ON PARAMETER ESTIMATION IN HIERARCHICAL CREDIBILITY

BY

HASSINE BELHADJ, VINCENT GOULET AND TOMMY OUELLET

ABSTRACT

One can find in the literature three main sets of estimators for the variance components in the hierarchical credibility model. This paper presents these estimators in a unified notation, studies some of their properties important for numerical evaluation and compares their relative performance by simulation. The paper also demonstrates how function `cm` of the `R` package **actuar** can be used to fit hierarchical models to insurance data.

KEYWORDS

Hierarchical, credibility, variance components, `R`, **actuar**.

1. INTRODUCTION

The hierarchical credibility model was introduced by Jewell (1975) as a means to use collateral data from the portfolio in the credibility premium of an insurance contract. Today, hierarchical credibility is mostly seen as a tool to distribute premiums fairly among a heterogeneous portfolio classified in a tree-like structure. Since fitting the model involves measuring the homogeneity of the various nodes, hierarchical credibility can also be used simply to judge of the quality of a classification structure.

This paper focuses on estimation of the structure parameters in the hierarchical credibility model. After a brief presentation of the model in the next section, we review, in Section 3, the estimators of the structure parameters with an emphasis on three sets of variance components estimators that have been proposed in the literature. In Sections 4 and 5, we study properties important for the numerical evaluation of these estimators and measure their relative performance by simulation. This research originated in our implementation of hierarchical credibility calculations in function `cm` of the `R` package **actuar**. We demonstrate this function in Section 6.

Astin Bulletin 39(2), 495-514. doi: 10.2143/AST.39.2.2044645 © 2009 by Astin Bulletin. All rights reserved.

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

2. The model

This section briefly outlines the main assumptions and results of the hierarchical credibility model. The interested reader will find more thorough introductions to hierarchical credibility in Jewell (1975), Goovaerts et al. (1990), Bühlmann and Gisler (2005), Goulet (1998b) and Dannenburg et al. (1996), among others. For simplicity, we would have liked to limit our discussion to a two-level model, but the "mechanics" of the model and the estimators show more clearly in a three-level model. Obviously, the results of the paper equally apply to models with any number of levels.

Consider a large portfolio of insurance contracts classified in a hierarchical structure. The data forms the lowest level, labeled 0 by convention. The data is grouped by contract (level 1), the contracts are grouped by class (level 2) and the classes are grouped by sector (level 3); see Figure 1(a) on page 504. With this numbering, the simpler credibility models of Bühlmann (1967) and Bühlmann and Straub (1970) are one-level hierarchical models. The actual names of the levels are immaterial.

The data is denoted X_{ijkt}, with subscript $i = 1, \ldots, I$ denoting the sector, subscript $j = 1, \ldots, J_i$ denoting the class, subscript $k = 1, \ldots, K_{ij}$ denoting the contract and subscript $t = 1, \ldots, n_{ijk}$ denoting the observation (usually one per period). To each observation corresponds a known, constant, "natural" weight w_{ijkt}. The random variables Ψ_i, Φ_{ij} and Θ_{ijk} represent the risk levels of the sectors, the classes and the contracts, respectively. The assumptions of the model are as follows.

(A1) The random variables $\Psi_1, \ldots, \Psi_I$ are independent and identically distributed (i.i.d.).

(A2) The random variables $\Phi_{i1}, \ldots, \Phi_{iJ_i}$ are conditionally i.i.d. given Ψ_i, $i = 1, \ldots, I$.

(A3) The random variables $\Theta_{ij1}, \ldots, \Theta_{ijK_{ij}}$ are conditionally i.i.d. given Φ_{ij} (and Ψ_i), $j = 1, \ldots, J_i$.

(A4) The random variables $X_{ijk1}, \ldots, X_{ijkn_{ij}}$ are conditionally i.i.d. given Θ_{ijk} (and Ψ_i, Φ_{ij}), $k = 1, \ldots, K_{ij}$. Their variance is finite.

(A5) For all $t, u = 1, \ldots, n_{ijk}$,

$$E[X_{ijkt} | \Theta_{ijk}, \Phi_{ij}, \Psi_i] = \mu(\Theta_{ijk}, \Phi_{ij}, \Psi_i)$$

$$\mathrm{Cov}(X_{ijkt}, X_{ijku} | \Theta_{ijk}, \Phi_{ij}, \Psi_i) = \begin{cases} \dfrac{\sigma^2(\Theta_{ijk}, \Phi_{ij}, \Psi_i)}{w_{ijkt}}, & t = u \\ 0, & t \neq u. \end{cases}$$

We define the following *structure parameters* of the portfolio: the collective mean

$$m = E[E[E[\mu(\Theta_{ijk}, \Phi_{ij}, \Psi_i) | \Phi_{ij}, \Psi_i] | \Psi_i]],$$

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

the average within contract variance

$$s^2 = E[E[E[\sigma^2(\Theta_{ijk}, \Phi_{ij}, \Psi_i) \,|\, \Phi_{ij}, \Psi_i] \,|\, \Psi_i]],$$

the average between contract (or within class) variance

$$a = E[E[\mathrm{Var}[\mu(\Theta_{ijk}, \Phi_{ij}, \Psi_i) \,|\, \Phi_{ij}, \Psi_i] \,|\, \Psi_i]],$$

the average between class (or within sector) variance

$$b = E[\mathrm{Var}[E[\mu(\Theta_{ijk}, \Phi_{ij}, \Psi_i) \,|\, \Phi_{ij}, \Psi_i] \,|\, \Psi_i]]$$

and, finally, the between sector variance

$$c = \mathrm{Var}[E[E[\mu(\Theta_{ijk}, \Phi_{ij}, \Psi_i) \,|\, \Phi_{ij}, \Psi_i] \,|\, \Psi_i]].$$

The goal is to estimate the contract *risk premiums* $\mu(\Theta_{ijk}, \Phi_{ij}, \Psi_i)$. In turn, this will require to successively estimate the class risk premiums $\mu(\Phi_{ij}, \Psi_i) = E[\mu(\Theta_{ijk}, \Phi_{ij}, \Psi_i) \,|\, \Phi_{ij}, \Psi_i]$ and the sector risk premiums $\mu(\Psi_i) = E[\mu(\Phi_{ij}, \Psi_i) \,|\, \Psi_i]$. Indeed, the best – in the mean square sense – inhomogeneous linear estimator of $\mu(\Theta_{ijk}, \Phi_{ij}, \Psi_i)$ is determined by the recursive system

$$\begin{aligned}
\hat{\pi}_{ijk} &= z_{ijk} X_{ijkw} + (1 - z_{ijk})\, \hat{\pi}_{ij}, \\
\hat{\pi}_{ij} &= z_{ij} X_{ijzw} + (1 - z_{ij})\, \hat{\pi}_{i}, \\
\hat{\pi}_{i} &= z_{i} X_{izzw} + (1 - z_{i})\, m
\end{aligned}$$

with the credibility factors

$$z_{ijk} = \frac{w_{ijk\Sigma}}{w_{ijk\Sigma} + s^2/a}, \qquad w_{ijk\Sigma} = \sum_{t=1}^{n_{ijk}} w_{ijkt}, \tag{2}$$

$$z_{ij} = \frac{z_{ij\Sigma}}{z_{ij\Sigma} + a/b}, \qquad z_{ij\Sigma} = \sum_{k=1}^{K_{ij}} z_{ijk}, \tag{3}$$

$$z_{i} = \frac{z_{i\Sigma}}{z_{i\Sigma} + b/c}, \qquad z_{i\Sigma} = \sum_{j=1}^{J_{i}} z_{ij} \tag{4}$$

and the weighted averages

$$X_{ijkw} = \sum_{t=1}^{n_{ijk}} \frac{w_{ijkt}}{w_{ijk\Sigma}}\, X_{ijkt}, \tag{5}$$

$$X_{ijzw} = \sum_{k=1}^{K_{ij}} \frac{z_{ijk}}{z_{ij\Sigma}}\, X_{ijkw}, \tag{6}$$

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

$$X_{izzw} = \sum_{j=1}^{J_i} \frac{z_{ij}}{z_{i\Sigma}} X_{ijzw}. \tag{7}$$

See Bühlmann and Jewell (1987) for an elegant proof using projections in Hilbert spaces.

One will note that, in the notation above, summation over an index is denoted by a Σ replacing the corresponding subscript. Furthermore, replacing a subscript with a *w* (resp. a *z*) denotes a weighted average using natural weights (resp. credibility factors).

The quantities in the second column of (2)-(4) are the *volumes* of the levels. The volume of a level is the sum of the node credibility factors of the level below, except for the contracts level where the natural weights w_{ijkt} are used. In addition, (5)-(7) are the *sufficient statistics* of the levels. Each is the weighted average of the node sufficient statistics of the previous level using the corresponding credibility factors as weights (or the natural weights for contracts). Again, see Bühlmann and Jewell (1987) or Bühlmann and Gisler (2005) for details.

In practice, evaluation of the credibility premium (1) requires to estimate the unknown structure parameters m, s^2, a, b and c. This is the topic of the next section.

3. Estimators of the structure parameters

We first discuss estimators of the collective mean m and the within contract variance s^2. These are well established in the actuarial literature and are not subject to much controversy. The references we give pertain to the Bühlmann-Straub model, but the results apply unchanged to the hierarchical model. On the other hand, we study in greater detail three sets of estimators that have been proposed over the years for the variance components a, b and c.

3.1. Estimator for *m*

The generally accepted estimator of the collective mean is

$$\widehat{m} = X_{zzzw} = \sum_{i=1}^{I} \frac{z_i}{z_\Sigma} X_{izzw}. \tag{8}$$

Actually, this credibility weighted average is a pseudo-estimator since it depends on the unknown parameters s^2, a, b and c. Assuming these parameters known, X_{zzzw} is the best linear unbiased estimator of m (DeVylder, 1978). In practice, the unknown parameters are replaced by their respective estimator. One can check by simulation that the bias of X_{zzzw} is negligible and that its variance is indeed smaller than the variance of the more intuitive weighted average

$$X_{wwww} = \sum_{i=1}^{I} \sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} \frac{w_{ijk\Sigma}}{w_{\Sigma\Sigma\Sigma\Sigma}} X_{ijkw}.$$

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

3.2. Estimator for s^2

The estimator of s^2 is

$$\hat{s}^2 = \frac{1}{\sum_i \sum_j \sum_k \left(n_{ijk} - 1\right)} \sum_{i=1}^{I} \sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} \sum_{t=1}^{n_{ijk}} w_{ijkt} (X_{ijkt} - X_{ijkw})^2. \tag{9}$$

It is straightforward to verify that the estimator is unbiased. Furthermore, DeVylder and Goovaerts (1992) and Goulet (1998a) showed that $\hat{s}^2$ is optimal within a large class of linear estimators under zero-excess assumptions. We do not discuss this estimator any further in this paper.

3.3. Estimators for a, b and c

It seems that the first estimators of the between (or heterogeneity) variance components in common use in hierarchical credibility were the iterative pseudo-estimators found in Goovaerts and Hoogstad (1987) and Goovaerts et al. (1990):

$$\tilde{a} = \frac{1}{\sum_i \sum_j \left(K_{ij} - 1\right)} \sum_{i=1}^{I} \sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} z_{ijk} (X_{ijkw} - X_{ijzw})^2 \tag{10}$$

$$\tilde{b} = \frac{1}{\sum_i \left(J_i - 1\right)} \sum_{i=1}^{I} \sum_{j=1}^{J_i} z_{ij} (X_{ijzw} - X_{izzw})^2 \tag{11}$$

$$\tilde{c} = \frac{1}{I - 1} \sum_{i=1}^{I} z_i (X_{izzw} - X_{zzzw})^2. \tag{12}$$

These estimators look simple and fairly intuitive. However, by using as weights in the estimator of the heterogeneity parameter of a level the credibility factors of that same level, parameters a, b and c end up on the right hand sides of (10)-(12). Consequently, the estimators must be evaluated iteratively by solving a three-dimensional fixed point problem.

To avoid this somewhat involved numerical evaluation scheme, Ohlsson (2005) and Bühlmann and Gisler (2005) proposed two sets of estimators that are small variations on the same idea. It appears these authors made their findings almost simultaneously and independently. However, when Ohlsson presented his work at the ASTIN Colloquium in Zurich, the book by Bühlmann and Gisler was just about to be released. Consequently, Ohlsson's paper was never published in a scientific journal.

The common idea of Ohlsson, Bühlmann and Gisler is to use weights that are fully known upon computing the estimator of a heterogeneity parameter and, yet, to still obtain simple estimators. Essentially, the trick consists in replacing the nodes credibility factors wherever they are used in the iterative

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

estimators (10)-(12) by the nodes volumes; that is, to replace the values in the first column of (2)-(4) by those in the second.

To define the alternative sets of estimators, we first let

$$A_{ij} = \sum_{k=1}^{K_{ij}} w_{ijk\Sigma}\,(X_{ijkw} - X_{ijww})^2 - (K_{ij} - 1)\,s^2, \tag{13}$$

$$B_i = \sum_{j=1}^{J_i} z_{ij\Sigma}\,(X_{ijzw} - \bar{X}_{izzw})^2 - (J_i - 1)\,a, \tag{14}$$

$$C = \sum_{i=1}^{I} z_{i\Sigma}\,(X_{izzw} - \bar{X}_{zzzw})^2 - (I - 1)\,b \tag{15}$$

and

$$d_{ij} = w_{ij\Sigma\Sigma} - \sum_{k=1}^{K_{ij}} \frac{w_{ijk\Sigma}^2}{w_{ij\Sigma\Sigma}},\quad e_i = z_{i\Sigma\Sigma} - \sum_{j=1}^{J_i} \frac{z_{ij\Sigma}^2}{z_{i\Sigma\Sigma}},\quad f = z_{\Sigma\Sigma} - \sum_{i=1}^{I} \frac{z_{i\Sigma}^2}{z_{\Sigma\Sigma}}, \tag{16}$$

with

$$X_{ijww} = \sum_{k=1}^{K_{ij}} \frac{w_{ijk\Sigma}}{w_{ij\Sigma\Sigma}}\,X_{ijkw}, \tag{17}$$

$$\bar{X}_{izzw} = \sum_{j=1}^{J_i} \frac{z_{ij\Sigma}}{z_{i\Sigma\Sigma}}\,X_{ijzw}, \tag{18}$$

$$\bar{X}_{zzzw} = \sum_{i=1}^{I} \frac{z_{i\Sigma}}{z_{\Sigma\Sigma}}\,X_{izzw}. \tag{19}$$

Notice how (18) and (19) differ from (7) and (8), respectively. The definitions of the terms $w_{ij\Sigma\Sigma}$, $z_{i\Sigma\Sigma}$ and $z_{\Sigma\Sigma}$ should be obvious from the notation introduced earlier.

One may verify that $E[A_{ij}] = d_{ij}a$, $E[B_i] = e_i b$ and $E[C] = fc$ when structure parameters are taken as known constants (see also Ohlsson, 2005). In other words, A_{ij}/d_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, J_i$ is an unbiased estimator of a, B_i/e_i, $i = 1, \ldots, I$ is an unbiased estimator of b and C/f is an unbiased estimator of c. However, each of these statistics can be negative. What differs in Ohlsson's and in Bühlmann and Gisler's approach is the solution to cope with negative variance estimators.

Ohlsson (2005) proposes the following unbiased estimators:

$$\hat{a} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J_i} A_{ij}}{\sum_{i=1}^{I}\sum_{j=1}^{J_i} d_{ij}} \tag{20}$$

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

$$\hat{b} = \frac{\sum_{i=1}^{I} B_i}{\sum_{i=1}^{I} e_i} \tag{21}$$

$$\hat{c} = \frac{C}{f}. \tag{22}$$

These estimators can be negative. In practice, the lower level variance components will be replaced by their estimates and $\hat{a}$, $\hat{b}$ and $\hat{c}$ will be truncated to zero, thereby introducing a bias.

As for Bühlmann and Gisler (2005), they truncate each unbiased statistic above *before* averaging them out. Therefore, their biased but always non-negative estimators are

$$\hat{a}' = \frac{1}{\sum_{i=1}^{I} J_i} \sum_{i=1}^{I} \sum_{j=1}^{J_i} \max\left(\frac{A_{ij}}{d_{ij}}, 0\right) \tag{23}$$

$$\hat{b}' = \frac{1}{I} \sum_{i=1}^{I} \max\left(\frac{B_i}{e_i}, 0\right) \tag{24}$$

$$\hat{c}' = \max\left(\frac{C}{f}, 0\right). \tag{25}$$

Of course, in computations one also needs to replace the lower level variance components by their estimates in the above formulas.

4. Numerical properties of the estimators

As one quickly realizes when trying to implement calculations for the hierarchical credibility model in a computer, the formulas of the previous section do not give the complete picture about the structure parameters estimators. Specifically, their numerical evaluation is problematic as soon as one of the (between) variance estimators is zero. Indeed, all the credibility factors of the affected level – and hence the volume of the level – are then also equal to zero. This will result in indeterminate forms ($\frac{0}{0}$) upon computing the sufficient statistics and the credibility factors of the levels higher in the tree.

In this section, we provide a few results and guidelines to help cope with such situations. Note that we do not discuss the case $s^2 = 0$ since it is of no real interest in practice. Indeed, contracts are very rarely *all* perfectly time homogeneous and, if they were, then predicting their future claims would be trivial.

A first lemma indicates how to compute the sufficient statistic of a level when the between variance of the previous level is zero.

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

Lemma 1. *For the hierarchical credibility model of Section 2, we have the following limits:*

$$\lim_{a \to 0} X_{ijzw} = X_{ijww}$$

$$\lim_{b \to 0} X_{izzw} = \bar{X}_{izzw}$$

$$\lim_{c \to 0} X_{zzzw} = \bar{X}_{zzzw}.$$

Proof. We prove only the first result; the others are symmetrical. If $a = 0$, then $z_{ijk} = 0$ for all i, j and k. Therefore, we have

$$\lim_{a \to 0} X_{ijzw} = \lim_{a \to 0} \sum_{k=1}^{K_{ij}} \frac{z_{ijk}}{z_{ij\Sigma}} X_{ijkw} = \left[\frac{0}{0}\right],$$

an indeterminate form. Now, since

$$\frac{d}{da} z_{ijk} = \frac{w_{ijk\Sigma} s^2}{\left(a w_{ijk\Sigma} + s^2\right)^2}, \tag{26}$$

applying l'Hôpital's rule to our original problem yields

$$\begin{aligned}
\lim_{a \to 0} X_{ijzw} &= \lim_{a \to 0} \sum_{k=1}^{K_{ij}} \frac{\frac{w_{ijk\Sigma} s^2}{\left(a w_{ijk\Sigma} + s^2\right)^2}}{\frac{w_{ijk\Sigma} s^2}{\left(a w_{ijk\Sigma} + s^2\right)^2}} X_{ijkw} \\
&= \sum_{k=1}^{K_{ij}} \frac{w_{ijk\Sigma}}{w_{ij\Sigma\Sigma}} X_{ijkw} \\
&= X_{ijww}.
\end{aligned}$$

□

The second lemma shows how to compute the credibility factors of a level if the between variance of the previous level – or within variance of the current level – is zero.

Lemma 2. *Under the same assumptions as in Lemma 1, we have*

$$\lim_{a \to 0} z_{ij} = \frac{w_{ij\Sigma\Sigma}}{w_{ij\Sigma\Sigma} + s^2/b}$$

$$\lim_{b \to 0} z_i = \frac{z_{i\Sigma\Sigma}}{z_{i\Sigma\Sigma} + a/c}.$$

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

Proof. The first result is immediate using l'Hôpital's rule with (26) and the second follows by symmetry. □

The results in the above two lemmas are fairly obvious when we look at the physical interpretation of a zero between variance. When a level in the classification tree is perfectly homogeneous, then this level may as well be omitted from the classification, as depicted in Figures 1(b)-1(d). This explains why the weights and within variances "skip" one level in the formulas for the sufficient statistics and the credibility factors.

Furthermore, we can apply the rules of Lemmas 1 and 2 repeatedly if more than one variance is zero. For example, if both $a = 0$ and $b = 0$, then the sufficient statistic for sector i is

$$X_{iwww} = \sum_{j=1}^{J_i} \frac{w_{ij\Sigma\Sigma}}{w_{i\Sigma\Sigma\Sigma}} X_{ijww} = \sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} \frac{w_{ijk\Sigma}}{w_{i\Sigma\Sigma\Sigma}} X_{ijkw}$$

and the credibility factor is

$$z_i = \frac{w_{i\Sigma\Sigma\Sigma}}{w_{i\Sigma\Sigma\Sigma} + s^2/c}.$$

Consider now the Ohlsson estimators (20)-(22) and the Bühlmann-Gisler estimators (23)-(25). These are computed from the top down in the order given. Again, a zero estimator will cause numerical difficulties when computing the next estimators. The following theorem shows how to correctly deal with such situations.

Theorem 1. *For the hierarchical credibility model of Section 2 and the quantities defined in* (13)-(15) *and* (16)*, we have:*

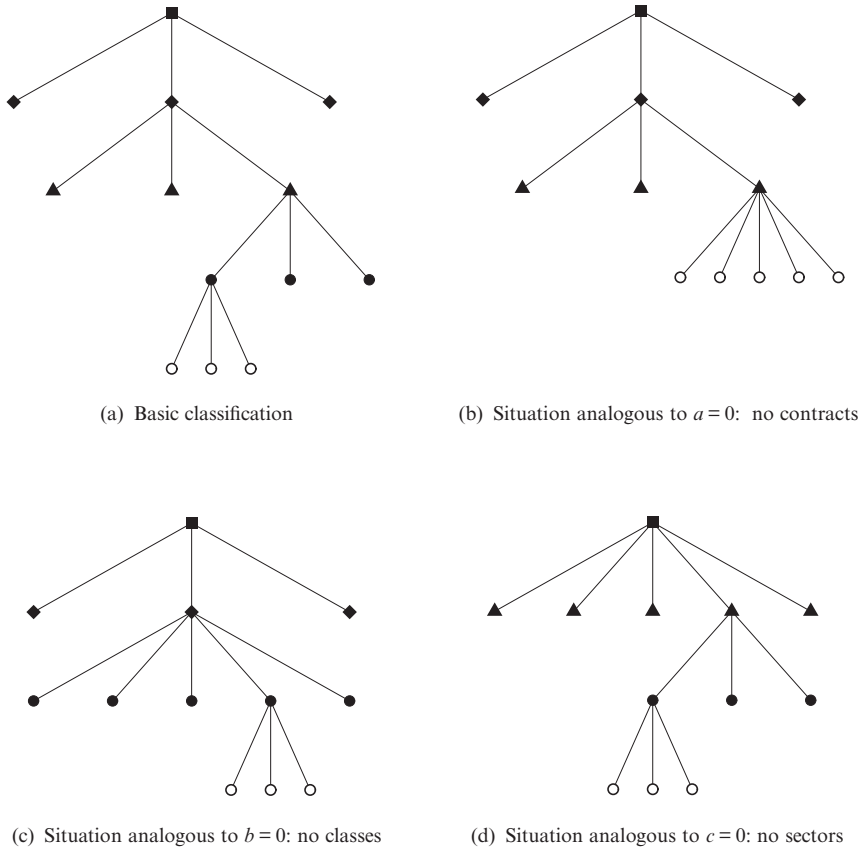
$$\lim_{a \to 0} \frac{B_i}{e_i} = \frac{\sum_{j=1}^{J_i} w_{ij\Sigma\Sigma} \left(X_{ijww} - X_{iwww} \right)^2 - (J_i - 1)\, s^2}{w_{i\Sigma\Sigma\Sigma} - \sum_{j=1}^{J_i} \frac{w_{ij\Sigma\Sigma}^2}{w_{i\Sigma\Sigma\Sigma}}} \tag{27}$$

$$\lim_{b \to 0} \frac{C}{f} = \frac{\sum_{i=1}^{I} z_{i\Sigma\Sigma} \left(\bar{X}_{izzw} - \bar{\bar{X}}_{zwww} \right)^2 - (I - 1)\, a}{z_{\Sigma\Sigma\Sigma} - \sum_{i=1}^{I} \frac{z_{i\Sigma\Sigma}^2}{z_{\Sigma\Sigma\Sigma}}}, \tag{28}$$

with

$$\bar{\bar{X}}_{zwww} = \sum_{i=1}^{I} \frac{z_{i\Sigma\Sigma}}{z_{\Sigma\Sigma\Sigma}} \bar{X}_{izzw}.$$

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press



(a) Basic classification

(b) Situation analogous to $a = 0$: no contracts

(c) Situation analogous to $b = 0$: no classes

(d) Situation analogous to $c = 0$: no sectors

FIGURE 1: Schematic representations of a three-level hierarchical classification structure. ■ Portfolio ♦ Sectors ▲ Classes ● Contracts ○ Data.

Proof. The results follow from lemmas 1 and 2 and some tedious but straightforward calculations. We merely outline the proof, here. For (27), we write the limit as

$$\lim_{a\to 0}\frac{B_i}{e_i} = \lim_{a\to 0}\frac{\sum_{j=1}^{J_i} z_{ij\Sigma}\left(X_{ijzw}-\bar{X}_{izzw}\right)^2}{z_{i\Sigma\Sigma}-\sum_{j=1}^{J_i}\frac{z_{ij\Sigma}^2}{z_{i\Sigma\Sigma}}} - \lim_{a\to 0}\frac{(J_i-1)\,a}{z_{i\Sigma\Sigma}-\sum_{j=1}^{J_i}\frac{z_{ij\Sigma}^2}{z_{i\Sigma\Sigma}}}$$

$$= \lim_{a\to 0}\frac{\sum_{j=1}^{J_i}\frac{z_{ij\Sigma}}{z_{i\Sigma\Sigma}}\left(X_{ijzw}-\bar{X}_{izzw}\right)^2}{1-\sum_{j=1}^{J_i}\left(\frac{z_{ij\Sigma}}{z_{i\Sigma\Sigma}}\right)^2} - \lim_{a\to 0}\frac{(J_i-1)\,a}{z_{i\Sigma\Sigma}-\sum_{j=1}^{J_i}\frac{z_{ij\Sigma}^2}{z_{i\Sigma\Sigma}}}.$$

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

After a lengthy application of l'Hôpital's rule, we obtain

$$\lim_{a \to 0} \frac{B_i}{e_i} = \frac{\sum_{j=1}^{J_i} w_{ij\Sigma\Sigma}\left(X_{ijww} - X_{iwww}\right)^2 - (J_i - 1)\, s^2}{w_{i\Sigma\Sigma\Sigma} - \sum_{j=1}^{J_i} \frac{w_{ij\Sigma\Sigma}^2}{w_{i\Sigma\Sigma\Sigma}}}. \tag{29}$$

The proof for (28) is symmetrical in every respect. □

From the above, it should be obvious that $\lim_{a \to 0} \hat{b} = \lim_{a \to 0} B_i / e_i$ is equal to the right hand side of (29) with an additional summation on i in both the numerator and the denominator. Moreover, as before, the results of the previous theorem can be used repeatedly if more than one variance component is zero. For example, if both $\hat{a} = 0$ and $\hat{b} = 0$, then

$$\hat{c} = \frac{\sum_{i=1}^{I} w_{i\Sigma\Sigma\Sigma}\left(X_{iwww} - X_{wwww}\right)^2 - (I - 1)\, \hat{s}^2}{w_{\Sigma\Sigma\Sigma\Sigma} - \sum_{i=1}^{I} \frac{w_{i\Sigma\Sigma\Sigma}^2}{w_{\Sigma\Sigma\Sigma\Sigma}}}.$$

This discussion extends naturally to the Bühlmann-Gisler estimators.

As a final result, the following theorem establishes a formal link between the Ohlsson estimators and the existence and uniqueness of a solution with the iterative estimators.

Theorem 2. *Let $\tilde{a}$ be the solution of $a = f_1(\mathbf{X}, s^2, a)$, where $\mathbf{X}$ is the vector of observations from the portfolio and f_1 is the right hand side of* (10)*. The sequence $\tilde{a}^{(n+1)} = f_1(\mathbf{X}, s^2, \tilde{a}^{(n)})$, $n = 0, 1, 2, \ldots$ converges to a unique, strictly positive solution if, and only if, $\Sigma_i \Sigma_j A_{ij} > 0$. When $\Sigma_i \Sigma_j A_{ij} \leq 0$, the sequence converges to* 0*.*

We have equivalent relationships between convergence of the sequence $\tilde{b}^{(n+1)} = f_2(\mathbf{X}, a, \tilde{b}^{(n)})$ and the sign of $\Sigma_i B_i$ as well as between convergence of the sequence $\tilde{c}^{(n+1)} = f_3(\mathbf{X}, b, \tilde{c}^{(n)})$ and the sign of C, where f_2 and f_3 are the right-hand sides of (11) *and* (12)*, respectively.*

Proof. This theorem is a generalization to the hierarchical model of the results previously proved by Dubey and Gisler (1981) for the Bühlmann-Straub model. Hence, the proof follows the same lines. We give the main ideas, but leave to the reader to work out the details.

For the first relationship, one establishes easily that $f_1(\mathbf{X}, s^2, 0) = 0$, $f_1'(\mathbf{X}, s^2, a) > 0$ and $f_1''(\mathbf{X}, s^2, a) < 0$ (where the derivatives are with respect to a). Therefore, f_1 is a monotone increasing, concave downward function of a. The function will have a unique, strictly positive fixed point if, and only if, it starts off from the origin above the line $y = x$, that is if $f_1'(\mathbf{X}, s^2, a) > 1$. One can show that this condition is equivalent to $\Sigma_i \Sigma_j A_{ij} > 0$. On the other hand, if $f_1'(\mathbf{X}, s^2, 0) < 1$ the function always remains below the line $y = x$ and, consequently,

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

does not have a fixed point other than $a = 0$. This condition is equivalent to $\Sigma\Sigma_i \Sigma_j A_{ij} \leq 0$.

The proof for the other relationships is similar. □

Note that the above theorem guarantees existence and uniqueness of the iterative estimators when computed *one at a time*. In practice, the estimators are computed simultaneously. Therefore, what is really needed is a statement on the convergence to a unique multivariate fixed point of the sequence

$$\begin{bmatrix} \tilde{a}^{(n+1)} \\ \tilde{b}^{(n+1)} \\ \tilde{c}^{(n+1)} \end{bmatrix} = \begin{bmatrix} f_1\left(\mathbf{X}, s^2, \tilde{a}^{(n)}\right) \\ f_2\left(\mathbf{X}, \tilde{a}^{(n)}, \tilde{b}^{(n)}\right) \\ f_3\left(\mathbf{X}, \tilde{b}^{(n)}, \tilde{c}^{(n)}\right) \end{bmatrix}$$

or, what is actually used in practice,

$$\begin{bmatrix} \tilde{a}^{(n+1)} \\ \tilde{b}^{(n+1)} \\ \tilde{c}^{(n+1)} \end{bmatrix} = \begin{bmatrix} f_1\left(\mathbf{X}, s^2, \tilde{a}^{(n)}\right) \\ f_2\left(\mathbf{X}, \tilde{a}^{(n+1)}, \tilde{b}^{(n)}\right) \\ f_3\left(\mathbf{X}, \tilde{b}^{(n+1)}, \tilde{c}^{(n)}\right) \end{bmatrix}.$$

Unfortunately, we could not prove existence and uniqueness of the fixed point for this much more complex problem. However, we may testify that obtaining convergence in our many numerical tests was never a problem.

This completes our study of the inner working of the hierarchical model and its structure parameters estimators. For the sake of simplicity and brevity, from now on we can limit ourselves to two-level models in numerical illustrations.

5. Relative performance of the estimators

We wanted to assess by simulation the relative performance of the three sets of estimators introduced in the previous section. From the outset, we can state that our conclusion will not be a very firm one. Indeed, no one set of estimators performed uniformly better or worse than the others for the criteria we considered. Actually, the results were often so close that small variations in two simulations with exactly the same parameters would sometimes yield different conclusions.

We simulated aggregate claim amounts S_{ijt} for a two-level hierarchical portfolio using the collective model of risk theory (Gerber, 1979; Daykin et al., 1994) and then we defined the observation in node (i, j, t) as the ratio

$$X_{ijt} = \frac{S_{ijt}}{w_{ijt}}.$$

See the appendix for the complete simulation model.

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

TABLE 1

BIAS AND VARIANCE OF STRUCTURE PARAMETERS ESTIMATORS IN A SIMULATION WITH A FREQUENCY COMPONENT ONLY. THE TRUE VALUES OF THE PARAMETERS ARE $m = 1$, $a = 0.75$ AND $b = 2$.

Estimators	*m*		*a*		*b*	
	bias	**variance**	**bias**	**variance**	**bias**	**variance**
Ohlsson	–0.005217	0.200233	–0.0015	0.6418	0.0032	6.0708
B.-Gisler	–0.005224	0.200228	0.0084	0.6396	0.0031	6.0711
Iterative	–0.005198	0.200236	–0.0015	0.6383	0.0020	5.9573

TABLE 2

BIAS AND VARIANCE OF STRUCTURE PARAMETERS ESTIMATORS IN A SIMULATION WITH FREQUENCY AND SEVERITY COMPONENTS. THE TRUE VALUES OF THE PARAMETERS ARE $m = 1$, $a = 19.55$ AND $b = 7.15$.

Estimators	*m*		*a*		*b*	
	bias	**variance**	**bias**	**variance**	**bias**	**variance**
Ohlsson	0.001831	0.7622	–0.07	18,548	0.20	2,224
B.-Gisler	0.001845	0.7620	0.85	20,509	0.20	2,220
Iterative	0.002413	0.7629	–0.41	13,776	0.23	2,252

We ran a number of simulations with different sets of parameters. Each simulation consisted of 10,000 runs with portfolios of 10 sectors, between 100 and 150 contracts per sector (determined randomly) and 5 years of experience. We used the (truncated) Ohlsson estimators as starting values for the iterative estimators. The latter always converged in less than 100 iterations. We report the results for two representative cases in Tables 1 and 2. The credibility model is applied to frequencies only in Table 1 and to aggregate claim amounts in Table 2; see the appendix for details.

First, we notice that the bias of the estimators in all three sets is not a major source of concern. In most cases, when one estimator is better than the others in this respect, it is by a very small margin. Similarly, the variance of the various estimators of the same parameter are usually comparable. In our tests, the Ohlsson estimators are rarely the best performers. Otherwise, the Bühlmann-Gisler and iterative estimators are essentially on par, perhaps with a small advantage for the latter when it comes to the variance of the estimators.

In the end, the key aspect for the actuary using hierarchical credibility for ratemaking purposes is how accurate the model is at predicting future claims. Table 3 gives the average over all simulation runs of the contract mean square errors

$$\frac{1}{\sum_i J_i} \sum_{i=1}^{I} \sum_{j=1}^{J_i} \left(\hat{\pi}_{ij} - X_{ij,n+1} \right)^2.$$

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

TABLE 3

AVERAGE MEAN SQUARE ERRORS (SCALED TO 1.0) FOR THE THREE SETS OF STRUCTURE PARAMETERS ESTIMATORS AND THE TWO MODELS OF TABLES 1 AND 2.

Estimators	Mean square errors	
	Frequency only model	Frequency-severity model
Ohlsson	1.000056	1.0000
B.-Gisler	1.000083	1.0004
Iterative	1.000000	1.0001

The results are scaled to 1.0 for comparison purposes. Here again, the differences are small. However, in our tests the iterative estimators had most often the smallest mean square error.

Now, if we are to draw any conclusion from our simulation study, it is that the Bühlmann-Gisler and iterative estimators seem generally superior to the Ohlsson estimators, but by a minuscule margin. The main advantage of the Bühlmann-Gisler (and Ohlsson) estimators is their numerical simplicity. They are easy to compute by hand – in case anyone would like to do it, which we doubt! – and to implement in a programming language or a spreadsheet application. This cannot be said of the iterative estimators. On the other hand, in our implementation (see below), calculations for the iterative estimators are coded in `C` versus in `R` for the other two sets. This results in almost no speed penalty for using the iterative estimators.

6. AN IMPLEMENTATION

In this section, we give a brief outline of our implementation of hierarchical credibility calculations in the `R` package **actuar** (Dutang et al., 2008). **actuar** is a package providing additional Actuarial Science functionality to the `R` statistical system (R Development Core Team, 2009). The software is distributed through the Comprehensive R Archive Network (http://cran.r-project.org).

The linear model fitting function of `R` is named `lm`. Since credibility models are very close in many respects to linear models, and since the credibility model fitting function of **actuar** borrows much of its interface from `lm`, we named the credibility function `cm`.

Function `cm` acts as a unified interface for all credibility models supported by **actuar**. Currently, these are the classical models of Bühlmann (1969) and Bühlmann and Straub (1970), the hierarchical model and the regression model of Hachemeister (1975). Furthermore, the function supports all three sets of estimators of the between variance structure parameters presented in Section 3.3.

The credibility modeling function assumes that data is available in the format most practical applications would use, namely a rectangular array (matrix

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

or data frame) with contract observations in the rows and with one or more classification index columns (numeric or character).

Then, function `cm` works much the same as `lm`. It takes in argument: a formula of the form `~ terms` describing the hierarchical interactions in a data set, the data set containing the variables referenced in the formula, and the names of the columns where the ratios and the weights are to be found in the data set. The latter should contain at least two nodes in each level and more than one period of experience for at least one contract. Missing values are represented by `NA`s. There can be contracts with no experience (complete lines of `NA`s).

In order to give a simple and easily reproducible example, we group states 1 and 3 of the Hachemeister (1975) data set (provided with the package) into one sector and states 2, 4 and 5 into another:

```
> library ("actuar")
> data(hachemeister)
> X <- cbind(sector = c(1,2,1,2,2), hachemeister)
```

The above shows that data does not have to be sorted by level. The fitted model using the iterative estimators is:

```
> fit <- cm(~sector + sector:state, data = X,
+   ratios = ratio.1:ratio.12, weights = weight.1:weight.12,
+   method = "iterative")
> fit

Call:
cm(formula = ~sector + sector:state, data = X,
   ratios = ratio.1:ratio.12, weights = weight.1:weight.12,
   method = "iterative")

Structure Parameters Estimators

   Collective premium: 1746

   Between sector variance: 88981
   Within sector/Between state variance: 10952
   Within state variance: 139120026
```

To fit with the Bühlmann-Gisler or the Ohlsson estimators, one merely specify `method = "Bühlmann-Gisler"` or `method = "Ohlsson"`, respectively, in the call to `cm`.

In the `R` expression above, we stored the model object returned by `cm` in an object named `fit` for further use. This object of class "`cm`" contains the estimators of the structure parameters. To compute the credibility premiums, one calls a method of `predict` for this class:

```
> predict(fit)
$sector

[1] 1949 1543
```

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

```
$state
[1] 2048 1524 1875 1497 1585
```

One can also obtain a nicely formatted view of the most important results of the fit with a call to `summary`:

```
> summary(fit)

Call:
cm(formula = ~sector + sector:state, data = X,
   ratios = ratio.1:ratio.12, weights = weight.1:weight.12,
   method = "iterative")

Structure Parameters Estimators

   Collective premium: 1746

   Between sector variance: 88981
   Within sector/Between state variance: 10952
   Within state variance: 139120026

Detailed premiums

Level: sector
 sector   mean Indiv.    Weight    Cred. factor    Cred. premium
    1         1967        1.407       0.9196           1949
    2         1528        1.596       0.9284           1543

Level: state
 sector state  mean Indiv.  Weight   Cred. factor  Cred. premium
    1      1       2061     100155     0.8874         2048
    2      2       1511      19895     0.6103         1524
    1      3       1806      13735     0.5195         1875
    2      4       1353       4152     0.2463         1497
    2      5       1600      36110     0.7398         1585
```

The methods of `predict` and `summary` can both report for a subset of the levels by means of an argument `level`. For example:

```
> predict(fit, levels = "sector")

sector [1]
1949 1543
```

The results above differ from those of Goovaerts and Hoogstad (1987) for the same example because the formulas for the credibility premiums are slightly different.

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

7. Conclusion

This paper discussed estimation of the structure parameters in the hierarchical credibility model, with an emphasis on between variance components. We reviewed in a unified notation three sets of estimators presented in the actuarial literature and we provided rules to compute credibility premiums when one or more of the variance component estimators are equal to zero. We also demonstrated a link between the values of the Ohlsson estimators and the convergence of the iterative ones.

A simulation study showed that all three sets of estimators perform very similarly. For their numerical simplicity, we recommend using the Bühlmann-Gisler estimators. If computing time is not much of an issue, then the iterative estimators are an excellent alternative.

Finally, we presented function `cm` of the `R` package **actuar**. This function is the package's unified interface to credibility modeling.

If you use `R` or **actuar** for actuarial analysis, please cite the software in publications. Use

```
> citation()
```

or

```
> citation("actuar")
```

at the `R` command prompt for information on how to cite the software.

Acknowledgments

The first implementation of the iterative estimators in `cm` was done by Louis-Philippe Pouliot. We thank Alois Gisler for his help and stimulating comments during the development of this project. We also thank two anonymous referees for improvements to the paper.

This research benefited from financial support from the Natural Sciences and Engineering Research Council of Canada and from the *Chaire d'actuariat* (Actuarial Science Chair) of Université Laval.

References

Bühlmann, H. (1967) Experience rating and credibility. *ASTIN Bulletin* **4**, 199-207.

Bühlmann, H. (1969) Experience rating and credibility. *ASTIN Bulletin* **5**, 157-165.

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

Bühlmann, H. and Jewell, W. S. (1987) Hierarchical credibility revisited. Bulletin of the Swiss *Association of Actuaries* **87**, 35-54.

Bühlmann, H. and Straub, E. (1970) Glaubwürdigkeit für [S]chadensätze. *Bulletin of the Swiss Association of Actuaries* **70**, 111-133.

Dannenburg, D.R., Kaas, R. and Goovaerts, M.J. (1996) *Practical Actuarial Credibility Models*. Ceuterick, Leuven.

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

DAYKIN, C., PENTIKÄINEN, T. and PESONEN, M. (1994) *Practical Risk Theory for Actuaries*. Chapman & Hall, London.
DE VYLDER, F. (1978) Parameter estimation in credibility theory. *ASTIN Bulletin* **10**, 99-112.
DE VYLDER, F. and GOOVAERTS, M.J. (1992) Optimal parameter estimation under zero-excess assumptions in the Bühlmann-Straub model. *Insurance: Mathematics and Economics* **11**, 167-171.
DUBEY, A. and GISLER, A. (1981) On parameter estimation in credibility. *Bulletin of the Swiss Association of Actuaries* **81**, 187-211.
DUTANG, C., GOULET, V. and PIGEON, M. (2008) **actuar**: An R package for actuarial science. *Journal of Statistical Software* **25(7)**.
URL http://www.jstatsoft.org/v25/i07
FORGUES, A., GOULET, V. and LU, J. (2006) Credibility for severity revisited. *North American Actuarial Journal* **10(1)**, 49-62.
GERBER, H.U. (1979) *An Introduction to Mathematical Risk Theory*. Huebner Foundation, Philadelphia.
GOOVAERTS, M.J. and HOOGSTAD, W.J. (1987) *Credibility Theory. No. 4 in Surveys of actuarial studies*. Nationale-Nederlanden N.V., Netherlands.
GOOVAERTS, M.J., KAAS, R., VAN HEERWAARDEN, A.E. and BAUWELINCKX, T. (1990) *Effective Actuarial Methods*. North-Holland, Amsterdam.
GOULET, V. (1998a) A note on optimal parameter estimation under zero-excess assumptions. *Insurance: Mathematics and Economics* **23**, 111-117.
GOULET, V. (1998b) Principles and application of credibility theory. *Journal of Actuarial Practice* **6**, 5-62.
GOULET, V. and POULIOT, L.-P. (2008) Simulation of compound hierarchical models in R. *North American Actuarial Journal* **12**, 401-412.
HACHEMEISTER, C.A. (1975) *Credibility for regression models with application to trend*. In: Credibility, theory and applications. Proceedings of the Berkeley actuarial research conference on credibility. Academic Press, New York.
HEWITT, C.C. (1971) Credibility for severity. *Proceedings of the Casualty Actuarial Society* **57**, 148-171.
JEWELL, WS. (1975) The use of collateral data in credibility theory: A hierarchical model. *Giornale dell'Istituto Italiano degli Attuari* **38**, 1-16.
OHLSSON, E., 2005. Simplified estimation of structure parameters in hierarchical credibility, presented at the Zurich ASTIN Colloquium.
URL http://www.actuaries.org/ASTIN/Colloquia/Zurich/Ohlsson.pdf
R DEVELOPMENT CORE TEAM (2009) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL http://www.r-project.org

APPENDIX

A. SIMULATION MODEL

The simulation model we use in the small study of Section 5 is the following. First, we simulate aggregate claim amounts S_{ijt} for a two-level hierarchical portfolio. Then, we define the observation of period $t = 1, \ldots, n_{ij}$ for contract $j = 1, \ldots, J_i$ of sector $i = 1, \ldots, I$ as the ratio

$$X_{ijt} = \frac{S_{ijt}}{w_{ijt}}. \tag{30}$$

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

We let $S_{ijt} = C_{ijt1} + \ldots + C_{ijtN_{ijt}}$, where N_{ijt} is the number of claims in node (i, j, t) and C_{ijtu} is the amount of claim u in that same node. As usual in the collective model of risk theory, the claim amounts are all mutually independent and independent from the number of claims.

If m, s^2, a and b are the structure parameters based on (30), then we have

$$\begin{aligned} m &= m_N m_C, \\ s^2 &= s_N^2(a_C + m_C^2) + s_C^2 m_N, \\ a &= a_N a_C + a_N m_C^2 + a_C m_N^2, \end{aligned}$$

where subscript N denotes the structure parameters based on N_{ijt}/w_{ijt} and subscript C denotes the structure parameters based on C_{ijtu} (see Forgues et al., 2006).

We use the following distribution assumptions:

$$\begin{aligned} N_{ijt}|\Lambda_{ij}, \Psi_i &\sim \text{Poisson}(w_{ijt}\Lambda_{ij}) \\ \Lambda_{ij}|\Psi_i &\sim \text{Gamma}(\alpha, 1/\Psi_i) \\ \Psi_i &\sim \text{Gamma}(\eta, \gamma) \end{aligned}$$

and

$$\begin{aligned} C_{ijtu}|\Theta_{ij}, \Phi_i &\sim \text{Lognormal}(\Theta_{ij}, \sigma_1^2) \\ \Theta_{ij}|\Phi_i &\sim \text{Normal}(\Phi_i, \sigma_2^2) \\ \Phi_i &\sim \text{Normal}(\mu, \sigma_3^2). \end{aligned}$$

The Poisson-gamma combination for frequencies is a typical choice in Bayesian modeling. The lognormal-normal combination for claim amounts was inspired by Hewitt (1971). These assumptions yield

$$\begin{aligned} m_N &= \frac{\alpha\eta}{\gamma}, & m_C &= e^{\mu + (\sigma_1^2 + \sigma_2^2 + \sigma_3^2)/2}, \\ s_N^2 &= \frac{\alpha\eta}{\gamma}, & s_C^2 &= e^{2\mu + \sigma_1^2 + 2(\sigma_2^2 + \sigma_3^2)}(e^{\sigma_1^2} - 1), \\ a_N &= \frac{\alpha\eta(n+1)}{\gamma^2}, & a_C &= e^{2\mu + \sigma_1^2 + \sigma_2^2 + 2\sigma_3^2}(e^{\sigma_2^2} - 1), \\ b_N &= \frac{\alpha^2\eta}{\gamma^2}, & b_C &= e^{2\mu + \sigma_1^2 + \sigma_2^2 + \sigma_3^2}(e^{\sigma_3^2} - 1). \end{aligned}$$

The weights are simulated with

$$\begin{aligned} w_{ijt} &\sim U(0.5\bar{w}_{ij}, 1.5\bar{w}_{ij}) \\ \bar{w}_{ij} &\sim U(2, 10). \end{aligned}$$

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press

This two-step procedure avoids too large fluctuations in weights within a contract.

The model of Table 1 has only a frequency component – that is $\mu = \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 0$ – and we simulated observations with parameters $\alpha = 4$, $\eta = \frac{1}{2}$ and $\gamma = 2$. This gives a fairly well classified portfolio, with sectors homogeneous within (small a) and heterogeneous between (large b).

For the model of Table 2, we used the same frequency parameters as above and severity parameters $\mu = \log 2 - \frac{3}{2}$ and $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$. In part, these parameters were chosen such that the collective mean is $m = 1$.

Data was simulated using function `simul` of the `R` package **actuar** (Goulet and Pouliot, 2008, see also Section 6).

HASSINE BELHADJ
École d'actuariat
Pavillon Alexandre-Vachon
1045, avenue de la Médecine
Université Laval
Québec (QC) G1V 0A6
Canada
Email: hassine.belhadj.1@ulaval.ca

VINCENT GOULET
École d'actuariat
Pavillon Alexandre-Vachon
1045, avenue de la Médecine
Université Laval
Québec (QC) G1V 0A6
Canada
Email: vincent.goulet@act.ulaval.ca

TOMMY OUELLET
Morneau Sobeco
Tour de la Bourse – Place Victoria
800 rue du Square-Victoria, bureau 4000
C.P. 211
Montréal (QC) - H4Z 0A4
Email: touellet@morneausobeco.com

https://doi.org/10.2143/AST.39.2.2044645 Published online by Cambridge University Press