistics calibration (LC)	6,14%	3,46%
Cubic logistics calibration (CLC)	5,75%	2,22%
Three-segment hyperbolic calibration (TSHC) without approximation $x(s)$	5,51%	0,73%

If there is applied the approximation of the quantile $x(s)$ according to the algorithm in Appendix 2, then the result of the minimum PD values for a three-segment calibration will deteriorate slightly (Table 6).

Table 6. The lowest calibration PD values for "green zone" of binomial test with quantile approximation $x(s)$ of score s (used some samples).

Calibration method	consumer loans segment	mortgage loans segment
Logistics calibration (LC)	6,14%	3,46%
Three-segment hyperbolic calibration (TSHC) with approximation $x(s)$	5,77%	1,50%

After all comparative tables implies that precision calibration is necessary for significant savings in capital and bank reserves for the mortgage loan segment.

Appendix 1. Calculation of additional parameters of ROC-curve

There are formulas for calculating some indicators below that were used for three-segment calibration [10]. The indicators μ_L and μ_R are calculated as solutions to the following equations:

$$LAR = \frac{AR}{\mu_L-1}\ln\left(\frac{AR}{\mu_L-1}\right) - \frac{\mu_L-1-AR}{\mu_L\cdot(1-AR)-1}\cdot\frac{AR\cdot\mu_L}{\mu_L-1}\ln\left(\frac{AR\cdot\mu_L}{\mu_L-1}\right), RAR = \frac{AR}{\mu_R-1}\ln\left(\frac{AR}{\mu_R-1}\right) - \frac{\mu_R-1-AR}{\mu_R\cdot(1-AR)-1}\cdot\frac{AR\cdot\mu_R}{\mu_R-1}\ln\left(\frac{AR\cdot\mu_R}{\mu_R-1}\right),$$

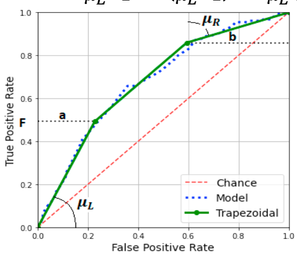

Figure 5. Trapezoidal triangulation of ROC curve.

where AR, LAR, RAR are calculated by formulas from (1), a and b are calculated by formulas below:

$$a \coloneqq \frac{AR}{(\mu_L-1)\cdot\left(1+\sqrt{1-AR\cdot\left(1+\frac{1}{\mu_L-1}+\frac{1}{\mu_R-1}\right)}\right)},$$

$$b \coloneqq \frac{\mu_R\cdot AR}{(\mu_R-1)\cdot\left(1+\sqrt{1-AR\cdot\left(1+\frac{1}{\mu_L-1}+\frac{1}{\mu_R-1}\right)}\right)}.$$

Appendix 2. Approximation of the distribution of scoring points

There are the values of scoring points $R_j \in R$ ordered in increasing order for k periods ($k \geq 5$) as sets R^k. Unified set of points is $R = \bigcup_k R^k$. The capacity of the set R is equal to N. The capacities of the sets R^k are equal to N^k thus $\sum_k N^k = N$.

The problem is to determine $X_i \in R$, as well as their minimum number n, $i = 0, \ldots, n$ such that the Kolmogorov–Smirnov (KS) metric for the resulting distribution $F_n(x)$ with respect to R is not more than the minimum of the KS metrics R^k with respect to R. Give an effective distribution function $F_X(x)$ such that $F_X(\min(R) - \varepsilon) = 0$, $F_X(\max(R) + \varepsilon) = 1$ for some $\varepsilon > 0$.

The simplest formula for approximating the distribution on R according to the values of $R_j, j = 1..N$ is as follows $F_R(x) \coloneqq \frac{1}{N}\sum_{j=1}^{N} I_{R_j}(x)$, where $I_{R_j}(x) = \begin{cases} 1 \; if \; x \geq R_j \\ 0 \; if \; x < R_j \end{cases}$. Similarly, the distribution is constructed for any of the k periods $F_{R^k}(x) \coloneqq \frac{1}{N^k}\sum_{j=1}^{N^k} I_{R^k{}_j}(x)$. Metrics of the KS-type which are the distance between the distributions on R^k and the reference R are defined as $M^k \coloneqq \max_{j=1\ldots N}\left|F_{R^k}(R_j) - \frac{j}{N}\right|$. In this case, the minimum metric of confidence is $M \coloneqq \min_k M^k$, it is also the threshold for determining $X_i \in R$ and n.

Definition

The approximating distribution $F_n(x)$ is determined by increasing values of $X_i \in R$ for $i = 0, \ldots, n$ as:

$$F_n(x) \coloneqq \frac{1}{n}\sum_{i=0}^{n-1} I(x, X_i, X_{i+1}), \text{ where } I(x, X_i, X_{i+1}) \coloneqq \begin{cases} 0 & , if \; x < X_i \\ 1 & , if \; x > X_i \; and \; x \geq X_{i+1} \\ \frac{x - X_i}{X_{i+1} - X_i} & , otherwise \end{cases}.$$

For a given n to select X_i as point with the minimum distance from the boundaries of R (i.e., $min(R)$, $max(R)$), which is assumed to be non-extremal. Let this be a natural m (i.e. $m-1$ of the first points and $m-1$ of the last points of R are not considered). Find the minimum L such that

$$\begin{cases}\dfrac{N-L}{n} = M \in \mathbb{N}, \\ L \geq 2m\end{cases}$$

then $X_0 := R_{m+1}, X_n := R_{N-m}, X_i := R_{i\cdot M+\left[\frac{L}{2}\right]}$ for $i = 1, \dots, n-1$. The metric $M_n := \max_{j=1\dots N}\left|F_n(R_j) - \frac{j}{N}\right|$ is compared with the metric M. There is selected the minimum n such that $M_n \leq M$.

Approximation testing. Examples:

1) Consumer loans segment: number of samples is $n = 23\,231\,154$, number of approximation points is $N = 16$, metric value is $M = 2.16\%$;
2) Mortgage loans segment: number of samples is $n = 2\,425\,852$, number of approximation points is $N = 26$, metric value is $M = 2.39\%$.

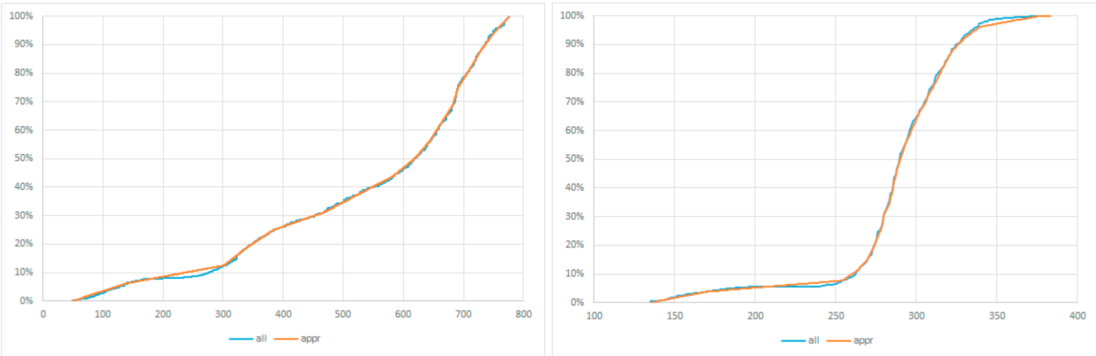

Figure 6. Testing the approximation of scoring distributions.

This work is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE University).

References

[1] B. Baesens, K.Smedts, *Boosting credit risk models*, The British Accounting Review, (2023), 101241.
[2] Basel Committee on Banking Supervision (BCBS), *Studies on the Validation of Internal Rating Systems(revised),* Working Paper No. 14. (2005b), (http://www.bis.org/publ/bcbs_wp14.htm)
[3] A. Brezigar-Masten, I. Masten, M. Volk, *Modeling credit risk with a Tobit model of days past due*, Journal of Banking & Finance, 122, (2021), 105984
[4] J. Hanley, B. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve*, Radiology, 143, 1 (1982) pp. 29–36
[5] Y.B. Huajian, *Smoothing algorithms by constrained maximum likelihood: Methodologies and implementations for Comprehensive Capital Analysis and Review stress testing and International Financial Reporting Standard 9 expected credit loss estimation,* Journal of Risk Model Validation. 12, (2018) pp. 89-102
[6] D. Joseph, *Estimating credit default probabilities using stochastic optimization* [J]. Data Science in Finance and Economics 1, 3 (2021), pp. 253-271.
[7] A.S. Khamalinsky, M. V. Pomazanov. Calibrating the rating model for sectors with low default rates. Financial risk management (Upravlenie finansovy'mi riskami) 2, (2012), pp. 82–94. [in Russian] https://grebennikon.ru/article-gaqy.html
[8] M. V. Pomazanov, *ROC analysis and calibration of scoring models based on second-order accuracy metrics,* Financial risk management (Upravlenie finansovy'mi riskami) 2, (2021), pp.100–121 [in Russian], https://doi.org/10.36627/2221-7541-2021-2-2-100-121
[9] M.V. Pomazanov, *Second-order accuracy metrics for scoring models and their practical use*, Procedia Computer Science 214 (2022), pp 565-572.
[10] M.V. Pomazanov, *Second-order accuracy metrics for scoring models and their practical use*, Preprint arXiv.2204.07989 (2022)
[11] N. Siddiqi, *Credit Risk Scorecards: Developing And Implementing Intelligent Credit Scoring*, Wiley and SAS Business Series, Wiley, 2005.
[12] D. Tasche, *The art of probability-of-default curve calibration*, Journal of Credit Risk 9, (2013). Pp. 63-103.
[13] M. Van der Burgt, *Calibrating low-default portfolios, using the cumulative accuracy profile*, Journal of Risk Model Validation, 1, 4(2008), pp. 17–33.