Credibility pseudo-estimators

STIG ROSENLUND*

Länsförsäkringar Alliance, Aktuarie Sak (retired), Stockholm, Sweden

We treat a model with independent claim numbers and claim amounts, conditional on stochastic parameters. Groups are categorized into a smaller number of classes, which likely differ in risk premium. The collective claim frequency and mean claim for a group are modeled as those of the class the group belongs to. For each group we find the Best Linear Predictor (BLP), also known as Credibility Estimator, in a generic model covering claim frequency and mean claim, as a weighted mean of the group's individual estimate and the collective estimate. Assuming Poisson distributed claim numbers and some distributional properties of claim amounts, we find estimators of variance components, estimation errors of the collective claim frequency and mean claim and covariances between observations, estimators and stochastic parameters. Pseudo-estimators, i.e. estimators which are defined by expressions that contain them and which must be solved numerically, are given for between-groups variance components. Simulation results, where some of the assumptions are violated, indicate when they are preferable over non-pseudo-estimators.

Keywords: Auxiliary argument; Best Linear Predictor; Credibility Estimator; Pseudo-estimator; Variance component

1. Introduction and summary of results

In tariff analysis we use the term multi-level factor (MLF) for a rating factor, where some classes have too few claims to admit basing the premium on the class alone. An example is geographical parish, when there are several thousands of those. Credibility analysis should be used for this argument. To distinguish the MLF from arguments (rating factors) with sufficiently many claims in each class, we call a class in it a group.

We have a prior categorizing of groups into a smaller number of classes by some property, which we have prior reasons to believe affects the risk premium. An example of groups is geographical parishes with the property population density in five classes, where higher population density likely implies higher risk premium. We call the classes of this property an auxiliary argument, or Auxiliary. We use the term of Ohlsson & Johansson (2010), Section 4.2.3, page 87. These authors treat there the same setting as we do.

The input is claims and exposures for some time period. Best Linear Predictors[1] (BLPs) are deduced. Under an essentially Compound Poisson assumption and some suitable distributional assumptions for claim amounts we also derive estimators of between-groups variance components.

The paper is organized as follows. Section 2 recapitulates credibility models found in the literature. Section 3 summarizes and gives reasons for our model. Section 4 states the notation. Section 5 sets up a generic model covering both claim frequency and mean claim. Section 5.1 gives the BLP. We take account of how estimators' variances and covariances with observations and stochastic parameters (i.e.

*Corresponding author. E-mail: stig.ingvar.rosenlund@gmail.com

[1] What is predicted is the expected claim frequency or mean claim of a group, conditional on a random effect that occurred in the past and will not change. Thus the word predictor seems disingenuous. The actuarial tradition is to call the BLP the Credibility Estimator, but the word estimator for a random variable, such as this conditional expectation, is also disingenuous. See Ohlsson & Johansson (2010), Section 4.1, page 75, for a discussion of terminology.

In Scandinavian Actuarial Journal online 2018, excluding Appendix 3.

random effects) affect the BLP, thereby arriving at a possibly new and more exact expression than has been known so far. Section 5.2 gives a pseudo-estimator for the variance component between groups, which is optimal under certain conditions. A pseudo-estimator is one that is defined by an expression that contains the estimator itself, which thus must be found by numerical root finding. In Section 6 the specifics of claim frequency are treated under the assumption of Poisson distributed claim numbers. Section 7 treats the specifics of mean claim under assumptions of some distributional properties of claim amounts. In Section 8 the separate claim frequency and mean claim results are combined to risk premium results. Section 9 describes simulation results for the goodness of estimators of between-groups variance components, when some assumptions are violated to test robustness. Appendix 1 gives proofs. Appendix 2 gives tables from the simulations.

The free language Rapp for credibility by this paper and other methods, GLM for non-life insurance pricing, claim reserve algorithms, bignum multiprecision computing, etc., is found here: `www.stigrosenlund.se/rapp.htm`.

2. Overview of some credibility models

Bühlmann & Straub (1970) give the classical Bühlmann-Straub estimator for a non-parametric credibility model with one MLF and no other rating factors. Norberg (1980) treats best linear unbiased prediction in empirical Bayes credibility. Campbell (1986) combines the MLF with groupings of it by auxiliary arguments which are functions of the MLF, e.g. median income and population density for parishes, and weight and power for car models. The grouping is made by cluster analysis. In the example rendered, exposure is normalized duration. For the Bühlmann-Straub model De Vylder (1996) III, Chapter 3, Section 3.4.7, gives pseudo-estimators credited to Bichsel & Straub. Frees (2003) treats credibility with a multivariate approach to groupings into different lines of business. Overviews of credibility are found in Bühlmann & Gisler (2005), including the Poisson model for claim numbers and a pseudo-estimator for Pareto credibility, and in Kaas et al. (2009). Ohlsson (2008) and Ohlsson & Johansson (2010), Chapter 4, treat a setting with multiplicative rating factors including the MLF and auxiliary arguments, without a Poisson assumption. We cite Ohlsson & Johansson (2010) frequently below due to its use as a textbook in actuarial education, although the results might have been found by other authors before them.

3. Summary of and reasons for model

For some of our results we assume that, conditional on stochastic parameters (random effects), the claim cost of any group of the MLF is Compound Poisson distributed, with the slight weakening that individual claim amounts are independent and have the same mean and variance, not necessarily identically distributed. Therefore we write essentially Compound Poisson above.

We have available a categorization of the groups, called the Auxiliary. Given independent random effects per group with mean 1 and the same variance, the expected claim frequency or mean claim is the product of a factor specific to the auxiliary and the random effect.

See Rosenlund (2010) for reasons to use pure Poisson in tariff analysis rather than Overdispersed Poisson for claim numbers, even with macroscopic fluctuations affecting large parts of the portfolio in the same way. It is also a mistake to assume Negative Binomial – Mixed Poisson with a gamma mixing distribution – for claim numbers. While the Negative Binomial distribution is appropriate if we draw a customer at random from the portfolio, it is inappropriate if used for tariff analysis and predictions of next year's claim numbers. We should instead condition with respect to the different customers' mixing variables, whether we regard them as stochastic or not. In this paper we condition with respect to stochastic parameters, which were realized in the past and do not change. So we get the pure Poisson distribution. Negative Binomial would only be right if all customers were to leave at year-end and be replaced by a random sample of customers, distributed as this year's sample and independent of it. Normally only a fraction of the customers leaves each year, making the claim number distribution somewhat more dispersed than Poisson, but not like Negative Binomial. This thought experiment suffices to clarify the matter.

4. Notation for observables

We define the following observables. Subscripts F and M denote claim Frequency and Mean claim, respectively.

$$J = \text{number of groups,} \tag{4.1}$$

$$e_j = \text{exposure in group } j, \text{ assumed } > 0 \text{ for all } j \in \{1, \ldots, J\}, \tag{4.2}$$

$$N_j = \text{number of claims in group } j, \tag{4.3}$$

$$Z_{jr} = \text{individual claim amounts in group } j \in \{1, \ldots, J\},\ r \in \{1, \ldots, N_j\}, \tag{4.4}$$

$$Y_{\mathsf{F}j} = N_j/e_j = \text{observed claim frequency in group } j, \tag{4.5}$$

$$Y_{\mathsf{M}j} = \sum_{r=1}^{N_j} Z_{jr}/N_j = \text{observed mean claim in group } j, \tag{4.6}$$

$$k_j = \text{the class of group } j \text{ in the Auxiliary,} \tag{4.7}$$

$$e_k^{\mathsf{A}} = \sum_{j:k_j=k} e_j = \text{sum of exposure over class } k \text{ of the Auxiliary,} \tag{4.8}$$

$$N_k^{\mathsf{A}} = \sum_{j:k_j=k} N_j = \text{sum of claim numbers over class } k \text{ of the Auxiliary.} \tag{4.9}$$

With no Auxiliary we set $k_j \equiv 1$. Then e_1^{A} is the total exposure and N_1^{A} is the total number of claims.

5. Generic BLP and pseudo-estimator

A generic notation will be employed to deduce a BLP and a pseudo-estimator. Certain functionals and their estimators will then be specified for claim frequency and mean claim. This notation is as follows.

Generic	w_j	w_k^{A}	Y_j
Claim frequency	e_j	e_k^{A}	$Y_{\mathsf{F}j}$
Mean claim	N_j	N_k^{A}	$Y_{\mathsf{M}j}$

We condition on Nj and on $N_j > 0$ for mean claim. This conditioning is implicit and is not written out below. See the remarks in Section 8 for a justification.

Here w_j and w_k^{A} are generic exposures. For mean claim the claim numbers take the role of exposures. Y_j is the generic claim rate. We make the following assumptions.

Assumption 1
Conditional on stochastic variables $\Theta_j\,(j = 1, \ldots, J)$, with expectation $\mathrm{E}[\Theta_j] = 1$ and variance $\mathrm{Var}[\Theta_j] = \tau^2$, Y_j has expectation $\mu_{k_j}\Theta_j$, where μ_k is the expected claim rate of Auxiliary class k.

Assumption 2
$(\Theta_1, Y_1), \ \ldots\ , (\Theta_J, Y_J)$ are independent.

Objective. To predict $\mu_{k_j}\Theta_j$ as well as possible.

Here $\mu_{k_j} = \mathrm{E}[Y_j]$ and τ^2 is a between-groups variance component.

Let

$$\Lambda_j = \mu_{k_j}\Theta_j, \ \text{claim rate group } j \text{ conditional on } \Theta_j. \tag{5.1}$$

Define these functionals. (5.2) and (5.3) are used both for the BLP and the pseudo-estimator, while (5.4) and (5.5) are used only for the pseudo-estimator.

$$\begin{aligned} \nu_k^2 &= \frac{1}{{w_k^{\mathsf{A}}}^2} \sum_{j:k_j=k} w_j^2(\sigma_j^2 + \mu_k^2\tau^2), &(5.2)\\ \sigma_j^2 &= \mathrm{E}[\mathrm{Var}[Y_j \mid \Theta_j]],\ \text{within-groups variance component}, &(5.3)\\ \rho_j(\tau^2) &= \mathrm{Var}[(Y_j/\mu_{k_j} - 1)^2], &(5.4)\\ \alpha_j(\tau^2) &= \mathbf{1}_{\{w_j>0\}}\left(\sigma_j^2/\mu_{k_j}^2 + \tau^2\right)^2/\rho_j(\tau^2). &(5.5) \end{aligned}$$

Estimators are named as the corresponding functionals with a $\widehat{}$ above them. They are obtained by plugging estimators into the expressions above.

The most laborious work in deducing a new pseudo-estimator is establishing an estimator for $\rho_j(\tau^2)$ by (5.4). This is done in sections 6.1 and 7.1.

Define the estimator

$$\widehat{\mu}_k = \frac{1}{w_k^{\mathsf{A}}} \sum_{j:k_j=k} w_j Y_j. \tag{5.6}$$

It is immediate that $\mathrm{E}[\widehat{\mu}_k] = \mu_k$. It will be shown in Appendix 1.1 that

$$\nu_k^2 = \mathrm{Var}[\widehat{\mu}_k]. \tag{5.7}$$

5.1. BLP

We establish first a non-observable predictor of Λ_j in the form

$$\Lambda_j^* = z_j Y_j + (1 - z_j)\widehat{\mu}_{k_j} \tag{5.8}$$

for an optimal z_j. Having established this predictor, we obtain an observable estimated predictor in the form

$$\widehat{\Lambda}_j = \widehat{z}_j Y_j + (1 - \widehat{z}_j)\widehat{\mu}_{k_j}, \tag{5.9}$$

where $\widehat{z}_j$ is a suitable estimator of z_j.

We seek the best linear combination of the observations Y_j to predict $\mu_{k_j}\Theta_j$ in L^2-norm, i.e. the BLP or Credibility Estimator. It is shown in Appendix 1.1 to be in the form (5.8). That is, z_j is to be determined so that $\mathrm{E}[(\Lambda_j^* - \mu_{k_j}\Theta_j)^2]$ is minimized. Then Λ_j^* is the BLP. The following is proved in Appendix 1.1.

Theorem 5.1. *If Assumptions 1 and 2 hold, then Λ_j^* by (5.8) is the BLP of Λ_j if we set $z_j = 0$ when $w_j = 0$, otherwise*

$$z_j = \frac{\mu_{k_j}^2\tau^2 - \frac{w_j}{w_{k_j}^{\mathsf{A}}}\left(\sigma_j^2 + 2\mu_{k_j}^2\tau^2\right) + \nu_{k_j}^2}{\left(\sigma_j^2 + \mu_{k_j}^2\tau^2\right)\left(1 - \frac{2w_j}{w_{k_j}^{\mathsf{A}}}\right) + \nu_{k_j}^2}. \tag{5.10}$$

We obtain $\widehat{z}_j$ from z_j by replacing unknown functionals with suitable estimators. We name such an estimator as the corresponding functional with a $\widehat{}$ above it. For τ^2 an estimator named with a $\widetilde{}$ or an $*$ above it can also be used. We find $\widehat{\mu}_k$ above in (5.6). The remaining ones will be specified for claim frequency and mean claim later.

Some terms of (5.10) are part of the classical Bühlmann & Straub (1970) estimator, while other terms might be new. See Remark A1.

The exposure-weighted total of the non-observable predictors is easily shown to be unbiased. The replacement with estimators in the observable predictors can cause a bias. It is corrected by first computing the simple expression $\sum_{i=1}^J w_i Y_i / \sum_{i=1}^J w_i \widehat{\Lambda}_i$ and then multiply that to each $\widehat{\Lambda}_j$.

5.2. Pseudo-estimator

We give a τ^2-estimator that contains itself, i.e. a pseudo-estimator in the sense of Bichsel & Straub. See De Vylder (1996) III, Chapter 3, Section 3.3.7, for a description of a pseudo-estimator in a simple setting. Equation (5.11) must be solved numerically, and the solution is the pseudo-estimator. We have to find a zero of the rather complicated function of $\widehat{\tau}^2$ given by the left side of (5.11) minus the right side.

The reason for a pseudo-estimator is to be able to use inverse variance weighting of squared deviations of observations from means, so that a minimum variance estimator can be obtained. These inverse variances require the variance component itself, hence the need for numerical root finding. This is laid out in Appendix 1.2. Non-pseudo estimators cannot make use of such weighting. Sufficiently many groups and observations are needed for a pseudo-estimator to be better than a non-pseudo one, however.

See Appendix 1.3 for the solution. We define $\widehat{\tau}^2$ as the largest solution. It can be 0, indicating no variance between groups. We give a sufficient condition for a positive solution, but could not prove its uniqueness or that the condition is necessary.

The iterative procedure for pseudo-estimators described in the literature is not suitable for finding the solution. Bisection (interval halvings) or some faster more complicated method should be used.

We need the concept of excess $\mathrm{e}(\cdot)$ as defined in Cramér (1946), Chapter 15, Section 15.8, equation (15.8.2), and in De Vylder (1996) III, Chapter 2, Section 2.1.2. Namely

$$\mathrm{e}(X) = \frac{\mathrm{E}[(X-\mu)^4]}{\mathrm{E}[(X-\mu)^2]^2} - 3 \ \text{ for a random variable } X \text{ with } \mathrm{E}[X]=\mu.$$

We will assume that the 3:rd central moment and the excess of Θ_j are 0. These assumptions admit a relatively simple and mathematically consistent estimator, and they imply approximate optimality of the estimator for a large number J of groups. However, in Section 9 we study by simulations how the estimator performs when J is not so large and when these moment assumptions are not true.

Theorem 5.2. *Let Assumptions 1 and 2 be true. Let $\widehat{\nu}^2_{k_j}$, $\widehat{\sigma}^2_j$ and $\widehat{\alpha}_j$ be estimators, to be specified later, of the corresponding functionals in (5.2), (5.4) and (5.5). Define the pseudo-estimator*

$$\widehat{\tau}^2 = \sum_{j=1}^{J} \frac{\widehat{\alpha}_j(\widehat{\tau}^2)}{\widehat{\alpha}_1(\widehat{\tau}^2)+\cdots+\widehat{\alpha}_J(\widehat{\tau}^2)}\, \widehat{\tau}^2 \left[\left(\widehat{\sigma}^2_j + \widehat{\mu}^2_{k_j}\widehat{\tau}^2\right)\left(1-\frac{2w_j}{w^{\mathsf{A}}_{k_j}}\right)+\widehat{\nu}^2_{k_j}\right]^{-1}(Y_j-\widehat{\mu}_{k_j})^2. \tag{5.11}$$

It holds $\mathrm{E}[\widehat{\tau}^2]\approx\tau^2$. If $\mathrm{E}[(\Theta_j-1)^3]=0$ and $\mathrm{e}(\Theta_j)=0$, then $\widehat{\tau}^2$ is approximately optimal in the sense of having the smallest mean square error of estimators in the form of a linear combination of $(Y_j-\widehat{\mu}_{k_j})^2$.

6. Claim frequency specifics

Here we have $w_j = e_j$, $w^{\mathsf{A}}_k = e^{\mathsf{A}}_k$ and $Y_j = Y_{\mathsf{F}j}$. We use the first two ones below, while retaining the generic notation for the rest. In addition a specific function $v_{\mathsf{F}}(,)$ is introduced.

Assumption 3
Conditional on Θ_j, N_j is Poisson distributed.

The properties of the Poisson distribution entail the following.

$$\nu^2_k = \frac{1}{{e^{\mathsf{A}}_k}^2}\sum_{j:k_j=k} e^2_j(\mu_k/e_j+\mu^2_k\tau^2), \tag{6.1}$$

$$\sigma^2_j = \mu_{k_j}/e_j, \tag{6.2}$$

$$\alpha_j(\tau^2) = \left(1/[\mu_{k_j}e_j]+\tau^2\right)^2/\rho_j(\tau^2). \tag{6.3}$$

6.1. Variance of squared observation deviations

To use (5.11) we must have an estimator $\widehat{\alpha}_j(\widehat{\tau}^2)$, which requires an estimator of $\rho_j(\tau^2)$.

Lemma 6.1. *Let Assumptions 1, 2 and 3 be true and assume that* $\mathrm{E}[(\Theta_j - 1)^3] = 0$ *and* $\mathrm{e}(\Theta_j) = 0$. *Let*

$$v_{\mathsf{F}}(x, y) = y^3 + (7x + 2)y^2 + 4xy + 2x^2. \tag{6.4}$$

Then $\rho_j(\tau^2)$ *by (5.4) is*

$$\rho_j(\tau^2) = v_{\mathsf{F}}(\tau^2, \frac{1}{\mu_{k_j} e_j}). \tag{6.5}$$

As before, plug in estimators in (6.5) to get $\widehat{\rho}_j(\widehat{\tau}^2)$ and then $\widehat{\alpha}_j(\widehat{\tau}^2)$, using (6.3).

6.2. Non-pseudo-estimator

With N_0 = the total number of claims, define a non-pseudo-estimator by

$$\widetilde{\tau}^2 = \max\left(0, \frac{\sum_{j=1}^{J} \widehat{\mu}_{k_j} e_j (Y_j / \widehat{\mu}_{k_j} - 1)^2 - (J-1)}{N_0 - \sum_{j=1}^{J} \widehat{\mu}_{k_j}^2 e_j^2 / N_0}\right). \tag{6.6}$$

The adaptation of (4.27) in Ohlsson & Johansson (2010) into $\widetilde{\tau}^2$ by (6.6) is this. On page 82, firstly we set $p = 1$ and $\sigma^2 = \mu$, in accordance with the Poisson assumption. Secondly, we divide the expression by the square of an estimate of the base factor μ. This is necessary, since we deal with claim frequency estimates without specifying them into base factors and argument factors. Thirdly we truncate the estimator from below to 0. This will decrease its mean square error. One might think that this would entail a positive bias, but our simulations nevertheless showed a negative bias.

7. Mean claim specifics

Here we have $w_j = N_j$, $w_k^{\mathsf{A}} = N_k^{\mathsf{A}}$ and $Y_j = Y_{\mathsf{M}j}$. We use the first two ones below, while retaining the generic notation for the rest. In addition a specific function $v_{\mathsf{M}}(,,,,)$ is introduced.

We need the following assumptions. Assumption 4 implies the generic Assumption 1.

Assumption 4
Conditional on stochastic variables Θ_j ($j = 1, \ldots, J$), with expectation $\mathrm{E}[\Theta_j] = 1$ and variance $\mathrm{Var}[\Theta_j] = \tau^2$, for any specific j the Z_{jr} are independent with expectation $\mu_{k_j}\Theta_j$, where μ_k is the expected mean claim of Auxiliary class k.

Assumption 5
$\mathrm{Var}[Z_{jr} \mid \Theta_j] = \phi(\mu_{k_j}\Theta_j)^p$ for some $\phi > 0$ and $1 \le p \le 2$, with $\mathrm{E}[\Theta_j^p]$ independent of j.

Here Assumption 5 implies $\mathrm{Var}[Y_j \mid \Theta_j] = \phi(\mu_{k_j}\Theta_j)^p / N_j$. In Rosenlund (2014) we argued against such an assumption in ordinary non-credibility multiplicative tariff analysis. But when many groups have few claims we risk overparametrization without the assumption.

We fix p in Assumption 5 initially. Normally $p = 2$ is suitable by virtue of giving claim amounts a constant coefficient of variation (CV), conditional on Θ_j.

We refer to Ohlsson & Johansson (2010) for the following three expressions.

Define the functional

$$\sigma^2 = \phi \mathrm{E}[\Theta_j^p]. \tag{7.1}$$

Then it holds

$$\sigma_j^2 = \sigma^2 \mu_{k_j}^p / N_j. \tag{7.2}$$

The assumptions entail the following estimator.

$$\widehat{\sigma}^2 = \frac{\sum_{j=1}^{J} \widehat{\mu}_{k_j}^{-p} \sum_{r=1}^{N_j} (Z_{jr} - Y_j)^2}{\sum_{j=1}^{J} \mathbf{1}_{\{N_j \geq 2\}} (N_j - 1)}. \tag{7.3}$$

7.1. Variance of squared observation deviations

To be able to use (5.11) we must have an estimator $\widehat{\alpha}_j(\widehat{\tau}^2)$. It is much more complicated for mean claim than for claim frequency.

We make this assumption. For $t = 2$ it follows from Assumption 5 if $p = 2$.

Assumption 6
For $t = 2, 3, 4$ it holds $\mu_{k_j}^{-t}\, \mathrm{E}[(Z_{jr} - \mu_{k_j}\Theta_j)^t \mid \Theta_j] = \phi_t \Theta_j^t$ for some $\phi_t > 0$, with $\mathrm{E}[\Theta_j^t]$ independent of j.

The assumption implies Assumption 5 with $p = 2$. It is true if Θ_j are only random IID scale factors for IID claim amounts W_{jr} with $\mathrm{E}[W_{jr}] = 1$ such that $Z_{jr} = \Theta_j \mu_{k_j} W_{jr}$.

Here $\phi_t \Theta_j^t$ are the central moments of Z_{jr}/μ_{k_j} conditional on Θ_j, and ϕ_2 is ϕ in Assumption 5. We use estimators of $\phi_t \mathrm{E}[\Theta_1^t]$ based on the sample central moments for $t = 2$, 3, 4. This can be done without first or simultaneously estimate τ^2, which is an advantage. The standard credibility procedure to estimate $\phi \mathrm{E}[\Theta_j^p]$ uses $\widehat{\sigma}^2$ by (7.3), which is a combination of sample central moments of order 2. The advantage of this procedure extends to higher order central moments.

The following lemma is a counterpart to Lemma 6.1 for claim frequency. It uses functions of x and y, where x will take the value $\widehat{\tau}^2$ and y the value N_j.

Lemma 7.1. *Let Assumptions 1, 4 and 6 be true and assume that* $\mathrm{E}[(\Theta_j - 1)^3] = 0$ *and* $\mathrm{e}(\Theta_j) = 0$. *Set*

$$\gamma_t = \phi_t E[\Theta_1^t]. \tag{7.4}$$

Let the sample central moments per j be

$$m_{tj} = \frac{1}{N_j} \sum_{r=1}^{N_j} (Z_{jr} - Y_j)^t, \ (t = 2, 3, 4). \tag{7.5}$$

Set

$$\begin{aligned}
\widehat{\gamma}_{2j} &= \frac{N_j}{N_j - 1} \frac{m_{2j}}{\widehat{\mu}_{k_j}^2}, \\
\widehat{\gamma}_{3j} &= \frac{N_j^2}{(N_j - 1)(N_j - 2)} \frac{m_{3j}}{\widehat{\mu}_{k_j}^3}, \\
\widehat{\gamma}_{4j} &= \frac{N_j(N_j^2 - 2N_j + 3)}{(N_j - 1)(N_j - 2)(N_j - 3)} \frac{m_{4j}}{\widehat{\mu}_{k_j}^4} - \frac{3N_j(2N_j - 3)}{(N_j - 1)(N_j - 2)(N_j - 3)} \frac{m_{2j}^2}{\widehat{\mu}_{k_j}^4}.
\end{aligned} \tag{7.6}$$

Then $E[\widehat{\gamma}_{tj}] \approx \gamma_t$. *In Appendix 1.6.1 are given unobservable estimators* $\widetilde{\gamma}_{tj}$, *where the true values* μ_{k_j} *are used. They are shown to be unbiased estimates of* γ_t.

Define the weighted totals of these expressions, the overall estimators of γ_t, to be

$$\widehat{\gamma}_t = \frac{\sum_{j=1}^{J} \mathbf{1}_{\{N_j \geq t\}}(N_j - t + 1)\widehat{\gamma}_{tj}}{\sum_{j=1}^{J} \mathbf{1}_{\{N_j \geq t\}}(N_j - t + 1)}. \tag{7.7}$$

Furthermore, let

$$\widehat{\phi}_2 = \frac{\widehat{\gamma}_2}{\widehat{\tau}^2 + 1}, \qquad \widehat{\phi}_3 = \frac{\widehat{\gamma}_3}{3\widehat{\tau}^2 + 1}, \qquad \widehat{\phi}_4 = \frac{\widehat{\gamma}_4}{3\widehat{\tau}^4 + 6\widehat{\tau}^2 + 1}. \tag{7.8}$$

In Appendix 1.6.1 we show that $\widehat{\phi}_t$ are suitable estimators of $\phi_t = \gamma_t/E[\Theta_1^t]$. They are not in general unbiased due to the substitution of $\widehat{\mu}_{k_j}$ for μ_{k_j} and of $\widehat{\tau}^2$ for τ^2.

Let also

$$\begin{aligned} f_2(x, y, \phi_2) &= y^{-1}(\phi_2 + y)(x + 1), \\ f_3(x, y, \phi_2, \phi_3) &= y^{-2}(\phi_3 + 3y\phi_2 + y^2)(3x + 1), \\ f_4(x, y, \phi_2, \phi_3, \phi_4) &= y^{-3}(\phi_4 - 3\phi_2^2 + 3y\phi_2^2 + 4y\phi_3 + 6y^2\phi_2 + y^3)(3x^2 + 6x + 1). \end{aligned} \tag{7.9}$$

Now define a counterpart to (6.4), namely

$$v_{\mathsf{M}}(x, y, \phi_2, \phi_3, \phi_4) = f_4(x, y, \phi_2, \phi_3, \phi_4) - 4f_3(x, y, \phi_2, \phi_3) + 8f_2(x, y, \phi_2) - f_2(x, y, \phi_2)^2 - 4. \tag{7.10}$$

Then $\rho_j(\tau^2)$ by (5.4) is

$$\rho_j(\tau^2) = v_{\mathsf{M}}(\tau^2, N_j, \phi_2, \phi_3, \phi_4). \tag{7.11}$$

As before, plug in estimators in (7.11) to get $\widehat{\rho}_j(\widehat{\tau}^2)$ and then $\widehat{\alpha}_j(\widehat{\tau}^2)$, using (5.5).

7.2. Pseudo-estimator for gamma-lognormal mixture

If specific parametric forms hold for the conditional claim amount distribution, we can get better τ^2-estimators.

We give an estimator $\overset{*}{\tau}{}^2$ under the following assumption, which covers a range of distributions between short-tailed and long-tailed ones.

Assumption 7
Assumption 5 is true and its exponent $p = 2$. Conditional on Θ_j, Z_{jr} is distributed as a mixture of a gamma distribution and a lognormal distribution with probability q for gamma, both with mean $\mu_{k_j}\Theta_j$.

Then Assumption 6 in Section 7.1 is true. For gamma and lognormal distributions the 3:rd and 4:th moments are determined by the 1:st and 2:nd ones. The idea is to use the empirical 3:rd central moment to estimate q, which then is used to estimate the 4:th central moment. We refer to Appendix 1.7 for the computations.

Corollary 7.1. *Let Assumptions 1, 4 and 7 be true. Let*

$$\widehat{q} \quad = \quad \min\left(1, \max\left(0, \frac{\widehat{\phi}_2^3 + 3\widehat{\phi}_2^2 - \widehat{\phi}_3}{(\widehat{\phi}_2 + 1)\widehat{\phi}_2^2}\right)\right), \tag{7.12}$$

$$\phi_3^* \quad = \quad \widehat{q}\,2\widehat{\phi}_2^2 + (1 - \widehat{q})(\widehat{\phi}_2^3 + 3\widehat{\phi}_2^2), \tag{7.13}$$

$$\phi_4^* = \widehat{q}(6\widehat{\phi}_2^3 + 3\widehat{\phi}_2^2) + (1-\widehat{q})\{(\widehat{\phi}_2+1)^3[(\widehat{\phi}_2+1)^3 - 4] + 6\widehat{\phi}_2 + 3\}, \tag{7.14}$$

$$\overset{*}{\rho}_j(\overset{*}{\tau}{}^2) = \upsilon_{\mathsf{M}}(\overset{*}{\tau}{}^2, N_j, \widehat{\phi}_2, \phi_3^*, \phi_4^*). \tag{7.15}$$

Define the pseudo-estimator $\overset{*}{\tau}{}^2$ *by replacing* $\widehat{\tau}^2$ *with* $\overset{*}{\tau}{}^2$ *and* $\widehat{\rho}_j(\widehat{\tau}^2)$ *with* $\overset{*}{\rho}_j(\overset{*}{\tau}{}^2)$*, as given in (7.15), in the estimator of* $\alpha_j(\tau^2)$ *in (5.5) and in (5.11). It holds* $\mathrm{E}[\overset{*}{\tau}{}^2] \approx \tau^2$*. If* $\mathrm{E}[(\Theta_j - 1)^3] = 0$ *and* $\mathrm{e}(\Theta_j) = 0$*, then* $\overset{*}{\tau}{}^2$ *is approximately optimal in the sense of having the smallest mean square error of estimators in the form of a linear combination of* $(Y_j - \widehat{\mu}_{k_j})^2$*.*

Remark 7.1. *The use of the 4:th central moment can cause unstable* $\widehat{\tau}^2$*, if there are too few claims. The simulation results of Section 9 illustrate this. Even if Assumption 7 is only remotely satisfied,* $\overset{*}{\tau}{}^2$ *can be preferable over* $\widehat{\tau}^2$ *if the latter is unstable.*

7.3. Non-pseudo-estimator

With N_0 = the total number of claims, as in the claim frequency non-pseudo-estimator, and J_0 = the number of groups with claims, we define this non-pseudo-estimator.

$$\widetilde{\tau}^2 = \max\left(0, \frac{\sum_{j=1}^{J} N_j (Y_j/\widehat{\mu}_{k_j} - 1)^2 - (J_0 - 1)\widehat{\sigma}^2}{N_0 - \sum_{j=1}^{J} N_j^2/N_0}\right). \tag{7.16}$$

The adaptation of (4.27) in Ohlsson & Johansson (2010) into $\widetilde{\tau}^2$ by (7.16) is this. Firstly, we specialize p on p. 82 to $p = 2$, in order to have a suitable classical estimator to compare the pseudo one to in simulations. Secondly, we divide the expression by the square of an estimate of the base factor μ in Ohlsson & Johansson (2010), like the claim frequency non-pseudo-estimator. Thirdly we truncate the estimator from below to 0. As for claim frequency, our simulations nevertheless showed a negative bias. Note that the denominator in (7.16) is similar to the one in (6.6).

8. Combining claim frequency and mean claim

The claim frequency and mean claim predictors are combined, if we wish to use these results for risk premium. To demonstrate this, we put subscripts F and M into the Θ_j and $\widehat{\Lambda}_j$. We then define

$$\widehat{\Lambda}_{\mathsf{FM}j} = \widehat{\Lambda}_{\mathsf{F}j}\widehat{\Lambda}_{\mathsf{M}j}. \tag{8.1}$$

It can be corrected by the simple factor $\sum_{i=1}^{J}\sum_{r=1}^{N_i} Z_{ir}/\sum_{i=1}^{J} e_i\widehat{\Lambda}_{\mathsf{FM}i}$ to make the exposure-weighted total of estimated predictors unbiased. It serves as final rating factor for risk premium.

We have to assume that $\{\Theta_{\mathsf{F}j}\}_1^J$ and $\{\Theta_{\mathsf{M}j}\}_1^J$ are independent. Claim numbers are S-ancillary for the mean claim parameters in the Compound Poisson model. This property implies that inference for $\{\Theta_{\mathsf{M}j}\}_1^J$ shall be made conditionally on the claim numbers. Together with independence it justifies (8.1).

9. Robustness test simulations

9.1. Setup of comparison

To get guidelines for choice of estimator, depending on the situation, we have compared our pseudo-estimators with the non-pseudo classical type ones. We used as the non-pseudo-estimator $\widetilde{\tau}^2$, given for claim frequency by (6.6) and for mean claim by (7.16).

The basic model stated in Assumptions 1–6 was obeyed. But, except for the case of zero τ^2, we did not let the distributions of Θ_j, for both claim frequency and mean claim, have zero excess or even mostly third central moment zero, as our pseudo-estimator theorems presuppose. From a practical viewpoint these are artificial assumptions, but ones that admit relatively simple and mathematically consistent pseudo-estimators. These estimators have to be reasonably robust against departures from the assumptions in order to be useful, though.

Three J-values 200, 1000 and 2000 were studied. For each value a fictitious insurance file was made with very varied exposure sizes per group j. An Auxiliary with five classes was assigned to each j.

Certain expected claim frequencies and mean claims per class were fixed. We set the base factor for claim frequency to 0.01, the base factor for mean claim to 2000, and the following class factors.

Class	1	2	3	4	5
Claim frequency factor	1	2	3	4	5
Mean claim factor	1.0	1.5	2.0	2.5	3.0

The Auxiliary was assigned to the groups successively with 1, 2, 3, 4, 5, 1, 2, 3, ...

Exposures per group j were assigned by the algorithm $k = 1 + (j-1)\%100$ and exposure $= 100k - 90$, where %100 gives the remainder after division by 100. I.e. in an arithmetic series 10, 110, 210, ... starting from the beginning at $j = 1, 101, 201, \ldots$.

One simulation generated about 30,200 claims for $J = 200$ and about 302,000 claims for $J = 2000$. We made as many simulations as were necessary to establish the best method, unless run times would have been too long.

As measure of the goodness of an estimate we used an estimate of expected mean square deviation of the estimate from the true parameter. These measures, in the form of 1000×(square root), are tabulated in Tables B1–B3. Let $\delta_1 = (\widehat{\tau}^2 - \tau^2)^2$ be the observed square deviation of $\widehat{\tau}^2$ and let $\delta_2 = (\widetilde{\tau}^2 - \tau^2)^2$. Set $\delta_0 = \delta_1 - \delta_2$ and let δ_{0t} be the value of the t:th simulation. We estimated $\mathrm{E}[\delta_0]$ by $\sum_{t=1}^{S} \delta_{0t}/S$, where S is the number of simulations. If this is negative, then $\widehat{\tau}^2$ is denoted as Best and vice versa. If the 99 % level confidence interval for $\mathrm{E}[\delta_0]$ contains 0, then a question mark is added.

9.2. Distributions and results

The results are given in Tables B1–B3 in Appendix 2. Our pseudo-estimator $\widehat{\tau}^2$ is denoted by Ps, the non-pseudo one $\widetilde{\tau}^2$ by Nps.

Let U(a,b) be a random variable having the uniform distribution on (a,b).

Let (α, β) be the usual gamma distribution parameter, such that the mean is α/β and the variance is α/β^2. Let Θ_k^0 have the gamma distribution of (k, k).

The table below lists Θ-distributions D1, . . ., D9 in ascending CV order.

Distribution of Θ_j	Meaning	τ^2
D1	1 always	0.000000
D2	U(0.875,1.125)	0.005208
D3	$0.25\Theta_4^0 + 0.75$	0.015625
D4	$0.25\Theta_2^0 + 0.75$	0.031250
D5	$0.25\Theta_1^0 + 0.75$	0.062500
D6	U(0.500,1.500)	0.083333
D7	Θ_4^0	0.250000
D8	Θ_2^0	0.500000
D9	Θ_1^0	1.000000

We let claim amounts be distributed as U(meanclaim/50.5,meanclaim×100/50.5) with CV 0.56592, or lognormally distributed with CV = 1, conditional on the Θs.

We give an estimate of $1000 \times \sqrt{\text{mean square deviation of estimate from true value}}$.

For $\tau^2 = 0$ the confidence intervals (confidence level 95 %) are for $10^5\times$parameter. This is marked with a †. Otherwise confidence intervals (95 %) are for the biases in percent of the Ps and Nps τ^2-estimates, i.e. for 100(estimate - truevalue)/truevalue.

9.3. Conclusions from simulations

The pseudo-estimators for mean claim have mixed positive and negative biases. Otherwise almost all estimators have negative bias, except of course for zero τ^2. This is remarkable for $\widetilde{\tau}^2$, since they were truncated from below to 0. The estimator with the smaller absolute bias has mostly the smaller mean square deviation.

Overall, the advantage of the pseudo-estimators over the classical estimators increases with increasing J, in line with Remark 7.1.

9.3.1. Claim frequency

In Table B1 it is seen that the pseudo-estimator $\widehat{\tau}^2$ is generally the best one, except for some cases with small τ^2. For the smaller number of groups $J = 200$ it is possibly worse than the $\widetilde{\tau}^2$ also for the large τ^2 in distribution D9. A guideline would be to recommend that $\widehat{\tau}^2$ is used, unless J is small and the suspected τ^2 is also small.

9.3.2. Mean claim

For the light-tailed uniform conditional claim amount distribution of Table B2, the classical estimator $\widetilde{\tau}^2$ is best for the smaller J-numbers 200 and 1000 when τ^2 is large. This holds also for the heavy-tailed lognormal distribution of Table B3 when $J = 200$. In a typical mass consumer credibility application J is 2000 or larger, and the conditional claim amount distribution is more heavy-tailed than the uniform distribution. For those applications the pseudo-estimators can be recommended, while for applications with a few large customers the case is not so clear.

10. Conclusion

We give sharp results for the BLP (Credibility Estimator) in a generic credibility model covering claim frequency and mean claim. The model has an auxiliary argument, which is a function of a multi-level factor. An optimal pseudo-estimator for the between-groups variance component is given under some moment conditions for random parameters. The pseudo-estimators are shown to be reasonably robust against departures from the moment assumptions.

Acknowledgements

Thanks are due to the referee for many valuable suggestions, which helped to improve the paper considerably. Thanks also to Dr. Adhitya Ronnie Effendie, Universitas Gadjah Mada, Yogyakarta, Indonesia, for valuable suggestions and encouragement.

References

Bichsel, F. & Straub, E. Erfahrungstarifierung in der Kollektif-Krankenversicherung. *Internal note of the Swiss-Re.* Unpublished. Year of note is unknown to the author.

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

Bühlmann, H. & Straub, E. (1970). Glaubwürdichkeit für Schadensätze. *Bulletin of the Swiss Association of Actuaries* **70**, 111–133.

Campbell, M. (1986). An integrated system for estimating the risk premium of individual car models in motor insurance. *ASTIN Bulletin* **16** (2), 165–184.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, Princeton.

De Vylder, F. E. (1996). *Advanced Risk Theory.* Editions de l'Université de Bruxelles.

Frees, E. (2003). Multivariate credibility for aggregate loss models. *North American Actuarial Journal.* **7** (1), 13–37.

Kaas R., Goovaerts M. & Dhaene, J. (2009). *Modern Actuarial Risk Theory Using R.* Springer, Berlin.
Norberg, R. (1980). Empirical Bayes Credibility. *Scandinavian Actuarial Journal*, **1980**, 177–194.
Ohlsson, E. (2008). Combining generalized linear models and credibility models in practice. *Scandinavian Actuarial Journal*, **2008** (4), 301–314.
Ohlsson, E. & Johansson, B. (2010). *Non-Life Insurance Pricing with Generalized Linear Models.* Springer, Berlin.
Rosenlund, S. (2010). Dispersion estimates for Poisson and Tweedie models. *ASTIN Bulletin* **40** (1), 271–279. doi: 10.2143/AST.40.1.2049229.
Rosenlund, S. (2014). Inference in multiplicative pricing. *Scandinavian Actuarial Journal* **2014** (8), 690-713. http://www.tandfonline.com/doi/abs/10.1080/03461238.2012.760885. Article is Open Access.

Appendix 1. Proofs

Below an estimator to be plugged into an expression is written with a $\widehat{}$ above it. This can also mean an estimator marked with a $\tilde{}$ or an $*$ in the previous sections, depending on assumptions and the suitability of different estimators in different situations.

Appendix 1.1. Proof of Theorem 5.1

The optimality of the BLP form of (5.8), as a linear combination of the individual and the collective mean, is a key result in basic credibility. For an extension to the case with Auxiliaries, see e.g. Ohlsson & Johansson (2010), Section 4.2, Theorem 4.3, which applies to the present model. Their final rating factor $\mu\gamma_i\widehat{U}_j$ for group j following (4.24) is in the form (5.8). The difference between methods is how to arrive at the z_j.

The resulting expression (A4) can be applied to total claim cost, letting $Y_j = \sum_{r=1}^{N_j} Z_{jr}/e_j$ and interpreting functionals accordingly.

We now compute z_j. Since only one j is treated at a time, we drop the subindex j in setting $z = z_j$, $Y = Y_j$, $\mu = \mu_{k_j}$, $\widehat{\mu} = \widehat{\mu}_{k_j}$, and $\Theta = \Theta_j$. From (5.8) we get

$$\begin{aligned}
&\mathrm{E}[(\Lambda_j^* - \mu_{k_j}\Theta_j)^2] = \mathrm{E}[(\Lambda_j^* - \mu\Theta)^2] \\
&= \mathrm{E}[\{zY + (1-z)\widehat{\mu} - \mu\Theta\}^2] \\
&= \mathrm{E}[\{z(Y-\mu\Theta) + (1-z)(\widehat{\mu} - \mu\Theta)\}^2] \\
&= z^2\mathrm{E}[(Y-\mu\Theta)^2] + 2(z-z^2)\mathrm{E}[(Y-\mu\Theta)(\widehat{\mu}-\mu\Theta)] + (1-z)^2\mathrm{E}[(\widehat{\mu}-\mu\Theta)^2].
\end{aligned}$$

Set 1/2 of the derivative of this expression with respect to z equal to 0. The resulting linear equation in z has only one solution, which gives the minimum. We obtain

$$z\mathrm{E}[(Y-\mu\Theta)^2] + (1-2z)\mathrm{E}[(Y-\mu\Theta)(\widehat{\mu}-\mu\Theta)] + (z-1)\mathrm{E}[(\widehat{\mu}-\mu\Theta)^2] = 0. \tag{A1}$$

To simplify (A1), note that from Assumption 1 we get

$$\mathrm{E}[Y \mid \Theta] = \mu\Theta, \qquad \mathrm{E}[Y] = \mu. \tag{A2}$$

For any stochastic variable X and σ-algebra $\mathcal{F}$ the following identities hold.

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}[\mathrm{E}[X \mid \mathcal{F}]], \\
\mathrm{Var}[X] &= \mathrm{E}[\mathrm{Var}[X \mid \mathcal{F}]] + \mathrm{Var}[\mathrm{E}[X \mid \mathcal{F}]].
\end{aligned}$$

Below we will use these identities with $\mathcal{F} = \sigma(\Theta)$, the σ-algebra induced by Θ.

First term of (A1)
We obtain

$$\begin{aligned}
&\mathrm{E}[(Y-\mu\Theta)^2] \\
&= \mathrm{Var}[Y-\mu\Theta] \\
&= \mathrm{E}[\mathrm{Var}[Y-\mu\Theta \mid \Theta]] + \mathrm{Var}[\mathrm{E}[Y-\mu\Theta \mid \Theta]] \\
&= \mathrm{E}[\mathrm{Var}[Y-\mu\Theta \mid \Theta]] \\
&= \mathrm{E}[\mathrm{Var}[Y \mid \Theta]] \\
&= \sigma_j^2.
\end{aligned}$$

Second term of (A1)
By using $\mathrm{E}[Y] = \mathrm{E}[\widehat{\mu}] = \mu$ we obtain

$$\begin{aligned}
&\mathrm{E}[(Y-\mu\Theta)(\widehat{\mu}-\mu\Theta)] \\
&= \mathrm{E}[(Y-\mu+\mu-\mu\Theta)(\widehat{\mu}-\mu+\mu-\mu\Theta)] \\
&= \mathrm{E}[(Y-\mu)(\widehat{\mu}-\mu)] + \mathrm{E}[(Y-\mu)(\mu-\mu\Theta)] + \mathrm{E}[(\mu-\mu\Theta)(\widehat{\mu}-\mu)] + \mathrm{E}[(\mu-\mu\Theta)^2] \\
&= \mathrm{Cov}(\widehat{\mu}, Y) - \mu\mathrm{Cov}(Y,\Theta) - \mu\mathrm{Cov}(\widehat{\mu},\Theta) + \mu^2\tau^2.
\end{aligned}$$

It holds

$$\mathrm{Cov}(Y,\Theta) = \mathrm{E}[\mathrm{E}[(Y-\mu)(\Theta-1)\mid\Theta] = \mathrm{E}[(\Theta-1)(\mu\Theta-\mu)] = \mu\tau^2. \tag{A3}$$

Therefore

$$\mathrm{E}[(Y-\mu\Theta)(\widehat{\mu}-\mu\Theta)] = \mathrm{Cov}(\widehat{\mu},Y) - \mu\mathrm{Cov}(\widehat{\mu},\Theta).$$

Third term of (A1)
We have

$$\begin{aligned}
&\mathrm{E}[(\widehat{\mu}-\mu\Theta)^2] \\
&= \mathrm{E}[(\widehat{\mu}-\mu+\mu-\mu\Theta)^2] \\
&= \mathrm{E}[(\widehat{\mu}-\mu)^2 + 2(\widehat{\mu}-\mu)(\mu-\mu\Theta) + (\mu-\mu\Theta)^2] \\
&= \mathrm{E}[(\widehat{\mu}-\mu)^2] - 2\mu\mathrm{E}[(\widehat{\mu}-\mu)(\Theta-1)] + \mu^2\mathrm{E}[(\Theta-1)^2] \\
&= \mathrm{Var}[\widehat{\mu}] - 2\mu\mathrm{Cov}(\widehat{\mu},\Theta) + \mu^2\tau^2.
\end{aligned}$$

Thus (A1) reduces to

$$z\sigma_j^2 + (1-2z)\left\{\mathrm{Cov}(\widehat{\mu},Y) - \mu\mathrm{Cov}(\widehat{\mu},\Theta)\right\} + (z-1)\left\{\mathrm{Var}[\widehat{\mu}] - 2\mu\mathrm{Cov}(\widehat{\mu},\Theta) + \mu^2\tau^2\right\} = 0$$

i.e.

$$\begin{aligned}
&z\left\{\sigma_j^2 - 2\mathrm{Cov}(\widehat{\mu},Y) + 2\mu\mathrm{Cov}(\widehat{\mu},\Theta) + \mathrm{Var}[\widehat{\mu}] - 2\mu\mathrm{Cov}(\widehat{\mu},\Theta) + \mu^2\tau^2\right\} \\
&= -\mathrm{Cov}(\widehat{\mu},Y) + \mu\mathrm{Cov}(\widehat{\mu},\Theta) + \mathrm{Var}[\widehat{\mu}] - 2\mu\mathrm{Cov}(\widehat{\mu},\Theta) + \mu^2\tau^2
\end{aligned}$$

i.e.

$$z\left\{\sigma_j^2 + \mathrm{Var}[\widehat{\mu}] + \mu^2\tau^2 - 2\mathrm{Cov}(\widehat{\mu},Y)\right\} = \mathrm{Var}[\widehat{\mu}] + \mu^2\tau^2 - \mathrm{Cov}(\widehat{\mu},Y) - \mu\mathrm{Cov}(\widehat{\mu},\Theta).$$

Reinstating the subindex j, again writing z_j for z etc., we get

$$z_j = \frac{\mathrm{Var}[\widehat{\mu}_{k_j}] + \mu_{k_j}^2\tau^2 - \mathrm{Cov}(\widehat{\mu}_{k_j},Y_j) - \mu_{k_j}\mathrm{Cov}(\widehat{\mu}_{k_j},\Theta_j)}{\sigma_j^2 + \mathrm{Var}[\widehat{\mu}_{k_j}] + \mu_{k_j}^2\tau^2 - 2\mathrm{Cov}(\widehat{\mu}_{k_j},Y_j)}. \tag{A4}$$

We can compute the variances and covariances. Assumption 1 gives

$$\mathrm{Var}[Y_j] = \mathrm{E}[\mathrm{Var}[Y_j\mid\Theta_j]] + \mathrm{Var}[\mathrm{E}[Y_j\mid\Theta_j]] = \sigma_j^2 + \mathrm{Var}[\mu_{k_j}\Theta_j] = \sigma_j^2 + \mu_{k_j}^2\tau^2. \tag{A5}$$

Hence by the independence condition Assumption 2

$$\mathrm{Var}[\widehat{\mu}_k] = \mathrm{Var}\left[\sum_{j:k_j=k}\frac{w_jY_j}{w_k^{\mathsf{A}}}\right] = \sum_{j:k_j=k}\frac{w_j^2}{{w_k^{\mathsf{A}}}^2}\mathrm{Var}[Y_j] = \frac{1}{{w_k^{\mathsf{A}}}^2}\sum_{j:k_j=k}w_j^2(\sigma_j^2+\mu_k^2\tau^2) = \nu_k^2. \tag{A6}$$

Thus we have proved (5.7).

Again by the independence condition Assumption 2, for $\mathrm{Cov}(\widehat{\mu}_{k_j},Y_j)$ we can eliminate all terms in $\widehat{\mu}_{k_j}$ not containing Y_j. Therefore

$$\mathrm{Cov}(\widehat{\mu}_{k_j},Y_j) = \mathrm{Cov}\left(\frac{w_jY_j}{w_{k_j}^{\mathsf{A}}},Y_j\right) = \frac{w_j}{w_{k_j}^{\mathsf{A}}}\mathrm{Var}[Y_j] = \frac{w_j}{w_{k_j}^{\mathsf{A}}}\sigma_j^2 + \frac{w_j}{w_{k_j}^{\mathsf{A}}}\mu_{k_j}^2\tau^2. \tag{A7}$$

In the same way we obtain from (A3)

$$\mathrm{Cov}(\widehat{\mu}_{k_j},\Theta_j) = \mathrm{Cov}\left(\frac{w_jY_j}{w_{k_j}^{\mathsf{A}}},\Theta_j\right) = \frac{w_j}{w_{k_j}^{\mathsf{A}}}\mu_{k_j}\tau^2. \tag{A8}$$

Inserting (A6), (A7) and (A8) in (A4) we obtain

$$z_j = \frac{\nu_{k_j}^2 + \mu_{k_j}^2\tau^2 - \dfrac{w_j}{w_{k_j}^{\mathsf{A}}}\sigma_j^2 - \dfrac{2w_j}{w_{k_j}^{\mathsf{A}}}\mu_{k_j}^2\tau^2}{\sigma_j^2 + \nu_{k_j}^2 + \mu_{k_j}^2\tau^2 - \dfrac{2w_j}{w_{k_j}^{\mathsf{A}}}\sigma_j^2 - \dfrac{2w_j}{w_{k_j}^{\mathsf{A}}}\mu_{k_j}^2\tau^2}. \tag{A9}$$

This gives (5.10) after some rearrangement.

Remark A1. *The Var- and Cov-terms in (A3) might be new. We could not find them in the literature, even with no Auxiliary. However, since the literature on credibility is so large, they are possibly already known. The Bühlmann & Straub (1970) estimator is retrieved by omitting them. The terms are often small in applications. The premise of credibility analysis is normally that the collective observed mean $\widehat{\mu}_{k_j}$ has so small variance, that it can be equated with the true mean μ_{k_j} for practical purposes. If j's Auxiliary class (the whole sample with no Auxiliary) comprises sufficiently many claims and the exposure w_i of each group in the Auxiliary class of j is sufficiently small relative to the total exposure $w_{k_j}^{\mathsf{A}}$ of its Auxiliary class, the premise is justified.*

Appendix 1.2. Proof of Theorem 5.2

For the optimal pseudo-estimator, we note that by (A2) and (A5)

$$\mathrm{E}[Y_j/\mu_{k_j}] = 1, \qquad \mathrm{Var}[Y_j/\mu_{k_j}] = \sigma_j^2/\mu_{k_j}^2 + \tau^2.$$

Define the random variables

$$T_j = \tau^2\left(\sigma_j^2/\mu_{k_j}^2 + \tau^2\right)^{-1}\left(\frac{Y_j}{\mu_{k_j}} - 1\right)^2, \text{ with expectation } \mathrm{E}[T_j] = \tau^2.$$

We seek the optimal estimator of τ^2 in the form of a linear combination

$$\sum_{j=1}^{J} a_j T_j, \text{ where } \sum_{j=1}^{J} a_j = 1.$$

Since T_j are independent the minimum variance standard solution is $a_j = \text{const}/\mathrm{Var}[T_j]$, for j with $w_j > 0$. (For mean claim we can have $w_j = N_j = 0$.) We obtain

$$\mathrm{Var}[T_j] = \tau^4\left(\sigma_j^2/\mu_{k_j}^2 + \tau^2\right)^{-2} \mathrm{Var}\left[\left(\frac{Y_j}{\mu_{k_j}} - 1\right)^2\right] = \tau^4\left(\sigma_j^2/\mu_{k_j}^2 + \tau^2\right)^{-2}\rho_j(\tau^2).$$

Here the factor τ^4 cancels out. Thus, with $\alpha_j(\tau^2)$ given by (5.5), we have

$$a_j = \alpha_j(\tau^2)/[\alpha_1(\tau^2) + \cdots + \alpha_J(\tau^2)].$$

The optimal estimator of τ^2 using unknown true parameters is then, with some rewriting,

$$\sum_{j=1}^{J} a_j T_j = \sum_{j=1}^{J} a_j \tau^2\left(\sigma_j^2/\mu_{k_j}^2 + \tau^2\right)^{-1}\left(\frac{Y_j}{\mu_{k_j}} - 1\right)^2 = \sum_{j=1}^{J} a_j \frac{\tau^2}{\sigma_j^2 + \mu_{k_j}^2\tau^2}(Y_j - \mu_{k_j})^2. \tag{A10}$$

It is unbiased. Substituting estimators for true values in the right sides of (5.5) and (A10) we can obtain $\widehat{\alpha}_j(\widehat{\tau}^2)$ from $\alpha_j(\tau^2)$ and an estimator $\widehat{\tau}^2$ of τ^2. This estimator will be biased.

One source of bias that can be dealt with is the use of $\widehat{\mu}_{k_j}$ in $(Y_j - \widehat{\mu}_{k_j})^2$ after plugging in estimators. We have

$$\begin{aligned}
&\mathrm{E}[(Y_j - \widehat{\mu}_{k_j})^2] = \mathrm{E}[(Y_j - \mu_{k_j} + \mu_{k_j} - \widehat{\mu}_{k_j})^2] \\
&= \mathrm{E}[(Y_j - \mu_{k_j})^2] + 2\mathrm{E}[(Y_j - \mu_{k_j})(\mu_{k_j} - \widehat{\mu}_{k_j})] + \mathrm{E}[(\widehat{\mu}_{k_j} - \mu_{k_j})^2] \\
&= \mathrm{Var}[Y_j] - 2\mathrm{Cov}(\widehat{\mu}_{k_j}, Y_j) + \mathrm{Var}[\widehat{\mu}_{k_j}] = \mathrm{Var}[Y_j] - \frac{2w_j}{w_{k_j}^{\mathsf{A}}}\mathrm{Var}[Y_j] + \nu_{k_j}^2 \\
&= (\sigma_j^2 + \mu_{k_j}^2\tau^2)\left(1 - \frac{2w_j}{w_{k_j}^{\mathsf{A}}}\right) + \nu_{k_j}^2,
\end{aligned}$$

where we used equations (A5), (A7) and (A6). Define

$$V_j = \frac{\tau^2}{(\sigma_j^2 + \mu_{k_j}^2 \tau^2)\left(1 - 2w_j/w_{k_j}^{\mathsf{A}}\right) + \nu_{k_j}^2}(Y_j - \widehat{\mu}_{k_j})^2.$$

Then from the above it holds $\mathrm{E}[V_j] = \tau^2$. Other bias effects on the expectation from plugging in estimators are not so easy to reduce.

Here V_j is similar to T_j, but uses an estimate $\widehat{\mu}_{k_j}$ in one place and has an extra factor and term in the denominator. An estimator based on V_j is likely to have normally less absolute bias and less mean square error than the estimator of τ^2 obtained by substituting estimators for true values in T_j.

Let $\widehat{\alpha}_j(\widehat{\tau}^2)$ be obtained from $\alpha_j(\tau^2)$ according to (5.5) by plugging in estimators. We are content to use these weights, since a more precise estimate of $\mathrm{Var}[V_j]$ than approximating it with $\mathrm{Var}[T_j]$ is difficult to compute, and since V_j are dependent within Auxiliary classes.

The final estimator will then be, after plugging in estimators,

$$\widehat{\tau}^2 = \sum_{j=1}^{J}\left(\frac{\widehat{\alpha}_j(\widehat{\tau}^2)}{\widehat{\alpha}_1(\widehat{\tau}^2) + \cdots + \widehat{\alpha}_J(\widehat{\tau}^2)}\right)\frac{\widehat{\tau}^2}{(\widehat{\sigma}_j^2 + \widehat{\mu}_{k_j}^2\widehat{\tau}^2)\left(1 - 2w_j/w_{k_j}^{\mathsf{A}}\right) + \widehat{\nu}_{k_j}^2}(Y_j - \widehat{\mu}_{k_j})^2. \tag{A11}$$

This gives Theorem 5.2. There we use the $\approx$ symbol and write approximately, due to the plugging in of estimators. These statements could be formulated as limit theorems as $J \to \infty$, provided some conditions were imposed to guarantee that the influence of any individual j vanishes in the limit.

For use in the next section we note that $\widehat{\nu}_{k_j}^2$ can be written as a linear expression in $\widehat{\tau}^2$ as

$$\widehat{\nu}_{k_j}^2 = \left[\frac{1}{{w_{k_j}^{\mathsf{A}}}^2}\sum_{i:k_i=k_j} w_i^2\widehat{\sigma}_i^2\right] + \left[\frac{1}{{w_{k_j}^{\mathsf{A}}}^2}\sum_{i:k_i=k_j} w_i^2\widehat{\mu}_{k_i}^2\right]\widehat{\tau}^2 = d_{1j} + d_{2j}\widehat{\tau}^2, \tag{A12}$$

as is seen in (A6).

Appendix 1.3. Solutions of pseudo-estimator equations

We seek solutions of (5.11).

We will rewrite equation (A11) in a way showing the dependence on $\widehat{\tau}^2$ in a simplified form. With d_{1j} and d_{2j} the coefficients of the partitioning of $\widehat{\nu}_{k_j}^2$ in equation (A12), let

$$h_{1j} = \widehat{\sigma}_j^2\left(1 - 2w_j/w_{k_j}^{\mathsf{A}}\right) + d_{1j}, \qquad h_{2j} = \widehat{\mu}_{k_j}^2\left(1 - 2w_j/w_{k_j}^{\mathsf{A}}\right) + d_{2j}.$$

Set

$$b_j(x) = \widehat{\alpha}_j(x)/[\widehat{\alpha}_1(x) + \cdots + \widehat{\alpha}_J(x)], \qquad c_j = h_{1j}/h_{2j}, \qquad U_j = (Y_j - \widehat{\mu}_{k_j})^2/h_{2j}.$$

With $x = \widehat{\tau}^2$ we have to solve the equation

$$f(x) = x - \sum_{j=1}^{J} b_j(x)\frac{x}{h_{1j} + h_{2j}x}(Y_j - \widehat{\mu}_{k_j})^2 = x - \sum_{j=1}^{J} b_j(x)\frac{x}{c_j + x}U_j = 0.$$

Here $x = 0$ is a solution. For possible positive solutions, let $g(x) = f(x)/x$, i.e.

$$g(x) = 1 - \sum_{j=1}^{J} b_j(x)\frac{U_j}{c_j + x}.$$

Let $c_{\min} = \min_{w_j>0}\{c_j\}$, $U_{\max} = \max_{w_j>0}\{U_j\}$ and $R = U_{\max} - c_{\min}$. Then $g(x) \geq 1 - U_{\max}/(c_{\min} + x) = (x - R)(c_{\min} + x)$. Hence $\lim_{x\to\infty} g(x) = 1$, and $g(x) > 0$ for $x > R$. If $R \leq 0$ no positive solution exists. If $R > 0$ we can take R as right endpoint of the interval where the solution is. The left endpoint is 0. The solution is in the closed interval [0,R].

If we can show that $g(x)$ is strictly increasing for $x > 0$, then $g(x) = 0$ has at most one positive solution. This is not obvious. All simulated cases in Appendix 2 have $g(x)$ strictly increasing. We challenge researchers to prove that this is always true, or else find a counter-example.

From the definition (5.5) and some calculations we get for claim frequency

$$g(0) = 1 - \left(\sum_{j=1}^{J} U_j c_j^{-1}(c_j+2)^{-1}\right)\left(\sum_{j=1}^{J}(c_j+2)^{-1}\right)^{-1}.$$

A more complicated expression for $g(0)$ holds for mean claim. If $g(0)$ is negative a positive solution exists. It remains to show that there is no positive solution if $g(0) \geq 0$. Also it remains to show that a positive solution is unique, or else that there are cases with several positive solutions.

Remark A2. *The upper limit $R = U_{max} - c_{min}$ might be too large for computation of $g(R)$ for numerical reasons. Instead we use the classical estimator $\widetilde{\tau}^2$. If $g(\widetilde{\tau}^2) \leq 0$ we search, by stepping up, for an upper limit x not too far from $\widetilde{\tau}^2$, where $g(x) > 0$. If $g(\widetilde{\tau}^2) > 0$ the upper limit is taken as $\widetilde{\tau}^2$.*

Appendix 1.4. Higher moments of stochastic parameters

If Θ_j has 3:rd central moment 0 and excess 0, then its moments of order 2 to 4 are easily shown to be these. The second order moment is always as stated.

$$\mathrm{E}[\Theta_j^2] = \tau^2 + 1, \qquad \mathrm{E}[\Theta_j^3] = 3\tau^2 + 1, \qquad \mathrm{E}[\Theta_j^4] = 3\tau^4 + 6\tau^2 + 1. \tag{A13}$$

We will in the sequel make frequent use of the Cramér (1946) formulas (15.10.4) and (15.10.5), connecting moments and central moments via semi-invariants, up to order 4, in order to establish estimators for $\rho_j(\tau^2)$, for both claim frequency and mean claim. The addition property of semi-invariants for sums of independent variables is here very useful.

Appendix 1.5. Proof of Lemma 6.1 for claim frequency

We show here that $\mathrm{Var}[(Y_j/\mu_{k_j} - 1)^2] = v_{\mathsf{F}}(\tau^2, \frac{1}{\mu_j e_j})$, with $v_{\mathsf{F}}(,)$ as defined in (6.4).

Suppressing j and introducing some expressions to simplify calculations, we set

$$N = N_j, \qquad m = \mu_{k_j} e_j, \qquad \Delta = m\Theta_j, \qquad a = m^2\tau^2.$$

From (A13) we obtain

$$\mathrm{E}[\Delta] = m, \qquad \mathrm{E}[\Delta^2] = a + m^2, \qquad \mathrm{E}[\Delta^3] = 3am + m^3, \qquad \mathrm{E}[\Delta^4] = 3a^2 + 6am^2 + m^4.$$

Now $N \mid \Delta \sim \mathrm{Poi}(\Delta)$. All Poisson semi-invariants are equal to the mean. From equation (15.10.4) in Cramér (1946), which gives moments in terms of semi-invariants, we get

$$\begin{aligned}
\mathrm{E}[N \mid \Delta] &= \Delta,\\
\mathrm{E}[N^2 \mid \Delta] &= \Delta + \Delta^2,\\
\mathrm{E}[N^3 \mid \Delta] &= \Delta + 3\Delta^2 + \Delta^3,\\
\mathrm{E}[N^4 \mid \Delta] &= \Delta + 7\Delta^2 + 6\Delta^3 + \Delta^4.
\end{aligned}$$

Since $\mathrm{E}[N^t] = \mathrm{E}[\mathrm{E}[N^t \mid \Delta]]$, we obtain

$$\begin{aligned}
\mathrm{E}[N] &= \mathrm{E}[\Delta] = m,\\
\mathrm{E}[N^2] &= \mathrm{E}[\Delta + \Delta^2] = m + a + m^2,\\
\mathrm{E}[N^3] &= \mathrm{E}[\Delta + 3\Delta^2 + \Delta^3] = m + 3(a + m^2) + 3am + m^3,\\
\mathrm{E}[N^4] &= \mathrm{E}[\Delta + 7\Delta^2 + 6\Delta^3 + \Delta^4] = m + 7(a + m^2) + 6(3am + m^3) + 3a^2 + 6am^2 + m^4.
\end{aligned}$$

For the 2:nd and 4:th central moments of N we have

$$\begin{aligned} \mathrm{E}[(N-m)^2] &= \mathrm{E}[N^2] - m^2, \\ \mathrm{E}[(N-m)^4] &= \mathrm{E}[N^4] - 4m\mathrm{E}[N^3] + 6m^2\mathrm{E}[N^2] - 3m^4. \end{aligned}$$

We are interested in

$$\mathrm{Var}[(N-m)^2] = \mathrm{E}[(N-m)^4] - \mathrm{E}[(N-m)^2]^2.$$

After some calculations we arrive at the comparatively simple expression

$$\mathrm{Var}[(N-m)^2] = m + 7a + 2m^2 + 4am + 2a^2.$$

Therefore we get, replacing a with $m^2\tau^2$,

$$\mathrm{Var}\left[\left(\frac{N}{m} - 1\right)^2\right] = \frac{1}{m^4}\mathrm{Var}[(N-m)^2] = \frac{1}{m^4}(m + m^2(7\tau^2+2) + 4m^3\tau^2 + 2m^4\tau^4).$$

Returning completely to the original notation, we have $N/m = Y_j/\mu_{k_j}$ and $m = \mu_{k_j}e_j$, which gives, with $v_{\mathsf{F}}(,)$ by (6.4),

$$\rho_j(\tau^2) = \mathrm{Var}[(Y_j/\mu_{k_j} - 1)^2] = \frac{1}{\mu_{k_j}^3 e_j^3} + \frac{7\tau^2+2}{\mu_{k_j}^2 e_j^2} + \frac{4\tau^2}{\mu_{k_j} e_j} + 2\tau^4 = v_{\mathsf{F}}(\tau^2, \frac{1}{\mu_{k_j} e_j}).$$

Appendix 1.6. Proof of Lemma 7.1 for mean claim

Appendix 1.6.1. Central moment estimators

We will develop estimators of ϕ_t, which are defined in Assumption 6. These are suitable for computing an estimator of $\rho_j(\tau^2)$. We have

$$\mathrm{E}[Z_{jr}/\mu_{k_j} \mid \Theta_j] = \Theta_j, \qquad \mathrm{E}[(Z_{jr}/\mu_{k_j} - \Theta_j)^t \mid \Theta_j] = \phi_t\Theta_j^t,\ (t = 2,3,4).$$

We use the sample central moments for $j \in \{1,\dots,J\}$. Let m_{tj} be as defined in expression (7.5). Then $m_{tj}/\mu_{k_j}^t$ are the sample central moments for Z_{jr}/μ_{k_j}, albeit containing the unknown functionals μ_{k_j}.

Cramér (1946), p. 352, gives unbiased central moment estimators of order $t \le 4$ for a sample of IID random variables. Using these we define

$$\begin{aligned} \widetilde{\gamma}_{2j} &= \frac{N_j}{N_j-1}\,\frac{m_{2j}}{\mu_{k_j}^2}, \\ \widetilde{\gamma}_{3j} &= \frac{N_j^2}{(N_j-1)(N_j-2)}\,\frac{m_{3j}}{\mu_{k_j}^3}, \\ \widetilde{\gamma}_{4j} &= \frac{N_j(N_j^2-2N_j+3)}{(N_j-1)(N_j-2)(N_j-3)}\,\frac{m_{4j}}{\mu_{k_j}^4} - \frac{3N_j(2N_j-3)}{(N_j-1)(N_j-2)(N_j-3)}\,\frac{m_{2j}^2}{\mu_{k_j}^4}. \end{aligned}$$

By Assumption 6 and Cramér (1946), p. 352, it holds

$$\mathrm{E}[\widetilde{\gamma}_{tj} \mid \Theta_j] = \phi_t\Theta_j^t$$

and hence

$$\mathrm{E}[\widetilde{\gamma}_{tj}] = \phi_t\mathrm{E}[\Theta_j^t] = \gamma_t,$$

where γ_t is defined by (7.4).

When weighting $\widetilde{\gamma}_{tj}$ together for total estimators $\widetilde{\gamma}_t$, albeit with unknown functionals in them, we observe that $\widetilde{\gamma}_{tj}$ is defined only for $N_j \ge t$. So we use weights $N_j - t + 1$, giving the estimators

$$\widetilde{\gamma}_t = \frac{\sum_{j=1}^{J} \mathbf{1}_{\{N_j \ge t\}}(N_j - t + 1)\widetilde{\gamma}_{tj}}{\sum_{j=1}^{J} \mathbf{1}_{\{N_j \ge t\}}(N_j - t + 1)},$$

with $\mathrm{E}[\widetilde{\gamma}_t] = \phi_t \mathrm{E}[\Theta_1^t]$ (recall that $\mathrm{E}[\Theta_j^t]$ is independent of j).

Since μ_{k_j} are unknown we use estimates $\widehat{\mu}_{k_j}$, giving $\widehat{\gamma}_{tj}$ by (7.6). They will be approximately unbiased estimators of $\phi_t \mathrm{E}[\Theta_j^t]$. We employ the weights $N_j - t + 1$ for $\widetilde{\gamma}_t$ also here, giving the estimators $\widehat{\gamma}_t$ by (7.7).

Using $N_j - 1$ as weight for $t = 2$ we obtain $\widehat{\gamma}_2 = \widehat{\sigma}^2$ by (7.3), which illustrates the feasibility of the weighting. We are, however, uncertain as to whether some other weighting, presumably equal to this one when specialized to $t = 2$, might be generally better without imposing more assumptions.

For ϕ_t we obtain the non-observable estimators $\widehat{\gamma}_t / \mathrm{E}[\Theta_j^t]$ and, for suitable estimators $\widehat{\mathrm{E}}[\Theta_j^t]$, the observable ones

$$\widehat{\phi}_t = \widehat{\gamma}_t / \widehat{\mathrm{E}}[\Theta_j^t]. \tag{A14}$$

The assumption of Lemma 7.1 is that $\mathrm{E}[(\Theta_j - 1)^3] = 0$ and $\mathrm{e}(\Theta_j) = 0$. Then we get estimates by using $\widehat{\tau}^2$ for τ^2 in (A13), e.g. $\widehat{\mathrm{E}}[\Theta_j^3] = 3\widehat{\tau}^2 + 1$. It follows that $\widehat{\phi}_t$ in (A14) will be those stated in (7.8).

Appendix 1.6.2. Variance estimator using semi-invariants

We shall compute an estimate $\widehat{\rho}_j(\widehat{\tau}^2)$ of $\rho_j(\tau^2) = \mathrm{Var}[(Y_j/\mu_{k_j} - 1)^2]$. For shortness we suppress indices etc. in the notation below. Let

$$\begin{aligned} \Theta &= \Theta_j, \\ N &= N_j, \\ V &= Y_j/\mu_{k_j} = N^{-1} \sum_{r=1}^{N} Z_{jr}/\mu_{k_j}. \end{aligned}$$

It holds $\mathrm{E}[V] = 1$. We seek

$$\mathrm{Var}[(V-1)^2] = \mathrm{E}[(V-1)^4] - \mathrm{E}[(V-1)^2]^2.$$

Now $(V-1)^4 = V^4 - 4V^3 + 6V^2 - 4V + 1$ and $(V-1)^2 = V^2 - 2V + 1$. This yields

$$\mathrm{Var}[(V-1)^2] = \mathrm{E}[V^4] - 4\mathrm{E}[V^3] + 8\mathrm{E}[V^2] - \mathrm{E}[V^2]^2 - 4. \tag{A15}$$

Let $\varkappa_t(\Theta)$ the semi-invariants of Z_{jr}/μ_{k_j} conditional on Θ. Cramér (1946) gives these in terms of central moments in (15.10.5). We obtain

$$\begin{aligned} \varkappa_1(\Theta) &= \Theta, \\ \varkappa_2(\Theta) &= \phi_2 \Theta^2, \\ \varkappa_3(\Theta) &= \phi_3 \Theta^3, \\ \varkappa_4(\Theta) &= \phi_4 \Theta^4 - 3\phi_2^2 \Theta^4. \end{aligned}$$

The semi-invariant of order t for NV conditional on Θ is $N\varkappa_t(\Theta)$, by virtue of the addition property of semi-invariants for sums of independent variables. We deduce the moments of NV conditional on Θ from its semi-invariants, again with the help of (15.10.4) in Cramér (1946). This gives

$$\begin{aligned} \mathrm{E}[NV \mid \Theta] &= N\Theta, \\ \mathrm{E}[(NV)^2 \mid \Theta] &= (N\phi_2 + N^2)\Theta^2, \\ \mathrm{E}[(NV)^3 \mid \Theta] &= (N\phi_3 + 3N^2\phi_2 + N^3)\Theta^3, \\ \mathrm{E}[(NV)^4 \mid \Theta] &= (N\phi_4 - 3N\phi_2^2 + 3N^2\phi_2^2 + 4N^2\phi_3 + 6N^3\phi_2 + N^4)\Theta^4. \end{aligned}$$

The unconditional moments of V are thus $N^{-t}\mathrm{E}[\mathrm{E}[(NV)^t \mid \Theta]]$.

To go further we need the moment assumptions for Θ, namely that $\mathrm{E}[(\Theta-1)^3] = 0$ and $\mathrm{e}(\Theta) = 0$. Then the moments of Θ are given by (A13). Hence, with f_t as defined in (7.9) with $x = \tau^2$ and $y = N$, we obtain

$$\mathrm{E}[V^2] = N^{-1}(\phi_2 + N)(\tau^2 + 1) = f_2(\tau^2, N, \phi_2),$$

$$\mathrm{E}[V^3] = N^{-2}(\phi_3 + 3N\phi_2 + N^2)(3\tau^2 + 1) = f_3(\tau^2, N, \phi_2, \phi_3),$$

$$\mathrm{E}[V^4] = N^{-3}(\phi_4 - 3\phi_2^2 + 3N\phi_2^2 + 4N\phi_3 + 6N^2\phi_2 + N^3)(3\tau^4 + 6\tau^2 + 1) = f_4(\tau^2, N, \phi_2, \phi_3, \phi_4).$$

We have from (7.10) the following expression, where we write N_j for N again.

$$v_{\mathsf{M}}(\tau^2,N_j,\phi_2,\phi_3,\phi_4) = f_4(\tau^2,N_j,\phi_2,\phi_3,\phi_4) - 4f_3(\tau^2,N_j,\phi_2,\phi_3) + 8f_2(\tau^2,N_j,\phi_2) - f_2(\tau^2,N_j,\phi_2)^2 - 4.$$

Then by (A15) we have

$$\rho_j(\tau^2) = \mathrm{Var}[(V-1)^2] = v_{\mathsf{M}}(\tau^2, N_j, \phi_2, \phi_3, \phi_4)$$

The estimator $\widehat{\rho}_j(\widehat{\tau}^2)$ of $\mathrm{Var}[(Y_j/\mu_{k_j}-1)^2]$ is obtained by plugging in estimators.

Remark A3. *There are high powers and mixed positive and negative terms in the expressions above. This will cause severe numerical problems in ordinary computer arithmetic. Multiprecision arithmetic must be used.*

Appendix 1.7. Proof of Corollary 7.1 for gamma-lognormal mixture

Assumption 7 implies that the distribution of $Z_{jr}/(\mu_{k_j}\Theta_j)$, conditional on Θ_j, is a gamma distribution with probability q and a lognormal distribution with probability $1-q$, both with mean 1. Let W_1 be a gamma distributed and let W_2 be a lognormally distributed random variable, with $\mathrm{E}[W_1] = \mathrm{E}[W_2] = 1$. Let W be distributed as $Z_{jr}/(\mu_{k_j}\Theta_j)$. We then have

$$\phi_t = \mathrm{E}[(W-1)^t] = q\mathrm{E}[(W_1-1)^t] + (1-q)\mathrm{E}[(W_2-1)^t]$$

The following expressions, giving the 3:d and 4:th moments as functions of the 2:nd one, can be deduced from the properties of the two distributions. The $\phi_t = \mathrm{E}[(W_1-1)^t]$ under Gamma are those valid for $q=1$, and the $\phi_t = \mathrm{E}[(W_2-1)^t]$ under Lognormal are those valid for $q=0$.

Gamma	Lognormal
$\phi_3 = 2\phi_2^2,$	$\phi_3 = \phi_2^3 + 3\phi_2^2,$
$\phi_4 = 6\phi_2^3 + 3\phi_2^2,$	$\phi_4 = (\phi_2+1)^3[(\phi_2+1)^3-4] + 6\phi_2 + 3.$

For the mixture we thus have

$$\phi_3 = q\,2\phi_2^2 + (1-q)(\phi_2^3 + 3\phi_2^2),$$
$$\phi_4 = q\,(6\phi_2^3 + 3\phi_2^2) + (1-q)[(\phi_2+1)^3[(\phi_2+1)^3-4] + 6\phi_2 + 3].$$

We can solve q from the expression for ϕ_3, namely

$$q = (\phi_2^3 + 3\phi_2^2 - \phi_3)/[(\phi_2+1)\phi_2^2].$$

The moment method, consisting of estimating a parameter, which is a function of the moments of the distribution, with the same function of the moment estimates, is here the most practical one. Using the estimates $\widehat{\phi}_2$ and $\widehat{\phi}_3$ in (A14) of Appendix 1.6.1 thus gives a q-estimate. It has to be truncated to at least 0 and at most 1. Thus we get estimates given by expressions (7.12), (7.13) and (7.14) in Corollary 7.1. Due to the truncation of $\widehat{q}$ to [0,1], ϕ_3^* is not always equal to $\widehat{\phi}_3$.

The desired estimate of $\rho_j(\tau^2) = \mathrm{Var}[(Y_j/\mu_{k_j}-1)^2]$ is then given by (7.15), to be used in the pseudo-estimator $\overset{*}{\tau}{}^2$ that follows from Corollary 7.1.

An estimate of $\rho_j(\tau^2)$ under the assumption that $p=2$ and Z_{jr} is purely gamma-distributed is obtained by setting $q=1$ identically regardless of the value of ϕ_3. The pure lognormal case is obtained by setting $q=0$.

Appendix 2. Simulation results

Pseudo-estimators are denoted by Ps, non-pseudo-estimators by Nps. The tables are further explained in Section 9.

Table B1. Claim frequency comparison of τ^2-estimates.

	Θ_j		Ps: $\widetilde{\tau}^2$				Nps: $\widetilde{\tau}^2$			
	distri-		1000	— Bias in percent —			1000	— Bias in percent —		
J	bution	Best	$\sqrt{\text{MSE}}$	Lo95	Point	Up95	$\sqrt{\text{MSE}}$	Lo95	Point	Up95
200	D1	Nps	0.54	28.4†	28.7†	29.0†	0.43	20.8†	21.0†	21.2†
200	D2	Nps	1.33	0.6	0.8	1.0	1.18	-5.2	-5.0	-4.9
200	D3	Ps	2.93	-0.3	-0.2	-0.1	2.97	-3.8	-3.7	-3.6
200	D4	Ps	5.52	-1.0	-0.9	-0.8	6.08	-3.9	-3.8	-3.7
200	D5	Ps	12.00	-2.1	-2.0	-1.9	13.94	-4.6	-4.4	-4.3
200	D6	Ps	8.54	0.8	0.9	0.9	8.89	-2.3	-2.3	-2.2
200	D7	Ps	32.12	-0.2	-0.1	0.0	35.85	-3.5	-3.4	-3.3
200	D8	Ps	70.79	-0.2	-0.1	-0.1	76.43	-4.2	-4.1	-4.0
200	D9	Ps?	169.30	-0.2	-0.2	-0.1	169.60	-5.6	-5.5	-5.4
1000	D1	Nps	0.22	12.0†	12.3†	12.5†	0.21	10.7†	11.0†	11.2†
1000	D2	Nps	0.59	-0.0	0.1	0.3	0.52	-1.1	-1.0	-0.9
1000	D3	Ps	1.31	-0.2	-0.1	0.0	1.33	-0.9	-0.8	-0.7
1000	D4	Ps	2.49	-0.3	-0.2	-0.1	2.77	-0.9	-0.7	-0.6
1000	D5	Ps	5.51	-0.5	-0.4	-0.3	6.51	-1.0	-0.9	-0.8
1000	D6	Ps	3.73	0.1	0.2	0.2	3.93	-0.5	-0.4	-0.4
1000	D7	Ps	14.14	-0.1	0.0	0.1	16.24	-0.7	-0.6	-0.5
1000	D8	Ps	31.28	-0.1	-0.1	0.0	35.12	-0.9	-0.8	-0.7
1000	D9	Ps	74.35	-0.1	-0.0	0.0	79.86	-1.3	-1.2	-1.1
2000	D1	Nps	0.16	8.5†	8.7†	9.0†	0.15	7.9†	8.2†	8.4†
2000	D2	Nps	0.41	-0.1	0.1	0.3	0.37	-0.6	-0.5	-0.3
2000	D3	Ps	0.94	-0.2	-0.1	-0.0	0.95	-0.6	-0.5	-0.4
2000	D4	Ps	1.77	-0.2	-0.1	0.0	1.97	-0.5	-0.3	-0.2
2000	D5	Ps	3.91	-0.4	-0.2	-0.1	4.56	-0.7	-0.5	-0.4
2000	D6	Ps	2.62	0.0	0.1	0.2	2.76	-0.3	-0.2	-0.2
2000	D7	Ps	9.96	-0.1	-0.0	0.0	11.43	-0.5	-0.4	-0.3
2000	D8	Ps	22.30	-0.1	0.0	0.1	25.14	-0.5	-0.4	-0.3
2000	D9	Ps	52.69	-0.1	-0.0	0.0	57.09	-0.7	-0.6	-0.5

Note: MSE is mean square deviation of estimate from true value.

Table B2. Mean claim comparison of τ^2-estimates.

Uniform conditional claim distribution

	Θ_j		Ps: $\widehat{\tau}^2$				Nps: $\widetilde{\tau}^2$			
	distri-		1000	— Bias in percent —			1000	— Bias in percent —		
J	bution	Best	$\sqrt{\text{MSE}}$	Lo95	Point	Up95	$\sqrt{\text{MSE}}$	Lo95	Point	Up95
200	D1	Nps	0.17	9.1†	9.2†	9.4†	0.13	6.2†	6.4†	6.5†
200	D2	Nps	0.77	0.2	0.3	0.5	0.69	-3.4	-3.3	-3.2
200	D3	Ps	2.38	-0.5	-0.3	-0.2	2.64	-3.6	-3.4	-3.2
200	D4	Ps	5.00	-1.2	-1.0	-0.8	5.83	-4.3	-4.1	-3.9
200	D5	Ps	11.61	-2.1	-1.9	-1.7	13.91	-5.6	-5.4	-5.2
200	D6	Ps	7.81	0.9	0.9	1.0	8.72	-2.0	-1.9	-1.8
200	D7	Ps	34.18	0.5	0.6	0.7	38.92	-4.2	-4.0	-3.9
200	D8	Nps?	87.61	2.1	2.2	2.4	86.91	-5.5	-5.3	-5.2
200	D9	Nps	295.43	8.5	8.7	8.9	209.15	-8.0	-7.8	-7.7
1000	D1	Nps	0.07	3.8†	3.9†	4.0†	0.06	3.2†	3.3†	3.4†
1000	D2	Nps	0.34	-0.1	0.0	0.1	0.30	-0.8	-0.7	-0.6
1000	D3	Ps	1.07	-0.2	-0.1	0.1	1.19	-0.9	-0.7	-0.6
1000	D4	Ps	2.26	-0.3	-0.2	-0.0	2.70	-1.0	-0.9	-0.7
1000	D5	Ps	5.25	-0.5	-0.4	-0.2	6.47	-1.4	-1.2	-1.0
1000	D6	Ps	3.35	0.1	0.2	0.2	3.83	-0.5	-0.4	-0.3
1000	D7	Ps	14.80	0.1	0.2	0.3	17.76	-1.0	-0.8	-0.7
1000	D8	Ps	36.96	0.3	0.4	0.5	41.60	-1.4	-1.3	-1.1
1000	D9	Nps	110.14	1.7	1.9	2.1	104.40	-2.1	-1.9	-1.7
2000	D1	Nps	0.05	2.6†	2.7†	2.8†	0.04	2.3†	2.4†	2.5†
2000	D2	Nps	0.24	-0.3	-0.1	0.0	0.21	-0.6	-0.5	-0.3
2000	D3	Ps	0.76	-0.1	0.0	0.1	0.86	-0.5	-0.3	-0.2
2000	D4	Ps	1.59	-0.4	-0.2	-0.1	1.87	-0.7	-0.5	-0.3
2000	D5	Ps	3.69	-0.4	-0.3	-0.1	4.57	-0.8	-0.6	-0.4
2000	D6	Ps	2.39	0.0	0.1	0.2	2.71	-0.3	-0.2	-0.1
2000	D7	Ps	10.41	-0.0	0.1	0.2	12.66	-0.5	-0.4	-0.2
2000	D8	Ps	25.99	0.1	0.2	0.4	30.13	-0.7	-0.6	-0.4
2000	D9	Ps	76.60	0.8	0.9	1.0	77.75	-1.1	-1.0	-1.0

Note: MSE is mean square deviation of estimate from true value.

Table B3. Mean claim comparison of τ^2-estimates.

Lognormal conditional claim distribution

J	Θ_j distri-bution	Best	Ps: $\widehat{\tau}^2$ 1000 $\sqrt{\text{MSE}}$	— Bias in percent — Lo95	Point	Up95	Nps: $\widetilde{\tau}^2$ 1000 $\sqrt{\text{MSE}}$	— Bias in percent — Lo95	Point	Up95
200	D1	Ps?	12.23	137.5†	145.9†	154.4†	24.36	248.5†	265.3†	282.1†
200	D2	Ps	4.08	3.0	3.7	4.5	12.26	0.6	2.9	5.2
200	D3	Ps	6.41	-1.1	-0.7	-0.3	14.02	-10.0	-9.2	-8.3
200	D4	Ps	11.02	-2.1	-1.8	-1.4	17.70	-8.0	-7.5	-6.9
200	D5	Ps	17.90	-3.5	-3.2	-3.0	24.32	-7.5	-7.2	-6.8
200	D6	Ps	13.63	0.4	0.5	0.7	18.19	-4.0	-3.8	-3.6
200	D7	Ps	44.04	-0.9	-0.7	-0.6	52.05	-5.0	-4.8	-4.6
200	D8	Ps?	98.03	-0.3	-0.2	-0.1	102.37	-6.0	-5.9	-5.8
200	D9	Nps?	469.63	2.5	2.8	3.1	237.64	-8.0	-7.8	-7.7
1000	D1	Ps	1.15	60.8†	62.7†	64.6†	7.47	156.8†	171.1†	185.3†
1000	D2	Ps	1.89	0.2	0.9	1.6	6.08	-3.3	-1.0	1.3
1000	D3	Ps	2.91	-0.5	-0.2	0.2	6.33	-3.5	-2.7	-2.0
1000	D4	Ps	4.55	-0.8	-0.5	-0.2	8.51	-2.2	-1.7	-1.2
1000	D5	Ps	8.88	-0.8	-0.6	-0.3	10.56	-1.9	-1.5	-1.2
1000	D6	Ps	5.98	-0.1	0.0	0.2	10.62	-1.0	-0.7	-0.5
1000	D7	Ps	20.27	-0.3	-0.2	-0.0	24.27	-1.3	-1.1	-0.9
1000	D8	Ps	44.47	-0.3	-0.1	0.0	51.28	-1.6	-1.4	-1.2
1000	D9	Ps?	116.23	0.5	0.6	0.7	121.14	-2.1	-2.0	-1.9
2000	D1	Ps	0.76	39.9†	41.7†	43.5†	4.24	115.4†	126.7†	137.9†
2000	D2	Ps	1.35	-0.6	0.2	0.9	7.28	-1.3	2.6	6.5
2000	D3	Ps	2.04	-0.5	-0.2	0.1	7.04	-2.2	-1.3	-0.3
2000	D4	Ps	3.24	-0.5	-0.2	0.1	5.48	-1.8	-1.3	-0.8
2000	D5	Ps	6.03	-0.7	-0.5	-0.2	7.94	-1.2	-0.8	-0.5
2000	D6	Ps	4.29	-0.0	0.1	0.3	8.70	-0.4	-0.1	0.1
2000	D7	Ps	14.46	-0.3	-0.1	0.1	16.82	-0.7	-0.5	-0.3
2000	D8	Ps	31.31	-0.2	0.0	0.2	34.18	-0.8	-0.7	-0.5
2000	D9	Ps	83.31	0.2	0.3	0.5	86.65	-1.2	-1.1	-0.9

Note: MSE is mean square deviation of estimate from true value.

Appendix 3. Use with GLM and some cautionary notes

This Appendix instructs in the use of Rapp, gives non-pseudo τ^2-estimators for several Auxiliaries, and is for the rest mostly common sense. It is not part of the published paper.

Rapp use

To compute $\widehat{\tau}^2$ for mean claim by Theorem 5.2, write Distfree within multimetod(s).

To compute $\overset{*}{\tau}{}^2$ by Corollary 7.1, write Mix within multimetod(s).

Write Gamma within multimetod(s) for setting $q = 1$ in Assumption 7.

Write Lognormal within multimetod(s) for setting $q = 0$ in Assumption 7.

As exposure e_j we can use normalized duration defined by multiplying ordinary duration with a risk premium estimate obtained from GLM analysis on some ordinary arguments like policyholder age. I.e.

$$e_j = (\text{some constant}) \times \sum(\text{estimated risk premium for ordinary arguments}) \times \text{duration},$$

where the sum is over all objects belonging to group j.

Normalized duration can be generalized to normalized sum insured under yearly risk.

See Rosenlund (2014) for an investigation into the suitability of different GLM methods.

The distribution on the two entities normalized duration and factor estimated collective risk premium is undetermined. One can multiply normalized duration with c and divide factor estimated risk premium with c for an arbitrary c – the end result will be the same. Pedagogically it is suitable to let summed normalized duration be equal to summed duration.

In Section 4 we defined k_j as the class of group j in the Auxiliary. Here we have several Auxiliaries. Hence k_j is multidimensional, the combination of Auxiliaries for group j.

We consider here risk premium. We use subscripts F and M for Θ_j, Λ_j and $\widehat{\Lambda}_j$, as in Section 8. We use these subscripts also for μ_k in Assumption 1 and its estimator $\widehat{\mu}_k$ in equation (5.6). Here k is multidimensional and μ_{k_j} is the collective mean of group j, as determined by its Auxiliaries in a multiplicative model. The $\widehat{\mu}_k$ are obtained from GLM simultaneously with factors for ordinary arguments.

Thus we define

$$Y_j = Y_{\mathsf{F}j} Y_{\mathsf{M}j}, \tag{C1}$$

$$\mu_{k_j} = \mu_{\mathsf{F}k_j} \mu_{\mathsf{M}k_j}, \tag{C2}$$

$$\widehat{\mu}_{k_j} = \widehat{\mu}_{\mathsf{F}k_j} \widehat{\mu}_{\mathsf{M}k_j}, \tag{C3}$$

$$\Theta_j = \Theta_{\mathsf{F}j} \Theta_{\mathsf{M}j}, \tag{C4}$$

$$\Lambda_j = \Lambda_{\mathsf{F}j} \Lambda_{\mathsf{M}j}, \tag{C5}$$

In the generic notation the left hand quantities pertain to either claim frequency or mean claim, but here they pertain to risk premium.

The bias-corrected risk premium factor is obtained from Section 8 the following way.

$$\widetilde{\Lambda}_j = \widehat{\Lambda}_{\mathsf{FM}j} \frac{\sum_{i=1}^{J} \sum_{r=1}^{N_i} Z_{ir}}{\sum_{i=1}^{J} e_i \widehat{\Lambda}_{\mathsf{FM}i}}. \tag{C6}$$

We use here the factor product estimates for ordinary arguments, that are part of normalized duration, as if they were deterministic constants, as in Ohlsson & Johansson (2010). This is often justifiable, since in typical applications normalized duration for a group is a sum of many approximately independent stochastic variables with a small CV compared to $\widehat{\mu}_{k_j}$.

For pricing we wish to use a predictor of $\mu_{k_j}\Theta_j$ as a risk premium factor per normalized duration for the value j of the credibility argument. It consists of the multiplicative risk premium μ_{k_j}, the same for all groups with the same values of the group properties, and the specific factor Θ_j. We have to multiply the predictor of $\mu_{k_j}\Theta_j$ with a factor product for ordinary arguments, e.g. policy holder age, in order to obtain the risk premium per ordinary duration for a specific object. The normalized duration e_j is a sum of such factor products, each multiplied with ordinary duration.

Appendix 3.1. Regression to the mean

When doing credibility analysis it is easy to succumb to the 'Regression Fallacy'. With that is meant that factor values $\widetilde{\Lambda}_j$ partly are affected by credibility levels, that really are rather normal, having randomly obtained lower or higher values than they deserve. How to correct for this is a matter of intuition. In our practice at WASA Insurance and Länsförsäkringar Alliance we have clearly found that there is a regression effect, since a subsequent analysis of credibility argument classes on new data independent of those that have been used for the grouping of classes by risk, has given flatter ladders for the factor estimates than the original analysis, if the latter has not been corrected with respect to this phenomenon. This problem with grouping by the outcome of a random variable is of course worse if credibility is not used.

To exemplify the phenomenon, assume that the groups' Λ_j take the discrete values $1, \ldots, 9$ with the distribution of exposures on Λ_j and discrete predictions $\widetilde{\Lambda}_j$ according to Table C1. The middle values of Λ_j have larger portfolio, which explains the phenomenon. Although the predicted value $\widetilde{\Lambda}_j$ is unbiased for any given Λ_j, the average Λ_j will be lower than $\widetilde{\Lambda}_j$ for high values of the latter and higher for low values.

Table C1. Portfolio distribution of predicted risk premium depending on real risk premium, and mean real risk premium per predicted risk premium value.

		Portfolio distribution of $\widetilde{\Lambda}_j$								
Λ_j	$\sum e_j$	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	0	0	0	0	0
2	60	15	30	15	0	0	0	0	0	0
3	80	0	25	30	25	0	0	0	0	0
4	100	0	0	30	40	30	0	0	0	0
5	140	0	0	0	45	50	45	0	0	0
6	100	0	0	0	0	30	40	30	0	0
7	80	0	0	0	0	0	25	30	25	0
8	60	0	0	0	0	0	0	15	30	15
9	0	0	0	0	0	0	0	0	0	0
Mean Λ_j		2.00	2.45	3.20	4.18	5.00	5.82	6.80	7.55	8.00

A way of correction is with a suitable exponent $\delta \in (0, 1)$, and k_0 such that the total sum property is kept, in the formula

$$\widetilde{\widetilde{\Lambda}}_j = k_0 \widetilde{\Lambda}_j^{\delta}. \tag{C7}$$

The order is kept, but the premium differences are smaller. Of course this is only ad hoc, but better than taking the $\widetilde{\Lambda}_j$ as they are for pricing.

Appendix 3.2. Model errors

Appendix 3.2.1. Iterations between factor estimation and credibility

As in Ohlsson & Johansson (2010), section 4.2.2, it is often suitable to iterate between the factors from the factor estimation and the credibility predictors, i.e. backfitting in the language of these authors.

It should be noted that this is due to model error. When we use normalized duration as e_j we treat factor product estimates for ordinary arguments, such as policy holder age, as deterministic constants, i.e. as having zero variance.

In practice, however, the ordinary factor product estimates often have non-zero covariances with the $\Theta_{\mathsf{F}j}$ and $\Theta_{\mathsf{M}j}$, so that their variances are positive to a degree that should be dealt with. E.g. if the groups j, where Θ_j happened to be large, have a disproportionate number of dangerous young customers, then the factor estimates for age will be affected. This can be a problem.

Now if every j has the same portfolio distribution of ordinary argument combinations the problem does not exist. Also if J is large and and every j has a small part of the portfolio the problem can be neglected, provided Assumption 2 holds for both claim frequency and mean claim. In the first case the ordinary factor estimates will be unaffected by the outcomes of Θ_j. In the second case the ordinary argument combinations will have approximately the same portfolio distribution of ordinary argument combinations for different intervals of Θ_j outcomes. This will work for the ordinary factor estimates in the same way as in the first case, approximately.

In these two cases iterations will have no (or little) effect, since ordinary factor estimates will not change (much). If neither of the two holds, we should iterate. The procedure is this.

Multiply all ordinary durations (not the normalized ones) with

$$\widehat{\Theta}_j = \widetilde{\Lambda}_j / \widehat{\mu}_{k_j}. \tag{C8}$$

for every individual object, where j is the value of the credibility argument for the object. $\widehat{\Theta}_j$ is the predicted specific factor for group j. Do a new GLM factor estimation with this duration and with all arguments, ordinary and group properties, that were part of the initial factor estimation. That gives new e_j and $\widehat{\mu}_{k_j}$ for credibility analysis, where we factor $\widehat{\mu}_{k_j}$ into $\widehat{\mu}_{\mathsf{F}k_j}$ and $\widehat{\mu}_{\mathsf{M}k_j}$ in a suitable way. Continue so e.g. five times for a total of six GLM and subsequent credibility analyses.

If the iterations have effect, this model error exists. Otherwise not. It is observable.

As well in the Ohlsson & Johansson (2010) method, model error motivates the need for iterations, since they also treat factor product estimates for ordinary arguments as deterministic.

Appendix 3.2.2. Errors in the multiplicative model

The multiplicative hypothesis of GLM log link analysis is an approximation to reality. With many arguments it is unavoidable that a multiplicative risk premium, that is fitted in the best way, will still differ, sometimes significantly, from the true risk premium for individual j. See Rosenlund (2014), section 2.1. That means that the weights $\widehat{z}_{\mathsf{F}j}$ and $\widehat{z}_{\mathsf{M}j}$ *really* should be larger, i.e. put more weight on the individual risk premium Y_j, than according to the algorithm given here. (That collides with the wish to keep down the variance of the risk premium factor predictor and hence increase its stability over time. Generally, though, it is advisable to try to get as correct risk premium estimates as possible, while stabilizing premiums with proper marketing considerations.)

This feature of the multiplicative design speaks for combining group property arguments to one. This is unless some group property values will have too few groups after combining, for then $\widehat{\mu}_{k_j}$ will be too unstable.

This model error is not observable.

Appendix 3.3. Non-pseudo estimators

When we have several Auxiliaries and a GLM model for them, the expressions for non-pseudo estimators are somewhat more involved. Set N_0 = the total number of claims and J_0 = the number of groups with claims, as in Section 7.3. For clarity we put subscripts F and M also on τ^2 and σ^2 and their estimators.

Then we have the following for claim frequency. Expression (6.6) is a special case.

$$\begin{aligned}
\widetilde{Y}_{\mathsf{F}} &= N_0 / \sum_{j=1}^{J} \widehat{\mu}_{\mathsf{F}k_j} e_j, \\
\widetilde{\tau}^2_{\mathsf{F}} &= \max\left(0, \frac{\sum_{j=1}^{J} \widehat{\mu}_{\mathsf{F}k_j} e_j (Y_{\mathsf{F}j}/\widehat{\mu}_{\mathsf{F}k_j} - \widetilde{Y}_{\mathsf{F}})^2 - (J-1)}{\sum_{j=1}^{J} \widehat{\mu}_{\mathsf{F}k_j} e_j - \sum_{j=1}^{J} \widehat{\mu}^2_{\mathsf{F}k_j} e_j^2 / \sum_{j=1}^{J} \widehat{\mu}_{\mathsf{F}k_j} e_j}\right), \text{ estimator of } \tau^2_{\mathsf{F}},
\end{aligned} \tag{C9}$$

an adaptation of (4.27) in Ohlsson & Johansson (2010) for $p = 1$.

The following, of which (7.16) is a special case, holds for mean claim.

$$\widetilde{Y}_{\mathsf{M}} = \frac{1}{N_0}\sum_{j=1}^{J} N_j Y_{\mathsf{M}j}/\widehat{\mu}_{\mathsf{M}k_j},$$

$$\widetilde{\tau}^2_{\mathsf{M}} = \max\left(0, \frac{\sum_{j=1}^{J} N_j (Y_{\mathsf{M}j}/\widehat{\mu}_{\mathsf{M}k_j} - \widetilde{Y}_{\mathsf{M}})^2 - (J_0-1)\widehat{\sigma}^2_{\mathsf{M}}}{N_0 - \sum_{j=1}^{J} N_j^2/N_0}\right), \text{ estimator of } \tau^2_{\mathsf{M}}, \qquad \text{(C10)}$$

an adaptation of (4.27) in Ohlsson & Johansson (2010) for $p = 2$.

Stig Rosenlund
Västmannagatan 93
S-113 43 Stockholm
Sweden
E-mail: stig.ingvar.rosenlund@gmail.com