THE BALANCE PROPERTY: FROM INSURANCE PRICING TO RISK-SHARING

Michel Denuit, Christian Y. Robert, Mario V. Wüthrich

LIDAM Discussion Paper ISBA
2026 / 30

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

The Balance Property: From Insurance Pricing to Risk-Sharing

Michel Denuit[*] Christian Y. Robert[†] Mario V. Wüthrich[‡]

August 19, 2026

Abstract

The balance property is an in-sample global calibration constraint requiring the total fitted actuarial price to be equal to the total observed loss. For generalized linear models (GLMs) fitted by maximum likelihood estimation (MLE), this property holds automatically under the canonical link choice, but it generally fails under non-canonical link choices. Existing balance correction methods include quasi-MLE, simple post-processing of the intercept, and constrained MLE. This note proposes a fourth approach, motivated by the linear regression approximation of the conditional mean risk-sharing (CMRS) rule in large heterogeneous pools. Starting from the correctly specified MLE, the proposed correction allocates the aggregate balance defect according to the fitted conditional variances. We compare the four methods through two complementary asymptotic criteria. First, a delta method expansion gives the first-order mean squared error (MSE) of the fitted pure premium. Under this criterion, the proposed correction shares the same MLE component as the simple post-processing and the constrained MLE corrections, but differs in the way it allocates the balance defect. Second, we derive its average Kullback–Leibler (KL) divergence and show that the variance-weighted correction is locally optimal among additive first-order balance corrections acting directly on the fitted mean function. Thus, the proposed method complements constrained MLE: the latter is optimal within the parametric GLM class, whereas the former is optimal in a larger class of mean function corrections and has a direct risk-sharing interpretation. The paper finally highlights the link between insurance pricing models possessing the balance property and risk-sharing rules. This view reconciles two seemingly disconnected core actuarial topics, ex-ante premium calculation and ex-post loss allocation to participants in a peer-to-peer insurance pool.

Keywords. Balance property, generalized linear model, risk-sharing, conditional mean risk-sharing, loss allocation, peer-to-peer insurance, delta method, post-processing.

1 Introduction

The balance property has its origins in the actuarial pricing method known as the method of total marginal sums, introduced independently by Bailey [1] and Jung [16]. This method

[*]Institute of Statistics, Biostatistics and Actuarial Science - ISBA, Louvain Institute of Data Analysis and Modeling - LIDAM, UCLouvain, Louvain-la-Neuve, Belgium; michel.denuit@uclouvain.be

[†]Laboratoire de Sciences Actuarielle et Financière - LSAF, Université Claude Bernard Lyon 1 - UCBL, France; christian.robert@univ-lyon1.fr

[‡]Department of Mathematics, ETH Zurich, Switzerland; mario.wuethrich@math.ethz.ch

is obtained by enforcing row and column balance in a cross-classified pricing scheme. It was later shown that the same pricing scheme arises naturally from a Poisson generalized linear model (GLM) with the canonical link (being the log-link). Within the GLM framework, the use of the canonical link ensures that the maximum likelihood estimates (MLEs) satisfy an in-sample balance condition: the fitted aggregate loss equals the total observed loss. In other words, the fitted premiums provide an ex-post allocation of the total observed claims across all policyholders. This is precisely the balance property. Bühlmann and Gisler [4, Theorem 4.5] recognized the importance of this property in insurance pricing and proved that it also holds for the homogeneous Bühlmann–Straub credibility estimator.

In the statistical literature, the balance property is less prominently represented because the statistical literature rather focuses on strong consistency and best asymptotic normality in estimation; see Gouriéroux et al. [13]. However, already in the original GLM paper of Nelder and Wedderburn [18, formula (10)], it has been verified that a MLE fitted GLM fulfills the balance property under the canonical link choice, but it generally fails to hold under non-canonical link choices; see also Wüthrich [22]. Actuaries are aware of this problem, and they usually apply a post correction method to the fitted model to rectify the balance property; this is done to ensure (in-sample) global calibration. The simple post-processing (SPP) method adjusts the intercept parameter of the GLM. The SPP method is the most commonly applied solution in industry practice. Lindholm and Wüthrich [17] and Wüthrich [23] discuss two alternative methods, a quasi maximum likelihood estimation (QMLE) method and a constrained MLE method. The former performs MLE under a distributional assumption that matches the selected link choice but not the properties of the claim observations, the latter enforces the balance property by MLE under a side constraint (SC). Wüthrich [23] concludes that the third method (SC, i.e., constrained MLE) outperforms that other two balance correction methods (SPP and QMLE) in terms of large sample accuracy.

This paper proposes a variance-weighted regression (REG) correction as a fourth balance correction method. The REG method is motivated by the large-pool approximation of the conditional mean risk-sharing (CMRS) rule derived by Denuit and Robert [9]. The first three methods (SPP, QMLE, SC) are parametric in the sense that they ultimately deliver a parameter estimate, or a corrected parameter estimate, in the original GLM parameter space. The fourth method is different. It starts from the MLE and then corrects the fitted pure premiums directly by allocating the aggregate balance defect in proportion of the fitted conditional variances to restore the balance property. The risk-sharing motivation for this correction is as follows. In a pool of independent heterogeneous losses, the CMRS rule introduced by Denuit and Dhaene [6] allocates to each participant the conditional expectation of the individual loss given the aggregate loss. In large pools, this generally non-linear rule is asymptotically approximated by a linear regression rule allocating deviations of the aggregate loss from its mean in proportion to individual variances. Replacing the unknown pure premium and variance by their GLM estimates gives the proposed REG correction.

Because the REG correction does not define a new estimator of the regression parameter, comparing all four methods through the estimation error is not appropriate. We instead compare them through the asymptotic error of the fitted mean function. The first comparison criterion is the pointwise mean squared error (MSE). A delta method expansion shows that SPP, SC and REG share the same unconstrained MLE component and differ only by the square of the balance correction allocation function. This makes the comparison transparent: SPP allocates the aggregate balance defect according to the local effect of an intercept shift, SC according to the optimal parametric likelihood direction, and REG according to the conditional variance contribution of each observation.

The second comparison criterion is the average Kullback–Leibler (KL) divergence to measure the difference of the true model from the estimated model belonging both to the same exponential dispersion family (EDF) but differ in their means. This criterion is closer to the likelihood-based comparisons used for parametric GLM estimators. This paper shows that REG outperforms SC in that respect. This result does not contradict the optimality of the SC method in the constrained parametric likelihood problem. Rather, it reflects that REG acts in a larger space as it corrects the fitted mean function directly and may leave the original GLM family. We further show that the variance-weighted allocation is locally KL-optimal among additive first-order balance corrections of the MLE satisfying the balance constraint. Thus, REG has two complementary interpretations: It is the large-pool linear approximation of CMRS, and it is the least costly first-order additive balance correction in average KL divergence.

The final part of the paper connects ex-ante premium calculation under the balance property and ex-post loss allocation in insurance pools. Insurance pools are a form of peer-to-peer insurance that allocate the total loss among its pool members; see Denuit et al. [7] or Feng [11]. Naturally, this allocation satisfies the balance property (otherwise it would not be an allocation). Indeed, every insurance pricing model fulfilling the (in-sample) balance property can be used to allocate total experienced losses among the participants in a peer-to-peer insurance pool. To this end, it suffices to fit the model ex-post, once individual losses have been observed. This paper thus illustrates the connection between pricing and risk sharing in both directions. First, it shows that risk sharing offers an alternative method to restore balance, complementing those developed by Lindholm and Wüthrich [17] and Wüthrich [23]. Second, it argues that any pricing technique fulfilling the balance property can be used to share losses within an insurance pool. These ideas are illustrated in the classical GLM setting but extend beyond.

The remainder of the paper is organized as follows. Section 2 introduces GLMs and we discuss how they align with the balance property. Section 3 presents the three parametric GLM balance corrections QMLE, SPP and SC considered in Lindholm and Wüthrich [17] and Wüthrich [23]. Section 4 introduces the variance-weighted REG correction, being motivated by the large-pool linear regression approximation of the CMRS rule. It acts directly on the fitted mean function rather than on the GLM parameter. Section 5 compares the four balance-corrected fitted models. It first studies integrated quadratic errors for the fitted pure premium, showing that the methods differ through their balance-allocation functions. It then turns to the average KL divergence, where the local MSE is weighted by the inverse conditional variance. This likelihood-induced geometry yields an explicit ranking of the four methods and establishes the local KL-optimality of REG among additive first-order balance corrections acting on the fitted mean function. Section 6 connects ex-ante premium calculation to ex-post loss allocation under the balance condition. Section 7 concludes with a summary of the main results and a discussion of possible extensions. Proofs of the main results are collected in the Appendix.

2 The balance property in generalized linear models

We start by revisiting the GLM framework and by discussing how it complies with the balance property; this section closely follows Wüthrich [23].

Let Y be the insurance claim, $W > 0$ be a strictly positive case weight, and $\boldsymbol{X}$ be the $(q+1)$-dimensional real-valued covariates with initial component being equal to one, that is,

$$\boldsymbol{X} = (X_0, X_1, \ldots, X_q)^\top = (1, X_1, \ldots, X_q)^\top \in \{1\} \times \mathbb{R}^q.$$

Assume that the insurance claim Y, given $(\boldsymbol{X}, W)$, has the following first two moments

$$\mathbb{E}[Y \mid \boldsymbol{X}, W] = \mu(\boldsymbol{X}) \qquad \text{and} \qquad \operatorname{Var}[Y \mid \boldsymbol{X}, W] = \frac{\varphi}{W}\, \sigma^2(\boldsymbol{X}) > 0, \tag{2.1}$$

with regression function $\boldsymbol{X} \mapsto \mu(\boldsymbol{X})$, variance function $\boldsymbol{X} \mapsto \sigma^2(\boldsymbol{X}) > 0$ and fixed dispersion parameter $\varphi > 0$. Note that the first moment in (2.1) includes the implicit assumption that the conditionally expected claim Y is independent of W, given $\boldsymbol{X}$. This property generally holds within the EDF; see Jørgensen [15] or Wüthrich–Merz [24].

Our goal is to estimate the (unknown) regression function $\mu(\cdot)$ from an i.i.d. learning sample $\mathcal{D} = (Y_i, \boldsymbol{X}_i, W_i)_{i=1}^n$ of size n and following the same distribution as $(Y, \boldsymbol{X}, W)$. Assume that the unknown regresssion function $\mu(\cdot)$ takes the following GLM form

$$\boldsymbol{X} \;\mapsto\; g(\mu(\boldsymbol{X}; \boldsymbol{\beta}^+)) = \langle \boldsymbol{\beta}^+, \boldsymbol{X} \rangle = \beta_0^+ + \sum_{j=1}^q \beta_j^+ X_j, \tag{2.2}$$

with strictly increasing and smooth link function $g(\cdot)$ and true regression parameter $\boldsymbol{\beta}^+ = (\beta_0^+, \ldots, \beta_q^+)^\top \in \mathbb{R}^{q+1}$. This GLM assumption (2.2) reduces the regression function estimation problem to a parametric problem in $\boldsymbol{\beta} \in \mathbb{R}^{q+1}$. If we furthermore assume that Y belongs to the EDF with cumulant function κ, given $(\boldsymbol{X}, W)$, this estimation problem can be solved by MLE; see Denuit et al. [8] and Wüthrich and Merz [24, Chapter 5]. Namely, the MLE of $\boldsymbol{\beta}^+$ is found by solving

$$\widehat{\boldsymbol{\beta}}_n \in \underset{\boldsymbol{\beta} \in \mathbb{R}^{q+1}}{\arg\min}\; \frac{1}{n} \sum_{i=1}^n \frac{W_i}{\varphi}\, L_\kappa\left(Y_i, \mu(\boldsymbol{X}_i; \boldsymbol{\beta})\right), \tag{2.3}$$

with deviance loss function

$$L_\kappa(y, m) = 2\left(y h(y) - \kappa\left(h(y)\right) - y h(m) + \kappa\left(h(m)\right)\right), \tag{2.4}$$

and $h = (\kappa')^{-1}$ denotes the *canonical link* of the selected EDF with cumulant function κ. This solves MLE of the GLM, and the fitted regression function takes the form

$$\boldsymbol{X} \;\mapsto\; \mu(\boldsymbol{X}; \widehat{\boldsymbol{\beta}}_n) = g^{-1}\big(\langle \widehat{\boldsymbol{\beta}}_n, \boldsymbol{X} \rangle\big). \tag{2.5}$$

Drawing on the statistical literature, under certain regularity conditions, the MLE $\widehat{\boldsymbol{\beta}}_n$ is strongly consistent for $\boldsymbol{\beta}^+$ as $n \to \infty$, and there is asymptotically normality

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^+\right) \;\Longrightarrow\; \mathcal{N}\left(\boldsymbol{0}, \mathcal{H}^{-1}\right) \qquad \text{for } n \to \infty,$$

where the following Fisher information matrix is assumed to exist and have full rank

$$\mathcal{H} = \mathbb{E}\left[\frac{W}{\varphi} \frac{1}{\sigma^2(\boldsymbol{X})}\, \dot{\mu}(\boldsymbol{X}) \dot{\mu}(\boldsymbol{X})^\top\right], \tag{2.6}$$

and with gradient in the true parameter $\boldsymbol{\beta}^+$

$$\dot{\mu}(\boldsymbol{X}) := \frac{\partial}{\partial \boldsymbol{\beta}}\, \mu(\boldsymbol{X};\boldsymbol{\beta})\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^+} = \frac{\boldsymbol{X}}{g'\big(\mu(\boldsymbol{X};\boldsymbol{\beta}^+)\big)}. \tag{2.7}$$

The aforementioned properties concern infinite sample limits. However, in insurance pricing, we are more concerned about finite sample biases as these directly related to correct price level specifications. We therefore introduce the balance property as follows: A regression fitting procedure satisfies the *balance property* if for almost every (a.e.) realization of the learning sample $\mathcal{D} = (Y_i, \boldsymbol{X}_i, W_i)_{i=1}^n$, the resulting estimated regression function $\widehat{\mu}(\cdot) = \widehat{\mu}_{\mathcal{D}}(\cdot)$ fulfills

$$\sum_{i=1}^n W_i\, \widehat{\mu}(\boldsymbol{X}_i) = \sum_{i=1}^n W_i\, Y_i. \tag{2.8}$$

The notation $\widehat{\mu}(\cdot) = \widehat{\mu}_{\mathcal{D}}(\cdot)$ indicates that the estimated model depends on the learning sample $\mathcal{D}$ through the optimization (2.3).

Proposition 2.1 (Nelder and Wedderburn [18, formula (10)]) *Consider the MLE fitted GLM regression function* (2.5) *under the canonical link choice $g = h = (\kappa')^{-1}$ within the chosen EDF with steep cumulant function κ. The MLE regression fitting procedure* (2.3) *fulfills the balance property* (2.8)*.*

Under the canonical link choice $g = h = (\kappa')^{-1}$, the MLE fitted GLM fulfills the balance property, and otherwise generally not. The latter situation makes it necessary in actuarial pricing applications to employ a balance correction method, otherwise the price level may be misspecified. The next section discusses such balance correction methods.

3 Estimation methods rectifying the balance property

Working under non-canonical links $g \neq h$ implies that the balance property generally fails to hold, e.g., in a gamma GLM with log-link it does not hold. This makes it necessary to employ a balance correction method to obtain global calibration on finite samples. The following three methods were studied in Lindholm and Wüthrich [17] and Wüthrich [23] (and we closely follow these two references):

- Quasi maximum likelihood estimation (QMLE) method;
- Simple post-processing (SPP) method (currently the industry standard);
- MLE under side constraint (SC).

These three methods have in common that they are parametric in the sense that they provide the actuary with an estimated parameter in the GLM parameter space fulfilling the balance property. Section 4 gives an optimal solution in a bigger model class.

This section presents the results from Lindholm and Wüthrich [17] and Wüthrich [23] of the three methods QMLE, SPP and SC. These results require certain technical assumptions as outlined in Gouriéroux et al. [13, Appendix 1], Gallant and Holly [12], Burguete et al. [5] and White [21]. We generally assume that these technical assumptions hold.

3.1 Quasi maximum likelihood estimation

Assume $g \neq h$, i.e., the selected link g is not the canonical one of the data generating mechanism being based on cumulant function κ. This motivates us to consider QMLE introduced by Wedderburn [20]. If the selected link g originates from the EDF member with cumulant function κ_g, i.e., if $g = (\kappa_g')^{-1}$ is the canonical link of κ_g, one can solve the QMLE problem

$$\widehat{\boldsymbol{\beta}}_n^{\mathrm{QMLE}} \in \underset{\boldsymbol{\beta} \in \mathbb{R}^{q+1}}{\arg\min} \; \frac{1}{n} \sum_{i=1}^n \frac{W_i}{\varphi} \, L_{\kappa_g}\left(Y_i, \mu(\boldsymbol{X}_i; \boldsymbol{\beta})\right). \tag{3.1}$$

Thus, the selected cumulant function κ_g aligns with the selected link g, but not with the data generating cumulant function κ that underlies the claims Y, given $(\boldsymbol{X}, W)$. Proposition 2.1 then immediately implies that the QMLE fitted model (3.1) fulfills the balance property (2.8); see also Wüthrich [23, Corollary 4.1]. Put differently, whether the balance property holds does not depend on the data generating mechanism (represented by κ), but rather on whether the chosen loss function L_{κ_g} is compatible with the chosen link function g. The price we pay by this misalignment with the data generating mechanism is that the estimation procedure is not best asymptotically normal, expressed by the inequality in (3.3), below.

We determine the (asymptotic) accuracy of the QMLE $\widehat{\boldsymbol{\beta}}_n^{\mathrm{QMLE}}$ for $\boldsymbol{\beta}^+$; its strong consistency was proved in Gouriéroux et al. [13, Theorem 1]. We introduce the matrices (and we assume their existence and positive definiteness)

$$\begin{aligned}
\mathcal{J}_g &= \mathbb{E}\left[\frac{W}{\varphi} \, \frac{1}{V_g(\mu(\boldsymbol{X}; \boldsymbol{\beta}^+))} \, \dot{\mu}(\boldsymbol{X}) \dot{\mu}(\boldsymbol{X})^\top\right], \\
\mathcal{I}_g &= \mathbb{E}\left[\frac{W}{\varphi} \, \frac{\sigma^2(\boldsymbol{X})}{V_g(\mu(\boldsymbol{X}; \boldsymbol{\beta}^+))^2} \, \dot{\mu}(\boldsymbol{X}) \dot{\mu}(\boldsymbol{X})^\top\right],
\end{aligned}$$

recall (2.7), and with variance function

$$m \mapsto V_g(m) = \kappa_g''((\kappa_g)^{-1}(m)) = \kappa_g''(g(m)) > 0. \tag{3.2}$$

Applying the results of Gouriéroux et al. [13, Theorem 3 and Property 5], see Wüthrich [23, formula (4.8)], we have the following asymptotic normality result for the QMLE

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n^{\mathrm{QMLE}} - \boldsymbol{\beta}^+\right) \implies \mathcal{N}\left(\mathbf{0}, \Sigma_{\mathrm{QMLE}}\right) \qquad \text{for } n \to \infty,$$

with asymptotic covariance matrix

$$\Sigma_{\mathrm{QMLE}} := \mathcal{J}_g^{-1} \mathcal{I}_g \mathcal{J}_g^{-1} \;\geq\; \mathcal{H}^{-1}. \tag{3.3}$$

The matrices $\mathcal{J}_g$ and $\mathcal{I}_g$ are the sensitivity and variability matrices associated with the quasi-likelihood estimating equations. Unless the quasi-family has the correct variance structure, i.e., $\sigma^2(\boldsymbol{X}) = V_g(\mu(\boldsymbol{X}; \boldsymbol{\beta}^+))$, one generally has $\Sigma_{\mathrm{QMLE}} > \mathcal{H}^{-1}$. Thus, QMLE enforces balance by changing the likelihood geometry.

Note that under the correct variance structure $\sigma^2(\boldsymbol{X}) = V_g(\mu(\boldsymbol{X}; \boldsymbol{\beta}^+))$, we have $\mathcal{J}_g = \mathcal{I}_g = \mathcal{H}$ and $\Sigma_{\mathrm{QMLE}} = \mathcal{H}^{-1}$, but this also implies that we work under the canonical link $g = h$, because within the EDF we have the identity for the variance function

$$\sigma^2(\boldsymbol{X}) = \kappa''\left((\kappa)^{-1}\left(\mu(\boldsymbol{X}; \boldsymbol{\beta}^+)\right)\right) = \kappa''\left(h\left(\mu(\boldsymbol{X}; \boldsymbol{\beta}^+)\right)\right) = V_h(\mu(\boldsymbol{X}; \boldsymbol{\beta}^+)). \tag{3.4}$$

We could already have mentioned this in (2.1).

3.2 Simple post-processing

The SPP method is the simplest correction method, and it is the one usually used in industry. The goal is to modify the intercept parameter of the MLE (2.3) such that the balance property is fulfilled. This is achieved by selecting the (unique) constant $\gamma \in \mathbb{R}$ that solves

$$\sum_{i=1}^{n} W_i \, \mu(\boldsymbol{X}_i; \widehat{\boldsymbol{\beta}}_n + \gamma \, \boldsymbol{e}_1) = \sum_{i=1}^{n} W_i \, Y_i, \tag{3.5}$$

with first basis vector $\boldsymbol{e}_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^{q+1}$. Denote the solution of (3.5) by $\widehat{\gamma}$. This gives the SPP estimate

$$\widehat{\boldsymbol{\beta}}_n^{\rm SPP} = \widehat{\boldsymbol{\beta}}_n + \widehat{\gamma} \, \boldsymbol{e}_1 \ \in \ \mathbb{R}^{q+1}.$$

Assume we work under the correct variance choice of the selected EDF, see (3.4). Then, the SPP estimate $\widehat{\boldsymbol{\beta}}_n^{\rm SPP}$ is strongly consistent for $\boldsymbol{\beta}^+$, and it has the following asymptotic normal behavior, see Lindholm and Wüthrich [17, Theorems 4.6-4.7],

$$\sqrt{n} \left(\widehat{\boldsymbol{\beta}}_n^{\rm SPP} - \boldsymbol{\beta}^+ \right) \implies \mathcal{N}\left(\boldsymbol{0}, \Sigma_{\rm SPP}\right) \qquad \text{for } n \to \infty,$$

with asymptotic covariance matrix

$$\Sigma_{\rm SPP} := \mathcal{H}^{-1} + \frac{\tau^2 - a}{m_0^2} \, \boldsymbol{e}_1 \boldsymbol{e}_1^\top,$$

with parameters, recall (2.7),

$$\boldsymbol{m} = \mathbb{E}\left[\frac{W}{\varphi} \, \dot{\mu}(\boldsymbol{X}) \right] \in \mathbb{R}^{q+1}, \qquad \tau^2 = \mathbb{E}\left[\frac{W}{\varphi} \sigma^2(\boldsymbol{X}) \right], \qquad a = \boldsymbol{m}^\top \mathcal{H}^{-1} \boldsymbol{m}, \tag{3.6}$$

and $m_0 > 0$ being the initial component of $\boldsymbol{m} = (m_0, m_1, \ldots, m_q)^\top$. Since $\mathcal{H}$ is positive definite and $\boldsymbol{m} \neq \boldsymbol{0}$, one has $a > 0$, and by the Cauchy–Schwarz inequality in the $\mathcal{H}$-metric, we have

$$0 < a \leq \tau^2. \tag{3.7}$$

3.3 Optimization under side constraint

Wüthrich [23] considers optimization under the side constraint (SC) enforcing the balance property. For this, introduce the balance side constraint

$$\mathcal{C}_n(\boldsymbol{\beta}) := \frac{1}{n} \sum_{i=1}^{n} \frac{W_i}{\varphi} \left(\mu(\boldsymbol{X}_i; \boldsymbol{\beta}) - Y_i \right) = 0. \tag{3.8}$$

This motivates the constrained MLE

$$\widehat{\boldsymbol{\beta}}_n^{\rm SC} \in \underset{\boldsymbol{\beta}:\, \mathcal{C}_n(\boldsymbol{\beta}) = 0}{\arg\min} \ \frac{1}{n} \sum_{i=1}^{n} \frac{W_i}{\varphi} \, L_\kappa\left(Y_i, \mu(\boldsymbol{X}_i; \boldsymbol{\beta}) \right).$$

Again, assume that we work under the correct variance choice (3.4), then Wüthrich [23, Theorem 4.5] proves the following asymptotic normality result

$$\sqrt{n} \left(\widehat{\boldsymbol{\beta}}_n^{\rm SC} - \boldsymbol{\beta}^+ \right) \implies \mathcal{N}\left(\boldsymbol{0}, \Sigma_{\rm SC}\right) \qquad \text{for } n \to \infty,$$

with asymptotic covariance matrix

$$\Sigma_{\rm SC} := \mathcal{H}^{-1} + \frac{\tau^2 - a}{a^2} \, \mathcal{H}^{-1} \boldsymbol{m} \boldsymbol{m}^\top \mathcal{H}^{-1},$$

and we also refer to (3.6) for the parameters involved. The SC method can be interpreted as the optimal likelihood-based correction of the unconstrained MLE in the local $\mathcal{H}$-geometry.

4 Variance-weighted regression correction

Contrary to the three methods considered in the preceding section, the fourth correction method proposed in this paper is not based on a new estimator of $\boldsymbol{\beta}^+$. It is a post-GLM correction of the fitted pure premiums. It may leave the GLM family, unlike QMLE, SPP and SC. It is rooted in risk-sharing: the aggregate imbalance is allocated in proportion to the fitted conditional variances. This approach is motivated by the large-pool approximation to the CMRS rule derived by Denuit and Robert [9]. More precisely, in a pool of independent heterogeneous losses $(Z_j)_{j=1}^n$, the linear regression approximation of the CMRS rule for contract $i \in \{1, \ldots, n\}$ is

$$h_{i,n}^{\text{reg}}(S_n) = \mathbb{E}[Z_i] + \frac{\text{Var}[Z_i]}{\sum_{j=1}^n \text{Var}[Z_j]} \left(S_n - \mathbb{E}[S_n] \right), \qquad \text{with total loss} \quad S_n = \sum_{j=1}^n Z_j.$$

Once the losses $(Z_j)_{j=1}^n$ are realized, they are aggregated into S_n and pool participant i is asked to contribute $h_{i,n}^{\text{reg}}(S_n)$ to the pool so that the total amount collected from all pool members exactly matches the total loss S_n.

Applied to GLM pricing, the monetary loss caused by contract i is $Z_i = W_i Y_i$. Its fitted conditional mean and variance are

$$\widehat{\mathbb{E}}[Z_i \mid \boldsymbol{X}_i, W_i] = W_i \, \mu(\boldsymbol{X}_i; \widehat{\boldsymbol{\beta}}_n) \qquad \text{and} \qquad \widehat{\text{Var}}[Z_i \mid \boldsymbol{X}_i, W_i] = \widehat{\varphi} \, W_i \, \widehat{\sigma}_i^2,$$

where $\widehat{\sigma}_i^2 = \sigma^2(\boldsymbol{X}_i; \widehat{\boldsymbol{\beta}}_n)$ is the fitted conditional variance function on the exposure-scaled scale, see (3.4). The variance-weighted correction allocates the aggregate imbalance according to the fitted conditional variances, that is,

$$\widehat{\mu}_i^{\text{REG}} = \mu(\boldsymbol{X}_i; \widehat{\boldsymbol{\beta}}_n) + \frac{\widehat{\sigma}_i^2}{\sum_{j=1}^n W_j \, \widehat{\sigma}_j^2} \left(\sum_{j=1}^n W_j Y_j - \sum_{j=1}^n W_j \, \mu(\boldsymbol{X}_j; \widehat{\boldsymbol{\beta}}_n) \right). \tag{4.1}$$

By construction, (4.1) satisfies the balance property (2.8). For a generic covariate vector $\boldsymbol{X}$, we define

$$\widehat{\mu}^{\text{REG}}(\boldsymbol{X}) := \widehat{\mu}_n^{\text{REG}}(\boldsymbol{X}) = \mu(\boldsymbol{X}; \widehat{\boldsymbol{\beta}}_n) + \frac{\widehat{\sigma}^2(\boldsymbol{X})}{\sum_{j=1}^n W_j \, \widehat{\sigma}_j^2} \left(\sum_{j=1}^n W_j Y_j - \sum_{j=1}^n W_j \, \mu(\boldsymbol{X}_j; \widehat{\boldsymbol{\beta}}_n) \right),$$

so that $\widehat{\mu}_i^{\text{REG}} = \widehat{\mu}^{\text{REG}}(\boldsymbol{X}_i)$. Notice that there is no positivity guarantee in (4.1). Typically, actuaries use the log-link for g in GLMs, and then the resulting solutions will be positive by design for QMLE, SPP, and SC. This issue is further discussed in the next sections. Also, REG loses the familiar multiplicative premium structure obtained under log-link.

In summary, the four methods QMLE, SPP, SC and REG correspond to four different ways of enforcing the balance property:

QMLE. Balance is enforced by changing the variance geometry so that the selected link g becomes canonical one. This may sacrifice efficiency if the quasi-family is not variance-correct, i.e., matches the conditional variance properties of the responses Y.

SPP. Balance is enforced by shifting the intercept. The aggregate balance defect is distributed according to the local effect of a common intercept perturbation.

SC. Balance is enforced directly in the likelihood optimization. The correction direction is optimal in the local Fisher (deviance) geometry.

REG. Balance is enforced by treating the aggregate imbalance as a pool residual and allocating it according to fitted conditional variances, as suggested by the large-pool linear approximation of the CMRS rule.

The REG correction investigated in this section is therefore not merely a numerical calibration adjustment. It has a risk-sharing interpretation: after estimating the systematic pure premium component $\widehat{\mu}(\boldsymbol{X}_i)$, the aggregate imbalance is allocated ex-post according to each contract's contribution to the conditional variance of the pool.

5 Comparative properties of balance-corrected fitted models

The preceding sections have introduced four ways of enforcing the balance property. The first three corrections, QMLE, SPP and SC, remain within the parametric GLM family, whereas REG acts directly on the fitted mean function and may therefore leave the original GLM class. This distinction is important for the comparison. A parameter-based comparison is natural for QMLE, SPP and SC, but not for REG as it does not live in the same parameter space. We therefore compare the four procedures at the level of the fitted pure premium function.

Two complementary criteria are considered below. The first one is based on quadratic errors of fitted pure premiums. It measures the absolute errors $\widehat{\mu}(\boldsymbol{x}) - \mu(\boldsymbol{x};\boldsymbol{\beta}^+)$ in the scale of the pure premium for fixed covariates $\boldsymbol{x}$. The second one is based on the average KL divergence between the fitted EDF model and the true EDF model. Locally, this criterion also depends on the squared error, but divided by the conditional variance $\sigma^2(\boldsymbol{x})$. Thus, at a fixed covariate value $\boldsymbol{x}$, the two criteria differ by the variance normalization $1/\sigma^2(\boldsymbol{x})$, up to the exposure-dispersion scaling. After averaging over the portfolio distribution, this normalization changes the geometry of the comparison and leads to explicit likelihood-based rankings.

5.1 Integrated quadratic errors

We first compare the methods through the first-order mean squared error (MSE) of the fitted pure premium. For a fixed covariate value $\boldsymbol{x}$, define

$$\Gamma_A(\boldsymbol{x}) = \lim_{n\to\infty} n\,\mathbb{E}\left[\left(\widehat{\mu}^A(\boldsymbol{x}) - \mu(\boldsymbol{x};\boldsymbol{\beta}^+)\right)^2\right], \qquad \text{for } A \in \{\mathrm{QMLE}, \mathrm{SPP}, \mathrm{SC}, \mathrm{REG}\}, \tag{5.1}$$

whenever the limit exists, and where we set $\widehat{\mu}^A(\boldsymbol{x}) = \widehat{\mu}^A_n(\boldsymbol{x}) = \mu(\boldsymbol{x};\widehat{\boldsymbol{\beta}}^A_n)$ for $A \in \{\mathrm{QMLE}, \mathrm{SPP}, \mathrm{SC}\}$. Then, there is the first-order MSE approximation

$$\mathrm{MSE}_A(\boldsymbol{x}) = \mathbb{E}\left[\left(\widehat{\mu}^A(\boldsymbol{x}) - \mu(\boldsymbol{x};\boldsymbol{\beta}^+)\right)^2\right] = \frac{1}{n}\,\Gamma_A(\boldsymbol{x}) + o(n^{-1}),$$

so that $\Gamma_A(\boldsymbol{x})$ measures the first-order MSE constant at covariate value $\boldsymbol{x}$.

The following theorem gives the first-order expansion of the pure premium error for the four balance corrections; recall the Fisher information $\mathcal{H}$, see (2.6), and vector $\boldsymbol{m}$, see (3.6).

Theorem 5.1 *Assume that $\mathcal{H}$ is positive definite and that the standard regularity conditions for asymptotic normality of M- and Z-estimators hold; see, e.g., van der Vaart and Wellner [19]. Then, for fixed covariate value $\boldsymbol{x}$*

$$\begin{aligned}
\Gamma_{\mathrm{QMLE}}(\boldsymbol{x}) &= \dot{\mu}(\boldsymbol{x})^\top \Sigma_{\mathrm{QMLE}}\, \dot{\mu}(\boldsymbol{x}), \\
\Gamma_{\mathrm{SPP}}(\boldsymbol{x}) &= \Gamma_{\mathrm{MLE}}(\boldsymbol{x}) + (\tau^2 - a)\left(\frac{\dot{\mu}(\boldsymbol{x})^\top \boldsymbol{e}_1}{m_0}\right)^2, \\
\Gamma_{\mathrm{SC}}(\boldsymbol{x}) &= \Gamma_{\mathrm{MLE}}(\boldsymbol{x}) + (\tau^2 - a)\left(\frac{\dot{\mu}(\boldsymbol{x})^\top \mathcal{H}^{-1}\boldsymbol{m}}{a}\right)^2, \\
\Gamma_{\mathrm{REG}}(\boldsymbol{x}) &= \Gamma_{\mathrm{MLE}}(\boldsymbol{x}) + (\tau^2 - a)\left(\frac{\sigma^2(\boldsymbol{x})}{\tau^2}\right)^2.
\end{aligned}$$

with $\Gamma_{\mathrm{MLE}}(\boldsymbol{x}) = \dot{\mu}(\boldsymbol{x})^\top \mathcal{H}^{-1} \dot{\mu}(\boldsymbol{x})$.

The proof of Theorem 5.1 is given in Appendix A. Theorem 5.1 separates the usual MLE contribution $\Gamma_{\mathrm{MLE}}(\boldsymbol{x})$ from the additional cost generated by enforcing balance. For SPP, the additional term is driven by the local sensitivity of the mean to an intercept shift, since

$$\dot{\mu}(\boldsymbol{x})^\top \boldsymbol{e}_1 = \frac{1}{g'\big(\mu(\boldsymbol{x};\boldsymbol{\beta}^+)\big)}.$$

For SC, the additional term is driven by the projection of the local mean sensitivity $\dot{\mu}(\boldsymbol{x})$ onto the likelihood-optimal balance-correction direction $\mathcal{H}^{-1}\boldsymbol{m}$. For REG, the additional term is driven by the conditional variance contribution $\sigma^2(\boldsymbol{x})$. Define

$$r_{\mathrm{SPP}}(\boldsymbol{x}) = \frac{\dot{\mu}(\boldsymbol{x})^\top \boldsymbol{e}_1}{m_0}, \qquad r_{\mathrm{SC}}(\boldsymbol{x}) = \frac{\dot{\mu}(\boldsymbol{x})^\top \mathcal{H}^{-1}\boldsymbol{m}}{a}, \qquad r_{\mathrm{REG}}(\boldsymbol{x}) = \frac{\sigma^2(\boldsymbol{x})}{\tau^2}.$$

Then, Theorem 5.1 shows that the structural behavior

$$\Gamma_A(\boldsymbol{x}) = \Gamma_{\mathrm{MLE}}(\boldsymbol{x}) + \big(\tau^2 - a\big)\, r_A(\boldsymbol{x})^2, \qquad \text{for } A \in \{\mathrm{SPP}, \mathrm{SC}, \mathrm{REG}\}.$$

Since $\tau^2 - a \geq 0$ by (3.7), we have for $A, B \in \{\mathrm{SPP}, \mathrm{SC}, \mathrm{REG}\}$

$$\Gamma_A(\boldsymbol{x}) \leq \Gamma_B(\boldsymbol{x}) \quad \Longleftrightarrow \quad |r_A(\boldsymbol{x})| \leq |r_B(\boldsymbol{x})|.$$

This immediately gives the following corollary.

Corollary 5.2 *We have the relations*

$$\begin{aligned}
\Gamma_{\mathrm{REG}}(\boldsymbol{x}) \leq \Gamma_{\mathrm{SPP}}(\boldsymbol{x}) \quad &\Longleftrightarrow \quad \left(\frac{\sigma^2(\boldsymbol{x})}{\tau^2}\right)^2 \leq \left(\frac{\dot{\mu}(\boldsymbol{x})^\top \boldsymbol{e}_1}{m_0}\right)^2, \\
\Gamma_{\mathrm{REG}}(\boldsymbol{x}) \leq \Gamma_{\mathrm{SC}}(\boldsymbol{x}) \quad &\Longleftrightarrow \quad \left(\frac{\sigma^2(\boldsymbol{x})}{\tau^2}\right)^2 \leq \left(\frac{\dot{\mu}(\boldsymbol{x})^\top \mathcal{H}^{-1}\boldsymbol{m}}{a}\right)^2.
\end{aligned}$$

The comparison with QMLE is different because QMLE changes the estimation equation itself. In view of Theorem 5.1, the extra variability induced by the quasi-likelihood geometry must dominate the balance-correction cost of REG. This gives the following corollary.

Corollary 5.3 *We have the following relation*

$$\Gamma_{\mathrm{REG}}(\boldsymbol{x}) \le \Gamma_{\mathrm{QMLE}}(\boldsymbol{x}) \quad \Longleftrightarrow \quad \dot{\mu}(\boldsymbol{x})^\top \left(\Sigma_{\mathrm{QMLE}} - \mathcal{H}^{-1}\right) \dot{\mu}(\boldsymbol{x}) \ge (\tau^2 - a) \left(\frac{\sigma^2(\boldsymbol{x})}{\tau^2}\right)^2.$$

There is no universal pointwise dominance under the first-order MSE criterion for $\mu(\boldsymbol{x})$. The preferred correction may depend on the covariate region under consideration, the variance function, the exposure distribution and the way errors are weighted.

For actuarial purposes, it is natural to aggregate pointwise errors over the portfolio distribution $(\boldsymbol{X}, W) \sim \mathbb{P}$. Let $\omega(\boldsymbol{X}, W) \ge 0$ be a weighting function and define the expected weighted first-order MSE constant by

$$\mathcal{R}_A(\omega) = \mathbb{E}\left[\omega(\boldsymbol{X}, W)\Gamma_A(\boldsymbol{X})\right]. \tag{5.2}$$

Two simple choices are $\omega(\boldsymbol{X}, W) = 1$, which gives an unweighted statistical criterion, and $\omega(\boldsymbol{X}, W) = W$, which gives an exposure-weighted actuarial criterion. From Theorem 5.1 we obtain for $A, B \in \{\mathrm{SPP}, \mathrm{SC}, \mathrm{REG}\}$

$$\mathcal{R}_A(\omega) - \mathcal{R}_B(\omega) = \left(\tau^2 - a\right) \mathbb{E}\left[\omega(\boldsymbol{X}, W)\left(r_A(\boldsymbol{X})^2 - r_B(\boldsymbol{X})^2\right)\right].$$

In particular,

$$\mathcal{R}_{\mathrm{REG}}(\omega) - \mathcal{R}_{\mathrm{SPP}}(\omega) = \left(\tau^2 - a\right) \mathbb{E}\left[\omega(\boldsymbol{X}, W)\left(\left(\frac{\sigma^2(\boldsymbol{X})}{\tau^2}\right)^2 - \left(\frac{\dot{\mu}(\boldsymbol{X})^\top \boldsymbol{e}_1}{m_0}\right)^2\right)\right],$$

$$\mathcal{R}_{\mathrm{REG}}(\omega) - \mathcal{R}_{\mathrm{SC}}(\omega) = \left(\tau^2 - a\right) \mathbb{E}\left[\omega(\boldsymbol{X}, W)\left(\left(\frac{\sigma^2(\boldsymbol{X})}{\tau^2}\right)^2 - \left(\frac{\dot{\mu}(\boldsymbol{X})^\top \mathcal{H}^{-1}\boldsymbol{m}}{a}\right)^2\right)\right],$$

and similarly, using Theorem 5.1,

$$\mathcal{R}_{\mathrm{REG}}(\omega) - \mathcal{R}_{\mathrm{QMLE}}(\omega) = \mathbb{E}\left[\omega(\boldsymbol{X}, W)\left(\dot{\mu}(\boldsymbol{X})^\top(\mathcal{H}^{-1} - \Sigma_{\mathrm{QMLE}})\dot{\mu}(\boldsymbol{X}) + (\tau^2 - a)\left(\frac{\sigma^2(\boldsymbol{X})}{\tau^2}\right)^2\right)\right].$$

These formulas are useful diagnostic criteria. They can be estimated empirically by replacing the population expectations, $\mathbb{E} = \mathbb{E}_{\mathbb{P}}$, by sample averages and using plug-in estimates of $\boldsymbol{\beta}^+$, φ, $\mathcal{H}$, $\boldsymbol{m}$, τ^2, a and Σ_{QMLE}. However, for a general weighting function ω, they do not provide a universal ordering of the four methods. The next subsection shows that the likelihood-induced weighting leads to a sharper comparison.

5.2 Average KL divergence

The integrated quadratic criteria (5.2) measure errors in the scale of the fitted pure premium. The average KL divergence uses a different local geometry. In an EDF model, a squared error in the mean has a larger distributional impact when the conditional variance is small, and a smaller distributional impact when the conditional variance is large. This is captured by the factor $1/\sigma^2(\boldsymbol{x})$ in the local KL expansion as the following proposition shows.

Proposition 5.4 *Let $P_{m,c}$ denote the EDF distribution with mean m and exposure-dispersion scaling $c = w/\varphi > 0$, and write for the KL divergence of $P_{m,c}$ from $P_{\widetilde{m},c}$*

$$D^{(c)}_{\mathrm{KL}}(m, \widetilde{m}) = D_{\mathrm{KL}}\left(P_{m,c} \,\middle\|\, P_{\widetilde{m},c}\right).$$

For an EDF with mean parameter μ and variance function $\sigma^2(\cdot)$, the local KL divergence of the true model with mean $\mu(\boldsymbol{x};\boldsymbol{\beta}^+)$ from a fitted model with mean $\widehat{\mu}(\boldsymbol{x})$ satisfies, for fixed $c>0$,

$$D_{\mathrm{KL}}^{(c)}\left(\mu(\boldsymbol{x};\boldsymbol{\beta}^+),\widehat{\mu}(\boldsymbol{x})\right) \;=\; \frac{c}{2}\,\frac{\left(\widehat{\mu}(\boldsymbol{x})-\mu(\boldsymbol{x};\boldsymbol{\beta}^+)\right)^2}{\sigma^2(\boldsymbol{x})}+o\left(\left(\widehat{\mu}(\boldsymbol{x})-\mu(\boldsymbol{x};\boldsymbol{\beta}^+)\right)^2\right). \tag{5.3}$$

The proof of Proposition 5.4 is given in Appendix B. For a fixed exposure $w>0$, the leading local KL constant associated with method A is therefore given by

$$\frac{w}{2\varphi}\,\frac{1}{\sigma^2(\boldsymbol{x})}\,\Gamma_A(\boldsymbol{x}).$$

Thus, at a fixed covariate value $\boldsymbol{x}$, the local KL comparison differs from the pointwise quadratic comparison (5.1) by the positive factor $1/\sigma^2(\boldsymbol{x})$, up to the exposure-dispersion scaling. The key difference appears after averaging over the portfolio, see (5.2), the KL criterion corresponds to the integrated quadratic criterion with the likelihood-induced weight

$$\omega_{\mathrm{KL}}(\boldsymbol{X},W)=\frac{W}{\varphi\sigma^2(\boldsymbol{X})}.$$

We now define the average KL divergence. For a given method A, let $\widehat{\mu}^A_{\mathcal{D}}(\boldsymbol{x})$ denote the fitted pure premium obtained on the i.i.d. learning sample $\mathcal{D}=(Y_i,\boldsymbol{X}_iW_i)_{i=1}^n$; we add $\mathcal{D}$ to the notation $\widehat{\mu}^A_{\mathcal{D}}(\boldsymbol{x})$ to indicated that the estimated models are learning sample dependent. For a new (independent) policy with covariates and exposure $(\boldsymbol{X},W)$, the true conditional distribution is the EDF with mean $\mu(\boldsymbol{X};\boldsymbol{\beta}^+)$ and dispersion scaling W/φ, whereas the fitted conditional distribution under method A is the same EDF with mean $\widehat{\mu}^A_{\mathcal{D}}(\boldsymbol{X})$ instead. We define

$$\mathcal{K}_A=\mathbb{E}_{\mathcal{D}}\left[\mathbb{E}_{\boldsymbol{X},W}\left[D_{\mathrm{KL}}^{(W/\varphi)}\left(\mu(\boldsymbol{X};\boldsymbol{\beta}^+),\widehat{\mu}^A_{\mathcal{D}}(\boldsymbol{X})\right)\right]\right],$$

the outer expectation is taken with respect to the learning sample $\mathcal{D}$, while the inner expectation averages over an independent generic policy $(\boldsymbol{X},W)$ drawn from the portfolio distribution. By Proposition 5.4, conditionally on $(\boldsymbol{X},W)$,

$$D_{\mathrm{KL}}^{(W/\varphi)}\left(\mu(\boldsymbol{X};\boldsymbol{\beta}^+),\widehat{\mu}^A_{\mathcal{D}}(\boldsymbol{X})\right) \;=\; \frac{W}{2\varphi}\,\frac{\left(\widehat{\mu}^A_{\mathcal{D}}(\boldsymbol{X})-\mu(\boldsymbol{X};\boldsymbol{\beta}^+)\right)^2}{\sigma^2(\boldsymbol{X})}+o_p(n^{-1}).$$

The following result gives both the average KL constants and the resulting ranking. It combines the KL expansion with the likelihood-based comparison of the parametric corrections.

Theorem 5.5 *For $A\in\{\mathrm{QMLE},\mathrm{SPP},\mathrm{SC},\mathrm{REG}\}$, we have*

$$\mathcal{K}_A=\frac{1}{2n}\,T_A+o(n^{-1})\qquad \text{with}\qquad T_A=\mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\,\Gamma_A(\boldsymbol{X})\right].$$

The constants are

$$\begin{aligned}
T_{\mathrm{QMLE}} &= \mathrm{trace}\,(\mathcal{H}\Sigma_{\mathrm{QMLE}}),\\
T_{\mathrm{SPP}} &= q+1+(\tau^2-a)\frac{\boldsymbol{e}_1^\top\mathcal{H}\boldsymbol{e}_1}{m_0^2},\\
T_{\mathrm{SC}} &= q+1+\frac{\tau^2-a}{a},\\
T_{\mathrm{REG}} &= q+1+\frac{\tau^2-a}{\tau^2}.
\end{aligned}$$

Moreover,

$$q+1 \;\leq\; T_{\mathrm{REG}} \;\leq\; T_{\mathrm{SC}} \;\leq\; \min\{T_{\mathrm{SPP}}, T_{\mathrm{QMLE}}\}. \tag{5.4}$$

The proof of Theorem 5.5 is given in Appendix C.

The interpretation of (5.4) requires some care. The inequality $T_{\mathrm{SC}} \leq \min\{T_{\mathrm{SPP}}, T_{\mathrm{QMLE}}\}$ compares three parametric corrections that remain in the GLM parameter space. It says that, in the local likelihood geometry, the constrained estimator uses the least costly parametric direction to restore balance. The additional inequality $T_{\mathrm{REG}} \leq T_{\mathrm{SC}}$ is of a different nature. REG is not a parametric GLM estimator. It corrects the fitted mean function directly and may leave the original GLM family. Hence this inequality does not contradict the parametric optimality of SC. It shows instead that, once balance corrections are allowed to act in the larger space of fitted mean functions, the variance-weighted allocation of the aggregate imbalance has a smaller local KL cost.

This last statement can be made explicit. Recall balance defect (3.8). Consider additive balance corrections of the unconstrained MLE of the form

$$\widehat{\mu}^r(\boldsymbol{X}) = \mu(\boldsymbol{X}; \widehat{\boldsymbol{\beta}}_n) - r(\boldsymbol{X})\, \mathcal{C}_n(\widehat{\boldsymbol{\beta}}_n).$$

At the population level, the first-order balance constraint is

$$\mathbb{E}[W r(\boldsymbol{X})] = \varphi. \tag{5.5}$$

Indeed, for the additive correction to satisfy the balance property, we must have

$$\sum_{i=1}^{n} W_i\, \widehat{\mu}^{\,r}(\boldsymbol{X}_i) = \sum_{i=1}^{n} W_i Y_i.$$

Using the definition of $\widehat{\mu}^{\,r}$, the difference between the two sides can be written as

$$\mathcal{C}_n(\widehat{\boldsymbol{\beta}}_n)\left(n\varphi - \sum_{i=1}^{n} W_i\, r(\boldsymbol{X}_i)\right).$$

Consequently, the empirical normalization

$$\frac{1}{n}\sum_{i=1}^{n} W_i\, r(\boldsymbol{X}_i) = \varphi$$

ensures exact balance. Its population counterpart, which is the relevant constraint for the first-order asymptotic analysis, is given by (5.5).

Next, from the proof of Theorem 5.1 in Appendix A, see formula (A.2), we have

$$\sqrt{n}\, \mathcal{C}_n(\widehat{\boldsymbol{\beta}}_n) \Longrightarrow R, \qquad \text{with } R \sim \mathcal{N}(0, \tau^2 - a),$$

and R is asymptotically independent of the first-order MLE error. Therefore, for fixed $\boldsymbol{x}$,

$$\sqrt{n}\left(\widehat{\mu}^r(\boldsymbol{x}) - \mu(\boldsymbol{x}; \boldsymbol{\beta}^+)\right) = \dot{\mu}(\boldsymbol{x})^\top \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^+\right) - r(\boldsymbol{x})\, \sqrt{n}\, \mathcal{C}_n(\widehat{\boldsymbol{\beta}}_n) + o_p(1).$$

The first-order MSE constant is thus

$$\Gamma_r(\boldsymbol{x}) = \dot{\mu}(\boldsymbol{x})^\top \mathcal{H}^{-1}\, \dot{\mu}(\boldsymbol{x}) + \left(\tau^2 - a\right) r(\boldsymbol{x})^2 = \Gamma_{\mathrm{MLE}}(\boldsymbol{x}) + \left(\tau^2 - a\right) r(\boldsymbol{x})^2.$$

Using the average KL weight $W/(\varphi\sigma^2(\boldsymbol{X}))$, the corresponding constant is

$$T_r = \mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\,\Gamma_{\mathrm{MLE}}(\boldsymbol{X})\right] + \left(\tau^2 - a\right)\mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\,r(\boldsymbol{X})^2\right].$$

The first term is independent of $r(\cdot)$ and is equal to $\mathrm{trace}\left(\mathcal{H}^{-1}\mathcal{H}\right) = q+1$; this is the term that corresponds to the pure MLE part $\Gamma_{\mathrm{MLE}}(\boldsymbol{X})$ under the correct variance specification, see Theorem 5.1. Hence minimizing the additional KL cost is equivalent to solving

$$\min_r \mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\,r(\boldsymbol{X})^2\right] \qquad \text{subject to} \quad \mathbb{E}[W r(\boldsymbol{X})] = \varphi.$$

A Lagrange multiplier argument gives the unique solution

$$r^\star(\boldsymbol{x}) = \frac{\sigma^2(\boldsymbol{x})}{\tau^2}.$$

This is precisely the first-order form of the REG correction. Thus, REG has two complementary interpretations. From the risk-sharing viewpoint, it is the large-pool linear regression approximation of the CMRS rule. From the local likelihood viewpoint, it is the least costly additive first-order balance correction of the MLE when corrections are allowed to act directly on the fitted mean function.

6 From balance to risk-sharing

The previous sections have considered the balance property mainly from a statistical perspective. We now turn to its actuarial interpretation and show that a balanced fitted model can also be viewed as an ex-post risk-sharing mechanism. In particular, the REG correction introduced in Section 4 admits three complementary interpretations: as a minimum-distance projection onto the balance constraint, as a linear approximation to the CMRS rule, and as a local KL projection. These interpretations provide a direct connection between ex-ante insurance pricing and ex-post loss allocation.

6.1 Ex-ante pricing versus ex-post allocation

Recall from (2.8) that a fitted model $\widehat{\mu}_{\mathcal{D}}(\cdot)$ satisfies the in-sample balance property if

$$\sum_{i=1}^n W_i\,\widehat{\mu}_{\mathcal{D}}(\boldsymbol{X}_i) = \sum_{i=1}^n W_i Y_i, \tag{6.1}$$

where $\mathcal{D} = (Y_i, \boldsymbol{X}_i, W_i)_{i=1}^n$ is the learning sample used to fit the model.
In classical insurance pricing, the fitted model is subsequently used ex ante to price new contracts. For a future portfolio $(Y_t, \boldsymbol{X}_t, W_t)_{t=1}^m$, the pure premiums charged to policyholders sum to

$$\sum_{t=1}^m W_t\widehat{\mu}_{\mathcal{D}}(\boldsymbol{X}_t).$$

There is no reason for this quantity to coincide with the total realized loss, and generally

$$\sum_{t=1}^m W_t\widehat{\mu}_{\mathcal{D}}(\boldsymbol{X}_t) \neq \sum_{t=1}^m W_t Y_t. \tag{6.2}$$

Thus, the balance property is an in-sample calibration property; it does not eliminate the aggregate risk associated with future claims. Insurance companies therefore need capital buffers to absorb fluctuations around the expected aggregate loss. Moreover, out-of-time calibration implicitly relies on sufficient stability of the claims process. In non-stationary environments, additional corrections may be required, for instance to account for inflation or concept drift; see Brauer et al. [3].

Risk-sharing pools and peer-to-peer insurance take a different perspective. Once the losses have been realized, the objective is to allocate the total experienced loss

$$S = \sum_{i=1}^{n} W_i Y_i \tag{6.3}$$

among the n participants. A risk-sharing rule consists of measurable functions $g_i(\cdot)$ satisfying

$$\sum_{i=1}^{n} g_i\left((Y_j, \boldsymbol{X}_j, W_j)_{j=1}^{n}\right) = S.$$

Hence, an ex-post risk-sharing rule satisfies a balance condition by construction.

Conversely, any fitted model satisfying (6.1) generates an ex-post allocation rule: participant i can simply be charged

$$g_i = W_i \widehat{\mu}_{\mathcal{D}}(\boldsymbol{X}_i).$$

The sum of these contributions then exactly matches the observed aggregate loss. This observation provides the first link between balance correction in insurance pricing and risk sharing.

6.2 Three interpretations of the REG correction

We now focus on the REG correction. Write, for shortness,

$$\widehat{\mu}_i = \mu(\boldsymbol{X}_i; \widehat{\boldsymbol{\beta}}_n) \text{ and } \widehat{\sigma}_i^2 = \sigma^2(\boldsymbol{X}_i; \widehat{\boldsymbol{\beta}}_n),$$

so that

$$\widehat{\mathrm{Var}}[Y_i \mid \boldsymbol{X}_i, W_i] = \frac{\varphi}{W_i} \widehat{\sigma}_i^2.$$

Recall from (4.1) that the REG fitted pure premium is

$$\widehat{\mu}_i^{\mathrm{REG}} = \widehat{\mu}_i + \frac{\widehat{\sigma}_i^2}{\sum_{j=1}^{n} W_j \widehat{\sigma}_j^2} \left(\sum_{j=1}^{n} W_j Y_j - \sum_{j=1}^{n} W_j \widehat{\mu}_j \right).$$

We next give three complementary interpretations of this expression.

REG as a minimum standardized adjustment. Consider all vectors of corrected pure premiums $(\mu_1^*, \ldots, \mu_n^*)$ satisfying the balance constraint

$$\sum_{i=1}^{n} W_i \mu_i^* = \sum_{i=1}^{n} W_i Y_i.$$

Among these balanced vectors, consider the one which minimizes the total squared adjustment from the original MLE fitted values, standardized by the estimated conditional variances:

$$\begin{aligned} \min_{\mu_1^*,\dots,\mu_n^*} \quad & \sum_{i=1}^n \frac{(\mu_i^* - \widehat{\mu}_i)^2}{(\varphi/W_i)\widehat{\sigma}_i^2} \\ \text{subject to} \quad & \sum_{i=1}^n W_i \mu_i^* = \sum_{i=1}^n W_i Y_i. \end{aligned} \tag{6.4}$$

Since $\widehat{\sigma}_i^2 > 0$, the objective function is strictly convex, and therefore (6.4) has a unique solution. Introducing a Lagrange multiplier λ, the first-order conditions give

$$\frac{W_i}{\varphi \widehat{\sigma}_i^2}(\mu_i^* - \widehat{\mu}_i) + \lambda W_i = 0,$$

and hence

$$\mu_i^* - \widehat{\mu}_i = -\lambda \varphi \widehat{\sigma}_i^2.$$

The balance constraint gives

$$-\lambda\varphi = \frac{\sum_{j=1}^n W_j Y_j - \sum_{j=1}^n W_j \widehat{\mu}_j}{\sum_{j=1}^n W_j \widehat{\sigma}_j^2}.$$

Consequently, the unique solution to (6.4) is precisely

$$\mu_i^* = \widehat{\mu}_i^{\mathrm{REG}}.$$

Thus, REG can be viewed as the projection of the vector of MLE fitted pure premiums onto the affine hyperplane defined by the balance property, using the inverse conditional variances as the local metric. Contracts whose losses have a larger conditional variance can absorb a larger correction, whereas departures from the MLE fit are more strongly penalized for contracts having a smaller conditional variance.

This characterization also makes clear that the basic REG correction does not impose positivity of the corrected pure premiums. If positivity is required, one may supplement (6.4) with the constraints $\mu_i^* \geq 0$. The resulting problem remains a convex quadratic program, but its solution is no longer given in general by the closed-form expression (4.1).

REG as a linear CMRS allocation. The second interpretation comes directly from risk sharing. Define the monetary loss brought to the pool by participant i as

$$Z_i = W_i Y_i.$$

Conditionally on $(\boldsymbol{X}_i, W_i)$, its fitted mean and variance are

$$\widehat{\mathbb{E}}[Z_i \mid \boldsymbol{X}_i, W_i] = W_i \widehat{\mu}_i,$$

and

$$\widehat{\mathrm{Var}}[Z_i \mid \boldsymbol{X}_i, W_i] = \varphi W_i \widehat{\sigma}_i^2.$$

The linear regression approximation to the CMRS rule derived by Denuit and Robert [9] allocates the deviation of the aggregate loss from its expected value proportionally to individual variances. Applying this principle with fitted conditional moments gives

$$\begin{aligned}\widehat{g}_i^{\mathrm{reg}}(S) &= W_i\widehat{\mu}_i + \frac{\varphi W_i\widehat{\sigma}_i^2}{\sum_{j=1}^n \varphi W_j\widehat{\sigma}_j^2}\left(S - \sum_{j=1}^n W_j\widehat{\mu}_j\right) \\ &= W_i\widehat{\mu}_i^{\mathrm{REG}}.\end{aligned} \tag{6.5}$$

Hence, the REG pure premium has an exact risk-sharing interpretation: $W_i\widehat{\mu}_i^{\mathrm{REG}}$ is the contribution assigned to participant i by the fitted linear-regression approximation to CMRS.
Formula (6.5) also provides a useful decomposition:

$$W_i\widehat{\mu}_i^{\mathrm{REG}} = \underbrace{W_i\widehat{\mu}_i}_{\text{systematic component}} + \underbrace{\frac{W_i\widehat{\sigma}_i^2}{\sum_{j=1}^n W_j\widehat{\sigma}_j^2}\left(S - \sum_{j=1}^n W_j\widehat{\mu}_j\right)}_{\text{allocation of the aggregate residual}}.$$

The first component reflects the individual risk characteristics through the GLM. The second component shares the aggregate deviation between realized and fitted losses according to the individual contributions to conditional pool variance. Summing over all participants gives exactly S.

REG as a local KL projection. The minimum-distance interpretation in (6.4) is also closely related to the likelihood geometry studied in Section 5.2. Indeed, Proposition 5.4 shows that, locally, the KL divergence between two EDF distributions whose means differ by a small amount is proportional to

$$\frac{W_i}{2\varphi\sigma_i^2}(\widetilde{\mu}_i - \mu_i)^2.$$

Replacing the unknown quantities by their MLE fitted counterparts, the objective function in (6.4) is, up to the factor 2, the sample analogue of the quadratic local KL cost of moving from the unconstrained MLE fitted means to a balanced vector of fitted means.
Consequently, REG may also be interpreted as the balanced vector which is locally closest to the original MLE fit in the likelihood-induced geometry. This finite-sample projection interpretation is consistent with the population first-order result obtained in Section 5.2: among additive first-order corrections acting directly on the fitted mean function, the variance-weighted allocation minimizes the additional average KL cost. Thus, the inverse-variance geometry appearing in (6.4) is not arbitrary; it is precisely the local geometry induced by the EDF likelihood.

6.3 Relation with the CMRS rule

Several risk-sharing rules have been proposed in the literature; see, e.g., Denuit et al. [7] and Feng [11]. Of particular relevance here is the CMRS rule introduced by Denuit and Dhaene [6]; see also Denuit and Robert [10]. For an axiomatic foundation of CMRS, see Jiao et al. [14], and for computational aspects, see Blier-Wong [2].
Conditionally on the individual risk characteristics, the exact CMRS contribution of participant i is

$$g_i(S, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, W_1, \ldots, W_n) = \mathbb{E}\left[W_iY_i\middle| S, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, W_1, \ldots, W_n\right]. \tag{6.6}$$

By the tower property,

$$\sum_{i=1}^{n} g_i(S, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, W_1, \ldots, W_n) = S,$$

so that CMRS is balanced by construction.

In some models, the exact CMRS rule has a particularly simple form. For instance, in the Poisson setting, conditioning independent Poisson losses on their sum leads to a multinomial allocation and

$$\begin{aligned} g_i(S, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, W_1, \ldots, W_n) &= \frac{\mathbb{E}[W_i Y_i \mid \boldsymbol{X}_i, W_i]}{\sum_{j=1}^{n} \mathbb{E}[W_j Y_j \mid \boldsymbol{X}_j, W_j]} S \\ &= W_i \mu(\boldsymbol{X}_i; \boldsymbol{\beta}^+) \frac{S}{\sum_{j=1}^{n} W_j \mu(\boldsymbol{X}_j; \boldsymbol{\beta}^+)}. \end{aligned}$$

The exact CMRS rule is therefore generally non-linear as a function of the individual loss distributions, whereas the approximation used to construct REG retains only their first two conditional moments. In large heterogeneous pools, Denuit and Robert [9] show that this leads to the linear regression allocation appearing in (6.5).

The role of REG can therefore be understood as follows. The original MLE fitted means provide the systematic ex-ante pricing component. Once the aggregate loss S has been observed, REG converts these fitted means into a balanced ex-post allocation by sharing the aggregate residual according to conditional variance contributions. Hence, the same quantities that describe heterogeneity in insurance pricing also determine how the remaining aggregate risk is shared among participants.

6.4 Reconciling pricing and risk sharing

The preceding interpretations highlight that $\widehat{\mu}_i^{\mathrm{REG}}$ is not merely an ad hoc correction designed to satisfy a calibration identity. The same expression arises from three different arguments:

- it is the minimum standardized adjustment of the MLE fitted pure premiums required to restore exact balance;
- multiplied by W_i, it is the fitted linear-regression approximation to the CMRS contribution of participant i;
- it is the local minimum-KL projection of the fitted mean vector onto the balance constraint.

These interpretations reconcile two actuarial problems that are usually treated separately. In ex-ante pricing, the regression model determines the systematic expected loss associated with each risk profile. In ex-post risk sharing, the same systematic component can be retained as a baseline, while the aggregate deviation between realized and expected losses is redistributed across the pool. For REG, this redistribution is governed by conditional variance contributions. The balance property therefore provides a direct bridge between statistical calibration of insurance prices and allocation of realized losses in a risk-sharing pool.

7 Conclusions

This paper revisited the balance property, a fundamental requirement in actuarial pricing. We first discussed three parametric balance corrections for generalized linear models (GLMs): quasi maximum likelihood estimation (QMLE), simple post-processing (SPP), and optimization under a side constraint (SC). Within the GLM framework, the SC estimator achieves the best performance under the asymptotic likelihood-based KL criterion. Consequently, if balance must be enforced while remaining within the GLM family, SC is the preferred approach.

We then introduced the variance-weighted regression (REG) correction, which starts from the maximum likelihood estimator and allocates the aggregate imbalance in proportion to the fitted conditional variances. This correction preserves the original likelihood fit as the systematic pricing component while enforcing exact portfolio balance. Beyond its computational simplicity, REG has a natural actuarial interpretation: it follows the asymptotic risk-sharing principle that allocates aggregate deviations according to individual risk contributions.

Unlike QMLE, SPP, and SC, REG corrects the fitted mean function directly rather than defining a new parameter estimator. Consequently, pointwise prediction error does not yield a universal ranking between these methods. Under the average KL criterion, however, REG attains a smaller leading asymptotic constant than SC because it is not restricted to the GLM parameter space. Thus, SC remains the optimal likelihood-based correction within the GLM family, whereas REG is locally KL-optimal among additive balance corrections at the mean-function level. The two methods therefore address complementary optimization problems rather than competing ones.

Finally, we linked balance correction to ex-post risk-sharing. Under the canonical GLM link, REG preserves the estimated systematic premiums while allocating the remaining idiosyncratic portfolio imbalance according to conditional variances, providing a principled and interpretable notion of fairness.

Several directions for future research remain. Positivity-preserving versions of REG deserve further investigation, as do finite-sample properties and robustness under variance misspecification. The variance-weighted principle may also be extended to more flexible predictive models, such as neural networks and gradient boosting, whenever conditional variances can be estimated. Further extensions include local or subgroup balance constraints and adjustments for non-stationary environments, where concept drift may require additional out-of-time calibration.

Acknowledgements

Michel Denuit gratefully acknowledges funding from the FWO and F.R.S.-FNRS under the Excellence of Science (EOS) programme, project ASTeRISK (40007517).

References

[1] Bailey, R.A. (1963). Insurance rates with minimum bias. *Proceedings CAS* **50**, 4-11.

[2] Blier-Wong, C. (2026). A Laplace-based perspective on conditional mean risk sharing. *arXiv*:2603.01434.

[3] Brauer, A., Menzel, P., Wüthrich, M.V. (2025). Model monitoring: A general framework with an application to non-life insurance pricing. *arXiv*:2510.04556.

[4] Bühlmann, H., Gisler, A. (2005). *A Course in Credibility Theory and its Applications*. Springer.

[5] Burguete, J., Gallant, R., Souza, G. (1982). On unification of the asymptotic theory of nonlinear econometric models. *Economic Review* **1**, 151-190.

[6] Denuit, M., Dhaene, J. (2012). Convex order and comonotonic conditional risk sharing. *Insurance: Mathematics and Economics* **51**, 265-270.

[7] Denuit, M., Dhaene, J., Robert, C.Y. (2022). Risk-sharing rules and their properties, with applications to peer-to-peer insurance. *Journal of Risk and Insurance* **89**, 615-667.

[8] Denuit, M., Hainaut, D., Trufin, J. (2019). *Effective Statistical Learning Methods for Actuaries. Volume 1: GLM and Extensions.* Springer.

[9] Denuit, M., Robert, C.Y. (2021). From risk sharing to pure premium for a large number of heterogeneous losses. *Insurance: Mathematics and Economics* **96**, 116-126.

[10] Denuit, M., Robert, C.Y. (2023). From risk reduction to risk elimination by conditional mean risk sharing of independent losses. *Insurance: Mathematics and Economics* **108**, 46-59.

[11] Feng, R. (2023). *Decentralized Insurance: Technical Foundation of Business Models*. Springer.

[12] Gallant, A.R., Holly, A. (1980). Statistical inference in an implicit, nonlinear, simultaneous equation model in the context of maximum likelihood estimation. *Econometrica* **48**, 697-720.

[13] Gouriéroux, C., Montfort, A., Trognon, A. (1984). Pseudo maximum likelihood methods: theory. *Econometrica* **52**, 681-700.

[14] Jiao, Z., Kou, S., Liu, Y., Wang, R. (2022). An axiomatic theory of anynomized risk sharing. *arXiv*:2208.07533.

[15] Jørgensen, B. (1997). *The Theory of Dispersion Models.* Chapman & Hall.

[16] Jung, J. (1968). On automobile insurance ratemaking. *ASTIN Bulletin - The Journal of the IAA* **5**, 41-48.

[17] Lindholm, M., Wüthrich, M.V. (2026). The balance property in insurance pricing. *Scandinavian Actuarial Journal* **2026**, 506-543.

[18] Nelder, J.A., Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A* **135**, 370-384.

[19] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

[20] Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss–Newton method. *Biometrika* **61/3**, 439-447.

[21] White, H. (1982). Maximum likelihood estimation in misspecified models. *Econometrica* **50**, 1-25.

[22] Wüthrich, M.V. (2020). Bias regularization in neural network models for general insurance pricing. *European Actuarial Journal* **10**, 179-202.

[23] Wüthrich, M.V. (2026). The balance property: The constrained case, with a view on risk sharing. *arXiv*:2606.07276.

[24] Wüthrich, M.V., Merz, M. (2023). *Statistical Foundations of Actuarial Learning and its Applications.* Springer.

Appendix

A Proof of Theorem 5.1

For a parametric method with asymptotic normality

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n^A - \boldsymbol{\beta}^+) \Longrightarrow \mathcal{N}(\mathbf{0}, \Sigma_A),$$

the delta method gives

$$\Gamma_A(\boldsymbol{x}) = \dot{\mu}(\boldsymbol{x})^\top \, \Sigma_A \, \dot{\mu}(\boldsymbol{x}). \tag{A.1}$$

Formula (A.1) gives the announced results for Γ_{QMLE}, Γ_{SPP}, and Γ_{SC}. We now turn to the REG correction. For shortness, write

$$\mu_i = \mu(\boldsymbol{X}_i; \boldsymbol{\beta}^+), \qquad \sigma_i^2 = \sigma^2(\boldsymbol{X}_i), \qquad \dot{\mu}_i = \dot{\mu}(\boldsymbol{X}_i).$$

Introduce the two empirical sums

$$\boldsymbol{A}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{W_i}{\varphi} \, \frac{Y_i - \mu_i}{\sigma_i^2} \, \dot{\mu}_i \qquad \text{and} \qquad B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{W_i}{\varphi} \left(Y_i - \mu_i \right).$$

Under the model assumptions, we get

$$\operatorname{Var}[B_n] = \mathbb{E}\left[\left(\frac{W}{\varphi} \right)^2 \operatorname{Var}[Y \mid \boldsymbol{X}, W] \right] = \mathbb{E}\left[\frac{W}{\varphi} \, \sigma^2(\boldsymbol{X}) \right] = \tau^2,$$

as well as

$$\begin{aligned} \operatorname{Var}[\boldsymbol{A}_n] &= \mathbb{E}\left[\left(\frac{W}{\varphi} \right)^2 \frac{\operatorname{Var}[Y \mid \boldsymbol{X}, W]}{\sigma^4(\boldsymbol{X})} \, \dot{\mu}(\boldsymbol{X}) \dot{\mu}(\boldsymbol{X})^\top \right] \\ &= \mathbb{E}\left[\frac{W}{\varphi} \frac{1}{\sigma^2(\boldsymbol{X})} \, \dot{\mu}(\boldsymbol{X}) \dot{\mu}(\boldsymbol{X})^\top \right] = \mathcal{H}, \end{aligned}$$

and

$$\operatorname{Cov}[B_n, \boldsymbol{A}_n] = \mathbb{E}\left[\left(\frac{W}{\varphi} \right)^2 \frac{\operatorname{Var}[Y \mid \boldsymbol{X}, W]}{\sigma^2(\boldsymbol{X})} \, \dot{\mu}(\boldsymbol{X}) \right] = \mathbb{E}\left[\frac{W}{\varphi} \, \dot{\mu}(\boldsymbol{X}) \right] = \boldsymbol{m}.$$

The score equation for the correctly specified MLE gives the usual first-order expansion

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^+) = \mathcal{H}^{-1} \boldsymbol{A}_n + o_p(1).$$

In accordance with (3.8), consider

$$\mathcal{C}_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \frac{W_i}{\varphi} \left(\mu(\boldsymbol{X}_i; \boldsymbol{\beta}) - Y_i \right).$$

A Taylor expansion around $\boldsymbol{\beta}^+$ yields

$$\mathcal{C}_n(\widehat{\boldsymbol{\beta}}_n) = \mathcal{C}_n(\boldsymbol{\beta}^+) + \left(\frac{1}{n} \sum_{i=1}^n \frac{W_i}{\varphi} \, \dot{\mu}_i \right)^\top \left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^+ \right) + o_p(n^{-1/2}).$$

Since

$$\sqrt{n}\,\mathcal{C}_n(\boldsymbol{\beta}^+) = -B_n \qquad \text{and} \qquad \frac{1}{n}\sum_{i=1}^{n} \frac{W_i}{\varphi}\,\dot{\mu}_i \overset{\text{a.s.}}{\to} \boldsymbol{m},$$

we obtain

$$R_n := \sqrt{n}\,\mathcal{C}_n(\widehat{\boldsymbol{\beta}}_n) = -B_n + \boldsymbol{m}^\top \mathcal{H}^{-1} \boldsymbol{A}_n + o_p(1).$$

Consequently,

$$R_n \Longrightarrow R, \qquad \text{with } R \sim \mathcal{N}(0, \tau^2 - a), \tag{A.2}$$

because

$$\operatorname{Var}[R] = \tau^2 + \boldsymbol{m}^\top \mathcal{H}^{-1}\mathcal{H}\mathcal{H}^{-1}\boldsymbol{m} - 2\boldsymbol{m}^\top \mathcal{H}^{-1}\boldsymbol{m} = \tau^2 - a.$$

Moreover,

$$\operatorname{Cov}\left[\mathcal{H}^{-1}\boldsymbol{A}_n, R_n\right] \to \mathcal{H}^{-1}\left(-\boldsymbol{m} + \mathcal{H}\mathcal{H}^{-1}\boldsymbol{m}\right) = \mathbf{0}.$$

Since the joint limit is Gaussian, R is asymptotically independent of the first-order MLE error. The aggregate imbalance satisfies

$$\sum_{j=1}^{n} W_j Y_j - \sum_{j=1}^{n} W_j\,\mu(\boldsymbol{X}_j; \widehat{\boldsymbol{\beta}}_n) = -n\,\varphi\,\mathcal{C}_n(\widehat{\boldsymbol{\beta}}_n) = -\sqrt{n}\,\varphi\,R_n.$$

Furthermore,

$$\frac{1}{n}\sum_{j=1}^{n} W_j\,\widehat{\sigma}_j^2 \to \mathbb{E}\left[W\sigma^2(\boldsymbol{X})\right] = \varphi\,\tau^2.$$

The plug-in effect of replacing $\sigma^2(\boldsymbol{x})$ by $\widehat{\sigma}^2(\boldsymbol{x})$ and $\sum_j W_j\sigma_j^2$ by $\sum_j W_j\widehat{\sigma}_j^2$ is second-order in the correction term, because it multiplies an imbalance of order $O_p(n^{1/2})$ while the denominator is of order n. Hence,

$$\sqrt{n}\left(\frac{\widehat{\sigma}^2(\boldsymbol{x})}{\sum_{j=1}^{n} W_j\,\widehat{\sigma}_j^2}\left(\sum_{j=1}^{n} W_j Y_j - \sum_{j=1}^{n} W_j\,\mu(\boldsymbol{X}_j; \widehat{\boldsymbol{\beta}}_n)\right)\right) = -\frac{\sigma^2(\boldsymbol{x})}{\tau^2}\,R_n + o_p(1).$$

Combining this with the delta method expansion

$$\sqrt{n}\left(\mu(\boldsymbol{x}; \widehat{\boldsymbol{\beta}}_n) - \mu(\boldsymbol{x}; \boldsymbol{\beta}^+)\right) = \dot{\mu}(\boldsymbol{x})^\top \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^+\right) + o_p(1),$$

we obtain

$$\sqrt{n}\left(\widehat{\mu}^{\text{REG}}(\boldsymbol{x}) - \mu(\boldsymbol{x}; \boldsymbol{\beta}^+)\right) = \dot{\mu}(\boldsymbol{x})^\top \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^+\right) - \frac{\sigma^2(\boldsymbol{x})}{\tau^2}\,R_n + o_p(1).$$

Using the asymptotic independence established above ends the proof.

B Proof of Proposition 5.4

Let us establish the local expansion (5.3). Fix a covariate value $\boldsymbol{x}$, and write

$$m = \mu(\boldsymbol{x}; \boldsymbol{\beta}^+), \qquad \widetilde{m} = \widehat{\mu}(\boldsymbol{x}).$$

Let $c > 0$ be fixed. Recall that

$$D_{\mathrm{KL}}^{(c)}(m, \widetilde{m}) = D_{\mathrm{KL}}\left(P_{m,c} \,\|\, P_{\widetilde{m},c}\right),$$

where $P_{m,c}$ denotes the EDF distribution with mean m and exposure-dispersion scaling c.
Let κ be the cumulant function of the EDF with canonical link $h = (\kappa')^{-1}$. Thus, this provides us with the true and fitted canonical parameters are $\theta = h(m)$ and $\widetilde{\theta} = h(\widetilde{m})$, respectively. Up to the carrier term, which cancels in the KL divergence, the EDF log-density with exposure-dispersion scaling c contains the contribution $c\big(y\theta - \kappa(\theta)\big)$. Therefore, using $\mathbb{E}_m[Y] = m$, the KL divergence from the model with mean m to the model with mean $\widetilde{m}$ is

$$D_{\mathrm{KL}}^{(c)}(m, \widetilde{m}) = c\Big(m\big(h(m) - h(\widetilde{m})\big) - \kappa(h(m)) + \kappa(h(\widetilde{m}))\Big) = \frac{c}{2}\, L_\kappa(m, \widetilde{m}),$$

where L_κ is the deviance loss function introduced in (2.4). Clearly, $L_\kappa(m, m) = 0$, and for the first derivative in the second argument we have

$$\frac{\partial}{\partial \widetilde{m}} L_\kappa(m, \widetilde{m})\bigg|_{\widetilde{m}=m} = 2\,(\widetilde{m} - m)h'(\widetilde{m})\big|_{\widetilde{m}=m} = 0.$$

Differentiating once more, see also Wüthrich–Merz [24, Lemma 2.22],

$$\frac{\partial^2}{\partial \widetilde{m}^2} L_\kappa(m, \widetilde{m})\bigg|_{\widetilde{m}=m} = 2\left(h'(\widetilde{m}) + (\widetilde{m} - m)h''(\widetilde{m})\right)\big|_{\widetilde{m}=m} = 2h'(m) = \frac{2}{\kappa''(h(m))} = \frac{2}{V_h(m)} > 0,$$

the latter being twice the variance function under the canonical link choice $g = h$, see (3.2). I.e., the EDF variance function is $V_h(m) = \kappa''(h(m))$, and under the correctly specified variance assumption used in this paper, see (3.4), we have $V_h(m) = \sigma^2(\boldsymbol{x})$. The second-order Taylor expansion in $\widetilde{m} = m$ therefore yields

$$D_{\mathrm{KL}}^{(c)}(m, \widetilde{m}) = \frac{c}{2}\frac{1}{\sigma^2(\boldsymbol{x})}\,(\widetilde{m} - m)^2 + o\left((\widetilde{m} - m)^2\right),$$

because $c > 0$ is fixed. Returning to $m = \mu(\boldsymbol{x}; \boldsymbol{\beta}^+)$ and $\widetilde{m} = \widehat{\mu}(\boldsymbol{x})$, gives the desired expansion (5.3).

C Proof of Theorem 5.5

We first derive the expansion of the average KL divergence. For $A \in \{\mathrm{QMLE}, \mathrm{SPP}, \mathrm{SC}, \mathrm{REG}\}$, recall that $\widehat{\mu}_{\mathcal{D}}^A(\boldsymbol{x})$ denotes the fitted pure premium obtained from the learning sample $\mathcal{D}$. Conditionally on a new policy $(\boldsymbol{X}, W)$, we apply Proposition 5.4 with

$$c = \frac{W}{\varphi}, \qquad m = \mu(\boldsymbol{X}; \boldsymbol{\beta}^+), \qquad \widetilde{m} = \widehat{\mu}_{\mathcal{D}}^A(\boldsymbol{X}).$$

Thus,

$$D_{\mathrm{KL}}^{(W/\varphi)}\left(\mu(\boldsymbol{X}; \boldsymbol{\beta}^+), \widehat{\mu}_{\mathcal{D}}^A(\boldsymbol{X})\right) = \frac{W}{2\varphi}\,\frac{\left(\widehat{\mu}_{\mathcal{D}}^A(\boldsymbol{X}) - \mu(\boldsymbol{X}; \boldsymbol{\beta}^+)\right)^2}{\sigma^2(\boldsymbol{X})} + o\left(\left(\widehat{\mu}_{\mathcal{D}}^A(\boldsymbol{X}) - \mu(\boldsymbol{X}; \boldsymbol{\beta}^+)\right)^2\right).$$

Taking first the expectation with respect to the learning sample $\mathcal{D}$ and then with respect to an independent generic policy $(\boldsymbol{X}, W)$, we obtain, under the integrability assumptions allowing the local expansion to be averaged,

$$\mathcal{K}_A = \frac{1}{2}\,\mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\,\mathbb{E}_{\mathcal{D}}\left[\left(\widehat{\mu}^A_{\mathcal{D}}(\boldsymbol{X}) - \mu(\boldsymbol{X};\boldsymbol{\beta}^+)\right)^2 \,\Big|\, \boldsymbol{X}, W\right]\right] + o(n^{-1}).$$

By definition of Γ_A,

$$\mathbb{E}_{\mathcal{D}}\left[\left(\widehat{\mu}^A_{\mathcal{D}}(\boldsymbol{x}) - \mu(\boldsymbol{x};\boldsymbol{\beta}^+)\right)^2\right] = \frac{1}{n}\,\Gamma_A(\boldsymbol{x}) + o(n^{-1}).$$

Therefore,

$$\mathcal{K}_A = \frac{1}{2n}\,\mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\,\Gamma_A(\boldsymbol{X})\right] + o(n^{-1}).$$

This proves the first assertion with

$$T_A = \mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\,\Gamma_A(\boldsymbol{X})\right].$$

We now compute the four constants. From Theorem 5.1,

$$\Gamma_{\mathrm{QMLE}}(\boldsymbol{x}) = \dot{\mu}(\boldsymbol{x})^\top\,\Sigma_{\mathrm{QMLE}}\,\dot{\mu}(\boldsymbol{x}).$$

Hence,

$$\begin{aligned}
T_{\mathrm{QMLE}} &= \mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\,\dot{\mu}(\boldsymbol{X})^\top\,\Sigma_{\mathrm{QMLE}}\,\dot{\mu}(\boldsymbol{X})\right] \\
&= \mathrm{trace}\left(\Sigma_{\mathrm{QMLE}}\,\mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\,\dot{\mu}(\boldsymbol{X})\dot{\mu}(\boldsymbol{X})^\top\right]\right) \\
&= \mathrm{trace}\left(\Sigma_{\mathrm{QMLE}}\mathcal{H}\right) = \mathrm{trace}\left(\mathcal{H}\Sigma_{\mathrm{QMLE}}\right).
\end{aligned}$$

For the MLE component $\Gamma_{\mathrm{MLE}}(\boldsymbol{X})$ common to SPP, SC and REG, we have

$$\begin{aligned}
\mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\dot{\mu}(\boldsymbol{X})^\top\mathcal{H}^{-1}\dot{\mu}(\boldsymbol{X})\right] &= \mathrm{trace}\left(\mathcal{H}^{-1}\mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\dot{\mu}(\boldsymbol{X})\dot{\mu}(\boldsymbol{X})^\top\right]\right) \\
&= \mathrm{trace}(\mathcal{H}^{-1}\mathcal{H}) = q+1.
\end{aligned}$$

For SPP, we recall from Theorem 5.1

$$\Gamma_{\mathrm{SPP}}(\boldsymbol{x}) = \dot{\mu}(\boldsymbol{x})^\top\,\mathcal{H}^{-1}\,\dot{\mu}(\boldsymbol{x}) + \left(\tau^2 - a\right)\left(\frac{\dot{\mu}(\boldsymbol{x})^\top\boldsymbol{e}_1}{m_0}\right)^2.$$

Thus,

$$\begin{aligned}
T_{\mathrm{SPP}} &= q + 1 + \frac{\tau^2 - a}{m_0^2}\,\mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\left(\dot{\mu}(\boldsymbol{X})^\top\boldsymbol{e}_1\right)^2\right] \\
&= q + 1 + \frac{\tau^2 - a}{m_0^2}\,\boldsymbol{e}_1^\top\,\mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\,\dot{\mu}(\boldsymbol{X})\dot{\mu}(\boldsymbol{X})^\top\right]\boldsymbol{e}_1 \\
&= q + 1 + \left(\tau^2 - a\right)\frac{\boldsymbol{e}_1^\top\mathcal{H}\boldsymbol{e}_1}{m_0^2};
\end{aligned}$$

this is precisely one of the results that can be found in the proof of Theorem 5.1 in Wüthrich [23]. The SC derivation is completely analogous, see formula (5.3) in Wüthrich [23]. For REG, Theorem 5.1 gives

$$\Gamma_{\mathrm{REG}}(\boldsymbol{x}) = \dot{\mu}(\boldsymbol{x})^{\top}\mathcal{H}^{-1}\dot{\mu}(\boldsymbol{x}) + (\tau^2 - a)\left(\frac{\sigma^2(\boldsymbol{x})}{\tau^2}\right)^2.$$

Hence,

$$\begin{aligned} T_{\mathrm{REG}} &= q + 1 + \frac{\tau^2 - a}{\tau^4}\,\mathbb{E}\left[\frac{W}{\varphi\sigma^2(\boldsymbol{X})}\,\sigma^4(\boldsymbol{X})\right] \\ &= q + 1 + \frac{\tau^2 - a}{\tau^4}\,\mathbb{E}\left[\frac{W}{\varphi}\sigma^2(\boldsymbol{X})\right] = q + 1 + \frac{\tau^2 - a}{\tau^2}. \end{aligned}$$

It remains to prove the announced ranking. Since $0 < a \leq \tau^2$, we have

$$0 \leq \frac{\tau^2 - a}{\tau^2} \leq \frac{\tau^2 - a}{a}.$$

Consequently, $q + 1 \leq T_{\mathrm{REG}} \leq T_{\mathrm{SC}}$. The remaining rankings were proved in Theorem 5.1 of Wüthrich [23]. This completes the proof.