Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 221 (2023) 237–244

Procedia Computer Science

www.elsevier.com/locate/procedia

10th International Conference on Information Technology and Quantitative Management

Continuous transformation of nonmonotonic factors of the scoring model increasing its discriminating power

Mikhail Pomazanov[a*]

[a] *School of Finance, Faculty of Economic Sciences, National Research University "Higher School of Economics", Moscow, Russia*

Abstract

When developing scoring models, Weight of Evidence (WOE) transformation is actively used, which is in demand to improve the discriminating power of the factors included in the model. For WOE transformation, various binning methods are used, the dimension of which is limited by the available statistics. As an alternative, the method of continuous transformation of model factors is proposed, which is applied when the target binary variable depends nonmonotonically on these factors. Two types of one-parameter families of transformations and strict conditions are proposed that determine the need for "healing" of factors, as well as determining the appropriate type of transformation (Φ or Ψ). The method for calculating transformation parameters is substantiated, and practical examples are presented. The use of the proposed transformation in the practice of developing scoring models in the banking sector confirms that the use of ΦΨ - transformations to correct the nonmonotonicity of the model factors gives a significant increase in the discriminating power at the level up to 10% of the Gini index of applied models for binary classification.

Keywords: ROC curve, scoring, binning, standardization, Gini index, discrimination power

1. Introduction

Let the scoring factor x be dimensionless and have passed the standardization stage convenient for the developer, the direction of its action on the rating is predominantly positive (from worse to better with growth of x) or neutral. This means that the Gini index is $\mathrm{Gini} = \mathrm{AR}\{x\} \geq 0$. Without loss of generality, we can assume that the factor is continuous, in the case of its finite sampling, the ROC curve can be linearly interpolated. In the case of a convex ROC-curve (Receiver Operating Characteristic), such Fig. 1 left, it is enough to carry out quantile normalization[2] of the factor, having

* Corresponding author. E-mail address: mhubble@yandex.ru

[2] In other words, this is the quantile transformation of the original factor X_F into its dimensionless quantile image $F = G(X_F)$, distributed uniformly on the segment [0,1]. Where $G(\cdot)$ is cumulative distribution function of X_F.

10.1016/j.procs.2023.07.033

achieved its uniform distribution on the segment [0,1]) and then introduce the factor into the scoring model. The question is more difficult, how to be in case of violation of the convexity of the ROC-curve, built on statistically significant objective binary observations (see Fig. 1 right)?

Traditionally, for such parameters, they resort to a WOE transformation, performing binning (data discrete binning), i.e. dividing (often empirically) the range of a variable into segments (bins) and setting the weights of each bin based on the statistics of the implementation of a binary variable [1]. Binning methods are widely used in exploratory data analysis and as an algorithm to speed up learning tasks [12]. In particular, binning is widely used in credit risk modeling, being an important tool for modeling credit performance to maximize the distinction between high and low risk observations, as well as for modeling expected credit losses.

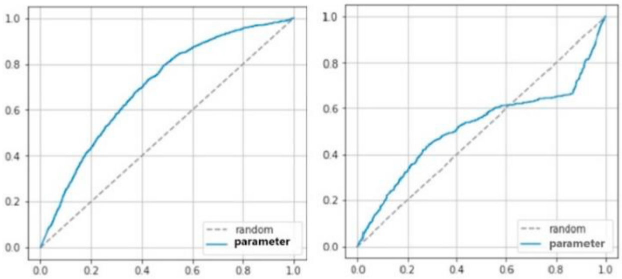

Fig. 1. Convex ROC curve (left) and Non-convex ROC curve (right). Left parameter-factor is Current liabilities/Revenue, Industry: manufacturing, Segment: Medium business, Gini=39%. Right parameter-factor is Current liabilities / EBITDA (inverse). Industry: trade. Segment: Medium business, Gini=3%

There are several unsupervised and supervised binning methods. Common unsupervised methods are tying intervals of equal width and the same size or equal frequency. On the other hand, well-known supervised fusion-based methods are Monotone Adjacent Pooling Algorithm (MAPA), also known as Maximum Likelihood Monotone Coarse Classifier (MLMCC) [20] and ChiMerge [10], whereas other methods based on decision trees are CART [3], Minimum Description Length Principle (MDLP) [6] and method of Conditional Inference Trees (CTREE) [8].

Binning that uses WOE transforms to prepare a non-convex factor (i.e., not monotonic with respect to the binary target variable) for transfer to the model has a number of disadvantages:

- the binning process requires compliance with certain restrictions. These limits can range from requiring a minimum number of entries per cell to monotonicity limits. This variant of the process is known as the optimal binning process. The optimal pooling is usually solved by iteratively pooling the initial fine-grained discretization until the imposed constraints are met. Performing the manual fine-tuning is likely to be unsatisfactory as the number of constraints increases leading to sub-optimal or even unworkable solutions. It should be noted that this manual setting has been encouraged by some authors [19], legitimizing the existing interaction of “art and science” in the binning process;
- statistical limit on the number of bins. As the number of bins increases, the statistical error of the WOE weights increases. With a small number of bins, a “rounding error” appears (this is when two essentially different values of an indicator fall into a common bin, for example, a profitability equal to 2% and 10%). This leads to the risk of losing the accuracy of the model;
- each bin weight is a separate model parameter with increased error due to outliers (low robustness), the more bins, the higher the error. This creates an increased risk of overtraining of the model;
- the interpretation of factor weights in the scoring model becomes more complicated. The uninterpretability arises because of the individual setting of each WOE-transformed variable. Understanding which factor is stronger or weaker is significantly blurred. A factor with a weight of 10% may have a stronger effect than one with a weight of 20%, depending on the distribution of bin weights and hit frequencies.

Results of some investigations (for example, [14]) suggest that using WOE transformation with logistic regression decreased the discriminatory power across a majority of the evaluation metrics compared to the models that did not use factors WOE transformed. Given the disadvantages of the approach based on the binning of WOE-transformed factors, including the risk of overlearning and instability, then such a model should be given increased attention in validation and reduce the life cycle (from validation to validation) by increasing the frequency of monitoring.

In the presented work, we propose an alternative approach to the WOE-transform tool. Bearing in mind that if the factor of model is convex, then the WOE transform does not give any increase in discriminating power at all (monotonic transformation does not change the discriminating power), and most likely, it will even give it a decrease due to the “rounding error”. For a non-monotonic (i.e., non-convex) factor, there is a means to "healing of dent" with only one additional parameter, which is determined statistically from the ROC curve (from the statistics of the development sample). The uniqueness of the transformation parameter and a clear algorithm for its calculation guarantees the minimum loss of robustness (resistance to noise) of the decision-making system. For the first time, the new transformations were proposed in [15] for practical use, but the proofs and conditions of application were not given in full.

2. Numerical criteria for determining the nonconvexity of the ROC-curve

In [16] new second-order accuracy metrics for scoring models were proposed that show the model's target preference for better diagnosing "good" objects or better diagnosing "bad" objects for a constant generally accepted discriminant determined by the well-known first-order metric AUROC (Gini index). There are two second-order metrics, and they have both an integral representation and a numerical one. The numerical representation based on point binary outcomes for the Gini index $AR = 2 \cdot AUC - 1$ has a well-studied relationship with the Wilcoxon statistic [7].

Let statistical binary data on a binary indicator (e.g., credit risk) be provided, among which: N – not defaults (binary indicator is equal zero) have ratings $S_n, n = 1 \dots N$, D – defaults (binary indicator is equal one) have ratings $\hat{S}_d, d = 1 \dots D$. All ratings are ordered from "bads" to "goods" $S_{n+1} \geq S_n$, $\hat{S}_{d+1} \geq \hat{S}_d$. The frequency of defaults in a given sample will be $PD = \frac{D}{D+N}$. Define the function $\delta_u(w) = \begin{cases} 1, if\ \ u > w \\ \frac{1}{2}, if\ u = w \\ 0, if\ u < w \end{cases}$.

Then the ROC curve at the points $x_n = 0 \dots \frac{n}{N} \dots 1$, will take on the values $ROC_0 = 0, ROC_n = \frac{1}{D} \cdot \sum_{d=1}^{D} \delta_{S_n}(\hat{S}_d)$. The first-order discriminating metric AR (Gini index) is calculated by the formula:

$$AR = 2 \cdot AUC - 1, \text{ where } AUC = \frac{1}{N \cdot D} \cdot \sum_{n=1}^{N} \sum_{d=1}^{D} \delta_{S_n}(\hat{S}_d) \tag{1}$$

Numerical representation second-order metrics LAR/RAR [16] has the form:

$$LAR = 2 \cdot LAUC - 1 \tag{2}$$

$$RAR = 2 \cdot RAUC - 1 \tag{3}$$

for the for the left second-order metric (LAR) and right one (RAR) respectively, where

$$LAUC = \frac{1}{N} \sum_{k=1}^{N} \left| \begin{matrix} \frac{\sum_{n=1}^{k} \sum_{d=1}^{D} \delta_{S_n}(\hat{S}_d)}{k \cdot \sum_{d=1}^{D} \delta_{S_k}(\hat{S}_d)}, if\ \sum_{d=1}^{D} \delta_{S_k}(\hat{S}_d) \neq 0 \\ 0, if, \sum_{d=1}^{D} \delta_{S_k}(\hat{S}_d) = 0 \end{matrix} \right., RAUC = \frac{1}{D} \sum_{d=1}^{D} \left| \begin{matrix} \frac{\sum_{n=d}^{D} \sum_{k=1}^{N} \delta_{S_k}(\hat{S}_n)}{(D-d+1) \cdot \sum_{k=1}^{N} \delta_{S_k}(\hat{S}_d)}, if\ \sum_{k=1}^{N} \delta_{S_k}(\hat{S}_d) \neq 0 \\ 0, if\ \sum_{k=1}^{N} \delta_{S_k}(\hat{S}_d) = 0 \end{matrix} \right. .$$

It is worth noting that the computational volume of the calculation of first-order metrics has the dimension $D \cdot N$ (for AR), the volume of calculations for second-order metrics is significantly larger: $D \cdot {N^2}/{2}$ for LAR and $N \cdot {D^2}/{2}$ for RAR. Consequently, in the case of large data, when calculating second-order accuracy metrics, it may be necessary to resort to thinning methods [9] to save computing resources.

The minimum value of LAR/RAR while maintaining the convexity of the ROC curve is

$$\min LR(AR) = AR + (1 - AR) \cdot \ln(1 - AR) \tag{4}$$

Given the limited number of measurements in the calculation of the metrics of the first and second order, it is necessary to take into account the statistical error. A fairly accurate estimate of the statistical error of the AUC (AR) metric was proposed in [4], but from a conservative point of view, it suffices to take the upper estimate [2]

$$\sigma_{AR} \cong \sqrt{\frac{(2N+1) \cdot (1 - AR^2) - (N-D) \cdot (1 - AR)^2}{3 \cdot N \cdot D}}. \tag{5}$$

On condition $\frac{D}{N} \ll 1, D \gg 1$, you can use the asymptotes (5) in the form

$$\sigma_{AR} \cong \sqrt{\frac{(1-AR)\cdot(1+3AR)}{3\cdot D}} \cdot \left(1 + \frac{1}{2N} \cdot \frac{1+AR}{1+3AR} + o\left(\frac{1}{N}\right)\right).$$

The statistical error in estimating the second-order metrics (2), (3) has not been studied to the extent necessary. But considering that $\min LR$ depends only on AR, it is possible to formulate a statistical rule, from the fulfillment of which follows the need for a nonmonotonic transformation of the scoring factor

$$\min(LAR, RAR) < minLR(AR) + t_\alpha \cdot \sigma_{AR} \cdot \ln(1 - AR)\ , \tag{6}$$

Where AR/LAR/RAR are calculated by formulas (1), (2), (3), $minLR(AR)$ – (4), σ_{AR} – (5), $t_\alpha = N^{-1}\left(\frac{1+\alpha}{2}\right)$ – coefficient for trust level α. Multiplier $-\ln(1 - AR)$ is the derivative (4) with respect to AR.

3. ΦΨ transformations that increase the discriminating power of non-convex factors of scoring model

Let a rating factor is quantile normalized and therefore uniformly distributed one $x \in [0,1]$, ranked from worst to best (Normalization №1), then the index $\text{Gini}\{x\} = AR\{x\} \geq 0$. A factor is convex if its ROC-curve is convex. Let ROC(x) be continuous, for example, prepared by applying linear or higher interpolation.

Definition

One-parameter transformations $\boldsymbol{\Phi}$ *and* $\boldsymbol{\Psi}$ *called:*

$\boldsymbol{\Phi}: y = \Phi(x, \varphi),\ \Phi(x, \varphi) = \frac{\left(x - \frac{1}{2}\varphi\right)^2}{\left(1 - \frac{1}{2}\varphi\right)^2}, \varphi \in (0,1], \boldsymbol{\Psi}: y = \Psi(x, \psi),\ \Psi(x, \psi) = (1 + \psi)(2 - (1 + \psi) \cdot x) \cdot x,$
$\psi \in (0,1],\ x \in [0,1]$.

Graphs Φ *and* Ψ transformations presented on Fig. 2 Second-order metrics are calculated: $LAR\{x\}$, $RAR\{x\}$ (2), (3). The minimum value of LAR, RAR for a convex ROC will be calculated by formula (4), considering (5). Denote

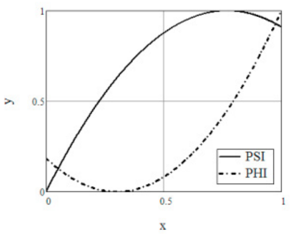

Fig. 2. Φ *and* Ψ transformations (PHI and PSI)

$\text{m}LR_\alpha = minLR(AR\{x\}) - \Delta_\alpha mL$, where $\Delta_\alpha mLR = -t_\alpha \cdot \sigma_{AR\{x\}} \cdot \ln(1 - AR\{x\})$.

Statistical criterion (6) selection of Φ *or* Ψ- transformations

At a given confidence level α, *the choice of the form of* Φ *or* Ψ*-transformation is made under the following criteria:*

- *If* $LAR\{x\} < \text{m}LR_\alpha$, *then the* $\boldsymbol{\Phi}$ *transformation is applied,* $\boldsymbol{\Phi}: y = \Phi(x, \varphi)$;
- *If* $RAR\{x\} < \text{m}LR_\alpha$, *then the* $\boldsymbol{\Psi}$ *transformation is applied,* $\boldsymbol{\Psi}: y = \Psi(x, \psi)$;
- *If* $LAR\{x\} > mLR_\alpha$ *and* $RAR\{x\} > mLR_\alpha$, *then the identity transformation remains valid y=x;*

If both upper criteria are true at the same time, then the factor is not allowed in the model.

Optimal φ^*, ψ^*:

$$\boldsymbol{\Phi}: y = \Phi(x, \varphi), \varphi^* = arg \max_{\varphi \in (0,1]} AR\{y\}, \boldsymbol{\Psi}: y = \Psi(x, \psi), \psi^* = arg \max_{\psi \in (0,1]} AR\{y\}. \tag{7}$$

Theorem

Let the ROC curve of an untransformed factor $\{x\}$ be twice continuously differentiable ($N, D \to \infty$) and the following conditions are met

1. equation (7) is holds,

2. there is a single inflection point x^b of the curve ROC(x), such as $|ROC''(x^b)| \geq \varepsilon > 0$,
3. there exist at least one point x^0 such that: $2 \cdot ROC\left({x^0}/{2}\right) > ROC(x^0)$ for Φ – transformation or $2 \cdot ROC\left(\frac{1}{1+x^0}\right) > \left(1 + ROC\left(\frac{1-x^0}{1+x^0}\right)\right)$ for Ψ – transformation.

Then there is a unique solution for parameters φ^*, ψ^* (7) according to the equations:

$$\mathbf{\Phi}:\quad ROC(\varphi^*) = 2 \cdot ROC\left(\frac{\varphi^*}{2}\right) \quad \mathbf{\Psi}:\quad 1 + ROC\left(\frac{1-\psi^*}{1+\psi^*}\right) = 2 \cdot ROC\left(\frac{1}{1+\psi^*}\right) \tag{8}$$

The proof is given in the Appendix. Initial approximation[3]: $\varphi^*, \psi^* = 1 - AR\{\text{x}\}$.

The order of operations with a transformable factor is as follows:

- for a scoring factor that satisfies the non-convexity condition (6), a quantile Normalization №1 is required;
- the position of the rating factor on the scale “bad – good” is considered on the scale $\boldsymbol{y}$ (after $\mathbf{\Phi\Psi}$- transformation);
- the distribution of the transformed factor y ceases to be uniform, and since the economic interpretation of the weight and value of this indicator is important, it must be additionally normalized (Normalization №2). This operation does not affect the discriminating power, i.e., the quality of the resulting model, since the transformation is monotonic and the variable y is already convex.

Normalization №2 is $y \to Y$, for this, a transformation (9) is applied;

$$\mathbf{\Phi}:\quad Y(y) = \begin{cases} (2-\varphi^*)\sqrt{y}, & if\ y \leq \left(\frac{\varphi^*}{2-\varphi^*}\right)^2 \\ \frac{\varphi^*}{2} + \left(1 - \frac{\varphi^*}{2}\right)\sqrt{y}, & if\ y > \left(\frac{\varphi^*}{2-\varphi^*}\right)^2 \end{cases},\ \mathbf{\Psi}:\quad Y(y) = \begin{cases} \frac{1-\sqrt{1-y}}{1+\psi^*}, & if\ y \leq 1 - \psi^{*2} \\ 1 - \frac{2\sqrt{1-y}}{1+\psi^*}, & if\ y > 1 - \psi^{*2} \end{cases}, \tag{9}$$

- the scoring model (for example, built by logistic regression based on the maximum Gini index) gets the transformed variable $\boldsymbol{Y}$, for which the weights are optimized together with the weights of other factors.

A demonstration on practical use of the "correction of dents" of the ROC curve using ΦΨ - transformation is shown in next Section.

4. Application of ΦΨ transformations

To demonstrate the preparation of factors for inclusion in the scoring model, calculations are presented based on publicly available data from Russian non-profit organizations (NPOs) (see Fig. 3). Table 1 presents the results of the analysis of condition (6), the necessary transformation is determined and the final result of the "healing" is given. Obviously, there is a significant increase in the Gini factors. The following Table 2 presents the intervals of real values of indicators (EBIT/Equity and Net Operating Income), distributed by deciles of the corresponding factor, which is included in the scoring model.

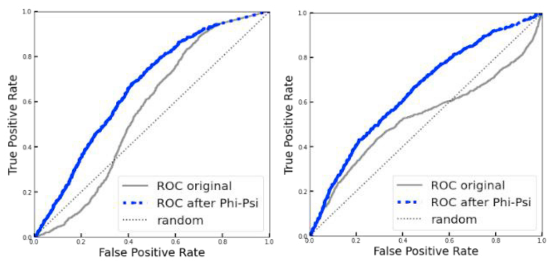

Fig. 3. An object: NPOs. Left graph: Φ-transformation, factor: Ratio EBIT/Equity. Right graph: Ψ-transformation, factor: Net operating income (Thousand RUR).

[3] In practice, it may be quite satisfactory

Table 1. Analysis of the fulfillment of the conditions for applying the ΦΨ-transformation and the result of the transformation

Factor	$AR\{x\}$ **original**	Second order metric	$minLR$	Error $\Delta_{95\%}mLR$	Condition (6)	Type $\Phi\Psi$	$AR\{y\}$ **after** $\Phi\Psi$
EBIT/Equity	**12.0%**	LAR=-15%	0.75%	0.05%	Yes	$\varphi = 0.79$	**32.4%**
Net operating income	**6.1%**	RAR=-13%	0.19%	0.05%	Yes	$\psi = 0.83$	**30.1%**

Table 2 shows that the first transformation (Normalization №1, i.e. quantile transformation) gives continuous intervals, which is obvious. The next transformation (ΦΨ-transformation plus Normalization №2) gives two intervals for most of the deciles of the factor. All intervals do not intersect. If the value of the real indicator falls within one of the intervals, then the factor is assigned the value of the corresponding decile. In the continuous case, intervals can be obtained with any precision greater than a decile.

Table 2. Intervals of real values of the factor corresponding to the decile of the transformed factor. (Net Operating Income in Thousand RUR)

EBIT/Equity decile (%)		**Interval of factor up to** $\Phi\Psi$		**Interval of factor after** $\Phi\Psi$			
0	10	$-\infty$	-42.4	-0.5	0.0	0.0	0.6
10	20	-42.4	-7.5	-1.1	-0.5	0.6	2.2
20	30	-7.5	-1.8	-2.1	-1.1	2.2	5.9
30	40	-1.8	-0.4	-4.2	-2.1	5.9	13.1
40	50	-0.4	0.8	-8.7	-4.2	13.1	36.1
50	60	0.8	7.1	-19.8	-8.7	36.1	130
60	70	7.1	46.9	-52.8	-19.8	130	6158
70	80	46.9	12343	$-\infty$	-52.8	6158	12943
80	90	12343	12806	12806	12943		
90	100	12806	$+\infty$	12943	$+\infty$		
Net Operating Income decile (%)		**Interval of factor up to** $\Phi\Psi$		**Interval of factor after** $\Phi\Psi$			
0	10	$-\infty$	-17 288	$-\infty$	-31 956	109 813	$+\infty$
10	20	-17 288	-4 039	-31 956	-13 497	22 398	109 813
20	30	-4 039	-1 113	-13 469	-6 656	7 953	22 398
30	40	-1 113	-310	-6 656	-3 274	3 463	7 953
40	50	-310	-54	-3 272	-1 685	1 469	3 459
50	60	-54	27	-1 685	-932	633	3 463
60	70	27	476	-932	-494	247	633
70	80	476	2 696	-494	-259	61	247
80	90	2 696	15 329	-259	-109	0	61
90	100	15 329	$+\infty$	-109	0		

The experience of applying the ΦΨ-transformation in model development practice shows that after transforming some of the long-list factors and improving their discriminating characteristics, it became possible to build a model with an updated short-list of variables. Versions of the models using ΦΨ-transformation improved the Gini index by 5-10%. This gave a noticeable annual economic profit to the credit business at the level of the obtained Gini delta multiplied by half of the amount of reserves for expected losses [17].

5. Conclusion

Let's list the recommendations how to prepare the factors for implementation in the scoring model using the ΦΨ-transformation:

1. Normalize all standardized model variables to a uniform distribution on a single interval (quantile normalization). This makes some economic sense, for example, in terms of limiting the range of default probabilities[4];
2. Calculate first- and second-order discriminant force metrics AR/LAR/RAR (1,2,3);
3. In case criterion (6) is valid, applying ΦΨ-transformation it is necessary to "healing" of the factor by estimating the optimal parameters $\varphi^* or\ \psi^*$ (7) or solving equation (8) and then moving to the adjusted variable y with an estimate of the transformed Gini index $AR\{\mathbf{y}\} > 0$;
4. Normalize again, i.e. Normalization №2 is applied $y \to Y$ (9), in order to subsequently accurately determine the weight coefficients of the model after identical standardization of the distributions of untransformed and transformed ("healinged") variables;

[4] With a logistic calibration type of $PD(R) = \frac{1}{1+e^{A \cdot R+B}}$, the probability of default $PD(R)$ will not appear as negligible or too high (as if the rating score "prophesies" default within a year)

5. In deciding whether to include a variable in the model, be guided by the transformed discriminant power indicator Gini-index, VIF, IV, etc. [18] for that part of the variables, that have been "healed" by ΦΨ-transformations;
6. Optimize scoring model factor weights with tools using best practices [5], [11], [13].

The ΦΨ transformation is especially relevant for use in the case of limited statistics of measurements of the target binary variable (1000-10000 measurements), for example, in the segment of building scoring maps for legal entities. The experience of implementing the approach in the development of credit risk assessment models in the banking sector has shown that the ΦΨ transformation provides an effective tool to improve the quality of scoring models. The transformations are explicit, highly stable, and transparent for validation. After improving the discriminating power of individual factors and including them in the short list of working-out, it had possible to increase the aggregate discriminating power by 5-10% over the Gini index original.

Appendix. Proof of the Theorem about of optimal parameters of the ΦΨ-transformation

Let the ROC curve be given by the integral $R(x) = \int_0^x p(\xi)d\xi,\ R(1) = 1$, then $AUC = \int_0^1 \int_0^x p(\xi)d\xi\, dx$.

Φ-transformation: $\Phi(x.\varphi) = \frac{\left(x-\frac{1}{2}\varphi\right)^2}{\left(1-\frac{1}{2}\varphi\right)^2}.\varphi \in (0.1]$ has roots

$$\begin{cases} 0 < t < \hat{t}(\varphi).\ \ y_1 = \frac{\varphi}{2} - \left(1-\frac{\varphi}{2}\right)\sqrt{t}.\ y_2 = \frac{\varphi}{2} + \left(1-\frac{\varphi}{2}\right)\sqrt{t} \\ \hat{t}(\varphi) \le t < 1.\ \ y_1 = 0.\ y_2 = \frac{\varphi}{2} + \left(1-\frac{\varphi}{2}\right)\sqrt{t} \end{cases}, \text{ where } \hat{t}(\varphi) = \left(\frac{\varphi/2}{1-\varphi/2}\right)^2.$$

The ROC-curve corrected after the Φ-transformation will be set parametrically $\left(\hat{x}(t).\hat{R}(t)\right)$, where $\hat{x}(t) = y_2(t) - y_1(t)$, $\hat{R}(t) = \int_{y_1(t)}^{y_2(t)} p(\xi)d\xi = R(y_2(t)) - R(y_1(t))$, $\widehat{AUC} = \int_0^1 \hat{R}(t)\cdot d\hat{x}(t)$, $\widehat{AUC} = \left(1-\frac{\phi}{2}\right)\cdot Q(\phi)$, where $Q(\phi) = \int_0^{\hat{t}(\varphi)} \frac{1}{\sqrt{t}}(R(y_2(t)) - R(y_1(t)))dt + \frac{1}{2}\int_{\hat{t}(\varphi)}^1 \frac{1}{\sqrt{t}} R(y_2(t))dt$. Considering $y_1(\hat{t}(\varphi)) = 0, y_2(\hat{t}(\varphi)) = \varphi$, $y_1(0) = y_2(0) = \varphi/2, y_2(1) = 1$, $dy_2(t) = \frac{1}{2\sqrt{t}}\left(1-\frac{\varphi}{2}\right)dt$, $dy_1(t) = -\frac{1}{2\sqrt{t}}\left(1-\frac{\varphi}{2}\right)dt$, we has $\widehat{AUC} = \int_{\varphi/2}^1 R(x)dx + \int_{\varphi/2}^{\varphi} R(x)dx + 2\int_{\varphi/2}^0 R(x)dx$.

Finally, $\widehat{AUC}(\varphi) = AUC + \int_0^{\varphi} R(x)dx - 4\int_0^{\varphi/2} R(x)dx$. Maximum value of $\widehat{AUC}(\varphi)$ can only reach in the point $\widehat{AUC}'(\varphi) = R(\varphi) - 2\cdot R(\frac{\varphi}{2}) = 0$, i.e. provided $R(\varphi^*) = 2\cdot R\left(\frac{\varphi^*}{2}\right)$.
Let $R(x)$ is continuous together with the second derivative, while it is convex, i.e. $R''(x) < 0$, with monotonicity of $R(x)$ (by ROC definition). Then, there is no solution to the equation $R(x) = 2\cdot R(x/2)$. Indeed, denoting $g(x) = R(x) - 2\cdot R(x/2)$, we get $g(0) = 0$, wherein $g'(x) = R'(x) - R'(x/2) < 0$, and since $R''(x) < 0$, it is clear that there are no roots. However, if there is an inflection point $\hat{x}$ such that $R''(x) > 0$ at $x \in (0,\hat{x})$ and $R''(x) < 0$ at $x \in (\hat{x}, 1)$, then on the interval $x \in (0,\hat{x})$ will be true $g'(x) > 0$, so $g(\hat{x}) > 0$. On the other hand, by the hypothesis of the theorem, there is x^0 such that $R(x^0) - 2\cdot R\left(x^0/2\right) = g(x^0) < 0$, hence, due to continuity, there exists a solution $R(\varphi) = 2\cdot R(\varphi/2)$. Moreover, this solution includes a maximum, since the sign of $\widehat{AUC}'(\varphi)$ changes from positive to negative on the segment $(0, x^0)$.

Let's prove the uniqueness φ. If φ is not unique, then there can be at least two roots, or three in the general case when $g(x)$ does not touch the zero axis. In any case, there are three extrema for the function $g(x)$. Moreover, all three extrema are located on the interval $(\hat{x}, \max\{2\hat{x}, 1\})$, because on the interval $(2\hat{x}, 1)$, if it exists, the function $g'(x) < 0$ is less than zero. The latter follows from the fact that $R'(x)$ decreases over the interval $(2\hat{x}, 1)$. Then, denoting two different extremums x_1, x_2 $\hat{x} < x_1 < x_2 < \max\{2\hat{x}, 1\}$, we have $g'(x_1) = g'(x_2) = 0$. This means $R'(x_2) - R'(x_1) = R'(x_2/2) - R'(x_1/2)$, but this is not possible with not equal x_1 *and* x_2, because x_1, x_2 and $x_1/2, x_2/2$ lie on opposite sides of the inflection point $\hat{x}$ of function $R(x)$. Hence the root φ of eq. (8) is unique for the Φ-transformation and the conditions of the Theorem.

Ψ- transformation: $\Psi(x.\psi) = (1+\psi)(2-(1+\psi)\cdot x)\cdot x.\ \ \psi \in (0.1]$ has roots
$\begin{cases} 0 < t < \hat{t}(\psi).\ \ y_1 = \frac{1}{1+\psi}\left(1-\sqrt{1-t}\right).\ y_2 = 1 \\ \hat{t}(\psi) \le t < 1.\ \ y_1 = \frac{1}{1+\psi}\left(1-\sqrt{1-t}\right).\ y_2 = \frac{1}{1+\psi}\left(1+\sqrt{1-t}\right) \end{cases}$, where $\hat{t}(\psi) = 1-\psi^2$.

The ROC-curve corrected after the Φ-transformation will be set parametrically $\left(\hat{x}(t).\hat{R}(t)\right)$, where $\hat{x}(t) = 1 + y_1(t) - y_2(t)$. $\hat{R}(t) = \int_0^{y_1(t)} p(\xi)d\xi + \int_{y_2(t)}^1 p(\xi)d\xi = R(y_1(t)) - R(\,y_2(t)) + 1$, $\widehat{AUC} = \int_0^1 \hat{R}(t)\cdot d\hat{x}(t), \widehat{AUC} = \frac{1}{1+\psi}\cdot P(\psi)$, where $P(\psi) = \int_0^{\hat{t}(\psi)} \frac{1}{2\sqrt{1-t}} R\left(y_1(t)\right)dt + \int_{\hat{t}(\psi)}^1 \frac{1}{\sqrt{1-t}}\left(R(y_1(t)) + 1 - R(y_2(t))\right)dt$. Applying active substitutions similar to the Φ-transformation, we obtain $\widehat{AUC}(\psi) = AUC + 2\frac{\psi}{1+\psi} + \int_{\frac{1-\psi}{1+\psi}}^1 R(x)dx - 4\int_{\frac{1}{1+\psi}}^1 R(x)dx$. The function $\widehat{AUC}(\psi)$ can take the maximum value at the point $\widehat{AUC}'(\psi^*) = 0$, thus $1 + R\left(\frac{1-\psi^*}{1+\psi^*}\right) = 2\cdot R\left(\frac{1}{1+\psi^*}\right)$. The existence of a solution is proved similarly to the Φ-transformation. Let $\hat{x}$ is inflection point, then $R''(x) < 0$ at $x \in (0,\hat{x})$, and $R''(\mathrm{x}) > 0$ on the interval $x \in (\hat{x}, 1)$. The function $g(x) = \frac{1}{2}R\left(\frac{1-x}{1+x}\right) - R\left(\frac{1}{1+x}\right)$ is introduced . Then $g'(x) = \frac{1}{(1+x)^2}\left(R'\left(\frac{1}{1+x}\right) - R'\left(\frac{1-x}{1+x}\right)\right) > 0$, at $x \in \left(0, \frac{1-\hat{x}}{1+\hat{x}}\right)$. It is means $\int_0^x g'(\theta)\,d\theta > 0$. Whence it follows that for $x \in \left(0, \frac{1-\tilde{x}}{1+\tilde{x}}\right)$ right inequality $2\cdot R\left(\frac{1}{1+x}\right) < \left(1 + R\left(\frac{1-x}{1+x}\right)\right)$. On the other hand, by the hypothesis of the Theorem, there exists a point x^0 such that $2\cdot R\left(\frac{1}{1+x^0}\right) > \left(1 + R\left(\frac{1-x^0}{1+x^0}\right)\right)$, whence it follows that under the conditions of the theorem the solution ψ^* for the Ψ-transformation exists. Uniqueness is proved similarly to the Φ-transformation.

References

[1] H. A. Abdou, Genetic programming for credit scoring: The case of Egyptian public sector banks, Expert Syst. Appl. 36, 9 (2009), pp. 11402–11417.
[2] D. Bamber, *The area above the ordinal dominance graph and the area below the receiver operating characteristic graph*, Journal of Mathematical Psychology 12, 4 (1975), pp. 387-415.
[3] L. Breiman, J. Friedman, R. Olshen and C. Stone, *Classification and Regression Trees*, Taylor & Francis, 1984.
[4] E. Demidenko, *Confidence intervals and bands for the binormal ROC curve revisited*, Journal of applied statistics 39, 1 (2012), pp. 67-79.
[5] S. Eguchi and J. Copas, *A class of logistic–type discriminant functions*, Biometrika 89, 1 (2002), pp. 1–22.
[6] U. M. Fayyad and K. B. Irani, *Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning*, International Joint Conferences on Artificial Intelligence Vol.13, pp.1022–1027, 1993.
[7] J. A. Hanley and B. J. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve*, Radiology 143, 1 (1982), pp. 29–36.
[8] T. Hothorn, K. Hornik and A. Zeileis, *Unbiased recursive partitioning: A conditional inference framework,* Journal of Computational and Graphical statistics 15,3 (2006), pp. 651–674.
[9] Iske, *Thinning Algorithms*, In: Multiresolution Methods in Scattered Data Modelling, Lecture Notes in Computational Science and Engineering, Springer, Berlin, Heidelberg Vol. 37, pp. 67-125, 2004.
[10] R. Kerber, *ChiMerge: Discretization of Numeric Attributes*, AAAI-92 Proceedings, 1992.
[11] K. Miura, S. Yamashita and S. Eguchi, *Area under the curve maximization method in credit scoring*, Journal of Risk Model Validation 4, 2 (2010), pp. 3–25.
[12] G. Navas-Palencia, *Optimal binning: mathematical programming formulation*, Preprint (2020), Available at *ArXiv. abs/2001.08025*
[13] M. S. Pepe, T. Cai and G. Longton, *Combining predictors for classification using the area under the receiver operating characteristic curve*, Biometrics 62, 1 (2006), pp. 221–229.
[14] R. Persson, *Weight of evidence transformation in credit scoring models: How does it affect the discriminatory power?* Preprint (2021), Available at http://lup.lub.lu.se/student-papers/record/9066332
[15] M.V. Pomazanov, *Φ i Ψ preobrazovaniia, povy`shaiushchie diskriminiruiushchuiu silu nemonotonny`kh peremenny`kh skoringovoi` modeli*, Upravlenie finansovy`mi riskami 2, 70 (2022), pp. 108-120 [in Russian]
[16] M.V. Pomazanov, *Second-order accuracy metrics for scoring models and their practical use*, Procedia Computer Science 214 (2022), pp 565-572.
[17] M. Pomazanov, *Risk Management Tools to Improve the Efficiency of Lending to Retail Segments*, Risk Management [Working Title] [Internet], 2022, Available at http://dx.doi.org/10.5772/intechopen.108527
[18] M. Rezac and F. Řezáč, *How to Measure the Quality of Credit Scoring Models*, Czech Journal of Economics and Finance 61 (2011), pp. 486-507.
[19] N. Siddiqi, *Credit Risk Scorecards: Developing And Implementing Intelligent Credit Scoring*, Wiley and SAS Business Series, Wiley, 2005.
[20] L. C. Thomas, D. B. Edelman and J. N. Crook, *Credit Scoring and Its Applications*, Society for Industrial and Applied Mathematics, 2002.