Effective *a Posteriori* Ratemaking with Large Insurance Portfolios via Surrogate Modeling

Sebastián Calcetero Vanegas*, Andrei L. Badescu, X. Sheldon Lin
Department of Statistical Sciences
University of Toronto
Toronto, Ontario
*sebastian.calcetero@mail.utoronto.ca

May 17, 2023

Abstract

A posteriori ratemaking in insurance uses a Bayesian credibility model to upgrade the current premiums of a contract by taking into account policyholders' attributes and their claim history. Most data-driven models used for this task are mathematically intractable, and premiums must be then obtained through numerical methods such as simulation such MCMC. However, these methods can be computationally expensive and prohibitive for large portfolios when applied at the policyholder level. Additionally, these computations become "black-box" procedures as there is no expression showing how the claim history of policyholders is used to upgrade their premiums. To address these challenges, this paper proposes the use of a surrogate modeling approach to inexpensively derive a closed-form expression for computing the Bayesian credibility premiums for any given model. As a part of the methodology, the paper introduces the "credibility index", which is a summary statistic of a policyholder's claim history that serves as the main input of the surrogate model and that is sufficient for several distribution families, including the exponential dispersion family. As a result, the computational burden of a posteriori ratemaking for large portfolios is therefore reduced through the direct evaluation of the closed-form expression, which additionally can provide a transparent and interpretable way of computing Bayesian premiums.

Keywords— Credibility, Surrogate modeling, Ratemaking, Bayesian Regression, Experience Rating

1 Introduction

Credibility, experience rating, or more recently the so-called "a posteriori" ratemaking are fundamental areas in actuarial science that enable actuaries to adjust premiums based on a policyholder's actual experience. Classical credibility theory, as formulated by Bühlmann and Gisler (2005), is based on the theory of Bayesian inference. The current understanding of a policyholder's risk behavior serves as the "a priori" information, which, when combined with the actual claims experience, results in "a posteriori" understanding of the policyholder's true risk behavior. This posterior knowledge can then be used to calculate upgraded premiums. The nature of the Bayesian credibility model depends on the goals of the modeling question and, more importantly, on the behavior of the data. Insurance data can be highly complex due to the size of the portfolio of policyholders and its heterogeneity. Consequently, several modeling approaches have been explored in the actuarial literature to calculate upgraded premiums.

As discussed by Norberg (2004), actuaries have primarily used simplistic assumptions under the classes of Bayesian conjugate prior families, best linear approximations or the so-called Bonus-Malus Systems (BMS) (see for e.g Asmussen et al. (2020)), to approach the problem of "a posteriori" ratemaking. Most commonly, they use the Bühlmann credibility formula and its variations (e.g., Bühlmann-Straub). The latter provides a non-parametric approach to credibility with a clear interpretation and inexpensive computations by approximating the predictive mean via a weighted average of the claim history's sample mean and the manual premium. However, this formula is extremely restrictive for "a posteriori" ratemaking applications because it only provides values for the predictive mean and not for other quantities of interest to actuaries (e.g., variance, quantiles, probabilities). Additionally, the resulting premium using Bühlmann's formula is not data-driven since it relies on strong assumptions such as linearity and lacks compatibility with different model structures (e.g., multivariate models, regression models, heavy-tail models). In fact, it is known that this expression is exact only for a limited modeling framework (see for e.g Jewell (1974)), and thus provides very poor performance when the true predictive mean is nonlinear, see for e.g Gómez-Déniz and Calderín-Ojeda (2014).

In today's insurance landscape, granular level information is increasingly available, and more complex insurance products, such as telematics-based insurance, have been developed. As a result, it is necessary to use tailored credibility models that account for the complexity of the data sets while still meeting regulatory requirements and standards, see for e.g Denuit et al. (2019). Such complicated structures that may appear in general insurance cannot be easily addressed by either the traditional conjugate prior credibility framework or simplistic closed-form approximation models. Furthermore, the inclusion of policyholder attributes is highly relevant, as noted by Ohlsson (2008), Xacur and Garrido (2018), and Diao and Weng (2019), and complicates the nature of the credibility model. Therefore, it is imperative to use general Bayesian modeling approaches that are data-driven and can provide a more flexible framework from which several quantities of interest can be derived for experience ratemaking, claim reserving, and risk management applications.

The general Bayesian workflow has not been widely used in insurance credibility applications due to the challenges of obtaining closed-form expressions. Many of the Bayesian models that provide a reasonable fit to real insurance data sets are mathematically intractable, and deriving Bayesian premiums requires computationally expensive numerical approximations, such as simulations via Markov Chain Monte Carlo (MCMC) methods (see for e.g Xacur and Garrido (2018), Zhang et al. (2018), Ahn et al. (2021)). Dealing with large and heterogeneous insurance portfolios is one of the significant challenges with these numerical methods, as a considerable number of simulations from the posterior predictive distribution of each policyholder are required. Additionally, if a large amount of experience is observed, these simulations must be repeatedly performed to update the premiums. Another issue with analytically intractable Bayesian credibility models is the "black-box" approach used to derive credibility premiums. These quantities are obtained through numerical approximations and not with closed-form expressions, which makes the upgrading premium system lack practical interpretation, thus becoming unappealing to practitioners as the results are not explainable to clients or regulators.

Our paper aims to facilitate and promote the use of data-driven Bayesian credibility models for a posteriori ratemaking in insurance by proposing an effective approach that aims to directly approximate the Bayesian credibility formula. There have been a limited number of studies in this regard in general insurance. For instance Virginia (1998) proposes to use B-Splines instead of the traditional linear credibility formula. Landsman (2002) uses a second-order Bayes estimator to approximate the predictive mean under non-conjugate models. Li et al. (2021) tackled this problem by approximating a dynamic Bayesian model via a mathematically tractable discrete Hidden Markov Model. In contrast, in the statistical literature, there have been some works that attempt to simplify the Bayesian inference problem of obtaining accurate approximations of the posterior or predictive distribution. Many of them are based on the use of summary statistics (a dimension reduction function) and so simplify the inference problem, see for e.g Sisson et al. (2018). Similar ideas can be used in *a posteriori* ratemaking in order to find a summary statistic of the policyholder's claim history that still provides the same information contained in that history (see for e.g. Taylor (1977), Sundt (1979)).

In this paper, we propose a surrogate modeling approach to address the challenges of computational burden and lack of closed-form expressions in Bayesian credibility models, specifically in performing efficient and transparent experience rating on complex insurance data sets, with a focus on large insurance portfolios. Our methodology derives an accurate analytical expression that approximates the Bayesian premiums resulting from any given Bayesian model. As part of the methodology, we introduce the credibility index, which is a summary statistic of the claim history for each policyholder and provides insights on the consistency of their experience with expected claims behavior. The credibility index is shown to be a sufficient statistic for a large class of distributional families, including the exponential dispersion family of distributions. Therefore, it can be used as the main input for the surrogate model, ensuring its accuracy in reproducing the credibility formula for non-tractable Bayesian models. The closed-form expression linking the credibility index to the credibility premium policy-wise allows for effective scaling of premium calculation to the portfolio level, particularly for large portfolios. In this scenario, we may compute true credibility premiums for a small percentage of policies, say 5%, from the portfolio and use the closed-form expression as a surrogate model to extrapolate the credibility premiums on the selected policies to the rest of the portfolio. Furthermore, the closed-form expression can be used for sensitivity analysis and provide some interpretation to the upgrading of premiums.

The paper is structured as follows. Section 2 provides an overview of the Bayesian framework for experience rating and highlights some of the primary challenges that will be addressed in the paper. Section 3 introduces the surrogate modeling approach to approximate any credibility formula and emphasizes the importance of sufficient summary statistics of the policyholder's claim history as inputs of such a model. In Section 4, we present the credibility index, its properties and its interpretation for insurance ratemaking. Section 5 explains the estimation of the surrogate model and how it simplifies the ratemaking process for large portfolios. Section 6 describes a simulation study that demonstrates the accuracy of the credibility formula for different Bayesian models. Section 7 presents a case study of experience rating using a real European automobile dataset, and finally, Section 8 concludes the paper.

2 The Bayesian Credibility Framework and Issues

The mathematical framework for classical credibility, as introduced by Bühlmann (see e.g Bühlmann and Gisler (2005)), is based on the Bayesian hierarchical model:

$$\mathbf{Y}_n|\Theta = \theta \sim_{iid} f(y|\theta, \mathcal{O})$$
$$\Theta \sim P(\theta)$$

where $\mathbf{Y}_n = (Y_1, \ldots, Y_n)$ and each $Y_j,\ i = 1, \ldots, n$ are the claim history (either claim size, frequency or other variables of interest) in period i of a policyholder, $f(y|\theta, \mathcal{O})$ is the density function of the *model distribution*, $\mathcal{O}$ is the set of parameters that contains other information available e.g policyholder attributes, Θ is the *latent variable* that represents the unobservable risk of each policyholder, and has a prior distribution $P(\theta)$. The variables Y_j and Θ can be either univariate or multivariate, however, we shall keep the notation as if they are univariate for reading convenience.

This formulation includes more general hierarchical models with several layers that are commonly used in credibility, see e.g Frees and Valdez (2008), Crevecoeur et al. (2022). For instance, a Bayesian model can have two (or more) layers of latent variables as follows:

$$\begin{aligned}\mathbf{Y}_n|\Theta_1 = \theta_1 &\sim_{iid} f(y|\theta_1, \mathcal{O})\\ \Theta_1|\Theta_2 = \theta_2 &\sim P(\theta_1|\theta_2)\\ \Theta_2 &\sim P(\theta_2)\end{aligned}$$

Do note that this is still embedded in the general formulation above by considering $\Theta = (\Theta_1, \Theta_2) \sim P(\theta) = P(\theta_1, \theta_2) = P(\theta_2)P(\theta_1|\theta_2)$. Similarly, the classical Bayesian setup also includes the case of regression-type credibility models such as panel data models widely used in experience rating or *a posteriori* ratemaking in insurance, see for e.g Tzougas and di Cerchiara (2021), Denuit et al. (2007), Bermúdez and Karlis (2017), Desjardins et al. (2023), Boucher and Guillén (2009). For instance, consider a Generalized Linear Mixed Model (GLMM) such as:

$$\begin{aligned}\mathbf{Y}_n|\Theta = \theta &\sim_{iid} f(y|\theta, \mathcal{O})\\ \eta(\theta; \langle \boldsymbol{x}, \boldsymbol{\beta}\rangle) &= \langle \boldsymbol{x}, \boldsymbol{\beta}\rangle + \varepsilon\\ \varepsilon &\sim P(\varepsilon)\end{aligned}$$

where $f(\cdot)$ is the model distribution, usually a member of the exponential family, $\eta(\cdot)$ would be the so-called link function that links the latent variable Θ to a regression on covariates $\boldsymbol{x}$ with coefficients $\boldsymbol{\beta}$, and ε is a random effect incorporated in the regression. Here we are using implicitly the set of parameters as $\mathcal{O} = \langle \boldsymbol{x}, \boldsymbol{\beta}\rangle$. To ease the reading, we will drop $\mathcal{O}$ in the notation, and only make explicit the dependence on covariates and their parameters $\mathcal{O}$ when necessary.

Given such a framework, the goal of *a posteriori* ratemaking consists of using the claim history $\mathbf{Y}_n$ to obtain a premium for the $(n+1)-th$ period. To do so, one first requires the predictive distribution of $Y_{n+1}|\mathbf{Y}_n$, which after simple manipulations can be computed as follows

$$f(Y_{n+1}|\mathbf{Y}_n) = \frac{E_\Theta\left[f(Y_{n+1}|\Theta)\exp(\ell(\mathbf{Y}_n|\Theta))\right]}{E_\Theta\left[\exp(\ell(\mathbf{Y}_n|\Theta))\right]} \tag{1}$$

where E_Θ is an expectation with respect to the prior distribution of Θ and $\ell(\mathbf{Y}_n|\Theta)$ is the conditional log-likelihood of the observed experience $\mathbf{Y}_n$ given by

$$\ell(\mathbf{Y}_n|\Theta) = \sum_{j=1}^{n} \log(f(Y_j|\Theta)).$$

The predictive distribution describes the risk behavior of a policyholder, and therefore it is the only quantity needed to compute any *premium principle* or *risk measure* that the actuary is considering. As defined by Kaas et al. (2008), a premium principle (or in general a risk measure) is an operator Π that assigns a given risk with a non-negative value, $\Pi(Y_{n+1}|\mathbf{Y}_n)$. Many premium principles have been studied in the actuarial literature as described in Dickson (2016) or Radtke et al. (2016). The most well-known examples are premiums based on operators defined through conditional expectations of the form $E(\pi(Y_{n+1})|\mathbf{Y}_n)$ for some weight function $\pi(y)$. Table 1 shows some examples.

Although this Bayesian framework is in theory sound, its implementation is not practical, especially when one needs to account for the claim history and policyholder's attributes in large and heterogeneous insurance portfolios. Indeed, the computation of the posterior predictive distribution

Premium principle	**Weight functions**	**Premium** $\Pi(Y_{n+1}\vert\mathbf{Y}_n)$
Net Premium	$\pi(y) = y$	$E(\pi(Y_{n+1})\vert\mathbf{Y}_n)$
Expected value	$\pi(y) = (1+\alpha)y$	$E(\pi(Y_{n+1})\vert\mathbf{Y}_n)$
Mean value / Utility function	$\pi(y) = \mathcal{U}(Y)$	$E(\pi(Y_{n+1})\vert\mathbf{Y}_n)$
Variance /Standard deviation	$\pi_1(y) = y, \pi_2(y) = y^2$	$E(\pi_1(Y_{n+1})\vert\mathbf{Y}_n) + \alpha E(\pi_2(Y_{n+1})\vert\mathbf{Y}_n)$
Exponential	$\pi(y) = \exp(\alpha y)$	$\log(E(\pi(Y_{n+1})\vert\mathbf{Y}_n))/\alpha$
Esscher	$\pi_1(y) = y\exp(\alpha y), \pi_2(y) = \exp(\alpha y)$	$E(\pi_1(Y_{n+1})\vert\mathbf{Y}_n)/E(\pi_2(Y_{n+1})\vert\mathbf{Y}_n)$

Table 1: Example of Premium Principles

in Equation (1) and any of the conditional expectations in Table 1 could be a challenging task under most practical considerations. It is known that closed-form solutions to these expressions can only be obtained under certain choices of the model distribution and the prior, which apply mostly to a limited number of simple parametric models such as the conjugate families of distributions. Moreover, most of the data-driven setups that commonly provide satisfactory fits to real insurance data sets do not belong to this limited class of models (see e.g Cheung et al. (2021), Czado et al. (2012), Yau et al. (2003)). As a result, performing Bayesian credibility is of much higher complexity due to the mathematical intractability of premium computations.

The most common approach used for the approximation of Bayesian credibility premiums is via Markov Chain Monte Carlo (MCMC) methods along the same lines as Bayesian inference procedures. These methods enable the actuary to draw samples from the posterior and predictive distribution in Equation (1) without the need of having a closed-form expression, and so the desired expectations in Table 1 can be obtained as the sample average of the simulated quantities. This process can be computationally expensive as it requires a considerable amount of simulation of the desired quantities before obtaining reliable estimates, and the fact that it must be performed individually for each policyholder when considering the heterogeneity of policyholders' attributes on a regression-type model. That said, when it comes to relatively large portfolios in non-life insurance, the repeated use of the MCMC process for each policyholder becomes almost prohibitive for practical applications because of the computational burden, see for e.g Ahn et al. (2021).

The issues mentioned earlier become more challenging when interpretability is at play. Ratemaking in insurance is a regulated activity and as a result, premiums must be transparent in terms of calculation so that policyholders are priced fairly without any type of discrimination, see for e.g Lindholm et al. (2022). However, when numerical methods are used to calculate premiums (as in the case of Bayesian premiums via MCMC), the ratemaking process becomes a "black box" due to the lack of a closed-form solution that links the policyholder's attributes and their claim history to the resulting premium. Therefore, the ratemaking process becomes unexplainable and so extremely unappealing for practitioners to find the true Bayesian premium. Therefore it is essential to be able to find a solution that reduces the computational burden and that is also interpretable. In the following sections, we propose a methodology that aims to address these challenges.

3 A Surrogate Modeling Approach

In order to address the computational burden of computing premiums in large insurance portfolios and the lack of transparency in the resulting premiums, we propose using a custom-built *surrogate modeling* approach. Surrogate models can help reduce the effort of computing the output of a function multiple times for different input values, especially when computing the output is computationally expensive. In such cases, a surrogate model approximates the output function in a more efficient way than the original computational process. The surrogate model is trained by using some known input-output pairs that are calculated using the true generating mechanism, and then extrapolates to new inputs to approximate their outputs. The training and extrapolation are assumed to be computationally inexpensive when compared to the real mechanism generating the outputs (see Sobester et al. (2008) for more on surrogate modeling). Some applications of such models have been explored in actuarial science in Lin and Yang (2020a) and Lin and Yang (2020b), and in ratemaking in Henckaerts et al. (2022).

In our context, the numerical challenge we aim to address with the surrogate model is the repeated use of numerical methods to compute the Bayesian credibility premiums. To introduce the main idea, note that any credibility formula under any premium principle can be generally expressed as:

$$\Pi(Y_{n+1}|\mathbf{Y}_n) = G_\Pi(\mathbf{Y}_n, n, \mathcal{O}) \tag{2}$$

where $G_\Pi(\cdot)$ is the theoretical or true functional form that links the claim history of the policyholder $\mathbf{Y}_n$ and the set of model parameters $\mathcal{O}$ to the credibility premium under the premium principle Π. Essentially, the function $G_\Pi(\cdot)$ is the integral of the premium principal with respect to the posterior predictive distribution, which, in turn, is the integration of the posterior distribution, as shown in Equation (1). The functional form of $G_\Pi(\cdot)$ entirely depends on the premium principle and the underlying Bayesian model. For instance, consider the net premium principle, a model distribution from the exponential family, and its conjugate prior distribution. Under some regularity conditions, the function G_Π admits an explicit solution as shown by Jewell (1974):

$$G_\Pi(\mathbf{Y}_n, n, \mathcal{O}) = Z_n \frac{\sum_{j=1}^n Y_j}{n} + (1 - Z_n)\Pi(Y_{n+1})$$

where Z_n is the so-called credibility factor, and $\Pi(Y_{n+1}) = E(Y_{n+1})$ is the so-called manual premium, which is the premium determined by the underlying Bayesian model without accounting for the claim history. Do note that this is one of the very few cases where there is an explicit expression.

That said, the target of the surrogate model is to approximate the function $G_\Pi(\cdot)$ in whole generality in an inexpensive fashion, say with a function $\hat{G}_\Pi(\cdot)$ so that the calculation of Bayesian credibility premiums for large portfolios becomes tractable. Indeed, once a fitted surrogate function $\hat{G}_\Pi(\cdot)$ is obtained, the experience rating process for a new policyholder is performed by direct evaluation of the surrogate function. Furthermore, note that the surrogate function provides an analytical expression that links the claim history of the policyholder with the resulting premium. Therefore, the ratemaking process becomes transparent as the function $\hat{G}_\Pi(\cdot)$ can be used for sensitivity analysis to interpret the upgrading of premiums and quantify the effect of the claim history and attributes, see for e.g Henckaerts et al. (2022), Lin and Yang (2020b) and Jones and Chen (2021) for examples in insurance. Unfortunately, the construction of such a surrogate function is not a trivial task in our context due to the heterogeneity of the portfolio of policyholders and the high and varying dimensionality of such a function. One of the contributions of this paper is to address this construction as follows.

In order to fit the surrogate model, a predefined grid of points is used where both inputs and outputs are known. In our context, this is equivalent to knowing the premiums associated with the underlying Bayesian model for some of the policyholders in the portfolio. However, the first main concern is how to select those representative policyholders to ensure proper extrapolation. Given the high level of heterogeneity of the portfolio, the inputs of the surrogate model have a wide variation that must be accurately reflected in the chosen points of the grid to avoid biases in the extrapolation.

In this paper, we approach this task by using techniques from population sampling (see, for example, Chambers and Clark (2012)) for the selection of representative policyholders. From this representative group, a minimal proportion of policyholders (i.e., 1% to 5%) can be selected from the portfolio while still properly reflecting the overall heterogeneity. The details are discussed in Section 5.1.

On the other hand, we have that the function $G_{\Pi}(\cdot)$ is characterized by high and varying dimensionality, which arises because each value of the claim history $\mathbf{Y}_n$, as well as all other attributes of the policyholder, constitutes an n-dimensional input for the function. The dimensionality is not fixed, as the number of observed claim history periods n may differ from one policyholder to another, ranging from a single period to many. Consequently, fitting an approximation function is challenging due to the curse of dimensionality, see for e.g Hou and Behdinan (2022).

A well-known approach to address the dimensionality problem is to use dimensionality reduction techniques through a non-linear transformation that summarizes the data, as discussed in Hou and Behdinan (2022). Specifically, this involves identifying a lower-dimensional function of the claim history and the policyholder's attributes that captures the same information contained in the entire set $\mathbf{Y}_n$. In theoretical statistics, this is accomplished through *sufficient statistics*, see for e.g Casella and Berger (2021). In Bayesian inference, for example, using a sufficient statistic, say $T(\mathbf{Y}_n)$, instead of the entire claim history $\mathbf{Y}_n$ as the basis for inference results in the posterior predictive distribution of the next period satisfying $f(Y_{n+1}|\mathbf{Y}_n) = f(Y_{n+1}|T(\mathbf{Y}_n))$, see for e.g Bernardo and Smith (2009), Section 4.5. Therefore, for a posteriori inference, it is unnecessary to consider the entire claim history $\mathbf{Y}_n$, but only the information contained in the sufficient statistic $T(\mathbf{Y}_n)$. As a result, the functional form of $G_{\Pi}(\cdot)$ can be simplified as follows:

$$\Pi(Y_{n+1}|\mathbf{Y}_n) = \tilde{G}_{\Pi}(T(\mathbf{Y}_n), n, \mathcal{O}) \tag{3}$$

for a different, yet very similar, function $\tilde{G}_{\Pi}(\cdot)$ that depends on the claim history only as a function of such sufficient statistic, which is of a much lower and fixed dimension than $\mathbf{Y}_n$. For illustration, consider again the example above. In such a case, note that for exponential dispersion families, the sufficient statistic is given by $T(\mathbf{Y}_n) = \frac{\sum_{j=1}^{n} Y_j}{n}$ and so the simplified form of the function $G_{\Pi}(\cdot)$ is:

$$\tilde{G}_{\Pi}(T(\mathbf{Y}_n), n, \mathcal{O}) = Z_n T(\mathbf{Y}_n) + (1 - Z_n)\Pi(Y_{n+1})$$

In practice, sufficient statistics might not be known depending on the complexity of the underlying Bayesian model and the heterogeneity of the data, however, one might still argue that if a summary statistic is able to summarize the information properly, then the above relationship between the posterior predictive distributions might hold approximately. This is indeed the idea of the so-called *approximately sufficient statistics* in the context of Approximate Bayesian Computation (ABC), where inference is based on summary statistics rather than of the whole data, see Joyce and Marjoram (2008), Sunnåker et al. (2013) and Fearnhead and Prangle (2012), for further references. Along those lines, then one can aim to estimate the desired surrogate function, say $\hat{\tilde{G}}_{\Pi}(\cdot)$, based on approximately sufficient summary statistics. However, in order to use this approach for our credibility application, we must select a suitable summary statistic for the claim history $\mathbf{Y}_n$.

In credibility theory, the selection of a proper summary statistic for dimension reduction is an open problem. As the aim of the ratemaking process is to be transparent and explainable, such a summary statistic should have a clear interpretation in describing the risk behavior of a policyholder's claim history while simultaneously capturing all the information it contains. However, the choice of such summary statistics remains under-explored in the experience rating and posterior ratemaking literature, despite some implicit reliance on the idea, such as in Künsch (1992), where the sample median is used instead of the mean, or Yan and Song (2022), which considers a general class of credibility formulas based on linear combinations of estimators of the mean. However, none of these approaches discuss the relevance of using such summary statistics, nor do they account for the heterogeneity of the portfolio.

In the next section, we introduce the "credibility index", a summary statistic of the claim history that can be effectively used in the surrogate model and that has desirable properties for its use in ratemaking. The construction, interpretation, and theoretical results are given in the next section. The details of how to use this summary statistic and how to fit a surrogate model in the context of large portfolios are provided in Section 5.

4 The Credibility Index

In this section, we introduce the concept of the "credibility index," which is a summary statistic that measures the likelihood of the claim history of a policyholder under a distorted probability measure. This index is tailor-made to the underlying Bayesian model, incorporates the policyholder attributes, and may be used by actuaries for interpretation of the experience rating process. Furthermore, it is designed to capture most of the information in the claim history of a policyholder.

As we discussed in the previous section, we seek a summary statistic that serves as a sufficient statistic, or that is very close to being one. In statistics, it is well-known that the likelihood, as a function of both the latent variable and the observed data, is always sufficient (see, for example, Lemma 1 in Mayo Wilson (2020) or Schweder and Hjort (2016)). Therefore, it is natural to consider the likelihood function itself (or the log-likelihood) as a reasonable choice for the summary statistic. However, the likelihood cannot be used as a statistic since the value of the latent variable is not fixed. Despite this, the likelihood function provides enough motivation to define our candidate summary statistic, the credibility index, as a modification of the likelihood function, as follows:

Definition 1 (Credibility index)**.** *The credibility index for a policyholder is defined as the following summary statistic:*

$$\mathcal{L}(\mathbf{Y}_n; \tilde{\theta}, \mathcal{O}) = \sum_{j=1}^{n} \log f(Y_j | \Theta = \tilde{\theta}, \mathcal{O}) \tag{4}$$

for a value $\tilde{\theta} \in \mathcal{R}_\Theta$, that is policyholder dependent, and that is determined in a data-driven fashion with the claim experience and the premium.

The modification in our approach involves selecting specific values $\tilde{\theta}$ from the latent variable domain $\mathcal{R}_\Theta$, one for each policyholder, instead of considering all possible values as it is done in the likelihood function. It's important to note that these $\tilde{\theta}$ values may not have a direct connection with the latent variable Θ associated with each policyholder. Rather, they can be viewed as some sort of discretization of the latent variable domain, providing a simplified representation of the overall likelihood function. To ensure the usefulness of this simplification for the surrogate model, proper tuning of these values is necessary. In practice, we adopt a data-driven approach where a model is trained to establish a link between each policyholder and its associated value $\tilde{\theta}$, so that it minimizes the overall approximation error of premiums. Further details regarding this tuning process can be found in Section 5.3. Once the $\tilde{\theta}$ values are determined, the credibility index can be considered as a summary statistic, as it solely relies on the claim history, and so can be interpreted as follows.

From a probabilistic viewpoint, the values $\tilde{\theta}$ may be viewed as the analogous of shape parameters that adjust the distribution function's structure to enhance the differentiation between claim histories. Consequently, the resulting summary statistic can be considered as a log-likelihood, but under a distorted probability measure utilized in the ratemaking process. We can think of this distorted measure in the same spirit as the pricing measure in finance, where premiums are obtained under the risk-neutral measure instead of the real-world probability measure. Analogous methods have also been applied in ratemaking in insurance, as demonstrated in Wang (2000). Nevertheless, we want to emphasize that the credibility index is a summary statistic that one uses to describe the data and it is not defining any kind of distributional model nor will be used along those lines.

The credibility index is not always a sufficient statistic, however, one can think of it as being close to one because of the similarities to the likelihood function, which is sufficient. In fact, we prove that the credibility index serves as a sufficient statistic for various families of distributions, including the extensively used exponential dispersion family, as discussed in Section 4.1.

We now proceed to present some desirable properties of the credibility index in the context of insurance ratemaking.

Interpretation of the index: The credibility index can be interpreted as a measure of the likeliness that a policyholder's claim history aligns with the expected behavior under parameters of the model $\mathcal{O}$, but under the distorted pricing measure. Specifically, the credibility index reflects the degree to which the observed claim behavior matches what the actuary anticipates based on the model, and so the index can be used to assess how "credible" is such anticipation when compared with the data. Intuitively, if these two match, then there is no reason to change the current premium, but if these don't, then the premium must be revised. The details of how to adjust the premium based on the credibility index will be explained in later sections.

We want to note that the credibility index has the same interpretation regardless of the nature of the variables $\mathbf{Y}_n$ (i.e. discrete or continuous) or the choice of the components in the underlying model (i.e., across model distributions). This is a desirable feature when considering models that account for frequency and severity simultaneously, as well as other types of information that are relevant to ratemaking, such as telematics.

Accounting for policyholder attributes: The credibility index's actual expression is tailored to the model distribution, taking into account all the actuarial considerations (e.g., policyholder information, tail heaviness, dependence structure) that have been incorporated into the model. This is in contrast to fixed-form summary statistics such as the sample mean or median, which do not account for any of these specifications. In models accounting for covariates, the set of parameters has the form $\mathcal{O} = \langle \boldsymbol{x}, \boldsymbol{\beta} \rangle$, where the regression function is introduced as another component in the summary statistic, independent of the value $\tilde{\theta}$.

$$\mathcal{L}(\mathbf{Y}_n; \mathcal{O}, \tilde{\theta}) = \sum_{j=1}^{n} \log f(Y_j |, \Theta = \tilde{\theta}, \langle \boldsymbol{x}, \boldsymbol{\beta} \rangle)$$

Additivity of the observed experience: The credibility index is an additive function of the policyholder's claim history, where each period of observed experience contributes a new term to the computation of the index. Thus, the overall information collected from a policyholder is the aggregation of information from each period. Consequently, to update the premium based on $n-1$ observations, $\mathbf{Y}_{n-1}$, we can use the already available credibility index, $\mathcal{L}(\mathbf{Y}_{n-1}; \tilde{\theta})$, add the contribution of the new observation via the relationship $\mathcal{L}(\mathbf{Y}_n; \tilde{\theta}) = \mathcal{L}(\mathbf{Y}_{n-1}; \tilde{\theta}) + \log f(Y_n|\tilde{\theta})$.

Sub-indexes for multivariate models: In insurance applications where policies have multiple coverages, the variable of interest Y_j is a vector of dimension D, which we shall denote as $Y_j = (Y_j^{(1)}, Y_j^{(2)}, \ldots, Y_j^{(D)})$, with $Y_j^{(d)}$ being the d-component of the vector. In such cases, the distribution function $f(Y_j|\Theta)$ is a multivariate conditional distribution. Most of the credibility models in the literature are constructed under a conditional independence structure in which the different components of the vector are conditionally independent given the latent variable Θ, see for e.g Frees et al. (2016), Tzougas and di Cerchiara (2021), Denuit et al. (2019) or Englund et al. (2008). In such a case, we can write the model distribution as the product of the individual conditional distributions of the vector, that is: $f(Y_j|\Theta) = \prod_{d=1}^{D} f(Y_j^{(d)}|\Theta)$. Therefore we could write the credibility index for a policyholder as

$$\mathcal{L}(\mathbf{Y}_n; \tilde{\theta}) = \sum_{d=1}^{D} \mathcal{L}^{(d)}(\mathbf{Y}_n^{(d)}; \tilde{\theta})$$

with

$$\mathcal{L}^{(d)}(\mathbf{Y}_n^{(d)}; \tilde{\theta}) = \sum_{j=1}^{n} \log f(Y_j^{(d)}|\Theta = \tilde{\theta}) \quad , d = 1, \ldots, D. \tag{5}$$

Hence, the credibility index can be broken down into D credibility sub-indices, with each sub-index corresponding to the claim history of a policyholder for the d-th component only. Each credibility sub-index can then be interpreted individually in the same way as the global index, i.e., the likeliness of the claim behavior for the d-th component of the policyholder, under the distorted measure. Thus, the likeliness of the whole claim history vector is obtained by summing the individual sub-indices. Consequently, the credibility index offers a lucid interpretation of the claim history for multivariate models, enabling the actuary to identify the components in the vector that drive most of the "likeness or unlikeness" of the claim behavior.

Accounting for partially observed claim history: In practice, claim information is often subject to policy limits and deductibles, which results in truncation and censoring, respectively. Therefore, the observed values Y_j for these types of claims cannot be used directly in traditional summary statistics. However, the credibility index can easily accommodate different types of partially observed claim information, making it a valuable tool for analyzing real-world datasets.

For instance, if the value Y_j is a right-censored observation, its contribution to the credibility index would be the term $\log\Big(1 - F(Y_j|\Theta = \tilde{\theta})\Big)$ instead of $\log\Big(f(Y_j|\Theta = \tilde{\theta})\Big)$, where F is the cumulative distribution function associated with the model distribution f. The same concept can be applied to the case of truncation, or to other types of modifications. An interesting case of censoring worth discussing is missing data, which can still be accommodated in the computation of the credibility index. If one of the D components of Y_j is entirely missing, then the contribution to the credibility index can be seen as $\log(1) = 0$. This is equivalent to not adding an observation to the associated credibility sub-index. This property is particularly useful in multivariate models when a policyholder has only an observed claim history in one line of business, and not in all.

4.1 On the Sufficiency of the Credibility Index

The motivation behind the introduction of the credibility index is to find a summary statistic of the claim history of a policyholder that captures almost all of the information on the claim history i.e. it is close to being a sufficient statistic. In this section, we show that a subtle simplification of the credibility index is a sufficient statistic for Θ (in the Bayesian sense) for many family distributions, including the exponential dispersion family of distributions that is commonly used for credibility models in insurance.

We first introduce the motivation of how to simplify the credibility index. Note that if we can decompose $\ell(\mathbf{Y}_n|\Theta) = l_1(\mathbf{Y}_n) + l_2(\mathbf{Y}_n, \Theta)$ for some functions l_1, l_2 (not necessarily probability functions), then we can omit the first term that depends only on $\mathbf{Y}_n$, as it simplifies both in the numerator and denominator of the predictive distribution in Equation (1). Therefore we can potentially find a simplified form of the credibility index without losing any information when it comes to performing inference on the posterior predictive distribution. With this in mind, we introduce a refined version of the credibility index and some of its properties.

Definition 2. *The refined credibility index denoted as $\tilde{\mathcal{L}}(\mathbf{Y}_n;\tilde{\theta})$, is defined as the "minimal" term in the following additive decomposition of the credibility index:*

$$\mathcal{L}(\mathbf{Y}_n;\tilde{\theta}) = L(\mathbf{Y}_n) + \tilde{\mathcal{L}}(\mathbf{Y}_n;\tilde{\theta}).$$

The term "minimal" is defined in the sense that if there is a further sub-decomposition of the form $\tilde{\mathcal{L}}(\mathbf{Y}_n;\tilde{\theta}) = L_1(\mathbf{Y}_n) + L_2(\mathbf{Y}_n;\tilde{\theta})$ for some functions L_1, L_2, then we must have $\tilde{\mathcal{L}}(\mathbf{Y}_n;\tilde{\theta}) = L_2(\mathbf{Y}_n;\tilde{\theta})$.

We now show that the refined credibility index may be a sufficient statistic for the latent variable Θ under certain conditions, and so guarantee the effectiveness of the credibility index in capturing the information of the claim history. To do so, let $T(\mathbf{Y}_n)$ be a sufficient statistic for the latent variable Θ. By the Fisher–Neyman factorization theorem, see for e.g Casella and Berger (2021), the conditional likelihood function can be factored into two non-negative functions, and so the log-likelihood function can be additively separated into two functions l_1 and l_2 as:

$$\ell(\mathbf{Y}_n|\Theta) = l_1(\mathbf{Y}_n) + l_2(T(\mathbf{Y}_n), \Theta) \quad \forall \mathbf{Y}_n, \Theta.$$

Proposition 1. *Suppose there exists a value $\tilde{\theta} \in R_\Theta$ at which $l_2(T(\mathbf{Y}_n), \tilde{\theta})$ is a one-to-one function when viewed as a function of the sufficient statistic $T(\mathbf{Y}_n)$. Then the refined credibility index $\tilde{\mathcal{L}}(\mathbf{Y}_n;\tilde{\theta})$ at the value $\tilde{\theta}$, is a sufficient statistic for the latent variable Θ.*

Proof. The decomposition of the log-likelihood above holds for every value of $\theta \in \mathcal{R}_\Theta$, in particular the one in the assumption. When fixing the value of Θ, the log-likelihood function at the left-hand side of the decomposition becomes the credibility index, and so at this particular value $\tilde{\theta}$ we must have the following relationship in terms of the refined credibility index:

$$\tilde{\mathcal{L}}(\mathbf{Y}_n;\tilde{\theta}) + L(\mathbf{Y}_n) = l_1(\mathbf{Y}_n) + l_2(T(\mathbf{Y}_n), \tilde{\theta})$$

And so we have:

$$\tilde{\mathcal{L}}(\mathbf{Y}_n;\tilde{\theta}) = (l_1(\mathbf{Y}_n) - L(\mathbf{Y}_n)) + l_2(T(\mathbf{Y}_n), \tilde{\theta})$$

The last expression provides an additive decomposition of the refined credibility index, and by the construction of its minimality, we must have that:

$$\tilde{\mathcal{L}}(\mathbf{Y}_n;\tilde{\theta}) = l_2(T(\mathbf{Y}_n), \tilde{\theta})$$

Now, by the assumption of the function $l_2(T(\mathbf{Y}_n), \tilde{\theta})$ at that specific value $\tilde{\theta}$, the right-hand side is a one-to-one function when viewed as a function of $T(\mathbf{Y}_n)$. Therefore, the left-hand side is a one-to-one function of a sufficient statistic and so the refined credibility index $\tilde{\mathcal{L}}(\mathbf{Y}_n;\tilde{\theta})$ is also a sufficient statistic for the latent variable Θ. Moreover, if $T(\mathbf{Y}_n)$ is also minimal sufficient, then by the one-to-one correspondence, $\tilde{\mathcal{L}}(\mathbf{Y}_n,\tilde{\theta})$ is also minimal sufficient.

□

Corollary 1. *Under the assumptions above, the following holds for any premium principle* Π

$$\Pi(Y_{n+1}|\mathbf{Y}_n) = G_\Pi(\mathbf{Y}_n, n, \mathcal{O}) = \tilde{G}_\Pi(\tilde{\mathcal{L}}(\mathbf{Y}_n;\tilde{\theta}), n, \mathcal{O}),$$

that is, one can replace the claim history vector with the refined credibility index without any loss of information for a posteriori *ratemaking purposes.*

Example 1 (Exponential dispersion family)**.** *In the particular case in which the model distribution $f(Y|\Theta,\mathcal{O})$ is given by a member of the exponential dispersion family of distributions with a dispersion parameter that does not depend on the latent variable (this is a construction widely used for insurance applications, see e.g Wuthrich and Merz (2022)), we have:*

$$f(Y_j|\Theta=\theta,\mathcal{O}) = \exp\left(\frac{\theta S(Y_j) - C(\theta)}{\varphi_j} - Q(Y_j;\varphi_j)\right)$$

for some functions $S(\cdot), C(\cdot), Q(\cdot)$ and the set of parameters $\mathcal{O}$ contains only the dispersion parameters φ_j. That said, the conditional log-likelihood takes the form:

$$\ell(\mathbf{Y}_n|\theta) = \sum_{j=1}^{n} \frac{\theta S(Y_j) - C(\theta)}{\varphi_j} - \sum_{j=1}^{n} Q(Y_j;\varphi_j)$$

And so the credibility index and the refined credibility index are respectively:

$$\mathcal{L}(\mathbf{Y}_n;\tilde{\theta}) = \tilde{\theta}\sum_{j=1}^{n} \frac{S(Y_j)}{\varphi_j} - \sum_{j=1}^{n} \frac{C(\tilde{\theta})}{\varphi_j} - \sum_{j=1}^{n} Q(Y_j;\varphi_j)$$

$$\tilde{\mathcal{L}}(\mathbf{Y}_n;\tilde{\theta}) = \tilde{\theta}\sum_{j=1}^{n} \frac{S(Y_j)}{\varphi_j} - \sum_{j=1}^{n} \frac{C(\tilde{\theta})}{\varphi_j}$$

It is known that $\sum_{j=1}^{n} \frac{S(Y_j)}{\varphi_j}$ is the minimal sufficient statistic for the exponential dispersion family and we can observe from above that the function linking the sufficient statistic and the refined credibility index is a one-to-one function for $\tilde{\theta} \neq 0$. Therefore, by the proposition above, the corrected credibility index is also a minimal sufficient statistic.

We note that the credibility index is one-dimensional in the sense that it reduces the information of the whole claim history of a policyholder to a single quantity. Therefore, the credibility index can potentially be a sufficient statistic when the latent variable Θ is one-dimensional. In the case of multivariate models for Y_j with D dimensions, this limitation can be weakened in the sense that now the credibility sub-indexes can also play a role to constitute a set of summary statistics that all together may be sufficient for a latent variable Θ. Indeed, the result of the proposition above extends directly if we consider a multivariate model for Y_j in which each of the components $Y_j^{(d)}$ is associated with at most one component of the latent variable $\tilde{\Theta}^{(d)}$ at a time. We illustrate this generalization with just an example as follows:

Example 2 (Multivariate exponential dispersion family). *Let's consider now the case of a particular multivariate exponential dispersion family of distributions with dispersion parameters that do not depend on the latent variable. This construction is also widely used for applications in insurance.*

$$f(Y_j|\Theta=\theta,\mathcal{O}) = \exp\left(\sum_{d=1}^{D} \frac{\theta^{(d)} S^{(d)}(Y_j^{(d)}) - C^{(d)}(\theta^{(d)})}{\varphi_j^{(d)}} - \sum_{d=1}^{D} Q^{(d)}(Y_j^{(d)};\varphi_j^{(d)})\right)$$

for some functions $S^{(d)}(\cdot), C^{(d)}(\cdot), Q^{(d)}(\cdot)$. Again, the set of parameters $\mathcal{O}$ is associated to the dispersion parameters. Observe that in this particular case the function $C(\theta)$ is split additively into each of the components, which is the case when the components of the vector Y_j are conditionally independent given Θ. The conditional log-likelihood takes the form

$$\ell(\mathbf{Y}_n|\theta) = \sum_{j=1}^{n}\sum_{d=1}^{D} \frac{\theta^{(d)} S^{(d)}(Y_j^{(d)}) - C^{(d)}(\theta^{(d)})}{\varphi_j^{(d)}} - \sum_{j=1}^{n}\sum_{d=1}^{D} Q^{(d)}(Y_j^{(d)};\varphi_j^{(d)}),$$

And so the credibility index and sub-indexes are respectively

$$\mathcal{L}(\mathbf{Y}_n;\tilde{\theta}) = \sum_{d=1}^{D}\tilde{\theta}^{(d)}\sum_{j=1}^{n}\frac{S^{(d)}(Y_j^{(d)})}{\varphi_j^{(d)}} - \sum_{d=1}^{D}\sum_{j=1}^{n}\frac{C^{(d)}(\tilde{\theta}^{(d)})}{\varphi_j^{(d)}} - \sum_{d=1}^{D}\sum_{j=1}^{n} Q^{(d)}(Y_j^{(d)};\varphi_j^{(d)}),$$

$$\mathcal{L}^{(d)}(\mathbf{Y}_n^{(d)};\tilde{\theta}^{(d)}) = \tilde{\theta}^{(d)}\sum_{j=1}^{n}\frac{S^{(d)}(Y_j^{(d)})}{\varphi_j^{(d)}} - \sum_{j=1}^{n}\frac{C^{(d)}(\tilde{\theta}^{(d)})}{\varphi_j^{(d)}} - \sum_{j=1}^{n} Q^{(d)}(Y_j^{(d)};\varphi_j^{(d)}),$$

$$\mathcal{L}(\mathbf{Y}_n;\tilde{\theta}) = \sum_{d=1}^{D}\mathcal{L}^{(d)}(\mathbf{Y}_n^{(d)};\tilde{\theta}^{(d)}).$$

The refined credibility index and corrected sub-indexes are respectively

$$\tilde{\mathcal{L}}(\mathbf{Y}_n;\tilde{\theta}) = \sum_{d=1}^{D}\tilde{\theta}^{(d)}\sum_{j=1}^{n}\frac{S^{(d)}(Y_j^{(d)})}{\varphi_j^{(d)}} - \sum_{d=1}^{D}\sum_{j=1}^{n}\frac{C^{(d)}(\tilde{\theta}^{(d)})}{\varphi_j^{(d)}},$$

$$\tilde{\mathcal{L}}^{(d)}(\mathbf{Y}_n^{(d)};\tilde{\theta}^{(d)}) = \tilde{\theta}^{(d)}\sum_{j=1}^{n}\frac{S^{(d)}(Y_j^{(d)})}{\varphi_j^{(d)}} - \sum_{j=1}^{n}\frac{C^{(d)}(\tilde{\theta}^{(d)})}{\varphi_j^{(d)}},$$

$$\tilde{\mathcal{L}}(\mathbf{Y}_n;\tilde{\theta}) = \sum_{d=1}^{D}\tilde{\mathcal{L}}^{(d)}(\mathbf{Y}_n^{(d)};\tilde{\theta}^{(d)}).$$

It is known that the set of statistics $\sum_{j=1}^{n}\frac{S^{(d)}(Y_j^{(d)})}{\varphi_j^{(d)}}$, $d=1,\ldots,D$ are the minimal sufficient statistic for this family, and the function linking each of these sufficient statistics and the associated corrected credibility sub-indexes are one-to-one functions for $\tilde{\theta}^{(d)} \neq 0$. Therefore, the set of credibility sub-indexes is also a minimal sufficient statistic.

5 Estimation of the Surrogate Model

In this section, we provide further details on how to fit the surrogate function $\hat{\tilde{G}}_{\Pi}(\cdot)$ and the tuning of the values $\tilde{\theta}$ to develop a surrogate function approach. The key steps to perform this task are presented below, and the remainder of this section discusses how each of these steps is achieved.

1. Selecting a sub-portfolio of representative policyholders.
2. Compute credibility premiums using simulation/numerical schemes only on such sub-portfolio.
3. Estimate the parameter $\tilde{\theta}$ and the nonlinear function $\hat{\tilde{G}}_{\Pi}(\cdot)$ using a method for interpolation, as chosen by the actuary.
4. Assess the accuracy of the fitted function and out-sample predictive behaviour.
5. Evaluate the fitted formula on the rest of the portfolio to obtain the premiums.

5.1 Selecting a representative sub-portfolio via population sampling

The selection of a representative sub-portfolio is a crucial step in obtaining accurate results through the surrogate function approach. This sub-portfolio should be small enough to be computationally efficient, yet large enough to exhibit similar properties to the original portfolio, ensuring reliable extrapolation. For large portfolios, a sample size between 1% and 10% of the whole portfolio may suffice for this purpose. It is recommended to start with a proportion of 1% and increase this number by 1 % at a time if the results are unsatisfactory, as we further discussed in Section 5.4.

Selecting representative policyholders is a well-studied problem in the statistical literature on population sampling, and several methodologies have been developed for this purpose, see for e.g., Chambers and Clark (2012). Population sampling is a statistical technique used to gather representative information about a larger group, known as the population. Instead of studying the entire population, a smaller subset called a sample, is selected using methods according to a sampling design. Such a design considers equal or unequal inclusion probabilities, which determine the likelihood of each element being included in the sample.

Given the particular goal of extrapolation, it is critical for the sub-portfolio (sample) to exhibit the same characteristics as the total portfolio concerning all the inputs involved in computing the credibility index, including the parameters of the model distributions (e.g covariates in a regression), the number of periods with observed exposure n, and the claim history $\mathbf{Y}_n$. Therefore, we recommend using model-assisted sampling methods that can sample from large populations while accounting for several attributes of the sub-portfolio to match those of the entire portfolio.

In this paper, we use the *cube method*, as described in Tillé (2011), which has been used in other applications of surrogate models in insurance, such as Lin and Yang (2020b). The cube method allows for the selection of samples from a population of any size, taking into account both equal and unequal inclusion probabilities. The samples are chosen in a way that aims to achieve *balance* across a specified set of attribute variables i.e. approximates equality or equilibrium in the distribution of the attributes variables. Briefly, the cube method consists of two phases: a flight phase and a landing phase. The flight phase defines inclusion probabilities that guarantee the balance property on the selected set of attribute variables in the sub-portfolio, and the landing phase converts these probabilities to either zero or one using linear programming, resulting in an approximately balanced random sample.

The algorithm has been implemented in several statistical packages, see for e.g the function `samplecube` in the sampling package in R, Tillé and Matei (2010). Therefore it is readily available for applications. We would like to remark that it is not computationally expensive as it is designed for large populations. The inputs are the data frame with the population of interest, including the attributes to keep balance in the sample, and starting point inclusion probability for each policyholder, which in our case is the sample size we want to sample from the population, e.g. 1%.

5.2 Computing credibility premiums for the sub-portfolio

The computation of credibility premiums can be achieved using numerical schemes based on either Markov chain Monte Carlo (MCMC) or quadrature methods. This problem is well-documented in the literature of computational Bayesian statistics. For further details, we refer the reader to works such as Sisson et al. (2018). We emphasize that this step represents the bottleneck process that we aim to mitigate as much as possible. Therefore, we remind the reader to use the most efficient algorithm available to them. For the sake of completeness, we present here a tailored-made setup that is quite efficient for the computation of premiums defined through the expectation operators, as motivated in Section 2.

We propose using a simulation approach via *importance sampling* in which the *biased distribution* of the posterior distribution of the latent variables is chosen to be the prior distribution. Therefore, expectations from the posterior predictive distribution can be estimated as

$$E(\pi(Y_{n+1})|\mathbf{Y}_n) \approx \hat{E}(\pi(Y_{n+1})|\mathbf{Y}_n) := \frac{\sum_{k=1}^{K} E\left[\pi(Y_{n+1})|\theta_k\right] \exp(\ell(\mathbf{Y}_n|\theta_k))}{\sum_{k=1}^{K} \exp(\ell(\mathbf{Y}_n|\theta_k))}, \tag{6}$$

where θ_k are iid samples drawn from the prior distribution $P(\theta)$. The value of K is selected such that reasonable estimates are guaranteed. Unlike the traditional MCMC approach that generates samples from the posterior, this setup generates samples from the prior distribution, making it easy to perform. Additionally, the prior distribution $P(\theta)$ is the same for every policyholder in the portfolio. Therefore, a single θ_k drawn can be used across policyholder simulations, resulting in a generation of samples that scales up efficiently to large sizes portfolios.

Finally, once these expectations are estimated, we are in a good position to find the associated credibility premium, according to the premium principle that is being considered (see e.g Table 1). Therefore, we end up with a reliable approximation of such premiums, which we denote by $\hat{\Pi}(Y_{n+1}|\mathbf{Y}_n)$ from now on.

5.3 Fitting of $\hat{\tilde{G}}_{\Pi}(\cdot)$ and the values $\tilde{\theta}$

To do this we use a non-parametric interpolation method in which the inputs are the set of parameters of the model, the credibility index and the number of periods of claim history of each policyholder. The outputs are the Bayesian credibility premiums for each policyholder in the sub-portfolio of representative policyholders. It should be noted that the literature on interpolation of general functions and surrogate modeling is extensive, see for e.g Mastroianni and Milovanovic (2008). Thus, the following description serves as a guideline, rather than a step-by-step recipe to be followed. The key points of this process are the selection of a structure for the surrogate model, followed by its estimation via least squares.

Selecting a structure for the surrogate function

In the literature, the most commonly used methods for fitting a surrogate function $\hat{\hat{G}}_\Pi(\cdot)$ are regression via Gaussian processes, as described by Sobester et al. (2008). However, other flexible nonlinear models, such as neural networks and spline-based methods, may also be used.

While these methods provide accurate estimates, the surrogate function $\hat{\hat{G}}_\Pi(\cdot)$ may have a non-clear structure of inputs, making it difficult to explain. To ensure transparency in the ratemaking process, it is necessary to select a particular structure for the surrogate function that allows for interpretation while also guaranteeing enough accuracy of the surrogate model. For example, one could choose a surrogate function of the form:

$$\hat{\hat{G}}_\Pi(\mathcal{L}(\mathbf{Y}_n;\tilde{\theta}), n, \mathcal{O}) = Z\mathcal{L}(\mathbf{Y}_n;\tilde{\theta}) + (1-Z)\Pi(Y_{n+1})$$

where the coefficient Z must be determined by the algorithm. This expression resembles the linear credibility formula proposed by Bühlmann, and in fact, the best linear approximation might be recovered as a particular case of the surrogate function under a proper parameterization and a proper summary statistic. This analogy provides another interpretation of the surrogate function as an empirical and nonlinear version of the regression problem proposed by Bühlmann. Other linear credibility formulas explored in the literature under premium principles different than the net premium might be recovered as well, as in Xie et al. (2018), Gómez-Déniz (2008), Najafabadi (2010) or Zhang (2022). However, we emphasize that the simplistic linear structure might not be able to provide accurate approximations in general, especially when the premium principle in consideration is not linear in nature and the credibility model is complex.

To obtain meaningful interpretations of the ratemaking process while maintaining accuracy when interpolating, we illustrate the methodology using a surrogate function with a rating factor functional form as shown in Equation (7) below. However, it is important to note that any other structure can be used.

$$\hat{\hat{G}}_\Pi(\mathcal{L}(\mathbf{Y}_n;\tilde{\theta}), n, \mathcal{O}) = \Pi(Y_{n+1}) \exp\Big(g(\mathcal{L}(\mathbf{Y}_n;\tilde{\theta}), n)\Big) \tag{7}$$

The expression given in Equation (7) resembles a process of rate upgrading, where $\Pi(Y_{n+1})$ represents the manual premium (i.e., the premium without experience), and the term on the right-hand side, $\exp\Big(g(\mathcal{L}(\mathbf{Y}_n;\tilde{\theta}), n)\Big)$, acts as a rating factor that adjusts the manual premium based on the claim history. The function $g(.)$ does not need to have any specific structure and can be non-linear to provide flexibility. Along those lines, the function $g(\mathcal{L}(\mathbf{Y}_n;\tilde{\theta}), n)$ describes how much of an adjustment should be made to have a premium that better reflects the actual experience. If the function $g(\mathcal{L}(\mathbf{Y}_n;\tilde{\theta}), n)$ equals 0, no adjustment is required; if $g(\mathcal{L}(\mathbf{Y}_n;\tilde{\theta}), n) > 0$, the premium must be increased; and if $g(\mathcal{L}(\mathbf{Y}_n;\tilde{\theta}), n) < 0$, the premium should be reduced. As a result, the ratemaking process acquires a more transparent view of the actuary, as the credibility premium can be interpreted and explained to both clients and regulators. Section 7 provides a clearer view of this interpretation in a case study.

Least squares estimation

Consider a sub-portfolio of M policyholders indexed by $i = 1, \ldots, M$. Let $\hat{\Pi}_i^p$ denote the credibility premium for the i-th policyholder obtained via simulation, represented by the $\hat{\Pi}(Y_{n+1}|\mathbf{Y}_n)$ values. Since these values were obtained as the sample mean of a large number of simulations, we can assume that $\hat{\Pi}_i^p \approx \text{Normal}(\Pi_i^p, se_i^2)$, where Π_i^p is the credibility premium $\Pi(Y_{n+1}|\mathbf{Y}_n)$ and se_i^2 is the standard error of the estimation. Let Π_i denote the manual premium of the i-th policyholder, i.e. the premium $\Pi(Y_{n+1})$ without incorporating claim history, and let $\mathcal{L}_i(\mathbf{Y}_n, \tilde{\theta}_i)$ denote the credibility index at the value $\tilde{\theta}_i$ for the i-th policyholder in the portfolio.

Assuming the rating factor functional form, our objective is to fit the relationship in Equation (7), where we need to estimate the function $g(\cdot)$ that depends on two inputs, the credibility index and the number of previously observed claim periods n. Since $g(\cdot)$ is an unstructured function, we can use a nonlinear formulation for its estimation, such as basis decomposition using B-splines.

Note that we can write such specific functional form as:

$$\hat{\Pi}_i^p \approx \text{Normal}(\Pi_i^p, se_i^2)$$
$$\log(\Pi_i^p) = \log(\Pi_i) + g(\mathcal{L}_i(\mathbf{Y}_n, \tilde{\theta}_i), n_i)$$

which can be identified as a generalized regression model with Gaussian response and a log-link function, where the simulated credibility premiums $\hat{\Pi}_i^p$ are the response variable, $\log(\Pi_i)$ is an offset, and both $\mathcal{L}_i(\mathbf{Y}_n, \tilde{\theta}_i)$ and n_i are features with a joint nonlinear effect.

To estimate the parameters of this model, it is standard to minimize the mean square error (MSE) using any statistical learning techniques such as Additive Models, Neural Networks, Trees-based methods, etc. The MSE is given by $MSE = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\Pi}_i^p - \Pi_i \exp\Big(g(\mathcal{L}_i(\tilde{\theta}_i), n_i)\Big)\right)^2$. Gradient descent methods are commonly used to solve this optimization problem, which is not complex as only two features are involved.

The quality of the interpolation can be evaluated by the coefficient of determination $R^2 = 1 - \frac{MSE}{MST}$, where $MST = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\Pi}_i^p - \bar{\hat{\Pi}}^p\right)^2$. This measure seems to be a standard approach to the assessment of surrogate models. The coefficient of determination is always in the interval $[0, 1]$, with 0 indicating poor interpolation and 1 indicating a perfect one. Therefore the fit can be easily interpreted in terms of percentages.

The values of $\tilde{\theta}_i$ are also determined by the algorithm. As discussed in Section 4.1, tuning these values allows the credibility index to better distinguish the claim history of a policyholder, leading to better approximation capability of the surrogate model. Therefore, we can set them as those values that provide the best approximation to the surrogate model. The values of $\tilde{\theta}_i$ are policyholder dependent and so we may view them as functions of the set of parameters of the model. Thus, a certain structure of the form $\tilde{\theta}_i = h(\mathcal{O}_i)$, for some unknown function $h(\cdot)$, can be assumed. The simpler the structure of such a function, the easier the identifiability of parameters. As these values are not intended to be interpreted on their own, any statistical learning technique can be used to find a data-driven function $h(\cdot)$. The estimation of $\tilde{\theta}_i$ must be achieved jointly with the function $g(\cdot)$ and incorporated as part of the optimization process when fitting the model for $g(\cdot)$. This can be achieved directly in a tailored algorithm specified by the user or in an iterative fashion. It may be easier to consider the latter approach to use some of the already implemented methods in software packages. The algorithm iteration stops when the interpolation error, as measured by the MSE, can no longer be decreased. Along these lines, the general scheme for an iterative estimation is shown in algorithm 1.

5.4 Assessment of the surrogate function

Before using the fitted formula for any analysis, it is important to evaluate the accuracy of the out-of-sample predictive power. In this section, we provide an overview of this task, but refer the interested reader to Sobester et al. (2008) for further details.

To assess the interpolation accuracy of the fitted formula, one can perform standard goodness of fit checks on the sub-portfolio of policyholders, akin to those performed in any linear regression

Algorithm 1 Fitting of $g(\cdot)$ and θ_i

$MSE \leftarrow \text{Tol} + 1$
$\tilde{\theta}_i \leftarrow$ Random number $\forall i = 1, \ldots, M$ ▷ Start with random values for $\tilde{\theta}_i$
while $MSE \geq$ Tol **do**
  $\mathcal{L}_i(\tilde{\theta}_i) \leftarrow \sum_{j=1}^{n} \log f(Y_{i,j}|\Theta = \tilde{\theta}_i)\ \forall i = 1, \ldots, M$ ▷ Compute the credibility index
  $g(\cdot) \leftarrow \arg\min_g \sum_{i=1}^{M} \left(\hat{\Pi}_i^p - \Pi_i \exp\Big(g(\mathcal{L}_i(\tilde{\theta}_i), n_i)\Big)\right)^2$ ▷ Fit the function $g(\cdot)$ via LS
  $\tilde{\theta}_i^o \leftarrow \arg\min_{\tilde{\theta}_i} \left(\hat{\Pi}_i^p - \Pi_i \exp\Big(g(\mathcal{L}_i(\tilde{\theta}_i), n_i)\Big)\right)^2\ \forall i = 1, \ldots, M$ ▷ Find pseudo observations $\tilde{\theta}_i$
  $h(\cdot) \leftarrow \arg\min_h \sum_{i=1}^{M} \left(\tilde{\theta}_i^o - h(\mathcal{O}_i)\right)^2$ ▷ Fit the model $h(\cdot)$ via LS
  $\tilde{\theta}_i \leftarrow h(\mathcal{O}_i)\ \ \forall i = 1, \ldots, M$ ▷ Upgrade the values $\tilde{\theta}_i$ using the fitted values of $h(\cdot)$
  $MSE \leftarrow \sum_{i=1}^{M} \left(\hat{\Pi}_i^p - \Pi_i \exp\Big(g(\mathcal{L}_i(\tilde{\theta}), n_i)\Big)\right)^2$ ▷ Upgrade current interpolation error
end while

model. This involves residual analysis (e.g. scatter plots, histograms with errors concentrated around zero, lack of bias, etc) and evaluation of error metrics (e.g. high coefficient of determination, small relative errors, small MSE etc.). If the formula does not perform satisfactorily under this assessment, the functional form of $\hat{\tilde{G}}_\Pi(\cdot)$ must be revised and changed to a more flexible structure to improve the accuracy.

Similarly, to evaluate the out-of-sample predictive power, one can use ideas from cross-validation methods in predictive modeling. For instance, the representative sub-portfolio of policyholders can be split into train and test samples. The surrogate model is fitted only on the former set and used to predict on the latter set. If the goodness of fit metrics on the train set are similar to those on the test set, then the out-of-sample predictive power of the formula is verified. However, if these two differ significantly, in the sense that the errors in the test set are considerably larger than in the train set, then the surrogate model is overfitting and therefore not reliable for extrapolation. In this case, the sample size of representative policyholders must be increased in order to fit such a flexible model. This can be achieved iteratively by gradually increasing the sample size, say 1% at a time, until both the in-sample and out-of-sample performance of the surrogate model is verified.

5.5 Extrapolation

After the surrogate model is evaluated, computing the credibility premiums for the remaining policyholders in the portfolio is a simple matter of evaluating the surrogate function for each policyholder. This approach significantly simplifies the computation workload for large portfolios, reducing it to only a minimal portion of the portfolio.

Certain surrogate models provide a way to construct confidence intervals for the prediction. For instance, if a Gaussian Process is used, confidence bands can be constructed based on the posterior predictive distribution. However, this approach is model-dependent. There is some research in the machine learning literature for constructing confidence bands that can be adapted to surrogate modeling. For example, works such as Kumar and Srivastava (2012) and Kabir et al. (2018) aim to construct confidence bands in a model-agnostic fashion. Nevertheless, we note that this task is still under research and refer the reader to those papers if there is a need to construct them.

6 A Simulation Study

In this section, we illustrate via a simulation study how our newly defined credibility index can indeed capture the information on the claim history of a policyholder. We do so by investigating the achievable accuracy of the surrogate model at approximating the premiums constructed on different choices of the model-prior distributions and premium principles commonly observed in actuarial modeling. We separate the analysis depending on whether the variable of interest is continuous or discrete. Results are summarized in Tables 2 and 3, in which we display the coefficient of determination R^2 obtained from the fitted surrogate model.

The simulation study's general setup involves generating synthetic portfolios of policyholders and their claim history from the Bayesian models listed in Tables 2 and 3. To emulate real insurance portfolios, we create a heterogeneous portfolio of 50,000 policyholders. The mean of the model distribution for each policyholder μ is linked to a synthetic systematic effect α (emulating effect of covariates) and a single latent variable Θ affecting the mean of the distribution as follows:

$$\log(\mu) = \alpha + \Theta$$

We assume non-randomness of the dispersion parameter in the model distribution in the cases where there is such a parameter. The values of α and the dispersion parameters are chosen to resemble those that are usually obtained when fitting a model. Therefore, these mimic the typical heterogeneity present in real datasets and the usual frequency and severity patterns observed in insurance.

To generate the claim history of each policyholder, we utilize the resulting value μ obtained from the previous step along with the chosen model distribution. In order to maintain simplicity, we consider a scenario where each policyholder has a claim history spanning $n = 5$ periods.

We compute the credibility premiums via the importance sampling algorithm discussed in Section 5.2, utilizing 20,000 samples for each policyholder so that the error of estimation is negligible. As our aim in this section is to evaluate the predictive power of the surrogate model, we fit $\hat{\tilde{G}}(\cdot)$ using an unstructured function that is estimated through multidimensional B-splines, which can be achieved via a generalized additive model (GAM) implementation. Additionally, we use a *random forest* structure for the function $h(\cdot)$ that connects the values of $\tilde{\theta}$ to the parameters of the model distribution and the claim history of policyholders. Specifically, we employ the systematic component α values of each policyholder as the features for the random forest so that $\tilde{\theta} = h(\alpha)$

We evaluate the precision of the surrogate function by comparing the premiums estimated by the model with the credibility premiums, using the coefficient of determination R^2. The results are presented in Tables 2 and 3.

	Premium Principle		
Model	Expected Value	Standard deviation	Exponential
Poisson-Gamma	99.63%	99.63%	99.63%
Poisson-LogNormal	99.89%	99.89%	99.89%
NegBinom-InvGaussian	99.59%	99.64%	99.99%
Logarithmic-Weibull	99.84%	99.85%	99.84%
GammaCount-Weibull	97.07%	96.98%	97.08%
Gen-Poisson-Lognormal	99.95%	99.96%	99.95%

Table 2: Coefficient of determination R^2 of the surrogate model on the entire portfolio for simulation study on discrete distributions

	Premium Principle	
Model	Expected Value	Standard deviation
Gamma-Gamma	96.11 %	96.12%
Lognormal-InvGaussian	94.52%	94.52%
LogLogistic-Lognormal	99.83%	99.83%
InvGaussian-Weibull	99.88%	99.90%
Pareto-Gamma	99.92%	99.92%
Burr-Lognormal	99.67%	99.67%

Table 3: Coefficient of determination R^2 of the surrogate model on the entire portfolio for simulation study on continuous distributions

Our results demonstrate that the R^2 values are close to 99% in almost all scenarios, indicating that the fitted premiums closely resemble the credibility premiums, regardless of the selected model distributions and premium principles. We observed in further simulations that the performance of the fitted formula fluctuates more when the portfolio exhibits different levels of heterogeneity with respect to the policyholder attributes and claim history. However, the surrogate model was still able to produce similar results as the ones above even in such scenarios. Some of the distributions in the study are not part of the exponential dispersion family, yet however, the surrogate model showed a desirable performance at reproducing such premiums. Therefore we conclude that the credibility index can effectively summarize almost all of the relevant information regarding a policyholder's claim history, enabling the creation of an appropriate surrogate model.

7 A Numerical Illustration with Real Data

In this section, we demonstrate the use of the surrogate model on a real data set obtained from a European automobile insurance company. The data set consists of policyholders' contract information spanning from January 2007 to December 2017, with claim frequencies from two business lines: Third Party Liability insurance (TPL) and Physical Damage (PD). The data set contains detailed information on the policyholder or their automobiles, such as car weight, engine displacement, engine power, fuel type (gasoline or diesel), car age, and age of the policyholder.

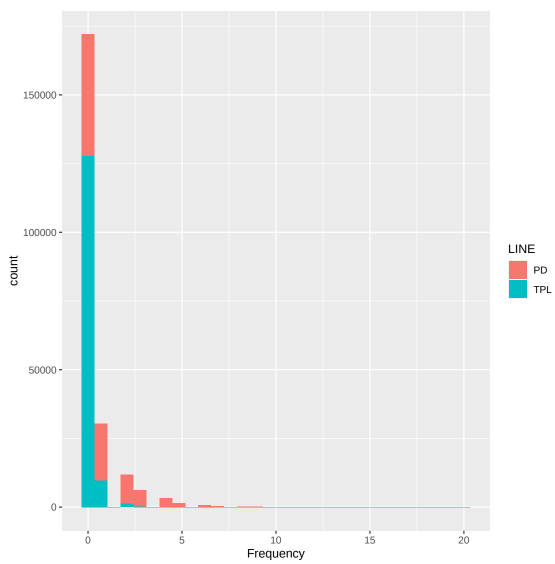
The contract of interest for this application is a policy with two coverages: TPL and PD. The number of claims in both lines may be dependent, as the same car accident can lead to claims in both lines of business. It should be noted that not all policyholders have a fully observed history in the two lines. Policyholders may initially start with a contract on one policy and then upgrade to obtain a policy that covers the two of them, or vice versa (see Figure 1). Therefore, the claim history for some policyholders can be considered partially observed in the sense that the number of claims for a particular line may not be available for certain periods.

Table 4 and Figure 1 present key summary statistics on the dataset analyzed in this study. The data related to the number of claims in both lines of business exhibit behavior consistent with established patterns in insurance, namely, over-dispersion, a high frequency of zero claims, and a significant correlation between the two lines of business. The large portfolio is composed of 184,848 policyholders, among which only 41,956 have fully observed exposure in both lines of business simultaneously.

As depicted in Figure 1, a significant proportion of policyholders renew their contract with the company, enabling the observation of long claim histories, with some policyholders having up to seven years of exposure. Therefore, it is essential to account for such observed experience in the ratemaking process.

TYPE	Number of Policyholders	Claim Frequency Mean	Median	Variance	Correlation
TPL	138,923	0.038	0.000	0.044	0.245
PD	87,517	0.306	0.000	0.512	
TOTAL	184,484	0.162	0.000	0.279	

Table 4: Summary statistics of the dataset



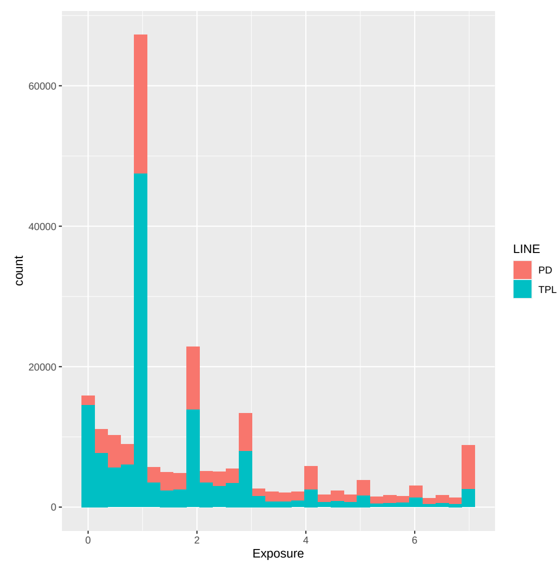


Figure 1: Distribution of claim frequency (left) and exposure (right) per claim type

Given the aforementioned observations, the insurance company intends to conduct an *a posteriori* ratemaking analysis of the claims experience of policyholders at the time of contract renewal. Specifically, we consider the application of the *Exponential* premium principle as an example, with a 5% surcharge, which is the method employed by the insurance company to establish actuarial premiums.

7.1 The Bayesian model and quantities of interest

We illustrate the use of the surrogate model with a bivariate mixed negative-binomial regression model proposed by Tzougas and di Cerchiara (2021), which is very flexible and can capture over-dispersion and dependent frequencies.

Let $Y_j^{(d)}$ be the number of claims from a given policyholder in year j, associated with the dth line of business, with $d = 1$ being PD and $d = 2$ being TPL. Let x be the associated information of covariates, and let $\beta^{(d)}$ be the associated vector of regression coefficients for the d-th coverage. Similarly, let's denote with $\omega^{(d)}$ the time exposure of the contract for each coverage. We consider the hierarchical model

$$Y_j = \begin{pmatrix} Y_j^{(1)} \\ Y_j^{(2)} \end{pmatrix} \sim_{iid} f(y|\Theta, \langle \boldsymbol{x}, \boldsymbol{\beta} \rangle) = \text{NegBinom}(y^{(1)}; \mu^{(1)}\Theta, r^{(1)}) * \text{NegBinom}(y^{(2)}; \mu^{(2)}\Theta, r^{(2)})$$

where, for $d = 1, 2$

$$\log \mu^{(d)} = \log \omega^{(d)} + \beta_0^{(d)} + \beta_1^{(d)}\text{CarWeight} + \beta_2^{(d)}\text{EngineDisplace} + \beta_3^{(d)}\text{CarAge} + \beta_4^{(d)}\text{Age} + \beta_5^{(d)}\text{EnginePower} + \beta_6^{(d)}\text{Fuel}$$

and $\Theta \sim P(\theta) = \text{InvGauss}(1, \sigma^2)$. We use the notation $\text{NegBinom}(y; \mu, r)$ to denote the probability mass function of a negative binomial with mean μ and dispersion r. Similarly, $\text{InvGauss}(1, \sigma^2)$ denotes an Inverse-Gaussian distribution with mean 1 and variance σ^2.

It is worth noting that the proposed model employs a shared latent variable for each line of business. This simplifies the task of working with partially observed data since it enables the estimation of Θ for both business lines even when a policyholder has observations in only one of them.

We emphasize that the model lacks a closed-form expression for both the posterior and the predictive distribution. As such, numerical methods are necessary to obtain any desired quantity of interest, as discussed in detail in Tzougas and di Cerchiara (2021). The estimation of the model parameters is conducted in R using generalized linear mixed models with a Negative-Binomial response. The resulting fitted parameters are presented below:

Variable	β_0	β_1	β_2	β_3	β_4	β_5	β_6	r	σ^2
$Y^{(1)}$	-2.78	$-4.48 * 10^{-5}$	$2.69 * 10^{-5}$	$-2.94 * 10^{-2}$	$-4.20 * 10^{-3}$	$2.24 * 10^{-3}$	$-8.23 * 10^{-2}$	0.86	0.58
$Y^{(2)}$	-1.397	$14.19 * 10^{-5}$	$6.10 * 10^{-5}$	$-0.20 * 10^{-2}$	$-7.94 * 10^{-3}$	$6.06 * 10^{-3}$	$-22.08 * 10^{-2}$	0.86	0.58

Table 5: Estimated Parameters of the Bivariate mixed NegBinomial regression model

7.2 Estimation of the surrogate model

Here we proceed to calculate the credibility premium using the exponential principle

$$\Pi(Y_{n+1}|\mathbf{Y_n}) = \frac{1}{0.05}\log\Big(E\left(\exp\Big(0.05 * (Y_{n+1}^{(1)} + Y_{n+1}^{(2)})\Big)|\mathbf{Y_n}\Big)\Big).$$

To handle the large size of the portfolio, we employ a surrogate model. Specifically, we utilize the *cube method* (implemented via the `samplecube()` function in `R`) to extract a sub-portfolio comprising approximately 5% of the overall portfolio, which amounts to roughly 9,224 policyholders. To ensure a balanced representation, we match the sub-portfolio with the overall portfolio in terms of the average number of claims for PD and TPL, which capture the claim history, as well as the average fitted values of $\mu^{(1)}$ and $\mu^{(2)}$, which reflect the policyholder's attributes.

We use the importance sampling method described in Section 5.2 to estimate premiums, with 50,000 samples to ensure the accuracy of these values. Note that this step is only necessary for the sub-portfolio of 5% policyholders, however, we perform the computationally expensive simulation for the entire portfolio to enable a comparison of the premiums obtained from the surrogate model. We would like to highlight in terms of the simulation time that a single replication on the entire portfolio takes approximately 18 times longer than a replication on the representative policyholder, as illustrated in Table 6 below. This is close to the empirical ratio of 20 associated with the proportion of 5% vs 100% of the representative portfolio. The simulation took place in a large server with 32 cores, 125GB RAM memory and CPU 2 x Intel E5-2683 v4 Broadwell @ 2.1GHz.

Portfolio	**CPU Time (per replica.)**	**Total CPU Time**	**Ratio (Total/Rep)**
Representative	1.02 sec	14.16 hours	18.02
Total	18.38 sec	255.28 hours	

Table 6: Comparison of the CPU Time required for the simulations

As previously discussed, the credibility index can be subdivided into sub-indexes that represent summary statistics, with one for each business line in the model. In this case, we can identify two credibility sub-indexes, denoted by $d = 1, 2$, and defined as follows:

$$\mathcal{L}^{(d)}(\mathbf{Y}_n; \tilde{\theta}) = \sum_{j=1}^{n} \log \left(\text{NegBinom} \left(Y_j^{(d)}; \mu^{(d)} \exp\left(\tilde{\theta}\right), r^{(d)} \right) \right).$$

Regarding the surrogate function, we place an emphasis on the rating factor functional as described in Equation (7) in order to achieve a transparent ratemaking process. Nonetheless, we also explore the use of an unstructured surrogate function via multidimensional B-Splines for comparison purposes. Since the model under consideration is multivariate, with $Y_j = (Y_j^{(1)}, \ldots, Y_j^{(D)})$, we can further decompose the function $g(\cdot)$ in Equation (7) to measure the effect of the credibility sub-indexes separately. Thus, it is possible to quantify the importance of the claim history of each component on the resulting credibility premium.

For instance, if we opt for an additive structure for the function $g(\mathcal{L}(\mathbf{Y}_n; \tilde{\theta}), n)$, our surrogate function can be expressed as:

$$\hat{\tilde{G}}_{\Pi}(\mathcal{L}(\mathbf{Y}_n; \tilde{\theta}), n, \mathcal{O}) = \Pi(Y_{n+1}) \exp\Big(g_1(\mathcal{L}^{(1)}(\mathbf{Y}_n; \tilde{\theta})) + g_2(\mathcal{L}^{(2)}(\mathbf{Y}_n; \tilde{\theta})) + g_3(n^{(1)}) + g_4(n^{(2)}) \Big).$$

The function $g_1(\cdot)$ provides a measure of the effects of the past experience in the number of PD claims in the credibility premium, $g_2(\cdot)$ provides a measure of the effects of the TPL number of claims in the credibility premium, and $g_3(\cdot)$ and $g_4(\cdot)$ provide the effect of the number of periods for which the policyholder has been observed for each line of business. It should be noted that some policyholders may have only partially observed information, meaning that their claim history may differ for each line of business. Therefore, separate effects for $n^{(1)}$ and $n^{(2)}$ are considered.

To estimate the functions $g_k(\cdot)$ and the tuning parameters $\tilde{\theta}$ for each individual, only 5% of the policyholders are used, as described previously. We use a B-Slines representation for the functions $g_j(\cdot)$, and the fit is accomplished via a generalized additive model (GAM) with Gaussian response and with the additive structure above. The values of $\tilde{\theta}$ are estimated using a random forest with $\mu^{(1)}$ and $\mu^{(2)}$, which depend on the policyholder attributes, as features. Such estimation of the functions $g_k(\cdot)$ and the values $\tilde{\theta}$ is performed iteratively, as described in Section 5. It is worth mentioning that both the GAM and random forest models exhibit flexibility and require minimal computational resources for training, making them suitable for surrogate modelling. In fact, the entire process of fitting the surrogate model and extrapolation was completed in less than 10 minutes.

7.3 Results and interpretation

We first illustrate the view of the credibility index as a summary statistic and its interpretation. In order to demonstrate the association between the credibility index and the risk behavior observed in the claim history of policyholders, we plotted the credibility index against the standardized claim frequency $\frac{\sum_{j=1}^{n} Y_j^{(d)}}{\sum_{j=1}^{n} \mu_j^{(d)}}$ for the entire portfolio. This measure intuitively provides a way to quantify the observed claim frequency, controlled by the heterogeneity of policyholder attributes. The resulting graphs for each of the two sub-indexes are displayed in Figure 2. We observe a strong correlation between the credibility index and the standardized frequency, indicating that the index is highly associated with the riskiness of policyholders as observed in their claim history. Specifically, we note an inverse relationship where the larger the standardized claim frequency, the lower the value of the credibility sub-index. However, we acknowledge that the association is not perfect since these two statistics do not capture the same amount of information from the claim history. The credibility index provides a wider range of values than the standardized claim frequency and is thus better able to distinguish policyholders' riskiness based on their covariates, as observed in the top left of Figure 2.

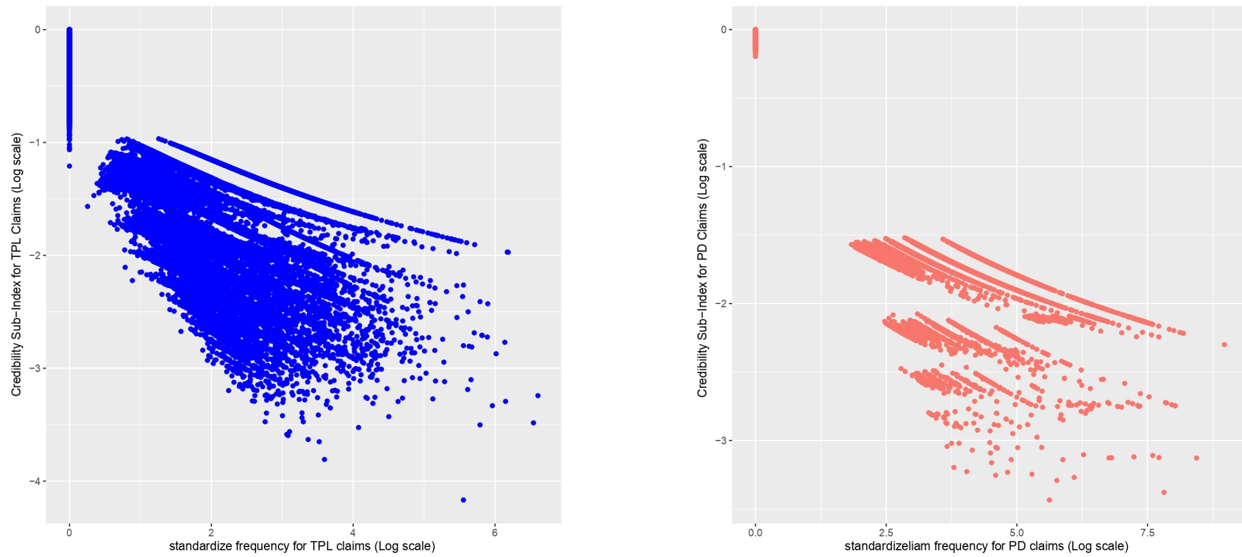


Figure 2: Credibility sub-indexes vs standardized claim frequency for the entire portfolio. Left PD claims / Right TPL claims.

Figure 3 displays the distribution of the credibility sub-index among policyholders, which exhibits a negative exponential-like behavior. Notably, the peak of the distribution is observed around 0, which indicates that a large proportion of policyholders exhibit a "low risk" behavior, which is evident as most of the policyholders have no claims. Conversely, relatively few policyholders exhibit "risky behavior," as observed in the left tail of the distribution. Thus, insurance companies can utilize the credibility index as a measure to evaluate the observed riskiness of the portfolio compared to what was originally anticipated. A concentration of the index distribution around 0 suggests that the current premiums are appropriately pricing the risk of the portfolio, thereby reflecting a sound assessment of risk. Conversely, a distribution with elongated left-tails indicates underestimation of portfolio risk, requiring corrective action.

We now proceed to show the fitted surrogate function. The estimated functions $g_k(\cdot)$ are presented in Figure 4. The results indicate that the greater the credibility indexes, the less the estimated credibility premium for a specific policyholder. This interpretation is consistent with the one we obtained in Figure 2.

We have that the function $\exp(g_1(\cdot))$ exhibits a null-effect at a value of $\mathcal{L}^{(1)}(\mathbf{Y}_n;\tilde{\theta})$ around -13, whereas for the function $\exp(g_2(\cdot))$, the null-effect is observed at around -15 for a value of $\mathcal{L}^{(2)}(\mathbf{Y}_n;\tilde{\theta})$. Thus, policyholders with credibility sub-index values less than -13 (for $\mathcal{L}^{(1)}(\mathbf{Y}_n;\tilde{\theta})$) and less than -15 (for $\mathcal{L}^{(2)}(\mathbf{Y}_n;\tilde{\theta})$ tend to have a higher than expected number of PD claims and TPL claims, respectively. Consequently, they require a revised premium that is larger than the manual premium. Similarly, policyholders with credibility sub-index values greater than -13 (for $\mathcal{L}^{(1)}(\mathbf{Y}_n;\tilde{\theta})$) and greater than -15 (for $\mathcal{L}^{(2)}(\mathbf{Y}_n;\tilde{\theta})$) tend to have a lower than expected number of PD claims and TPL claims, respectively, and may receive a revised premium that is lower than the manual premium.

The estimated effect range is larger for the function $\exp(g_2(\cdot))$ (i.e., from 0.3 to 1.6) than for the function $\exp(g_1(\cdot))$ (i.e., from 0.5 to 1.4), indicating that the experience on the TPL number of

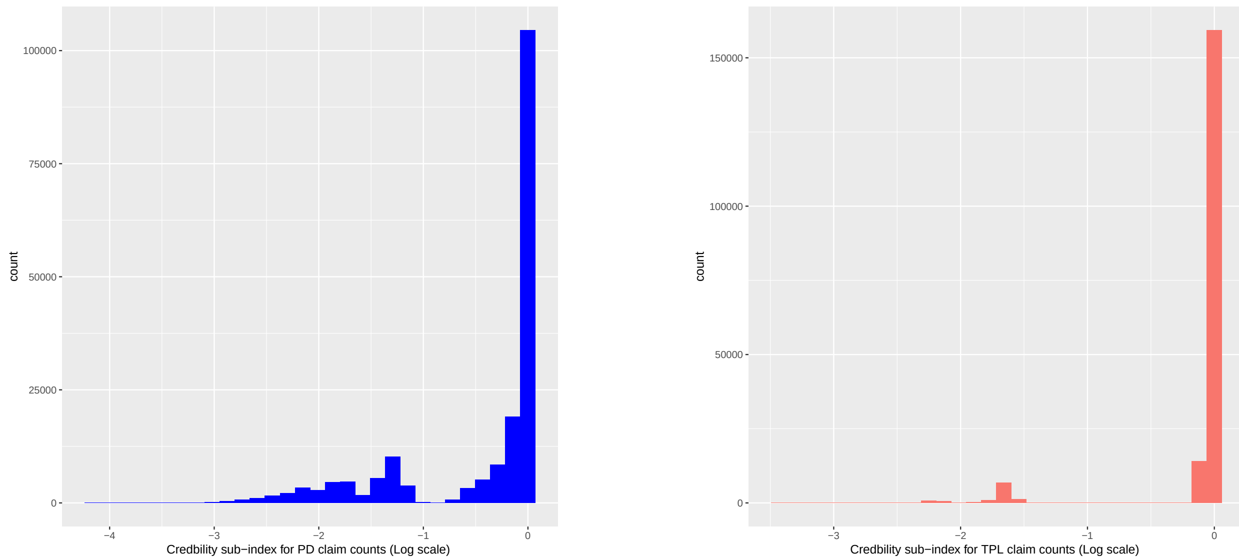


Figure 3: Distribution of the credibility sub-indexes. Left is for PD number of claims and Right for TPL number of claims

claims is more critical in determining the credibility premiums than the experience in the number of PD claims. The last is in the sense that the former can produce a larger deviation in the resulting premium than the latter. This result is intuitive as TPL claims have a significantly lower frequency than PD claims (see Table 4), and thus, having a TPL claim would have a more substantial impact on the credibility premium than a PD claim.

The two plots at the bottom of Figure 4 depict the effect of the number of periods of exposure on the resulting credibility premium. While the effect of these quantities is not directly interpretable, they provide insights into the impact of the number of periods of exposure. The results indicate that the effect from these quantities tends to stabilize as the number of periods increases. This is intuitive as the credibility premium stabilizes when enough history from a policyholder is observed. Similarly, note that this stabilization is a bit longer for the TPL line of business. This may be the case as these types of claims have less frequency, and so the stabilization might take a longer time (i.e more experience must be observed).

Finally, note that based on the ranges defined above for the overall function $\hat{\tilde{G}}(\cdot)$, the credibility index can be utilized to identify three groups of policyholders: less risky than expected, as risky as expected, and riskier than expected. The policyholders in the third group could be considered “bad risks” as their risk is not appropriately captured in the current premiums, requiring extra attention during the remaking process. However, it is worth noting that this grouping does not aim to perform a risk classification as typically done in ratemaking applications, but rather identifies policyholders riskier than expected.

7.4 Goodness of fit

We now proceed to evaluate the suitability of the surrogate function as presented in Section 5.4. The evaluation involves two main tasks: first, we verify the accuracy of the surrogate function in reproducing the credibility premiums, and second, we assess its reliability for extrapolation when only 5% of the portfolio is utilized.

To assess the accuracy of the surrogate function in reproducing the credibility premiums, we

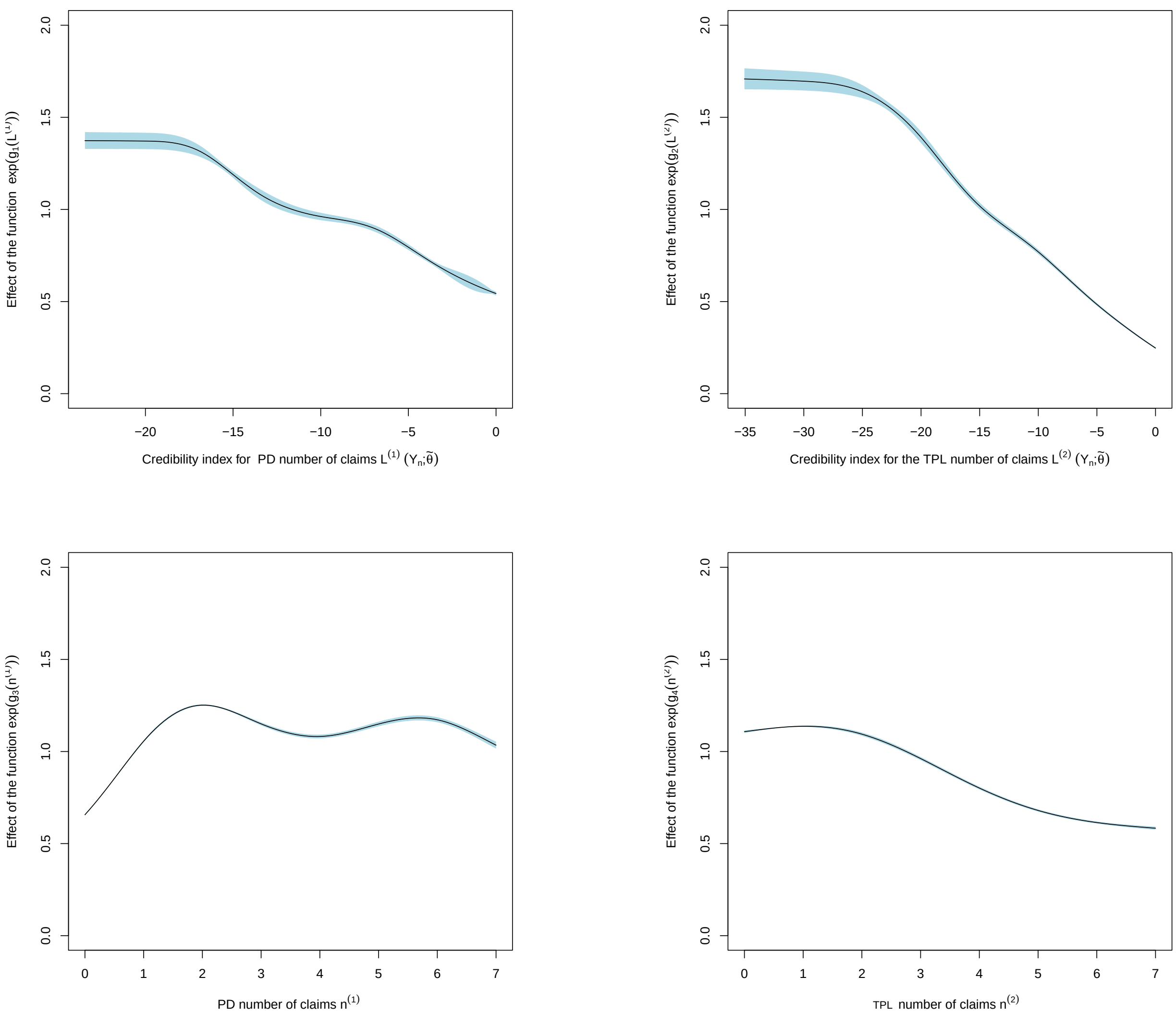


Figure 4: Estimation of the functions $g(\cdot)$. Top-left is $g_1(\cdot)$, Top-right is $g_2(\cdot)$, Bottom-left is $g_3(\cdot)$ and Bottom-right is $g_4(\cdot)$

compare the fitted values of the surrogate model with the credibility premiums. Figure 5 on the left-hand side presents a dispersion plot of the two premiums. The plot indicates that our credibility formula produces premiums that are a close approximation to the credibility premiums, with the points closely clustered around the 45-degree line. To further evaluate the accuracy of the surrogate function, the table on the left-hand side of Figure 6 presents error metrics of the fitted premium versus the credibility premiums. The interpolation results in an overall coefficient of determination of $R^2 = 0.97$, indicating that the credibility formula reproduces around 97% of the variance in the premiums, which is an overall descent level of interpolation. The fitted premiums deviate from the credibility premiums on average by only 3%, and the mean error of interpolation is almost zero, implying no bias in the resulting estimation.

Similarly, from an insurance company's perspective, it is also important to ensure that the distribution of the premiums resulting from the surrogate model is similar to that obtained from the credibility premiums. To evaluate this, we compare the two distributions using a QQ plot, which is presented on the right-hand side of Figure 5. The plot shows that the distribution of the fitted premiums is essentially the same as that of the credibility premiums. Therefore, we conclude that the credibility formula provides an accurate approximation of the credibility premiums, and the insurance company can rely on the fitted premiums without significant differences in further portfolio-level metrics, such as total premium earned across the portfolio, loss ratios, and others.

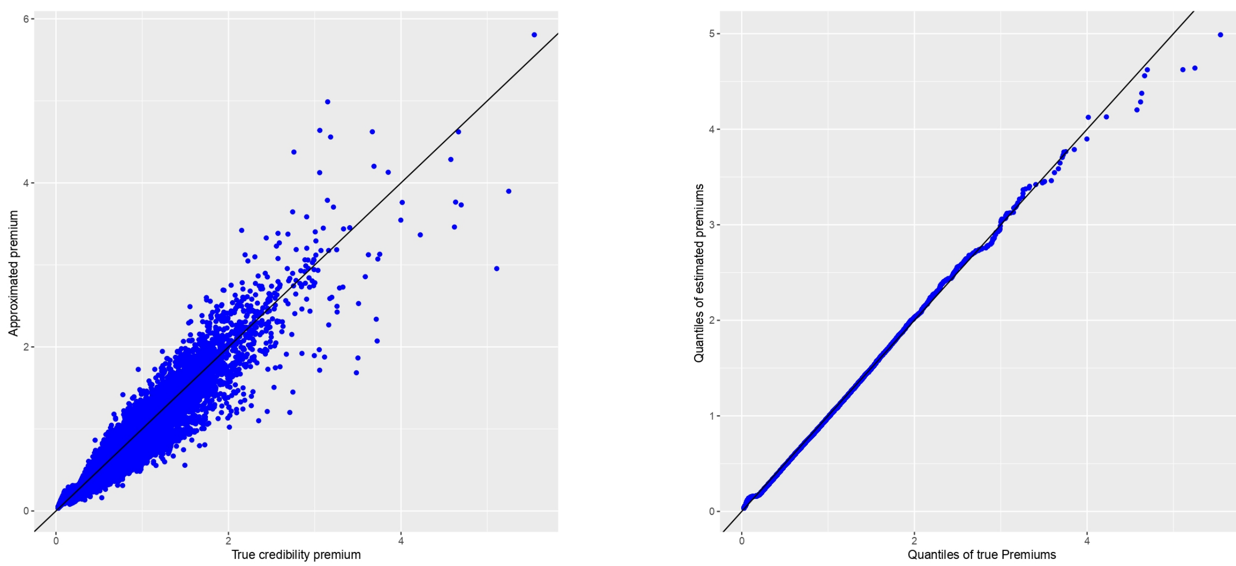


Figure 5: Comparison of approximated premiums vs true premiums. Left-hand side dispersion plot and Right-hand side QQ-plot.

Now, to asses the reliability for extrapolation, we need to evaluate the out-of-sample performance of the predictive model on a test data set and ensure that it exhibits behavior consistent with the in-sample data. Specifically, in this application, as we computed the premiums for the entire portfolio, we use the remaining 95% as our test data. The table on the left-hand side of Figure 6 displays the error metrics for both the in-sample and out-of-sample policyholders. In this case, we observe that the error metrics for the out-of-sample policyholders are comparable to those obtained for the in-sample policyholders, with no appreciable differences. Hence, extrapolating the predictive power of the model from only 5% of the portfolio is sufficient to provide reliable predictions for the entire portfolio using the surrogate model.

To further verify this claim, we estimated the surrogate function using the entire portfolio

(100%) rather than just the 5% sample and compared the two estimates. The dispersion plot on the right-hand side of Figure 6 compares these two sets of fitted values. It can be observed that the premiums computed using the selected policyholders are almost the same as those using the entire portfolio. Therefore, the fitting of the surrogate function using only 5% of the portfolio is indeed reliable.

Sub-portfolio	R^2	ME	MAE	MAPE
Out of Sample	0.973	0.0089	0.0293	0.129
In Sample	0.972	0.0087	0.0294	0.129

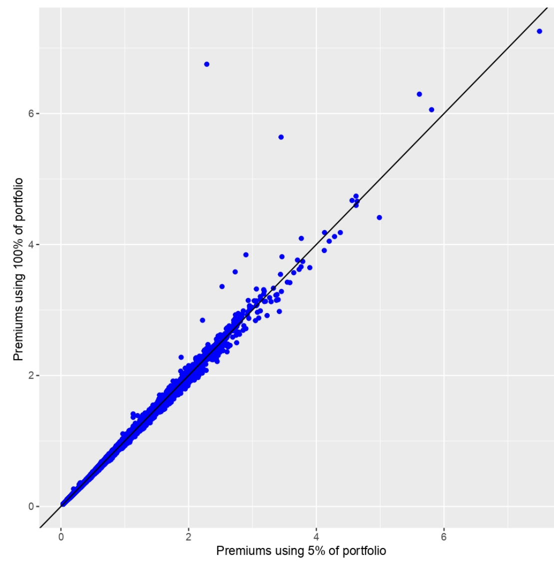


Figure 6: Assessment of the accuracy of the surrogate model. Left-hand side Error Metrics and Right-hand side Premiums using 5% vs 100% of the portfolio

Lastly, we evaluate the results of the surrogate model using an unstructured functional form in Figure 7 and Table 7. The surrogate model is fitted using multidimensional B-Splines representation for the function $\hat{\hat{G}}_{\Pi}$, and using as inputs the manual premium, the two credibility sub-indexes, and the number of known periods in each policy.

Sub-portfolio	R^2	ME	MAE	MAPE
Out of Sample	0.990	-0.0005	0.0036	0.022
In Sample	0.990	-0.0006	0.0036	0.022

Table 7: Error Metrics of the surrogate model with unstructured surrogate function

Our results, presented in Table 7, demonstrate that the unstructured functional form provides a relatively better fit than the rating factor functional form, with lower error metrics and a higher coefficient of determination of 99%. This implies an almost perfect interpolation, indicating that the fitted premiums are much closer to the true premiums than the previous surrogate. Similarly, the graph on the left-hand side of Figure 7 shows that the fitted premiums are much closer to the true premiums than the rating functional form, in the sense of much lower fluctuation around the 45-degree line. Similarly, on the right-hand side of Figure 7, the QQplot shows that the distribution of the fitted premiums still resembles the same as the true premiums, so the insurance company can rely on them.

We would like to note that the surrogate model using the rating factor functional form provides comparable results to the unstructured form, with the added benefit of providing a transparent ratemaking process. Therefore, the choice between these two models will depend on the specific needs of the insurance company. Finally, we want to emphasize that the overall methodology via surrogate modeling provides desirable results in an inexpensive fashion.

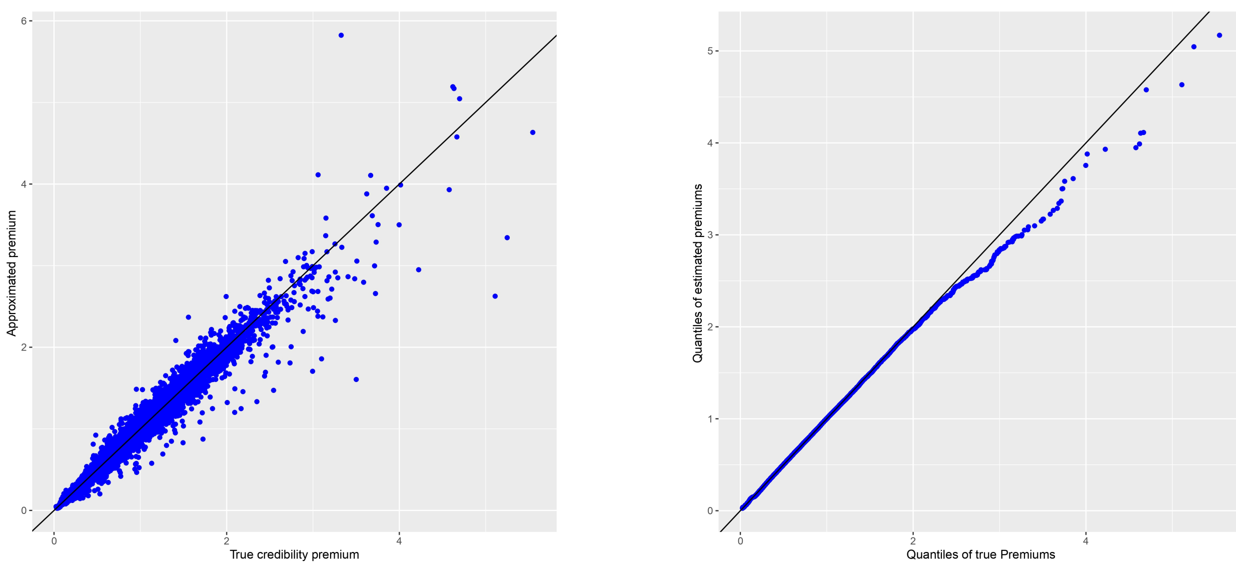


Figure 7: Comparison of approximated premiums vs true premiums for the unstructured model. Left-hand side dispersion plot and Right-hand side QQ-plot.

7.5 An example with two policyholders

This subsection illustrates how the surrogate function is utilized for the experience rating process. We examine two policyholders who exhibit similar risk behavior, as indicated by their attributes, but possess different claim histories. These two policyholders have nearly identical predicted mean numbers of claims, as estimated by the model, and have held policies with the insurance company for a duration of $n = 4$ years. The individual attributes of these policyholders are presented below.

Policyholder	**CarWeight**	**EngDisp**	**CarAge**	**Age**	**EngPow**	**Fuel type**	**Expected Claims**
A	1515	1248	1	82	63	Gasoline	0.245 (per year)
B	1475	1360	6	72	55	Gasoline	0.246 (per year)

Table 8: Covariates of the Policyholders

Policyholder A has not filed any claims in the past 4 years, neither in PD nor in TPL, while Policyholder B has filed two claims - one in PD and one in TPL. Consequently, it can be inferred that Policyholder A is less risky than Policyholder B. Intuitively, the premium for Policyholder A should decrease since no claims have been observed in 4 years, even though 1 claim is anticipated on average in this time period (i.e., 4*0.25 = 1 claim). Conversely, the premium for Policyholder B should increase considerably as the total number of claims is twice the number expected from the model for this period (i.e., 2 claims vs. 4*0.25 = 1 claim).

Keeping this in mind, we further examine how the credibility index and the fitted credibility formula reflect these two risk behaviors. Table 9 presents the credibility index. However, note the difference in the magnitude of the credibility sub-indexes. Specifically, for Policyholder A, who has not filed any claims, the sub-indexes exhibit relatively minimal values closer to 0, indicating a claim history that is less risky than expected, as per our interpretation of Figure 4. On the other hand, for Policyholder B, who has filed multiple claims, the sub-indexes exhibit the opposite behavior, with a large negative value indicating a claim history that is riskier than expected. Consequently, the index captures the distinct nature of the claim histories of these two policyholders.

Policyholder	$\mathcal{L}^{(1)}(\mathbf{Y}_n; \tilde{\theta})$	$\mathcal{L}^{(2)}(\mathbf{Y}_n; \tilde{\theta})$
A	$-2.11 * 10^{-5}$	$-1.27 * 10^{-4}$
B	-12.1	-10.3

Table 9: Credibility Sub-Indexes for Policyholders A and B

Now we proceed to look at the estimated credibility premiums under the exponential principle in Table 10. The premiums are computed as the product of the manual premium and the rating factor given by the credibility formula. We observe that the resulting premiums align with our intuitive analysis. Specifically, policyholder B, who has a claim history with two claims, is expected to pay almost twice as much in premiums as policyholder A, who has no claims in the past four years. Therefore, the claim history is the main driver of the differences in premiums, and the credibility index is the measure that captures such discrepancies.

Policyholder	**Manual Premium**	$\exp\Big(g(\mathcal{L}(\mathbf{Y}_n; \tilde{\theta}), n)\Big)$	**Credibility Premium**
A	0.253	0.854	0.216
B	0.253	1.696	0.429

Table 10: Calculation of credibility premiums

8 Conclusions

Performing accurate experience rating on large insurance portfolios is a challenging task due to two major problems: 1) accounting for the heterogeneity of the policyholders requires flexible, and possibly not mathematically tractable models that can fit complexity in the risk behavior of policyholders and 2) it is necessary to have large computational power in order to deal with extremely large sized insurance portfolios, especially when no closed-form solutions are available. The first issue can be partially addressed by the use of general Bayesian models beyond the simplistic assumptions commonly used in insurance ratemaking. However, these methods heavily rely on computational techniques which may be affected due to the second problem. Therefore, it is of big importance for actuaries to address effectively the computational issues and to have an effective ratemaking system that is transparent and suited to actuarial standards.

In this paper, we propose a methodology that addresses these challenges by computing the credibility premium through a surrogate modeling approach based on a tailored-made summary statistic that we term the credibility index. It measures how likely it is for a policyholder to experience a certain claim history but in a distorted measure, and it is a sufficient statistic for several distribution families. This approach provides an analytical expression for the Bayesian premium, i.e. a credibility formula, which can be used to alleviate the expensive calculation of Bayesian premiums in large portfolios and enable the actuary to interpret the results. The actuary can rely on the credibility formula to approximate premiums for any policyholder for any given parametric model. Additionally, this expression enables the actuary to provide a transparent picture of the ratemaking process to both clients and regulators and provides a reliable way of performing risk classification among the policyholders.

Future research can explore the application of the surrogate model and the credibility index in the context of evolutionary credibility. Indeed, more recent claims may provide a better assessment of the current risk behavior of a policyholder, and so should have a larger impact than very old claims when upgrading the premiums. This is usually achieved in insurance by means of State-Space models (see for e.g Ahn et al. (2021)), which can still be embedded into the Bayesian model in Section 2 under a high dimensional latent variable vector. Therefore the methodology here described still applies, but much more research must be performed in future case studies.

Another future research direction is the construction of other summary statistics that complement the credibility index. The credibility index may not be the only quantity on which the claim history affects the predictive distribution of policyholders. Therefore, it may be possible to improve the accuracy of the credibility formula by considering other types of credibility indexes in the model. Another possible research direction is to consider the development of a credibility index that is "non-parametric." The credibility index here proposed is model-dependent, as it relies on a given parametric model. Thus, it is desirable to develop a credibility index based solely on empirical data sets.

Finally, it is worth mentioning that the surrogate modeling approach here described can find applications in other areas not directly related to experience rating. For instance, this could be applied on predictive models for claim reserving to account for both policyholders' experiences and the current claims on development. Similarly, the applications of the credibility index can be extended to the generality of Bayesian inference. For instance, the idea of a surrogate model for the approximation of expectations can be used in the context of approximate Bayesian computations, particularly in EM algorithms to make them more efficient.

Acknowledgment

This work was partly supported by Natural Sciences and Engineering Research Council of Canada [RGPIN 284246, RGPIN-2017-06684].

References

Ahn, J. Y., Jeong, H., and Lu, Y. (2021). On the ordering of credibility factors. *Insurance: Mathematics and Economics*, 101:626–638.

Asmussen, S., Steffensen, M., Asmussen, S., and Steffensen, M. (2020). *Risk and Insurance: A Graduate Text*. Springer.

Bermúdez, L. and Karlis, D. (2017). A posteriori ratemaking using bivariate poisson models. *Scandinavian Actuarial Journal*, 2017(2):148–158.

Bernardo, J. M. and Smith, A. F. (2009). *Bayesian Theory*, volume 405. John Wiley & Sons.

Boucher, J.-P. and Guillén, M. (2009). A survey on models for panel count data with applications to insurance. *RACSAM-Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, 103(2):277–294.

Bühlmann, H. and Gisler, A. (2005). *A Course in Credibility Theory and its Applications*, volume 317. Springer.

Casella, G. and Berger, R. L. (2021). *Statistical Inference*. Cengage Learning.

Chambers, R. and Clark, R. (2012). *An Introduction to Model-based Survey Sampling with Applications*. Oxford University Press.

Cheung, E. C., Ni, W., Oh, R., and Woo, J.-K. (2021). Bayesian credibility under a bivariate prior on the frequency and the severity of claims. *Insurance: Mathematics and Economics*, 100:274–295.

Crevecoeur, J., Robben, J., and Antonio, K. (2022). A hierarchical reserving model for reported non-life insurance claims. *Insurance: Mathematics and Economics*, 104:158–184.

Czado, C., Kastenmeier, R., Brechmann, E. C., and Min, A. (2012). A mixed copula model for insurance claims and claim sizes. *Scandinavian Actuarial Journal*, 2012(4):278–305.

Denuit, M., Guillen, M., and Trufin, J. (2019). Multivariate credibility modelling for usage-based motor insurance pricing with behavioural data. *Annals of Actuarial Science*, 13(2):378–399.

Denuit, M., Maréchal, X., Pitrebois, S., and Walhin, J.-F. (2007). *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. John Wiley & Sons.

Desjardins, D., Dionne, G., and Lu, Y. (2023). Hierarchical random-effects model for the insurance pricing of vehicles belonging to a fleet. *Journal of Applied Econometrics*, 38(2):242–259.

Diao, L. and Weng, C. (2019). Regression tree credibility model. *North American Actuarial Journal*, 23(2):169–196.

Dickson, D. C. M. (2016). *Insurance Risk and Ruin*. International Series on Actuarial Science. Cambridge University Press, 2 edition.

Englund, M., Guillén, M., Gustafsson, J., Nielsen, L. H., and Nielsen, J. P. (2008). Multivariate latent risk: A credibility approach. *ASTIN Bulletin: The Journal of the IAA*, 38(1):137–146.

Fearnhead, P. and Prangle, D. (2012). Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474.

Frees, E. W., Lee, G., and Yang, L. (2016). Multivariate frequency-severity regression models in insurance. *Risks*, 4(1):4.

Frees, E. W. and Valdez, E. A. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, 103(484):1457–1469.

Gómez-Déniz, E. (2008). A generalization of the credibility theory obtained by using the weighted balanced loss function. *Insurance: Mathematics and Economics*, 42(2):850–854.

Gómez-Déniz, E. and Calderín-Ojeda, E. (2014). A suitable alternative to the pareto distribution. *Hacettepe Journal of Mathematics and Statistics*, 43(5):843–860.

Henckaerts, R., Antonio, K., and Côté, M.-P. (2022). When stakes are high: Balancing accuracy and transparency with model-agnostic interpretable data-driven surrogates. *Expert Systems with Applications*, 202:117230.

Hou, C. K. J. and Behdinan, K. (2022). Dimensionality reduction in surrogate modeling: A review of combined methods. *Data Science and Engineering*, 7(4):402–427.

Jewell, W. S. (1974). Credible means are exact bayesian for exponential families. *ASTIN Bulletin: The Journal of the IAA*, 8(1):77–90.

Jones, H. and Chen, S. (2021). Surrogate models: A comfortable middle ground? `https://www.soa.org/sections/pred-analytics-futurism/pred-analytics-futurism-newsletter/2021/august/paf-2021-08-jones/`. [Online; accessed 17-May-2023].

Joyce, P. and Marjoram, P. (2008). Approximately sufficient statistics and bayesian computation. *Statistical Applications in Genetics and Molecular Biology*, 7(1).

Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008). *Modern Actuarial Risk Theory: Using R*, volume 128. Springer Science & Business Media.

Kabir, H. D., Khosravi, A., Hosen, M. A., and Nahavandi, S. (2018). Neural network-based uncertainty quantification: A survey of methodologies and applications. *IEEE Access*, 6:36218–36234.

Kumar, S. and Srivastava, A. (2012). Bootstrap prediction intervals in non-parametric regression with applications to anomaly detection. In *The 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

Künsch, H. R. (1992). Robust methods for credibility. *ASTIN Bulletin: The Journal of the IAA*, 22(1):33–49.

Landsman, Z. (2002). Credibility theory: A new view from the theory of second order optimal statistics. *Insurance: Mathematics and Economics*, 30(3):351–362.

Li, H., Lu, Y., and Zhu, W. (2021). Dynamic bayesian ratemaking: a markov chain approximation approach. *North American Actuarial Journal*, 25(2):186–205.

Lin, X. S. and Yang, S. (2020a). Efficient dynamic hedging for large variable annuity portfolios with multiple underlying assets. *ASTIN Bulletin: The Journal of the IAA*, 50(3):913–957.

Lin, X. S. and Yang, S. (2020b). Fast and efficient nested simulation for large variable annuity portfolios: A surrogate modeling approach. *Insurance: Mathematics and Economics*, 91:85–103.

Lindholm, M., Richman, R., Tsanakas, A., and Wüthrich, M. V. (2022). Discrimination-free insurance pricing. *ASTIN Bulletin: The Journal of the IAA*, 52(1):55–89.

Mastroianni, G. and Milovanovic, G. (2008). *Interpolation Processes: Basic Theory and Applications*. Springer Science & Business Media.

Mayo Wilson, C. (2020). A qualitative generalization of birnbaum's theorem. `https://faculty.washington.edu/conormw/Papers/Mayo-Wilson-Qualitative_Birnbaum_Preprint.pdf`. [Online in author's webpage; accessed 17-May-2023].

Najafabadi, A. T. P. (2010). A new approach to the credibility formula. *Insurance: Mathematics and Economics*, 46(2):334–338.

Norberg, R. (2004). Credibility theory. *Encyclopedia of Actuarial Science*, 1:398–406.

Ohlsson, E. (2008). Combining generalized linear models and credibility models in practice. *Scandinavian Actuarial Journal*, 2008(4):301–314.

Radtke, M., Schmidt, K. D., and Schnaus, A. (2016). *Handbook on Loss Reserving*. Springer.

Schweder, T. and Hjort, N. L. (2016). *Confidence, Likelihood, Probability*, volume 41. Cambridge University Press.

Sisson, S. A., Fan, Y., and Beaumont, M. (2018). *Handbook of Approximate Bayesian Computation*. CRC Press.

Sobester, A., Forrester, A., and Keane, A. (2008). *Engineering Design via Surrogate Modelling: A Practical Guide*. John Wiley & Sons.

Sundt, B. (1979). On choice of statistics in credibility estimation. *Scandinavian Actuarial Journal*, 1979(2-3):115–123.

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., and Dessimoz, C. (2013). Approximate bayesian computation. *PLoS Computational Biology*, 9(1):e1002803.

Taylor, G. C. (1977). Abstract credibility. *Scandinavian Actuarial Journal*, 1977(3):149–168.

Tillé, Y. (2011). Ten years of balanced sampling with the cube method: an appraisal. *Survey Methodology*, 37(2):215–226.

Tillé, Y. and Matei, A. (2010). Teaching survey sampling with the 'sampling'r package. In *ICOTS-8 Conference Proceedings, Data and Context in Statistics Education: Towards an Evidence-Based Society*. International Association for Statistical Education, Kuala Lumpur, Malaysia.

Tzougas, G. and di Cerchiara, A. P. (2021). The multivariate mixed negative binomial regression model with an application to insurance a posteriori ratemaking. *Insurance: Mathematics and Economics*, 101:602–625.

Virginia, R. (1998). Credibility using a loss function from spline theory. *North American Actuarial Journal*, 2(1):101–111.

Wang, S. S. (2000). A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance*, 67(1):15–36.

Wuthrich, M. V. and Merz, M. (2022). Statistical foundations of actuarial learning and its applications. *Available at SSRN 3822407*.

Xacur, O. A. Q. and Garrido, J. (2018). Bayesian credibility for glms. *Insurance: Mathematics and Economics*, 83:180–189.

Xie, Y. T., Li, Z. X., and Parsa, R. A. (2018). Extension and application of credibility models in predicting claim frequency. *Mathematical Problems in Engineering*, 2018:1–8.

Yan, Y. and Song, K.-S. (2022). A general optimal approach to bühlmann credibility theory. *Insurance: Mathematics and Economics*, 104:262–282.

Yau, K., Yip, K., and Yuen, H. (2003). Modelling repeated insurance claim frequency data using the generalized linear mixed model. *Journal of Applied Statistics*, 30(8):857–865.

Zhang, J., Qiu, C., and Wu, X. (2018). Bayesian ratemaking with common effects modeled by mixture of polya tree processes. *Insurance: Mathematics and Economics*, 82:87–94.

Zhang, Y. (2022). Experience rating of risk premium for esscher premium principle. *Available at SSRN 4170554*.