AUTOCALIBRATION BY BALANCE CORRECTION IN NONLIFE INSURANCE PRICING

Michel Denuit, Julien Trufin

LIDAM Discussion Paper ISBA
2022 / 41

UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication.html

AUTOCALIBRATION BY BALANCE CORRECTION IN NONLIFE INSURANCE PRICING

Michel Denuit
Institute of Statistics, Biostatistics and Actuarial Science - ISBA
Louvain Institute of Data Analysis and Modeling - LIDAM
UCLouvain
Louvain-la-Neuve, Belgium

Julien Trufin
Department of Mathematics
Université Libre de Bruxelles (ULB)
Brussels, Belgium

December 15, 2022

Abstract

By exploiting massive amounts of data, machine learning techniques provide actuaries with predictors exhibiting high correlation with claim frequencies and severities. However, these predictors generally fail to achieve financial equilibrium and thus do not qualify as pure premiums. Autocalibration effectively addresses this issue since it ensures that every group of policyholders paying the same premium is on average self-financing, as demonstrated by Denuit et al. (2021), Ciatto et al. (2022), Lindholm et al. (2022) and Wüthrich (2022). These authors proposed balance correction as a way to make any candidate premium autocalibrated. The present paper further studies the effect of balance correction on resulting pure premiums. It is shown that this method is also beneficial in terms of out-of-sample, or predictive Tweedie deviance, Bregman divergence as well as concentration curves. The paper then derives conditions ensuring that the initial predictor and its balance-corrected version are ordered in Lorenz order. Finally, criteria are proposed to rank the balance-corrected versions of two competing predictors in the convex order.

Keywords: Tweedie deviance, Bregman divergence, financial equilibrium, convex order, Lorenz order.

1 Introduction and motivation

Numerical evidence suggests that candidate premiums produced by machine learning tools, like boosted regression trees or neural networks, can depart a lot from observed claim totals. See e.g. Denuit et al. (2019) and the references therein. Autocalibration has been proposed as a remedy by Denuit et al. (2021), Denuit and Trufin (2021, 2022), Ciatto et al. (2022), Lindholm et al. (2022) and Wüthrich (2022). The reason is as follows: if premiums are autocalibrated then the amounts collected from policyholders paying the same premium match on average the corresponding claim totals. Subsidizing transfers are thus avoided, helping to prevent adverse selection effects. Balance correction has been proposed by Denuit et al. (2021) as a practical way to make a given predictor autocalibrated. This simple procedure consists in replacing the initial predictor with the conditional expectation of the response given the predictor.

The present paper is motivated by the systematic findings in the numerical results obtained by Ciatto et al. (2022). It turns out on the data sets under consideration that deviance is always reduced when switching to autocalibrated predictors with the help of the balance-correction method. This suggests that balance correction is not only successful in restoring financial equilibrium but also improves the performances of the predictor under consideration.

In this paper, we establish that balance correction indeed decreases out-of-sample, or predictive Tweedie deviance, as well as Bregman divergence and concentration curves. We also derive criteria ensuring that a predictor is dominated by its balance-corrected version in the Lorenz order, and that the balance-corrected versions of two predictors are ordered in the convex order. These results demonstrate the benefits of balance correction, making it an appealing post-processing technique for insurance pricing studies.

The developments in this paper apply in any setting where balance is desirable, that is, where it is important that the sum of estimates do not deviate too much from the sum of actual observations. This naturally translates into a global balance constraint: considering that the total claim figures are representative of next-year's experience, it is important that the sum of fitted premiums match the sum of responses taken as proxy for the total premium income, as closely as possible. In addition to global balance, balance is equally important locally in order to limit premium transfers between different categories of policyholders. Autocalibration appears to be the appropriate tool to translate financial equilibrium into the price list.

The remainder of this paper is organized as follows. Section 2 recalls the basics of insurance pricing and introduces the notation used throughout the paper. Eligible loss functions are discussed as well as corresponding dominance rules. Section 3 recalls the financial equilibrium concept, central to insurance pricing, and the way it translates into balance properties at different granularity levels. This naturally leads to the notion of autocalibration recalled in Section 4. Section 5 is devoted to the balance correction approach for making a given predictor autocalibrated. The merits of this approach are established there, confirming empirical evidence from the previous literature. Sections 6 and 7 derive criteria to compare a predictor and its balance-corrected version in the Lorenz order and two balance-corrected predictors in the convex order.

2 Nonlife insurance pricing

2.1 Pure premiums

Consider a response Y and a set of features $X_1, \ldots, X_p$ gathered in the vector $\boldsymbol{X}$. The dependence structure inside the random vector $(Y, X_1, \ldots, X_p)$ is exploited to extract the information contained in $\boldsymbol{X}$ about Y. In actuarial pricing, the aim is to evaluate the pure premium as accurately as possible. This means that the target is the conditional expectation $\mu(\boldsymbol{X}) = \mathrm{E}[Y|\boldsymbol{X}]$ of the response Y (claim number or claim amount) given the available information $\boldsymbol{X}$. Henceforth, $\mu(\boldsymbol{X})$ is referred to as the true (pure) premium. Notice that in some applications, $\mu(\boldsymbol{X})$ only refers to one component of the pure premium. For instance, working in the frequency-severity decomposition of insurance losses, $\mu(\boldsymbol{X})$ can be either the expected number of insured events or the expected claim size, or severity.

The function $\boldsymbol{x} \mapsto \mu(\boldsymbol{x}) = \mathrm{E}[Y|\boldsymbol{X} = \boldsymbol{x}]$ is unknown to the actuary, and may exhibit a complex behavior in $\boldsymbol{x}$. This is why this function is approximated by a (working, or actual) premium $\boldsymbol{x} \mapsto \pi(\boldsymbol{x})$ with a relatively simple structure compared to the unknown regression function $\boldsymbol{x} \mapsto \mu(\boldsymbol{x})$. When the analyst is working in the frequency-severity decomposition of insurance losses, $\pi(\boldsymbol{x})$ targets the expected number of insured events or the mean claim severity, separately. Once fitted on the training data set using an appropriate learning procedure, this produces estimates, or fitted values $\widehat{\pi}(\boldsymbol{x})$ for $\mu(\boldsymbol{x})$.

The developments in this paper hold the training set as fixed. Precisely, $\widehat{\pi}$ is assumed to have been obtained by minimizing the empirical loss on some training set and all formulas in the text are meant given this training set. The respective performances of competing models can then be assessed on the basis of a validation set $\{(Y_i, \boldsymbol{X}_i),\ i = 1, 2, \ldots, n\}$, that has not been used to obtain $\widehat{\pi}$. Provided the validation set comprises a large number of independent and identically distributed pairs $(Y_i, \boldsymbol{X}_i)$, this allows us to perform calculations with a generic random vector $(Y, \boldsymbol{X})$, independent of, and distributed as the observations contained in the training set. The merits of a given pricing tool can thus be assessed using the pair $\big(\mu(\boldsymbol{X}), \widehat{\pi}(\boldsymbol{X})\big)$.

Henceforth, we consider $Y \geq 0$. We assume that $\widehat{\pi}(\boldsymbol{X})$ as well as the conditional expectation $\mu(\boldsymbol{X})$ are continuous random variables admitting positive probability density functions over $(0, \infty)$. This is generally the case when there is at least one continuous feature contained in the available information $\boldsymbol{X}$ and the function $\widehat{\pi}$ is a continuously increasing function of a real score built from $\boldsymbol{X}$. We denote as

$$F_{\widehat{\pi}}(t) = \mathrm{P}[\widehat{\pi}(\boldsymbol{X}) \leq t],\ \ t \geq 0,$$

the distribution function of $\widehat{\pi}(\boldsymbol{X})$, as $f_{\widehat{\pi}}$ the corresponding probability density function, and as $F_{\widehat{\pi}}^{-1}$ the associated quantile function defined as the inverse of $F_{\widehat{\pi}}$.

2.2 Eligible loss functions for insurance pricing

2.2.1 Consistent loss functions for the mean

The concept of Bregman divergence defines a broad and important class of loss functions when the analyst targets the mean response. For a convex function $\ell(\cdot)$, define the error

measure for the mean as

$$L(y,m) = \ell(y) - \ell(m) - \ell'(m)(y-m). \tag{2.1}$$

The loss function $L(\cdot,\cdot)$ in (2.1) is called Bregman loss function. It is also referred to as q-loss function after Efron (1986). Bregman loss function measures the discrepancy between y and m on two distinct scales: on the ℓ-scale with the difference $\ell(y) - \ell(m)$ and on the original scale with the difference $y - m$. Both differences are then combined using the relative weight $\ell'(m)$. Sometimes, (2.1) is defined in terms of the concave function $-\ell(\cdot)$.

Pure premiums corresponding to conditional expectations, they can be consistently estimated only if the expected loss is minimum for the mean response. Savage (1971) showed, subject to weak regularity conditions, that for any loss function (2.1) and response Y, $\mathrm{E}[L(Y,\mathrm{E}[Y])] \leq \mathrm{E}[L(Y,m)]$ for any m in the support of Y. The class of Bregman loss functions is then said to be consistent for the (conditional) mean functional. It is worth to stress that Savage (1971) refers to insurance to motivate his proposed approach.

Notice that the term $\ell(y)$ is sometimes subtracted in (2.1). This does not modify the predictions and ensures that the loss function is defined over the whole range of the response. Let us also mention that there are classes of loss functions that are consistent with other indicators, such as quantiles or expectiles for instance. We refer the interested reader to Gneiting (2011) for a general overview. Loss functions are sometimes called scoring functions in the statistical literature but we do not adopt this terminology here in order to avoid any confusion with the score involved in the predictor.

2.2.2 Some Bregman loss function of particular interest

In actuarial application targeting the pure premium, every Bregman loss function would be eligible to produce $\widehat{\pi}$. Depending on the situation, the actuary can select a particular function $\ell(\cdot)$ that appears to be meaningful for the problem under consideration and adopts (2.1) as objective function. For appropriate choice of $\ell(\cdot)$, we recover classical loss functions. See e.g. Table 1 in Krüger and Ziegel (2021) as well as Patton (2020) for an overview. Let us mention a few examples that are particularly relevant for insurance studies.

Considering $\ell(m) = m^2$ yields the quadratic loss $L(y,m) = (y-m)^2$. This loss function is widely used in insurance studies. It nicely expresses that the pure premium must match observed losses as accurately as possible. Interpreting $y - m$ as the deficit or surplus on the insurance contract under consideration, squaring it means that positive or negative departures should be avoided and that they are both treated similarly (since the loss function is symmetric).

Quadratic loss corresponds to Normal log-likelihood functions. More generally, the function $\ell(m) = 2\big(m\theta_m - a(\theta_m)\big)$ where the function $a(\cdot)$ is non-decreasing and convex, with $a'(\theta_m) = m$ results in the deviance loss

$$L(y,m) = 2\big(y(\theta_y - \theta_m) - a(\theta_y) + a(\theta_m)\big) \tag{2.2}$$

where $a'(\theta_y) = y$. For example, the QLIKE loss defined by Patton (2011) corresponds to the Negative Exponential distribution. It only depends on the ratio of the response and the prediction, that is, on the loss ratio. It can be shown that the quadratic and QLIKE loss

functions are particular in the sense that they are the only two Bregman loss functions that only depend on the difference (for quadratic) or the ratio (for QLIKE) of the response and the estimated mean.

Bregman loss functions can take a wide variety of shapes: these loss functions can be asymmetric, with either under-predictions or over-predictions being more heavily penalized, and they can be strictly convex or have concave segments. Thus, restricting attention to loss functions that generate the mean as the optimum value does not require imposing symmetry or other assumptions on the loss function. In fact, out of the infinite number of Bregman loss functions, only one is symmetric: the quadratic loss function.

2.3 Tweedie loss functions

Generalized Linear Models (GLMs) with power variance function have been successfully used in many applications. Poisson, Gamma and Inverse Gaussian distributions belong to this family, as well as compound Poisson distributions with Gamma-distributed summands. Within the Tweedie subclass of the Exponential Dispersion family, the logarithm of the probability mass function for a discrete response, or of the probability density function for a continuous response, is of the form $(y\theta - a(\theta))/\phi$ up to a constant term, for some known dispersion parameter ϕ and non-decreasing and convex cumulant function $a(\cdot)$ corresponding to a variance function of the form $V(\mu) = \mu^\xi$ for some power parameter $\xi \geq 1$.

Tweedie deviance is then recovered from (2.2), resulting in

$$L(y,m) \;=\; \begin{cases} 2\left(y \ln \frac{y}{m} - (y-m)\right) \text{ for } \xi = 1 \\ 2\left(-\ln \frac{y}{m} + \frac{y}{m} - 1\right) \text{ for } \xi = 2 \\ 2\left(\frac{y^{2-\xi}}{(1-\xi)(2-\xi)} - \frac{y m^{1-\xi}}{1-\xi} + \frac{m^{2-\xi}}{2-\xi}\right) \text{ for } \xi > 1 \text{ and } \xi \neq 2. \end{cases} \tag{2.3}$$

The deviance loss function (2.3) is thus a Bregman loss function. Patton (2011) proposed family of homogeneous Bregman loss functions on the positive half line corresponding to Tweedie deviances with power parameter $\xi = b + 1$. The QLIKE loss corresponds to $b = 0$.

Performances of $\widehat{\pi}$ can be assessed with the help of the predictive Tweedie deviance corresponding to (2.3) given by

$$D(\xi, \widehat{\pi}) = \begin{cases} \mathrm{E}\left[\widehat{\pi}(\boldsymbol{X}) - Y \ln \widehat{\pi}(\boldsymbol{X})\right] \text{ for } \xi = 1 \\ \mathrm{E}\left[\ln \widehat{\pi}(\boldsymbol{X}) + \frac{Y}{\widehat{\pi}(\boldsymbol{X})}\right] \text{ for } \xi = 2 \\ \mathrm{E}\left[\frac{\widehat{\pi}(\boldsymbol{X})^{2-\xi}}{2-\xi} - \frac{Y \widehat{\pi}(\boldsymbol{X})^{1-\xi}}{1-\xi}\right] \text{ for } \xi > 1 \text{ and } \xi \neq 2 \end{cases} \tag{2.4}$$

where the terms not involving $\widehat{\pi}(\boldsymbol{X})$ have been discarded.

2.4 Bregman and Tweedie dominances

Bregman dominance is defined as dominance for every Bregman loss function (2.1). Precisely, $\widehat{\pi}_2$ outperforms $\widehat{\pi}_1$ in terms of Bregman dominance if the inequality $\mathrm{E}[L(Y, \widehat{\pi}_2(\boldsymbol{X}))] \leq$

$\mathrm{E}[L(Y,\widehat{\pi}_1(\boldsymbol{X}))]$ holds true for every loss function L of the form given in (2.1). Bregman dominance is thus a particular stochastic order relation used to compare the performances of two estimators $\widehat{\pi}_1$ and $\widehat{\pi}_2$ for the conditional means. We refer the interested reader to Krüger and Ziegel (2021) and the references therein for an extensive presentation of this concept, as well as to Shaked and Shanthikumar (2007) for a general presentation of stochastic order relations and to Denuit et al. (2005) for applications to insurance.

Krüger and Ziegel (2021) established in their Theorem 2.1 that $\widehat{\pi}_2$ outperforms $\widehat{\pi}_1$ in terms of Bregman dominance if, and only if, the inequality

$$\mathrm{E}[(\widehat{\pi}_2(\boldsymbol{X})-t)_+]+\mathrm{E}\big[(Y-\widehat{\pi}_2(\boldsymbol{X}))\mathrm{I}[\widehat{\pi}_2(\boldsymbol{X})>t]\big] \geq \mathrm{E}[(\widehat{\pi}_1(\boldsymbol{X})-t)_+]+\mathrm{E}\big[(Y-\widehat{\pi}_1(\boldsymbol{X}))\mathrm{I}[\widehat{\pi}_1(\boldsymbol{X})>t]\big] \tag{2.5}$$

holds for all t, where $\mathrm{I}[\cdot]$ is the indicator function (equal to 1 if the event appearing within the brackets is realized and to 0 otherwise). Both sides of the latter inequality involve stop-loss premiums $\mathrm{E}[(\widehat{\pi}_k(\boldsymbol{X})-t)_+]$ of the predictors $\widehat{\pi}_k$ under consideration as well as expected deficits $Y-\widehat{\pi}_k(\boldsymbol{X})$ on the contracts with premium exceeding the threshold t.

Tweedie dominance is defined as dominance for every Tweedie deviances (2.4). Precisely, $\widehat{\pi}_2$ outperforms $\widehat{\pi}_1$ in terms of Tweedie dominance if the inequality $D(\xi,\widehat{\pi}_2) \leq D(\xi,\widehat{\pi}_1)$ holds true for every power parameter $\xi \geq 1$. Tweedie dominance appears to be a particular case of Bregman dominance. Tweedie dominance is thus another particular stochastic order relation used to compare the performances of two estimators $\widehat{\pi}_1$ and $\widehat{\pi}_2$ for the conditional means.

3 Financial equilibrium and balance properties

Financial equilibrium is an important property in insurance pricing. Since risk classification is meant to vary the amounts of premium charged to policyholders, with little to no effect on claims, it means that the insurer must collect the same amount of premiums for the entire portfolio. Furthermore, to prevent adverse selection effects, the same property must hold locally, for meaningful groups of policyholders.

Financial equilibrium must therefore be imposed at different levels:

global balance ensures that individual fluctuations $Y-\widehat{\pi}(\boldsymbol{X})$ cancel out in a sufficiently large portfolio comprising independent risks. This translates into the global balance equation

$$\mathrm{E}[Y-\widehat{\pi}(\boldsymbol{X})]=0.$$

Considering the deficit $Y-\widehat{\pi}(\boldsymbol{X})$ on the contract, this guarantees financial equilibrium of insurer's operations. Stated alternatively, this means that pricing errors $\mu(\boldsymbol{X})-\widehat{\pi}(\boldsymbol{X})$ also cancel out, that is,

$$\mathrm{E}[\mu(\boldsymbol{X})-\widehat{\pi}(\boldsymbol{X})]=0.$$

local balance prevents premium transfers from one group of policyholders to another, ensuring that premiums are competitive and avoid subsidizing transfers, helping to limit adverse selection effects. Depending on the groups over which local balance is required, this leads to different notions:

marginal balance corresponds to the method of marginal totals as well as to likelihood equations in GLMs with canonical link functions and binary features X_j. If an intercept is included in the GLM score then global balance is automatically fulfilled. Furthermore, local balance also holds among policyholders with the same value for X_j, since maximum likelihood equations write

$$\mathrm{E}[Y - \widehat{\pi}_{\mathrm{glm}}(\boldsymbol{X})|X_j] = 0 \text{ for } j = 1, 2, \ldots, p.$$

Now,

$$\mathrm{E}[Y - \widehat{\pi}_{\mathrm{glm}}(\boldsymbol{X})|X_j] = \mathrm{E}[\mu(\boldsymbol{X}) - \widehat{\pi}_{\mathrm{glm}}(\boldsymbol{X})|X_j],$$

so that the same reasoning applies to pricing errors.

autocalibration It would be natural to impose that every group of policyholders paying the same premium is on average self-financing. This would ensure that there is no systematic cross-financing between groups paying different amounts of premium, preventing any subsidizing solidarity and adverse selection effects. Translated into mathematical terms, this gives

$$\mathrm{E}[Y - \widehat{\pi}(\boldsymbol{X})|\widehat{\pi}(\boldsymbol{X})] = 0.$$

Stated alternatively, this means that pricing errors $\mu(\boldsymbol{X}) - \widehat{\pi}(\boldsymbol{X})$ also cancel out, that is,

$$\mathrm{E}[\mu(\boldsymbol{X}) - \widehat{\pi}(\boldsymbol{X})|\widehat{\pi}(\boldsymbol{X})] = 0.$$

This requirement corresponds to autocalibration. This concept is reviewed in the next section.

4 Autocalibration

4.1 Definition

Recall that a predictor $\widehat{\pi}$ is said to be autocalibrated if $\widehat{\pi}(\boldsymbol{X}) = \mathrm{E}[Y|\widehat{\pi}(\boldsymbol{X})]$. Autocalibration thus ensures that the average response in a neighborhood defined with the help of the autocalibrated predictor under consideration matches the latter. See, e.g., Krüger and Ziegel (2021). As explained previously, autocalibration is an important property in insurance pricing because it ensures that every group of policyholders paying the same premium is on average self-financing: the premium paid exactly covers the expected claim of that group.

4.2 Properties

Clearly, the more $\widehat{\pi}(\boldsymbol{X})$ is dispersed, the more information it contains about the true premium. Thus, comparing the underlying variability appears to be important in the problem under study. The convex order is an effective probabilistic tool to assess the dispersion of random variables, beyond simple indicators such as standard deviations. Recall from Denuit et al. (2005, Section 3.4) that given two random variables Z and T, T is said to be smaller than Z in the convex order (denoted as $T \preceq_{\mathrm{CX}} Z$) if

$$\mathrm{E}[g(T)] \leq \mathrm{E}[g(Z)] \text{ for all convex functions } g, \tag{4.1}$$

provided the expectations exist. It is well known that $\preceq_{\mathrm{CX}}$ can be characterized with the help of stop-loss premiums. Precisely,

$$T \preceq_{\mathrm{CX}} Z \Leftrightarrow \mathrm{E}[T] = \mathrm{E}[Z] \text{ and } \mathrm{E}[(T-d)_+] \leq \mathrm{E}[(Z-d)_+] \text{ for all } d. \tag{4.2}$$

Autocalibrated predictors are such that

$$\widehat{\pi}(\boldsymbol{X}) \preceq_{\mathrm{CX}} \mu(\boldsymbol{X}) \preceq_{\mathrm{CX}} Y.$$

The first stochastic inequality has been obtained in Lemma 1 from Wüthrich (2022). The second one is well known and directly follows from Jensen inequality for conditional expectations. Thus, autocalibration implies that the predictor is less variable than the true premium $\mu(\boldsymbol{X})$, and a fortiori than the response Y, in the sense of the convex order. This important property means that premiums charged by the insurance company always induce softer relativities compared to the true ones.

Autocalibrated predictors fulfill the following useful property.

Property 4.1. *For any autocalibrated predictor $\widehat{\pi}$ and measurable function $g : \mathbb{R}^+ \to \mathbb{R}$, we have*

$$\mathrm{E}\big[Y g\big(\widehat{\pi}(\boldsymbol{X})\big)\big] = \mathrm{E}\Big[\widehat{\pi}(\boldsymbol{X}) g\big(\widehat{\pi}(\boldsymbol{X})\big)\Big].$$

Proof. This result follows from the tower property of conditional expectation, which gives

$$\begin{aligned}
\mathrm{E}\big[Y g\big(\widehat{\pi}(\boldsymbol{X})\big)\big] &= \mathrm{E}\Big[\mathrm{E}\big[Y g\big(\widehat{\pi}(\boldsymbol{X})\big)\big|\widehat{\pi}(\boldsymbol{X})\big]\Big] \\
&= \mathrm{E}\Big[\mathrm{E}\big[Y|\widehat{\pi}(\boldsymbol{X})\big] g\big(\widehat{\pi}(\boldsymbol{X})\big)\Big] \\
&= \mathrm{E}\Big[\widehat{\pi}(\boldsymbol{X}) g\big(\widehat{\pi}(\boldsymbol{X})\big)\Big],
\end{aligned}$$

where the last step follows from autocalibration. This ends the proof. □

Property 4.1 then gives the following corollaries.

Corollary 4.2. *Bregman dominance reduces to the convex order for autocalibrated predictors, as pointed out by Krüger and Ziegel (2021) in their Theorem 3.1. This is easily deduced from Property 4.1 applied to* (2.5) *since*

$$\mathrm{E}\big[Y \mathrm{I}[\widehat{\pi}(\boldsymbol{X}) > t]\big] = \mathrm{E}\big[\widehat{\pi}(\boldsymbol{X}) \mathrm{I}[\widehat{\pi}(\boldsymbol{X}) > t]\big]$$

so that

$$\mathrm{E}\big[(Y - \widehat{\pi}(\boldsymbol{X})) \mathrm{I}[\widehat{\pi}(\boldsymbol{X}) > t]\big] = 0$$

for every autocalibrated predictor. Only the stop-loss premiums remain in (2.5) *so that we recover the stated result from* (4.2).

Corollary 4.3. *The concentration curve of the response Y with respect to the predictor $\widehat{\pi}$ based on the information contained in the vector $\boldsymbol{X}$ is defined as*

$$\mathrm{CC}[Y, \widehat{\pi}(\boldsymbol{X}); \alpha] = \frac{\mathrm{E}\big[Y \mathrm{I}[\widehat{\pi}(\boldsymbol{X}) \leq F_{\widehat{\pi}}^{-1}(\alpha)]\big]}{\mathrm{E}[Y]}$$

for a probability level α. These curves turn out to be useful to assess the performances of $\widehat{\pi}$, as demonstrated in Denuit et al. (2019). We refer the interested reader to Yitzhaki and Schechtman (2013) for a thorough description of these curves. Up to sign switches, concentration curves are referred to as cumulative accuracy profiles (CAP) in binary classification, see Wüthrich (2022). The Lorenz curve associated to the predictor $\widehat{\pi}$ is defined by

$$\begin{aligned}\mathrm{LC}[\widehat{\pi}(\boldsymbol{X});\alpha] &= \frac{1}{\mathrm{E}[\widehat{\pi}(\boldsymbol{X})]}\int_0^{\alpha} F_{\widehat{\pi}}^{-1}(u)\,du \\ &= \frac{\mathrm{E}\big[\widehat{\pi}(\boldsymbol{X})\mathrm{I}[\widehat{\pi}(\boldsymbol{X})\leq F_{\widehat{\pi}}^{-1}(\alpha)]\big]}{\mathrm{E}[\widehat{\pi}(\boldsymbol{X})]}\end{aligned}$$

for a given probability level α. Now, Property 4.1 gives

$$\mathrm{CC}[Y,\widehat{\pi}(\boldsymbol{X});\alpha] = \frac{\mathrm{E}\big[\widehat{\pi}(\boldsymbol{X})\mathrm{I}[\widehat{\pi}(\boldsymbol{X})\leq F_{\widehat{\pi}}^{-1}(\alpha)]\big]}{\mathrm{E}[Y]} = \mathrm{LC}[Y,\widehat{\pi}(\boldsymbol{X});\alpha]$$

when $\widehat{\pi}$ is autocalibrated. Hence, concentration and Lorenz curves coincide for autocalibrated predictors as established by Denuit and Trufin (2021).

5 Balance correction

5.1 Balance correction as a way to restore financial equilibrium

Predictors obtained by modern machine learning tools such as boosted regression trees or neural networks are generally not autocalibrated. It is therefore useful to have a practical way to make a given predictor autocalibrated. An effective method has been proposed by Denuit et al. (2021), referred to as balance corrections.

The candidate premiums $\widehat{\pi}$ can be auto-calibrated by switching to its balance-corrected version $\widehat{\pi}_{\mathrm{bc}}$ defined as

$$\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X}) = \mathrm{E}[Y|\widehat{\pi}(\boldsymbol{X})].$$

The predictor $\widehat{\pi}_{\mathrm{bc}}$ obtained in this way is always autocalibrated as shown in Proposition 3.3 in Wüthrich (2022).

Let us present two examples of balance-correction techniques.

Example 5.1 (Lift predictor)**.** Lift charts are often used in practice to assess the performances of a candidate premium. To draw such graphs, the data set is sorted based on the values of $\widehat{\pi}(\boldsymbol{X})$. The data are then bucketed into equally populated classes based on quantiles. Within each bucket, the average expected response is calculated with the help of $\widehat{\pi}(\boldsymbol{X})$ as well as the actual loss Y. The average predictions and average costs are then graphed for each class. Lift charts allow the analyst to check whether the actual response monotonically increases as we move to higher buckets. Property 3.5 in Denuit et al. (2019) identifies the condition required to observe an increasing trend in such a lift chart (that is, positive regression dependence).

Let $a_0 = 0 < a_1 < a_2 < \ldots$ and partition $(0,\infty)$ into intervals $I_j = (a_{j-1}, a_j)$. Define

$$\widehat{\pi}_{\mathrm{lift}}(\boldsymbol{x}) = \sum_{j\geq 1} \mathrm{E}[Y|\widehat{\pi}(\boldsymbol{x})\in I_j]\mathrm{I}[\widehat{\pi}(\boldsymbol{x})\in I_j].$$

If the conditional expectations $\mathrm{E}[Y|\widehat{\pi}(\boldsymbol{x}) \in I_j]$ are increasing in j, which appears to be a reasonable requirement for any candidate premium $\widehat{\pi}$, we then have

$$\mathrm{E}[Y|\widehat{\pi}_{\text{lift}}(\boldsymbol{X})] = \widehat{\pi}_{\text{lift}}(\boldsymbol{X})$$

so that $\widehat{\pi}_{\text{lift}}$ is autocalibrated. Notice that this autocalibrated predictor is essentially the same as the one given in Definition 1 of Lindholm et al. (2022).

Note that this example considers a discrete predictor, departing from the continuity assumption made throughout the text. Despite assuming a finite number of values, the construction relates to lift diagrams and justifies the approach proposed in the next example, leading to continuous predictors.

Example 5.2 (Moving average predictor)**.** A locally-constant regression, or moving average approach allows the analyst to implement auto-calibration, that is, to switch from $\widehat{\pi}$ to $\widehat{\pi}_{\text{bc}}$. See e.g. Loader (1999) for a detailed account. The difference compared to Example 5.1 is that the disjoint intervals I_j are replaced with a moving window over which responses are averaged. In fact, the score of the model is enough to compute $\widehat{\pi}_{\text{bc}}$ so that the learning model is essentially used to assess proximity among individuals, before performing local averaging.

The balance-corrected $\widehat{\pi}_{\text{bc}}$ can be obtained by local regression in the appropriate GLM, recognizing the nature of the response Y: local Poisson GLM for claim counts, local Binomial GLM for propensities, local Gamma GLM for average claim sizes and local Tweedie GLM for claim totals. The choice of the response distribution is thus dictated by the format of the response. The canonical link function is adopted so that maximum-likelihood estimates replicate observed totals. An intercept-only GLM is fitted locally, on a set of observations close to the one of interest with proximity assessed with the help of $\widehat{\pi}$. Precisely, $\widehat{\pi}_{\text{bc}}(\boldsymbol{x})$ is obtained by averaging a percentage of the observed responses y_i whose predictions $\widehat{\pi}(\boldsymbol{x}_i)$ are the closest to $\widehat{\pi}(\boldsymbol{x})$. This percentage is obtained by likelihood cross-validation in the appropriate GLM.

The following result turns out to be useful to derive the results contained in the next sections.

Property 5.3. *For any predictor $\widehat{\pi}$ and measurable function $g : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$, we have*

$$\mathrm{E}\big[Y g\big(\widehat{\pi}(\boldsymbol{X}), \widehat{\pi}_{\text{bc}}(\boldsymbol{X})\big)\big] = \mathrm{E}\left[\widehat{\pi}_{\text{bc}}(\boldsymbol{X}) g\big(\widehat{\pi}(\boldsymbol{X}), \widehat{\pi}_{\text{bc}}(\boldsymbol{X})\big)\right].$$

Proof. The tower property of the conditional expectation gives

$$\begin{aligned}
\mathrm{E}\big[Y g\big(\widehat{\pi}(\boldsymbol{X}), \widehat{\pi}_{\text{bc}}(\boldsymbol{X})\big)\big] &= \mathrm{E}\Big[\mathrm{E}\big[Y g\big(\widehat{\pi}(\boldsymbol{X}), \widehat{\pi}_{\text{bc}}(\boldsymbol{X})\big)\big|\widehat{\pi}(\boldsymbol{X})\big]\Big] \\
&= \mathrm{E}\Big[\mathrm{E}\big[Y|\widehat{\pi}(\boldsymbol{X})\big] g\big(\widehat{\pi}(\boldsymbol{X}), \widehat{\pi}_{\text{bc}}(\boldsymbol{X})\big)\Big] \\
&= \mathrm{E}\left[\widehat{\pi}_{\text{bc}}(\boldsymbol{X}) g\big(\widehat{\pi}(\boldsymbol{X}), \widehat{\pi}_{\text{bc}}(\boldsymbol{X})\big)\right],
\end{aligned}$$

which ends the proof. □

5.2 Impact of balance correction on predictive Tweedie deviance

The numerical illustrations in Ciatto et al. (2022) suggest that predictive deviance may be smaller for $\widehat{\pi}_{\text{bc}}$ compared to $\widehat{\pi}$. This is indeed the case. The next result shows that switching from $\widehat{\pi}$ to $\widehat{\pi}_{\text{bc}}$ reduces the predictive Tweedie deviance.

Proposition 5.4. *For any power parameter $\xi \geq 1$, we have $D(\xi, \widehat{\pi}) \geq D(\xi, \widehat{\pi}_{\text{bc}})$.*

Proof. For $\xi = 1$, we get from Property 5.3 that

$$\begin{aligned}
D(\xi,\widehat{\pi}) &= \mathrm{E}\left[\widehat{\pi}(\boldsymbol{X}) - Y \ln \widehat{\pi}(\boldsymbol{X})\right] \\
&= \mathrm{E}\left[\widehat{\pi}_{\text{bc}}(\boldsymbol{X}) - Y \ln \widehat{\pi}_{\text{bc}}(\boldsymbol{X})\right] + \mathrm{E}\left[\widehat{\pi}(\boldsymbol{X}) - \widehat{\pi}_{\text{bc}}(\boldsymbol{X}) - Y \ln\left(\frac{\widehat{\pi}(\boldsymbol{X})}{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})}\right)\right] \\
&= D(\xi,\widehat{\pi}_{\text{bc}}) + \mathrm{E}\left[\widehat{\pi}(\boldsymbol{X}) - \widehat{\pi}_{\text{bc}}(\boldsymbol{X}) - \widehat{\pi}_{\text{bc}}(\boldsymbol{X}) \ln\left(\frac{\widehat{\pi}(\boldsymbol{X})}{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})}\right)\right] \\
&= D(\xi,\widehat{\pi}_{\text{bc}}) + \mathrm{E}\left[\widehat{\pi}_{\text{bc}}(\boldsymbol{X})\left(\frac{\widehat{\pi}(\boldsymbol{X})}{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})} - 1 - \ln\left(\frac{\widehat{\pi}(\boldsymbol{X})}{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})}\right)\right)\right] \\
&\geq D(\xi,\widehat{\pi}_{\text{bc}})
\end{aligned}$$

since $y \to y - 1 - \ln y$ is positive on $\mathbb{R}^+$.

Turning to the case $\xi = 2$, Property 5.3 leads to

$$\begin{aligned}
D(\xi,\widehat{\pi}) &= \mathrm{E}\left[\ln \widehat{\pi}(\boldsymbol{X}) + \frac{Y}{\widehat{\pi}(\boldsymbol{X})}\right] \\
&= \mathrm{E}\left[\ln \widehat{\pi}_{\text{bc}}(\boldsymbol{X}) + \frac{Y}{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})}\right] + \mathrm{E}\left[\ln\left(\frac{\widehat{\pi}(\boldsymbol{X})}{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})}\right) + \frac{Y}{\widehat{\pi}(\boldsymbol{X})} - \frac{Y}{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})}\right] \\
&= D(\xi,\widehat{\pi}_{\text{bc}}) + \mathrm{E}\left[\frac{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})}{\widehat{\pi}(\boldsymbol{X})} - 1 - \ln\left(\frac{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})}{\widehat{\pi}(\boldsymbol{X})}\right)\right] \\
&\geq D(\xi,\widehat{\pi}_{\text{bc}})
\end{aligned}$$

since $y \to y - 1 - \ln y$ is positive on $\mathbb{R}^+$.

Finally for the remaining cases for ξ, we get from Property 5.3 that

$$\begin{aligned}
D(\xi,\widehat{\pi}) &= \mathrm{E}\left[\frac{\widehat{\pi}(\boldsymbol{X})^{2-\xi}}{2-\xi} - \frac{Y\widehat{\pi}(\boldsymbol{X})^{1-\xi}}{1-\xi}\right] \\
&= \mathrm{E}\left[\frac{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})^{2-\xi}}{2-\xi} - \frac{Y\widehat{\pi}_{\text{bc}}(\boldsymbol{X})^{1-\xi}}{1-\xi}\right] \\
&\quad + \mathrm{E}\left[\frac{\widehat{\pi}(\boldsymbol{X})^{2-\xi} - \widehat{\pi}_{\text{bc}}(\boldsymbol{X})^{2-\xi}}{2-\xi} - \frac{Y(\widehat{\pi}(\boldsymbol{X})^{1-\xi} - \widehat{\pi}_{\text{bc}}(\boldsymbol{X})^{1-\xi})}{1-\xi}\right] \\
&= D(\xi,\widehat{\pi}_{\text{bc}}) + \mathrm{E}\left[\frac{\widehat{\pi}(\boldsymbol{X})^{2-\xi} - \widehat{\pi}_{\text{bc}}(\boldsymbol{X})^{2-\xi}}{2-\xi} - \frac{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})\widehat{\pi}(\boldsymbol{X})^{1-\xi} - \widehat{\pi}_{\text{bc}}(\boldsymbol{X})^{2-\xi}}{1-\xi}\right] \\
&= D(\xi,\widehat{\pi}_{\text{bc}}) + \mathrm{E}\left[\widehat{\pi}_{\text{bc}}(\boldsymbol{X})^{2-\xi}\left(\frac{\left(\frac{\widehat{\pi}(\boldsymbol{X})}{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})}\right)^{2-\xi} - 1}{2-\xi} - \frac{\left(\frac{\widehat{\pi}(\boldsymbol{X})}{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})}\right)^{1-\xi} - 1}{1-\xi}\right)\right] \\
&= D(\xi,\widehat{\pi}_{\text{bc}}) + \mathrm{E}\left[\widehat{\pi}_{\text{bc}}(\boldsymbol{X})^{2-\xi} h\left(\frac{\widehat{\pi}(\boldsymbol{X})}{\widehat{\pi}_{\text{bc}}(\boldsymbol{X})}\right)\right]
\end{aligned}$$

where the function $h(\cdot)$ is defined for $z > 0$ as $h(z) = \frac{z^{2-\xi}}{2-\xi} - \frac{z^{1-\xi}}{1-\xi} + \frac{1}{(2-\xi)(1-\xi)}$. We have

$$h'(z) = z^{-\xi}(z-1) \text{ and } h''(z) = z^{-\xi}(1-\xi+\xi z^{-1}).$$

Therefore, for $z > 0$, $h'(z) = 0 \Leftrightarrow z = 1$ and $h''(1) = 1$, so that $h(z) \geq h(1) = 0$ for all $z > 0$ and thus

$$\mathrm{E}\left[\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})^{2-\xi} h\left(\frac{\widehat{\pi}(\boldsymbol{X})}{\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})}\right)\right] \geq 0,$$

which completes the proof. □

Proposition 5.4 extends to the whole class of Tweedie deviances the result obtained for the quadratic loss function in Proposition 1 from Lindholm et al. (2022).

5.3 Impact of balance correction on Bregman divergence

The results derived for Tweedie deviance in Proposition 5.4 extends to the whole class of Bregman loss functions, as shown next

Proposition 5.5. *For any Bregman loss function L, we have* $\mathrm{E}\big[L\big(Y, \widehat{\pi}(\boldsymbol{X})\big)\big] \geq \mathrm{E}\big[L\big(Y, \widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})\big)\big]$.

Proof. We have

$$\begin{aligned}
\mathrm{E}\big[L\big(Y, \widehat{\pi}(\boldsymbol{X})\big)\big] &= \mathrm{E}\big[\ell(Y) - \ell(\widehat{\pi}(\boldsymbol{X})) - \ell'(\widehat{\pi}(\boldsymbol{X}))\big(Y - \widehat{\pi}(\boldsymbol{X})\big)\big] \\
&= \mathrm{E}\big[\ell(Y) - \ell(\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})) - \ell'(\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X}))\big(Y - \widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})\big)\big] \\
&\quad + \mathrm{E}\big[\ell(\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})) - \ell(\widehat{\pi}(\boldsymbol{X})) - \ell'(\widehat{\pi}(\boldsymbol{X}))\big(Y - \widehat{\pi}(\boldsymbol{X})\big)\big] \\
&\quad + \mathrm{E}\big[\ell'(\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X}))\big(Y - \widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})\big)\big] \\
&= \mathrm{E}\big[L\big(Y, \widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})\big)\big] + \mathrm{E}\big[\ell(\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})) - \ell(\widehat{\pi}(\boldsymbol{X})) - \ell'(\widehat{\pi}(\boldsymbol{X}))\big(Y - \widehat{\pi}(\boldsymbol{X})\big)\big]
\end{aligned}$$

where the last equality follows from Property 5.3. Now, since $\ell(\cdot)$ is a convex function, we know that $\ell(z_1) - \ell(z_2) - \ell'(z_2)(z_1 - z_2) \geq 0$ for any real values z_1 and z_2, so that we have from Property 5.3

$$\begin{aligned}
&\mathrm{E}\big[\ell(\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})) - \ell(\widehat{\pi}(\boldsymbol{X})) - \ell'(\widehat{\pi}(\boldsymbol{X}))\big(Y - \widehat{\pi}(\boldsymbol{X})\big)\big] \\
&= \mathrm{E}\big[\ell(\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})) - \ell(\widehat{\pi}(\boldsymbol{X})) - \ell'(\widehat{\pi}(\boldsymbol{X}))\big(\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X}) - \widehat{\pi}(\boldsymbol{X})\big)\big] \geq 0,
\end{aligned}$$

which completes the proof. □

5.4 Impact of balance correction on concentration curves

Balance correction is also beneficial in terms of concentration curves, as shown next.

Proposition 5.6. *We have* $\mathrm{CC}[Y, \widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X}); \alpha] = \mathrm{LC}[\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X}); \alpha] \leq \mathrm{CC}[Y, \widehat{\pi}(\boldsymbol{X}); \alpha]$ *for any probability level* α.

Proof. Let us first establish that $\mathrm{E}[Z|A] \geq \mathrm{E}[Z|Z \leq z]$ for any continous random variable Z and event A such that $\mathrm{P}[Z \leq z] = \mathrm{P}[A]$. This comes from

$$\begin{aligned}
\mathrm{E}[Z|Z \leq z] &= z + \mathrm{E}[Z - z|Z \leq z, A]\mathrm{P}[A|Z \leq z] \\
&\quad + \mathrm{E}[Z - z|Z \leq z, \overline{A}]\mathrm{P}[\overline{A}|Z \leq z] \\
&\leq z + \mathrm{E}[Z - z|Z \leq z, A]\mathrm{P}[A|Z \leq z] \\
&= z + \mathrm{E}[Z - z|Z \leq z, A]\mathrm{P}[Z \leq z|A] \\
&\leq z + \mathrm{E}[Z - z|Z \leq z, A]\mathrm{P}[Z \leq z|A] \\
&\quad + \mathrm{E}[Z - z|Z > z, A]\mathrm{P}[Z > z|A] = \mathrm{E}[Z|A].
\end{aligned}$$

We then also have $\mathrm{E}\big[Z\mathrm{I}[A]\big] \geq \mathrm{E}\big[Z\mathrm{I}[Z \leq z]\big]$. Applying this result with $Z = \widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})$, $\widehat{\pi}(\boldsymbol{X}) \leq F_{\widehat{\pi}}^{-1}(\alpha)$ for A and $z = F_{\widehat{\pi}_{\mathrm{bc}}}^{-1}(\alpha)$, we get

$$\mathrm{E}\Big[\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})\mathrm{I}\big[\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X}) \leq F_{\widehat{\pi}_{\mathrm{bc}}}^{-1}(\alpha)\big]\Big] \leq \mathrm{E}\Big[\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})\mathrm{I}\big[\widehat{\pi}(\boldsymbol{X}) \leq F_{\widehat{\pi}}^{-1}(\alpha)\big]\Big].$$

Property 5.3 allows us to replace $\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})$ with Y in the latter inequality, so that we end up with the announced result. □

6 Comparing a predictor with its balance-corrected version

Balance correction is desirable in insurance pricing because it implements financial equilibrium, as explained before. In this section, we examine whether balance correction also makes the initial predictor more informative. Since expected values of $\widehat{\pi}(\boldsymbol{X})$ and $\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})$ generally differ, we cannot compare them with the help of $\preceq_{\mathrm{CX}}$ but $\frac{\mathrm{E}[Y]}{\mathrm{E}[\widehat{\pi}(\boldsymbol{X})]}\widehat{\pi}(\boldsymbol{X})$ and $\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})$ could possibly be ranked in the convex order. Here, the factor $\frac{\mathrm{E}[Y]}{\mathrm{E}[\widehat{\pi}(\boldsymbol{X})]}$ implements global balance. This corresponds to the Lorenz order $\preceq_{\mathrm{Lorenz}}$. Formally,

$$\widehat{\pi}(\boldsymbol{X}) \preceq_{\mathrm{Lorenz}} \widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X}) \Leftrightarrow \frac{\mathrm{E}[Y]}{\mathrm{E}[\widehat{\pi}(\boldsymbol{X})]}\widehat{\pi}(\boldsymbol{X}) \preceq_{\mathrm{CX}} \widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X}) \Leftrightarrow \frac{\widehat{\pi}(\boldsymbol{X})}{\mathrm{E}[\widehat{\pi}(\boldsymbol{X})]} \preceq_{\mathrm{CX}} \frac{\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})}{\mathrm{E}[\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})]}.$$

Lorenz order can be equivalently defined by pointwise comparison of Lorenz curves.

Balance correction consists in replacing the predictor $\widehat{\pi}$ with a function of it, corresponding to the conditional expectation of the response given the predictor. The class of functions inducing Lorenz dominance is known, leading to the following result which is deduced from Theorem 3.A.25 in Shaked and Shanthikumar (2007).

Proposition 6.1. *Define the function $m(\cdot)$ as $m(t) = \mathrm{E}[Y|\widehat{\pi}(\boldsymbol{X}) = t]$. Then,*

$$t \mapsto \frac{m(t)}{t} \text{ non-decreasing} \Rightarrow \widehat{\pi}(\boldsymbol{X}) \preceq_{\mathrm{Lorenz}} \widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X}).$$

The sufficient condition in Proposition 6.1 is best explained using loss ratios. Indeed, it means that

$$\frac{m(t)}{t} = \mathrm{E}\left[\frac{Y}{\widehat{\pi}(\boldsymbol{X})}\Big|\widehat{\pi}(\boldsymbol{X}) = t\right]$$

is non-decreasing in t. Higher values of the predictor $\widehat{\pi}$ thus identify profiles with larger loss ratios.

The sufficient condition in Proposition 6.1 implies that $m(\cdot)$ is non-decreasing. This property is referred to as regression dependence. We know from Property 3.5 in Denuit et al. (2019) that this is equivalent to the convexity of the concentration curve. If $m(\cdot)$ is continuous and strictly increasing then

$$\mathrm{CC}[Y,\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X});\alpha]=\mathrm{CC}[Y,\widehat{\pi}(\boldsymbol{X});\alpha] \text{ for all probability levels } \alpha$$

since $\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})\leq F^{-1}_{\widehat{\pi}_{\mathrm{bc}}}(\alpha)$ if, and only if, $\widehat{\pi}(\boldsymbol{X})\leq F^{-1}_{\widehat{\pi}}(\alpha)$ for any probability level α in this case. Moreover, $\widehat{\pi}_{\mathrm{bc}}$ being autocalibrated, its concentration curve is equal to its Lorenz curve. This leads to the following result.

Proposition 6.2. *If $m(\cdot)$ is continuous and strictly increasing then*

$$\widehat{\pi}(\boldsymbol{X})\preceq_{\mathrm{Lorenz}}\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})\Leftrightarrow \mathrm{LC}[\widehat{\pi}(\boldsymbol{X});\alpha]\geq \mathrm{CC}[Y,\widehat{\pi}(\boldsymbol{X});\alpha] \textit{ for all probability levels } \alpha.$$

Notice that in the unlikely case where the opposite inequality holds in Proposition 6.2, we end up with $\widehat{\pi}_{\mathrm{bc}}(\boldsymbol{X})\preceq_{\mathrm{Lorenz}}\widehat{\pi}(\boldsymbol{X})$.

7 Comparing balance-corrected predictors

Assume that we have two predictors $\widehat{\pi}_1$ and $\widehat{\pi}_2$. Both attempt to predict the unknown pure premium. These predictors may differ in their functional form and/or in the information on which they are based. Since balance corrections essentially consists in using the candidate premium to rank risk profiles from the cheapest one to the most expensive one, it is clear that if $\widehat{\pi}_1$ and $\widehat{\pi}_2$ are comonotonic then their ranks coincide and $\widehat{\pi}_{\mathrm{bc},1}=\widehat{\pi}_{\mathrm{bc},2}$.

For any reasonable candidate premium, regression dependence must hold true, that is, the conditional expectation $m(\cdot)$ must be non-decreasing. The next result then identifies a situation where the balance-corrected version of two predictors are ordered in the convex sense (so that Bregman dominance holds since they are both autocalibrated).

Proposition 7.1. *Let $m_k(t)=\mathrm{E}[Y|\widehat{\pi}_k(\boldsymbol{X})=t]$ for $k\in\{1,2\}$. If $m_1(\cdot)$ and $m_2(\cdot)$ are continuous and strictly increasing then*

$$\mathrm{CC}[Y,\widehat{\pi}_2(\boldsymbol{X});\alpha]\leq \mathrm{CC}[Y,\widehat{\pi}_1(\boldsymbol{X});\alpha] \textit{ for all } \alpha \Rightarrow \widehat{\pi}_{\mathrm{bc},1}(\boldsymbol{X})\preceq_{\mathrm{CX}}\widehat{\pi}_{\mathrm{bc},2}(\boldsymbol{X}).$$

Proof. By the same reasoning as for Proposition 6.2, we have

$$\mathrm{CC}[Y,\widehat{\pi}_k(\boldsymbol{X});\alpha]=\mathrm{CC}[Y,\widehat{\pi}_{\mathrm{bc},k}(\boldsymbol{X});\alpha]=\mathrm{LC}[\widehat{\pi}_{\mathrm{bc},k}(\boldsymbol{X});\alpha] \text{ for all probability levels } \alpha,$$

from which it follows that $\widehat{\pi}_{\mathrm{bc},1}(\boldsymbol{X})\preceq_{\mathrm{Lorenz}}\widehat{\pi}_{\mathrm{bc},2}(\boldsymbol{X})$ holds true. The announced result then follows since $\mathrm{E}[\widehat{\pi}_{\mathrm{bc},1}(\boldsymbol{X})]=\mathrm{E}[\widehat{\pi}_{\mathrm{bc},2}(\boldsymbol{X})]=\mathrm{E}[Y]$. □

Proposition 7.1 can also be seen as a consequence of Proposition 3.1 in Denuit (2010). A single sign-change property is enough to ensure that the balance-corrected versions of $\widehat{\pi}_1$ an $\widehat{\pi}_2$ are ordered in the convex sense. See Proposition 2.1 in Denuit (2010).

Since

$$\begin{aligned}\mathrm{Var}[Y] &= \mathrm{E}\Big[\mathrm{Var}\Big[Y|\widehat{\pi}_k(\boldsymbol{X})\Big]\Big] + \mathrm{Var}\Big[\mathrm{E}\Big[Y|\widehat{\pi}_k(\boldsymbol{X})\Big]\Big] \\ &= \mathrm{Var}\Big[\widehat{\pi}_{\mathrm{bc},k}(\boldsymbol{X})\Big] + \mathrm{E}\Big[\mathrm{Var}\Big[Y|\widehat{\pi}_k(\boldsymbol{X})\Big]\Big]\end{aligned}$$

we see that if $\pi_{\mathrm{bc},1}(\boldsymbol{X}) \preceq_{\mathrm{CX}} \pi_{\mathrm{bc},2}(\boldsymbol{X})$ then

$$\mathrm{E}\Big[\mathrm{Var}\Big[Y|\widehat{\pi}_{\mathrm{bc},2}(\boldsymbol{X})\Big]\Big] \leq \mathrm{E}\Big[\mathrm{Var}\Big[Y|\widehat{\pi}_{\mathrm{bc},1}(\boldsymbol{X})\Big]\Big]$$

so that $\widehat{\pi}_{\mathrm{bc},2}(\boldsymbol{X})$ conveys more information about Y compared to $\widehat{\pi}_{\mathrm{bc},1}(\boldsymbol{X})$.

References

- Ciatto, N., Verelst, H., Trufin, J., Denuit, M. (2022). Does autocalibration improve goodness of lift? European Actuarial Journal, in press.

- Denuit, M. (2010). Positive dependence of signals. Journal of Applied Probability 47, 893-897.

- Denuit, M., Charpentier, A., Trufin, J. (2021). Autocalibration and Tweedie-dominance for insurance pricing with machine learning. Insurance: Mathematics and Economics 101, 485-497.

- Denuit, M., Dhaene, J., Goovaerts, M.J., Kaas, R. (2005). Actuarial Theory for Dependent Risks: Measures, Orders and Models. Wiley, New York.

- Denuit, M., Sznajder, D., Trufin, J. (2019). Model selection based on Lorenz and concentration curves, Gini indices and convex order. Insurance: Mathematics and Economics 89, 128-139.

- Denuit, M., Trufin, J. (2021). Lorenz curve, Gini coefficient, and Tweedie dominance for autocalibrated predictors. Working Paper available from https://dial.uclouvain.be/pr/boreal/object/boreal%3A254535/datastream/PDF_01/view

- Denuit, M., Trufin, J. (2022). Model selection with Pearson's correlation, concentration and Lorenz curves under autocalibration. Working Paper available from https://dial.uclouvain.be/pr/boreal/object/boreal%3A266744/datastream/PDF_01/view

- Efron, B. (1986). How biased is the apparent error rate of a prediction rule? Journal of the American Statistical Association, 81(394), 461–470.

- Gneiting, T. (2011). Making and evaluating point forecasts. Journal of the American Statistical Association 106, 746–762.

- Krüger, F., Ziegel, J.F. (2021). Generic conditions for forecast dominance. Journal of Business & Economic Statistics 39, 972-983.

- Lindholm, M., Lindskog, F., Palmquist, J. (2022). Local bias adjustment, duration-weighted probabilities, and automatic construction of tariff cells. Available at SSRN: https://ssrn.com/abstract=4256876 or http://dx.doi.org/10.2139/ssrn.4256876

- Loader, C. (1999). Local Regression and Likelihood. Springer, New York.

- Patton, A.J. (2011). Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246–256.

- Patton, A.J. (2020). Comparing possibly misspecified forecasts. Journal of Business & Economic Statistics 38, 796-809.

- Savage, L.J. (1971). Elicitation of personal probabilities and expectations. Journal of the American Statistical Association 66, 783-810.

- Shaked, M., Shanthikumar, J.G. (2007). Stochastic Orders. Springer, New York.

- Wüthrich, M. V. (2022). Model selection with Gini indices under auto-calibration. arXiv preprint arXiv:2207.14372.

- Yitzhaki, S., Schechtman, E. (2013). The Gini Methodology: A Primer on Statistical Methodology. Springer.