INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0217

ON THE STOCHASTIC INCREASINGNESS OF FUTURE CLAIMS IN THE BÜHLMANN LINEAR CREDIBILITY PREMIUM

I. OANA PURCARU & MICHEL DENUIT

http://www.stat.ucl.ac.be

Abstract

Since Nelder & Verrall (1997), the connection between Generalized Linear Models (GLM's) and credibility theory has been recognized in actuarial science. Specifically, the credibility construction amounts to add a random effect on the same scale as the fixed effects to model unexplained heterogeneity. The present paper aims to examine the dependence existing between future claims (severity or frequency components) and the Bühlmann linear credibility premium. As expected, future claims are shown to increase with the amount of Bühlmann premium.

Key words and phrases: GLM, random effects, stochastic dominance, conditional increasingness, comonotonicity, linear credibility.

1 Introduction and Motivation

1.1 Credibility theory

Credibility theory can be seen as the art to combine different collections of data to obtain an accurate overall estimate. In many cases, a compromise estimator is derived from a convex combination of a prior mean and the mean of the current observations. The weight given to the observed mean is called the credibility factor (since it fixes the extent to which the actuary may be confident with the data). An excellent introduction to credibility theory may be found in Herzog (1994).

Bayesian statistics offer an intelectually acceptable approach to credibility theory. The book by Klugman (1992) provides an in-depth treatment of the question. Nevertheless, practical applications involve computer-intensive methods of MCMC-type (to perform integration with respect to a posteriori distribution), making more elementary approaches desirable (at least to get a first easy to compute approximation of the result). In that respect, the greatest accuracy credibility approach initiated by Bühlmann (1970) is of prime interest. Basically, the actuary still resorts to a quadratic loss function but the shape of the credibility predictor is constrained ex ante to be linear. This paper explores some nice features of the linear credibility approach, complementing some recent works by Purcaru & Denuit (2001, 2002) and Brouhns & Denuit (2002).

1.2 GLM's

Generalized linear models (GLM's) unify regression methods for a variety of discrete and continuous outcomes; a standard reference for GLM's is Mc Cullagh & Nelder (1989). The basis of GLM's is the assumption that the data are sampled from a one-parameter exponential family of distributions. This happens for instance with data conforming to Binomial law, Poisson law, Normal law with a known variance or Gamma law with a known dispersion parameter. Specifically, consider a single observation y; GLM's admit a log-likelihood of the form

$$\ell(\theta, \phi; y) = \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \tag{1.1}$$

where θ denotes the canonical parameter and $\phi > 0$ is the dispersion parameter (assumed known). It is easily seen that

$$\mu = \mathbb{E}[Y] = b'(\theta) \text{ and } \mathbb{V}\mathrm{ar}[Y] = b''(\theta)\phi = v(\mu)\phi$$

where $v(\cdot)$ is called the variance function. A link function relating the mean μ to the linear predictor is then specified. More precisely, given a vector $\boldsymbol{x}$ of explanatory variables and a vector $\boldsymbol{\beta}$ of regression coefficients, the linear predictor takes the form $\eta = \boldsymbol{x}^t\boldsymbol{\beta}$; η is then related to the mean by $\eta = a(\mu)$. The function $a(\cdot)$ is called the link function and the special case $a(\mu) = \theta$ is called the canonical link function.

Let us now examine three examples of interest to actuaries.

Claim frequencies: Poisson law is particularly suitable for modelling claim counts. In this case, the discrete probability density function writes

$$\exp(-\lambda)\frac{\lambda^y}{y!} = \exp\{-\lambda + y\ln\lambda - \ln y!\},\ \ y \in \mathbb{N},\ \ \lambda > 0.$$

Identifying the latter expression to the canonical log-likelihood (1.1) yields $\theta = \ln\lambda$, $b(\theta) = \exp(\theta)$, $c(y,\phi) = -\ln y!$, $\phi = 1$, $a(\xi) = \ln\xi$, $v(\mu) = \mu$.

Loss ratios: Normal law often provides an appropriate model for loss ratios (i.e. the total claims relating to a risk class or to the whole portfolio, divided by the corresponding pure premium). The associated density function is

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right) = \exp\left\{\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{1}{2\sigma^2}y^2 - \ln(\sigma\sqrt{2\pi})\right\}$$

and we have $\theta = \mu$, $b(\theta) = \theta^2/2$, $\phi = \sigma^2$, $a(\xi) = \xi$, $v(\mu) = 1, c(y,\phi) = -\frac{1}{2\sigma^2}y^2 - \ln(\sigma\sqrt{2\pi})$.

Claim severities: when the distribution of individual claim severities is of interest, Gamma distribution may provide an acceptable model. In a regression approach, this model is desirable since it falls beyond the scope of GLM's. Moreover, it takes into account the non-negativity and the dissymetry of empirical data. The associated pdf writes

$$\begin{aligned} & \frac{1}{\Gamma(\vartheta)}\left(\frac{\vartheta y}{\mu}\right)^{\vartheta}\exp\left(-\frac{\vartheta y}{\mu}\right)d\ln y,\ \ y \in \mathbb{R}^+,\ \ \vartheta, \mu > 0, \\ = \ & \exp\left\{\vartheta\left\{-\frac{y}{\mu} - \ln\mu\right\} + (\vartheta - 1)\ln y + \vartheta\ln\vartheta - \ln\Gamma(\vartheta)\right\} \end{aligned}$$

whence we get $\theta = -1/\mu$, $b(\theta) = \ln(-1/\theta)$, $\phi = 1/\vartheta$, $a(\xi) = -1/\xi$, $v(\mu) = \mu^2, c(y,\phi) = (\vartheta-1)\ln y + \vartheta\ln\vartheta - \ln\Gamma(\vartheta)$. The canonical link induces the relation $\eta = 1/\mu$ between the linear predictor η and the mean μ. This imposes restrictions on the regressors $\boldsymbol{\beta}$ to ensure the positivity of the mean μ. Therefore, a log-link $\eta = \ln\mu$ is often advocated in practice. Moreover, when combined with the Poisson model to represent the total burden of claims relating to an individual policy, the Gamma specification with the log-link produces a multiplicative pure premium, which is appreciated in practice.

1.3 Credibility theory and GLMM's

Several extensions of GLM's involve models with random terms in the linear predictor; in such a case, we speak of Generalized Linear Mixed Models (GLMM's). Given an unobserved random effect (Λ_i, say), the ν_i observations $Y_{i1}, Y_{i2}, \ldots, Y_{i\nu_i}$ relating to the ith policy in years $1, 2, \ldots, \nu_i$, respectively, are assumed to be conditionally independent with means that depend on the linear predictor through a specified link function, conditional variances that are specified by a variance function, and a scale factor, in the spirit of (1.1).

Assume now that n random vectors $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$ of respective dimensions $\nu_1, \nu_2, \ldots, \nu_n$ are observed. Each of the vectors relates to one of the policies in the portfolio and summarizes past claims history of the contract. The $\boldsymbol{Y}_i$'s are mutually independent, but their components may be correlated (because of repeated measures on the same policies). GLMM's rely on the two following assumptions:

A1 Given $\Lambda_i = \lambda_i$ the responses Y_{it}, $t = 1, \ldots, \nu_i$, are mutually independent and admit a log-likelihood $\ell(\theta_{it}, \phi; y_{it})$ of the form (1.1) where

$$\mu_{it} = \mu_{it}(\lambda_i) = \mathbb{E}[Y_{it}|\Lambda_i = \lambda_i] \text{ and } v_{it} = v_{it}(\lambda_i) = \mathbb{V}\text{ar}[Y_{it}|\Lambda_i = \lambda_i]$$

satisfy

$$a(\mu_{it}(\lambda_i)) = \boldsymbol{\beta}^t \boldsymbol{x}_{it} + \lambda_i = \eta_{it} + \lambda_i \text{ and } v_{it} = v(\mu_{it})\phi.$$

The $\boldsymbol{x}_{it}$ are know design vectors, $\boldsymbol{\beta}$ is the vector of unknown regression coefficients, η_{it} is the linear predictor, $a(\cdot)$ and $v(\cdot)$ are known link and variance functions, respectively.

Henceforth, we restrict our study to the special case of the canonical link function, that is, $a(\mu_{it}(\lambda_{it})) = \theta_{it}$. This does not represent a real loss of generality (since most results in fact rely on the monotonicity of the function $a(\cdot)$) but facilitates the exposition. Moreover, we treat the η_{it}'s as constants.

A2 The random effects $\Lambda_1, \ldots, \Lambda_n$ are mutually independent with a common underlying distribution.

Moreover the random effects are taken such that

$$\mathbb{E}[\mu_{it}(\Lambda_i)] = a^{-1}(\boldsymbol{\beta}^t \boldsymbol{x}_{it})$$

(in particular, if $a(\cdot)$ is the log-link, then for the Poisson and the Gamma distribution, we obtain $\mathbb{E}[\exp(\Lambda_i)] = 1$). This condition is important in actuarial practice. It ensures that the *a priori* ratemaking is correct on average: in every risk class, some policyholders have to pay too much and subsidize other policyholders who are undercharged but the company gets enough premium for the whole risk class. Statistically speaking such a condition is often necessary for identifiability.

GLMM's are widely used by actuaries, since they form the basis of credibility theory and bonus-malus systems. Building on Lee & Nelder's (1996) work, Nelder & Verrall (1997) showed how credibility theory can be encompassed within the theory of GLMM's. In this context, Y_{it} represents the number or amount of claims reported to the insurance company by policyholder i in year t. The random effect Λ_i represents hidden features influencing the risk covered by the insurer.

1.4 Scope of the work

The present paper aims to investigate some features of the Bühlmann credibility premium. Specifically, by establishing various probabilistic inequalities we show that Bühlmann's premium is indeed a good predictor of the future claims characteristics (number or amount) in model A1-A2. Basically, we prove that increasing the linear credibility premium (i.e. deteriorating the claim record of the policyolder) makes larger the probability to observe more important losses in the future.

The main theoretical vehicles of the work are stochastic dominance and conditional increasingness, which proved their usefulness in various actuarial problems. In our context, these probabilistic tools translate in mathematical terms the intuitive concepts of "being larger than" and "being positively dependent".

Let us briefly detail the content of the paper. In Section 2, we recall the definitions of stochastic dominance and we derive elementary monotonicity results in the credibility model A1-A2. In Section 3, we show that because of the time invariance of the random effect modelling residual heterogeneity, the past claims history is exhaustively summarized in the total claims. Section 4 establishes the stochastic increasingness of the random effect in the total claims reported in the past. Section 5 considers predictive distributions and extends the results of Section 4 to observed claims. The final Section 6 deals with linear credibility model.

2 Preliminary results and concepts

2.1 Univariate stochastic dominance

Since the random effects have been added on the scale of the fixed effects, we must investigate whether an increase of the canonical parameter makes the GLM outcome "larger" or "smaller". In order to formalize this monotonicity, we need stochastic orderings; for a thorough study of these probabilistic tools, we refer the reader e.g. to Kaas, Van Heerwaarden & Goovaerts (1994) or to Shaked & Shanthikumar (1994).

Stochastic dominance is a widely accepted method for comparing random variables in the economic and statistical theories of risk. It is defined as follows:

Definition 2.1. Let X and Y be two random variables. X is said to be smaller than Y in the stochastic dominance, denoted as $X \preceq_{\text{st}} Y$, if any of the following equivalent conditions is verified:

(i) for every decision-maker with an increasing utility function u, the loss X is preferred over the loss Y, that is $\mathbb{E}u(-X) \geq \mathbb{E}u(-Y)$;

(ii) $\Pr[X > \xi] \leq \Pr[Y > \xi]$ for all $\xi \in \mathbb{R}$.

The inequality $\Pr[X > \xi] \leq \Pr[Y > \xi]$ shows that Y is indeed "larger" than X since the probability for Y to be large (i.e. to exceed the treshold ξ) is bigger than the probability for X to be large.

A useful sufficient condition for $X \preceq_{st} Y$ is that the ratio of their respective (discrete or continuous) densities f_X/f_Y decreases over the union of their supports (taking $a/0$ as $+\infty$). Indeed, this is equivalent to

$$f_X(u)f_Y(v) \geq f_X(v)f_Y(u) \text{ for all } u \leq v \in \mathbb{R}. \tag{2.1}$$

Integrating (or summing in the discrete case) both sides of (2.1) over $u \in (-\infty, \xi]$ and $v \in (\xi, +\infty)$ yields

$$\begin{aligned} & \Pr[X \leq \xi] \Pr[Y > \xi] \geq \Pr[X > \xi] \Pr[Y \leq \xi] \\ \Leftrightarrow \quad & \big(1 - \Pr[X > \xi]\big) \Pr[Y > \xi] \geq \Pr[X > \xi]\big(1 - \Pr[Y > \xi]\big) \\ \Leftrightarrow \quad & \Pr[Y > \xi] \geq \Pr[X > \xi] \end{aligned}$$

so that (2.1) indeed ensures $X \preceq_{st} Y$. In the literature devoted to stochastic orderings, random variables whose probability density functions satisfy (2.1) are said to be ordered in the likelihood ratio order. For such a ranking it can be given the following nice interpretation. Provided X and Y are independent (which can be assumed without loss of generality), the left-hand side of (2.1) can be regarded as the likelihood of the event "X is small and Y is large" whereas the right-hand side of this relation reads "X is large and Y is small". Then, (2.1) expresses the fact that the latter event is less likely to occur than the former.

2.2 Stochastic increasingness of Y_{it} in the canonical parameter

The following property will be extremely useful in the remainder of our work. It states that in the model **A1-A2**, all covariates being observed (fixed), the response is stochastically increasing in the random parameter.

Property 2.2. *In the model **A1-A2**, $[Y_{it}|\Lambda_i = \lambda_i]$ is increasing in λ_i in the $\preceq_{st}$-sense, that is,*

$$\lambda_i \leq \lambda_i' \Rightarrow [Y_{it}|\Lambda_i = \lambda_i] \preceq_{st} [Y_{it}|\Lambda_i = \lambda_i'] \text{ for any } i \text{ and } t.$$

Proof. Since $\eta_{it} + \lambda_i = \theta_{it}$, the announced implication is the same as

$$\theta_{it} \leq \theta_{it}' \Rightarrow [Y_{it}|\Theta_{it} = \theta_{it}] \preceq_{\mathrm{st}} [Y_{it}|\Theta_{it} = \theta_{it}']$$

where $\Theta_{it} = \eta_{it} + \Lambda_i$. The ratio of (discrete or continuous) probability density functions of the response has the form:

$$\frac{f(y; \theta_{it}, \phi)}{f(y; \theta_{it}', \phi)} = \exp\left(\frac{y(\theta_{it} - \theta_{it}') - (b(\theta_{it}) - b(\theta_{it}'))}{\phi}\right)$$

which is clearly decreasing in y provided $\theta_{it} \leq \theta_{it}'$. We then get the announced result from (2.1). □

Note that the result of Property 2.2 is also valid in the gamma case with a log-link.

2.3 Multivariate stochastic dominance

In case we want to compare random vectors, an intuitively acceptable strategy consists in transforming those vectors into random variables using increasing mappings, and then to compare the resulting outcomes with $\preceq_{\mathrm{st}}$. This yields the following definition.

Definition 2.3. Given two n-dimensional random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, $\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ in the stochastic dominance, written as $\boldsymbol{X} \preceq_{\mathrm{st}} \boldsymbol{Y}$, if $\psi(\boldsymbol{X}) \preceq_{\mathrm{st}} \psi(\boldsymbol{Y})$ for every non-decreasing function $\psi : \mathbb{R}^n \to \mathbb{R}$.

2.4 Stochastic increasingness of $\boldsymbol{Y}_i$ in the canonical parameter

The next property is the multivariate counterpart of Property 2.2. It applies in particular to GLMM's. Henceforth, an inequality between two real vectors has to be interpreted componentwise.

Property 2.4. *In the model* ***A1-A2****,* $[\boldsymbol{Y}_i|\Lambda_i = \lambda_i]$ *is increasing in* λ_i *in the* $\preceq_{st}$*-sense, that is,*

$$\lambda_i \leq \lambda_i' \Rightarrow [\boldsymbol{Y}_i|\Lambda_i = \lambda_i] \preceq_{st} [\boldsymbol{Y}_i|\Lambda_i = \lambda_i'].$$

Proof. Under **A1-A2**, $[\boldsymbol{Y}_i|\Lambda_i = \lambda_i]$ is a random vector with independent components. Moreover, by virtue of Property 2.2, the stochastic inequalities

$$\begin{array}{rcl} [Y_{i1}|\Lambda_i = \lambda_i] & \preceq_{\rm st} & [Y_{i1}|\Lambda_i = \lambda_i'] \\ [Y_{i2}|\Lambda_i = \lambda_i] & \preceq_{\rm st} & [Y_{i2}|\Lambda_i = \lambda_i'] \\ \vdots & \vdots & \vdots \\ [Y_{i\nu_i}|\Lambda_i = \lambda_i] & \preceq_{\rm st} & [Y_{i\nu_i}|\Lambda_i = \lambda_i'] \end{array}$$

hold true for any $\lambda_i \leq \lambda_i'$. Since a comparison in the multivariate $\preceq_{\rm st}$-sense reduces to a componentwise $\preceq_{\rm st}$-ranking when both random vectors have independent component, we get that $[\boldsymbol{Y}_i|\Lambda_i = \lambda_i] \preceq_{\rm st} [\boldsymbol{Y}_i|\Lambda_i = \lambda_i']$ holds true provided $\lambda_i \leq \lambda_i'$, as announced. □

3 Exhaustive summary of past claims

Since the random effect is constant over time, the age of the claims is not taken into account in the prediction. Technically, this is expressed by the fact that the predictive distribution of Λ_i given the past only depends on the sum $Y_{i\bullet}$ of the Y_{it}'s, and not on the particular values of the Y_{it}'s.

Property 3.1. *The predictive distribution of* Λ_i *only depends on* $Y_{i\bullet} = \sum_{t=1}^{\nu_i} Y_{it}$.

Proof. Denote by $g(\cdot)$ the a priori probability density function of Λ_i. Then the a posteriori probability density function $g(\cdot|\boldsymbol{y}_i)$ of Λ_i given $\boldsymbol{Y}_i = \boldsymbol{y}_i$ can be written as:

$$\begin{aligned} g(\lambda_i|\boldsymbol{y}_i) &= \frac{\left\{\prod_{t=1}^{\nu_i} f_{it}(y_{it}|\lambda_i)\right\} g(\lambda_i)}{\int_{\xi\geq 0} \left\{\prod_{t=1}^{\nu_i} f_{it}(y_{it}|\xi)\right\} g(\xi) d\xi} \\ &= \frac{\exp\left(\frac{\sum_{t=1}^{\nu_i}\{y_{it}(\eta_{it}+\lambda_i)-b(\eta_{it}+\lambda_i)\}}{\phi}\right) g(\lambda_i)}{\int_{\xi\geq 0} \exp\left(\frac{\sum_{t=1}^{\nu_i}\{y_{it}(\eta_{it}+\xi)-b(\eta_{it}+\xi)\}}{\phi}\right) g(\xi) d\xi} \\ &= \frac{\exp\left(\frac{\lambda_i y_{i\bullet}-\sum_{t=1}^{\nu_i} b(\eta_{it}+\lambda_i)}{\phi}\right) g(\lambda_i)}{\int_{\xi\geq 0} \exp\left(\frac{\xi y_{i\bullet}-\sum_{t=1}^{\nu_i} b(\eta_{it}+\xi)}{\phi}\right) g(\xi) d\xi} \\ &\equiv g(\lambda_i|y_{i\bullet}). \end{aligned}$$

The fact that $g(\cdot|\boldsymbol{y}_i)$ only depends on $y_{i\bullet}$ ends the proof. □

4 A posteriori distribution of the random effects

Let us now consider the posterior distribution of Λ_i given the observations $Y_{i\bullet} = y_{i\bullet}$. Let us prove the following result, which is in the vein of WHITT (1979) and FAHMY ET AL. (1982). It basically states that observing large outcomes $Y_{i\bullet}$ increases the unobservable latent variables Λ_i (in the $\preceq_{\text{st}}$-sense).

Proposition 4.1. *In the model **A1**-**A2**, $[\Lambda_i|Y_{i\bullet} = y_{i\bullet}] \preceq_{st} [\Lambda_i|Y_{i\bullet} = y'_{i\bullet}]$ whenever $y_{i\bullet} \leq y'_{i\bullet}$.*

Proof. We have that

$$
\begin{aligned}
f(\boldsymbol{y}_i|\lambda_i) &= \prod_{t=1}^{\nu_i} f_{it}(y_{it}|\lambda_i) \\
&= \prod_{t=1}^{\nu_i} \exp\left(\frac{y_{it}\theta_{it} - b(\theta_{it})}{\phi}\right) \exp\left(c(y_{it}, \phi)\right) \\
&= \left\{\prod_{t=1}^{\nu_i} \exp\left(c(y_{it}, \phi)\right) \exp\left(\frac{y_{it}\eta_{it}}{\phi}\right)\right\} \left\{\prod_{t=1}^{\nu_i} \exp\left(\frac{-b(\lambda_i + \eta_{it})}{\phi}\right)\right\} \left\{\prod_{t=1}^{\nu_i} \exp\left(\frac{y_{it}\lambda_i}{\phi}\right)\right\} \\
&\equiv h_1(\boldsymbol{y}_i) h_2(\lambda_i) \exp\left(\frac{y_{i\bullet}\lambda_i}{\phi}\right).
\end{aligned}
$$

Hence, we can express the posterior probability density function of Λ_i given $Y_{i\bullet} = y_{i\bullet}$ as

$$
\begin{aligned}
g(\lambda_i|\boldsymbol{y}_i) &= \frac{f(\boldsymbol{y}_i|\lambda_i) g(\lambda_i)}{f(\boldsymbol{y}_i)} \\
&= \frac{h_1(\boldsymbol{y}_i) h_2(\lambda_i) \exp\left(\frac{y_{i\bullet}\lambda_i}{\phi}\right) g(\lambda_i)}{f(\boldsymbol{y}_i)}.
\end{aligned}
$$

For $y_{i\bullet} \leq y'_{i\bullet}$ and $\lambda_i \leq \lambda'_i$, the inequality

$$
g(\lambda_i|\boldsymbol{y}_i) g(\lambda'_i|\boldsymbol{y}'_i) \geq g(\lambda'_i|\boldsymbol{y}_i) g(\lambda_i|\boldsymbol{y}'_i)
$$

(i.e. the sufficient condition (2.1) for the $\preceq_{\text{st}}$-order) holds true. Indeed, it reduces to

$$
\exp\left(\frac{y_{i\bullet}\lambda_i}{\phi} + \frac{y'_{i\bullet}\lambda'_i}{\phi}\right) \geq \exp\left(\frac{y_{i\bullet}\lambda'_i}{\phi} + \frac{y'_{i\bullet}\lambda_i}{\phi}\right) \Leftrightarrow (\lambda'_i - \lambda_i)(y'_{i\bullet} - y_{i\bullet}) \geq 0
$$

which is obviously true. □

In a Bayesian framework, it is common to predict Λ_i by means of the posterior mean, that is

$$
\widehat{\Lambda}_i(\boldsymbol{y}_i) = \mathbb{E}[\Lambda_i|Y_{i\bullet} = y_{i\bullet}].
$$

Proposition 4.1 ensures in particular that $\widehat{\Lambda}_i(y_{i\bullet}) \leq \widehat{\Lambda}_i(y'_{i\bullet})$ whenever $y_{i\bullet} \leq y'_{i\bullet}$.

5 Predictive distribution

The very aim of credibility theory is to predict future claim behaviour. In that respect, predictive distributions are of prime interest: these are the distributions of claim characteristics for next year, given past observations. Looking at $Y_{i\nu_i+1}$, we get the following result.

Proposition 5.1. *In the model A1-A2,* $[Y_{i\nu_i+1}|Y_{i\bullet}=y_{i\bullet}] \preceq_{st} [Y_{i\nu_i+1}|Y_{i\bullet}=y'_{i\bullet}]$ *whenever* $y_{i\bullet} \leq y'_{i\bullet}$.

Proof. It suffices to write

$$\begin{aligned}\Pr[Y_{i\nu_i+1}>\xi|Y_{i\bullet}=y_{i\bullet}] &= \int_{\lambda_i\geq 0}\Pr[Y_{i\nu_i+1}>\xi|\Lambda_i=\lambda_i]g(\lambda_i|y_{i\bullet})d\lambda_i \\ &= \int_{\lambda_i\geq 0}\phi(\lambda_i)g(\lambda_i|y_{i\bullet})d\lambda_i\end{aligned}$$

where

$$\phi(\lambda_i)=\Pr[Y_{i\nu_i+1}>\xi|\Lambda_i=\lambda_i]$$

is non-decreasing in λ_i by virtue of Property 2.2. Therefore,

$$\begin{aligned}\Pr[Y_{i\nu_i+1}>\xi|Y_{i\bullet}=y_{i\bullet}] &= \mathbb{E}[\phi(\Lambda_i)|Y_{i\bullet}=y_{i\bullet}] \\ &\leq \mathbb{E}[\phi(\Lambda_i)|Y_{i\bullet}=y'_{i\bullet}] \\ &= \Pr[Y_{i\nu_i+1}>\xi|Y_{i\bullet}=y'_{i\bullet}]\end{aligned}$$

where the inequality follows from Property 4.1 together with Definition 2.1(i). This ends the proof. □

6 Linear credibility premium

Bayes revision $\mathbb{E}[\Lambda_i|Y_{i\bullet}]$ of the heterogeneity component is theoretically very satisfying but is often difficult to compute (except for conjugate distributions). Computer intensive methods of MCMC-type can be used to evaluate it, but they are sometimes rather time-consuming. An alternative is furnished by the Bühlmann linear credibility model. In this case, the actuary still uses a quadratic loss function but the shape of the predictor is constrained ex ante to be linear in past observations, i.e. the predictor $\hat{Y}_{i\nu_i+1}$ of $Y_{i\nu_i+1}$ is of the form

$$\hat{Y}_{i\nu_i+1}=c_{i0}+\sum_{t=1}^{\nu_i}c_{it}Y_{it}$$

Specifically, we look for c_{i0} and c_{it}'s, $t=1,\ldots,\nu_i$, such that the expected square difference between $Y_{i\nu_i+1}$ and its prediction $\hat{Y}_{i\nu_i+1}$ is minimum, i.e. such that

$$\Psi(\boldsymbol{c})=\mathbb{E}\left[\left(Y_{i\nu_i+1}-c_{i0}-\sum_{t=1}^{\nu_i}c_{it}Y_{it}\right)^2\right]$$

is minimum. Equating $\partial\Psi(\boldsymbol{c})/\partial c_{i0}$ to 0 yields

$$c_{i0} = \mathbb{E}[Y_{i\nu_i+1}] - \sum_{t=1}^{\nu_i} c_{it}\mathbb{E}[Y_{it}]. \tag{6.1}$$

Next, $\partial\Psi(\boldsymbol{c})/\partial c_{is} = 0$ gives

$$\mathbb{E}[Y_{i\nu_i+1}Y_{is}] = c_{i0}\mathbb{E}[Y_{is}] + \sum_{t=1}^{\nu_i} c_{it}\mathbb{E}[Y_{it}Y_{is}]$$

where we insert (6.1) to get

$$\mathbb{C}\text{ov}[Y_{i\nu_i+1}, Y_{is}] = \sum_{t=1}^{\nu_i} c_{it}\mathbb{C}\text{ov}[Y_{it}, Y_{is}]. \tag{6.2}$$

In the model A1-A2, for $s \neq t$

$$\begin{aligned}\mathbb{C}\text{ov}[Y_{it}, Y_{is}] &= \mathbb{E}\Big[\mathbb{C}\text{ov}[Y_{it}, Y_{is}|\Lambda_i]\Big] + \mathbb{C}\text{ov}\Big[\mathbb{E}[Y_{it}|\Lambda_i], \mathbb{E}[Y_{is}|\Lambda_i]\Big] \\ &= \mathbb{C}\text{ov}[\mu_{it}(\Lambda_i), \mu_{is}(\Lambda_i)]\end{aligned}$$

whereas

$$\begin{aligned}\mathbb{V}\text{ar}[Y_{is}] &= \mathbb{E}\Big[\mathbb{V}\text{ar}[Y_{is}|\Lambda_i]\Big] + \mathbb{V}\text{ar}\Big[\mathbb{E}[Y_{is}|\Lambda_i]\Big] \\ &= \phi\mathbb{E}[v_{is}(\Lambda_i)] + \mathbb{V}\text{ar}\Big[\mu_{is}(\Lambda_i)\Big]\end{aligned}$$

so that (6.2) becomes

$$\mathbb{C}\text{ov}[\mu_{i\nu_i+1}(\Lambda_i), \mu_{is}(\Lambda_i)] = \phi c_{is}\mathbb{E}[v_{is}(\Lambda_i)] + \sum_{t=1}^{\nu_i} c_{it}\mathbb{C}\text{ov}[\mu_{it}(\Lambda_i), \mu_{is}(\Lambda_i)]. \tag{6.3}$$

Let us now examine what (6.3) gives in the three examples of interest:

(i) In the Poisson case, we have

$$\mu_{it}(\lambda_i) = \exp(\eta_{it} + \lambda_i), \;\; \phi = 1 \text{ and } v_{is}(\lambda_i) = \exp(\eta_{is} + \lambda_i)$$

so that the system (6.3) becomes

$$\begin{aligned}&\exp(\eta_{i\nu_i+1})\exp(\eta_{is})\mathbb{V}\text{ar}[\exp(\Lambda_i)] \\ = \;& c_{is}\mathbb{E}[\exp(\eta_{is} + \Lambda_i)] + \sum_{t=1}^{\nu_i} c_{it}\mathbb{C}\text{ov}[\exp(\eta_{it} + \Lambda_i), \exp(\eta_{is} + \Lambda_i)] \\ = \;& c_{is}\exp(\eta_{is}) + \left\{\sum_{t=1}^{\nu_i} c_{it}\exp(\eta_{it})\right\}\exp(\eta_{is})\mathbb{V}\text{ar}[\exp(\Lambda_i)]\end{aligned}$$

$$\Rightarrow \exp(\eta_{i\nu_i+1})\mathbb{V}\mathrm{ar}[\exp(\Lambda_i)] = c_{is} + \mathbb{V}\mathrm{ar}[\exp(\Lambda_i)]\sum_{t=1}^{\nu_i} c_{it}\exp(\eta_{it}).$$

This yields $c_{is} = c_i$ for all $s = 1, \ldots, \nu_i$ and then

$$\begin{aligned}
c_i &= \frac{\exp(\eta_{i\nu_i+1})\mathbb{V}\mathrm{ar}[\exp(\Lambda_i)]}{1 + \mathbb{V}\mathrm{ar}[\exp(\Lambda_i)]\sum_{t=1}^{\nu_i}\exp(\eta_{it})} \\
c_{i0} &= \exp(\eta_{i\nu_i+1}) - \frac{\exp(\eta_{i\nu_i+1})\mathbb{V}\mathrm{ar}[\exp(\Lambda_i)]}{1 + \mathbb{V}\mathrm{ar}[\exp(\Lambda_i)]\sum_{t=1}^{\nu_i}\exp(\eta_{it})}\sum_{t=1}^{\nu_i}\exp(\eta_{it}) \\
&= \frac{\exp(\eta_{i\nu_i+1})}{1 + \mathbb{V}\mathrm{ar}[\exp(\Lambda_i)]\sum_{t=1}^{\nu_i}\exp(\eta_{it})}.
\end{aligned}$$

Bühlmann linear credibility premium for year $\nu_i + 1$ thus equals

$$\exp(\eta_{i\nu_i+1})\frac{1 + \mathbb{V}\mathrm{ar}[\exp(\Lambda_i)]\sum_{t=1}^{\nu_i} Y_{it}}{1 + \mathbb{V}\mathrm{ar}[\exp(\Lambda_i)]\sum_{t=1}^{\nu_i}\exp(\eta_{it})}$$

which appears as the product of the a priori premium $\exp(\eta_{i\nu_i+1})$ times an approximation of the theoretical bonus-malus coefficient $\mathbb{E}[\exp(\Lambda_i)|Y_{i1}, \ldots, Y_{i\nu_i}]$. The latter approximation possesses a particularly simple interpretation since it entails a malus when $\sum_{t=1}^{\nu_i} Y_{it} > \sum_{t=1}^{\nu_i}\exp(\eta_{it})$, that is, if the policyholder reported more claims than expected a priori.

(ii) In the normal case, we get

$$\mu_{it}(\lambda_i) = \eta_{it} + \lambda_i, \;\; \phi = \sigma^2 \text{ and } v(\lambda_i) = 1$$

so that the system (6.3) becomes

$$\mathbb{V}\mathrm{ar}[\Lambda_i] = \sigma^2 c_{is} + \mathbb{V}\mathrm{ar}[\Lambda_i]\sum_{t=1}^{\nu_i} c_{it}$$

whence it follows that $c_{is} = c_i$ for all $s = 1, \ldots, \nu_i$ and

$$\begin{aligned}
c_i &= \frac{\mathbb{V}\mathrm{ar}[\Lambda_i]}{\sigma^2 + \mathbb{V}\mathrm{ar}[\Lambda_i]\nu_i} \\
c_{i0} &= \eta_{i\nu_i+1} - \frac{\mathbb{V}\mathrm{ar}[\Lambda_i]}{\sigma^2 + \mathbb{V}\mathrm{ar}[\Lambda_i]\nu_i}\sum_{t=1}^{\nu_i}\eta_{it}.
\end{aligned}$$

The Bühlmann credibility linear premium for year $\nu_i + 1$ now equals

$$\eta_{i\nu_i+1} + \frac{\mathbb{V}\mathrm{ar}[\Lambda_i]}{\sigma^2 + \nu_i\mathbb{V}\mathrm{ar}[\Lambda_i]}\sum_{t=1}^{\nu_i}\big(Y_{it} - \eta_{it}\big).$$

This can be seen as a correction of the a priori premium $\eta_{i\nu_i+1}$ according to the value of the difference $Y_{it} - \eta_{it}$ between the observed outcome and its a priori expectation.

(iii) Finally, in the Gamma case with log-link, we have $\ln \mu_{it} = \eta_{it} + \lambda_i$, $\phi = 1/\vartheta$ and $v_{is}(\lambda_i) = \exp(2\eta_{is} + 2\lambda_i)$.
Therefore the system (6.3) becomes:

$$\begin{aligned}
&\exp(\eta_{i\nu_i+1})\exp(\eta_{is})\mathbb{V}\text{ar}[\exp(\Lambda_i)] \\
=\ & \vartheta^{-1}c_{is}\exp(2\eta_{is})\mathbb{E}[\exp(2\Lambda_i)] + \sum_{t=1}^{\nu_i} c_{it}\exp(\eta_{it})\exp(\eta_{is})\mathbb{V}\text{ar}[\exp(\Lambda_i)] \\
=\ & \vartheta^{-1}c_{is}\exp(2\eta_{is})\big\{\mathbb{V}\text{ar}[\exp(\Lambda_i)] + 1\big\} + \left\{\sum_{t=1}^{\nu_i} c_{it}\exp(\eta_{it})\right\}\exp(\eta_{is})\mathbb{V}\text{ar}[\exp(\Lambda_i)]
\end{aligned}$$

It follows that

$$\begin{aligned}
&\exp(\eta_{i\nu_i+1})\mathbb{V}\text{ar}[\exp(\Lambda_i)] \\
=\ & \vartheta^{-1}c_{is}\exp(\eta_{is})\big\{\mathbb{V}\text{ar}[\exp(\Lambda_i)] + 1\big\} + \left\{\sum_{t=1}^{\nu_i} c_{it}\exp(\eta_{it})\right\}\mathbb{V}\text{ar}[\exp(\Lambda_i)].
\end{aligned}$$

By summing both sides over s we obtain:

$$\sum_{t=1}^{\nu_i} c_{it}\exp(\eta_{it}) = \frac{\nu_i\exp(\eta_{it})\mathbb{V}\text{ar}[\exp(\Lambda_i)]}{\vartheta^{-1}\big\{\mathbb{V}\text{ar}[\exp(\Lambda_i)] + 1\big\} + \nu_i\mathbb{V}\text{ar}[\exp(\Lambda_i)]}$$

This yields to :

$$c_{is} = \frac{\exp(\eta_{i\nu_i+1})\mathbb{V}\text{ar}[\exp(\Lambda_i)]}{\exp(\eta_{is})\big\{\vartheta^{-1}\big\{\mathbb{V}\text{ar}[\exp(\Lambda_i)] + 1\big\} + \nu_i\mathbb{V}\text{ar}[\exp(\Lambda_i)]\big\}}$$

$$\begin{aligned}
c_{i0} &= \exp(\eta_{i\nu_i+1}) - \frac{\nu_i\exp(\eta_{i\nu_i+1})\mathbb{V}\text{ar}[\exp(\Lambda_i)]}{\vartheta^{-1}\big\{\mathbb{V}\text{ar}[\exp(\Lambda_i)] + 1\big\} + \nu_i\mathbb{V}\text{ar}[\exp(\Lambda_i)} \\
&= \frac{\vartheta^{-1}\exp(\eta_{i\nu_i+1})\big\{\mathbb{V}\text{ar}[\exp(\Lambda_i)] + 1\big\}}{\vartheta^{-1}\big\{\mathbb{V}\text{ar}[\exp(\Lambda_i)] + 1\big\} + \nu_i\mathbb{V}\text{ar}[\exp(\Lambda_i)]}.
\end{aligned}$$

Thus the Bühlmann credibility linear premium for year $\nu_i + 1$ equals

$$\frac{\exp(\eta_{i\nu_i+1})}{\vartheta^{-1}\big\{\mathbb{V}\text{ar}[\exp(\Lambda_i)] + 1\big\} + \nu_i\mathbb{V}\text{ar}[\exp(\Lambda_i)]}\left[\vartheta^{-1}\big\{\mathbb{V}\text{ar}[\exp(\Lambda_i)]+1\big\}+\mathbb{V}\text{ar}[\exp(\Lambda_i)]\sum_{t=1}^{\nu_i}\frac{Y_{it}}{\exp(\eta_{it})}\right]$$

$$= \exp(\eta_{i\nu_i+1})\left\{1 + \frac{\mathbb{V}\text{ar}[\exp(\Lambda_i)]}{\vartheta^{-1}\big\{\mathbb{V}\text{ar}[\exp(\Lambda_i)] + 1\big\} + \nu_i\mathbb{V}\text{ar}[\exp(\Lambda_i)]}\sum_{t=1}^{\nu_i}\left(\frac{Y_{it}}{\exp(\eta_{it})} - 1\right)\right\}$$

Again this can be seen as a correction of the a priori premium $\exp(\eta_{i\nu_i+1})$ according to residuals $\frac{Y_{it}}{\exp(\eta_{it})} - 1$.

It is worth mentioning that in the three cases examined above, we have $c_i \geq 0$ and a linear credibility premium of the form $\pi_{\text{cred}} = c_{i0} + c_i Y_{i\bullet}$ (in the Gamma case, we work with the auxiliary variable $\frac{Y_{it}}{\mathbb{E}(Y_{it})}$).

Let us now prove that $Y_{i\nu_i+1}$ is indeed increasing in π_{cred}.

Proposition 6.1. *For $p \leq p'$, $[Y_{i\nu_i+1}|\pi_{cred} = p] \preceq_{st} [Y_{i\nu_i+1}|\pi_{cred} = p']$, that is*

$$\Pr[Y_{i\nu_i+1} > y|\pi_{cred} = p] \leq \Pr[Y_{i\nu_i+1} > y|\pi_{cred} = p'] \text{ for all } y \in \mathbb{R}.$$

Proof. The result straightly follows from Proposition 5.1 since

$$\begin{aligned}
\Pr[Y_{i\nu_i+1} > y|\pi_{\text{cred}} = p] &= \Pr\left[Y_{i\nu_i+1} > y\middle|Y_{i\bullet} = \frac{p - c_{i0}}{c_i}\right] \\
&\leq \Pr\left[Y_{i\nu_i+1} > y\middle|Y_{i\bullet} = \frac{p' - c_{i0}}{c_i}\right] \\
&= \Pr[Y_{i\nu_i+1} > y|\pi_{\text{cred}} = p']
\end{aligned}$$

as announced. □

Acknowledgements

The support of the Belgian Government under contract "Projet d'Actions de Recherche Concertées" ARC 98/03-217 is gratefully acknowledged.

References

[1] Brouhns, N., & Denuit, M. (2002). On the dependence induced by GLMMs. In preparation.

[2] Bühlmann, H. (1970). *Mathematical Methods in Risk Theory*. Springer Verlag, New York.

[3] Fahmy, S., Proschan, F., Pereira, C., & Shaked, M. (1982). The influence of the sample on the posterior distribution. *Communications in Statististics - Theory and Methods* **11**, 1757-1768.

[4] Herzog, T.N. (1994). *Introduction to Credibility Theory*. Actex Publications, Winsted.

[5] Joe H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

[6] Kaas, R., van Heerwaarden, A., & Goovaerts, M.J. (1994). *Ordering of Actuarial Risks*. CAIRE Education Series 1, Brussels.

[7] Klugman, S.A. (1992). *Bayesian Statistics in Actuarial Science (with Emphasis on Credibility)*. Kluwer Academix Publishers, London.

[8] Lee, Y., & Nelder, J.A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society - Series B* **58**, 619-678.

[9] Mc Cullagh P., Nelder J.A. (1989). *Generalized Linear Models.* Chapman & Hall, London.

[10] Nelder, J.A., & Verrall, R.J. (1997). Credibility theory and generalized linear models. *ASTIN Bulletin* **27**, 71-82.

[11] Purcaru, O., & Denuit, M. (2001). On the dependence induced by frequency credibility models. I. Static random effects. Discussion Paper 01-34, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

[12] Purcaru, O., & Denuit, M. (2002). On the dependence induced by frequency credibility models. II. Time-dependent random effects. Discussion Paper 02-03, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

[13] Shaked, M., & Shanthikumar, J.G. (1994). *Stochastic Orders and Their Applications.* Academic Press, New York.

[14] Whitt, W. (1979). A note on the influence of the sample on the posterior distribution. *Journal of the American Statistical Association* **74**, 424-426.