Compound Poisson credibility

STIG ROSENLUND

Länsförsäkringar Alliance, Stockholm, Sweden (retired)
Västmannagatan 93, S-113 43 Stockholm, Sweden

We develop a model in multiplicative pricing integrating ordinary GLM (Generalized Linear Model) with credibility. The collective risk premium is a multiplicative expression in discrete properties of the credibility argument. Exposure is normalized duration, defined for an individual object as a number proportional to duration multiplied with estimated risk premium for ordinary arguments. The Compound Poisson distribution is assumed for claim costs, conditional on stochastic credibility parameters. Estimators of variance components within groups and between groups are developed and used, for risk premium as one entity and separately for claim frequency and mean claim. Also we use the estimation error of the collective risk premium, in a way that is an improvement at least for only one property. Pseudo-estimators are given for between-groups variance components. Simulation results indicate in which situations they are preferable over non-pseudo-estimators. We describe a pragmatic method for correction of the "regression to the mean" effect, i.e. that new analyses with independent data of groupings by risk premium mostly give flatter factor ladders than those initially obtained. Outside the model, we describe iterations between ordinary GLM analysis and credibility and how to deal with errors in the multiplicative model.

Keywords: generalized linear model; pseudo-estimator; variance component

1. Introduction

We treat here credibility in the multiplicative pricing model, where the collective risk premium is a multiplicative expression in factors for group properties. Indata are claims and exposures for some time period. Changes over time are not treated. We intend to give improved estimates under the Compound Poisson assumption.

The paper is organized as follows. Section 2 gives an overview of non-credibility multiplicative pricing models. Section 3 recapitulates credibility models found in the literature. Section 4 summarizes our model and describes the difference between it and previous credibility models. Section 5 states the notation. Section 6 treats risk premium, where Section 6.1 gives the best linear predictor BLP, defined in the L^2-norm, of the credibility factor. Sections 6.2.1 and 6.2.2 gives estimators of variance components within groups and between groups, respectively. In Section 6.3 these results are used for a credibility factor predictor. A bias correction method is also given. Section 7 gives a pseudo-estimator of the variance component between groups for claim frequency, which is optimal under certain conditions. Section 8 is devoted to mean claim analysis, where Section 8.3.2 gives a pseudo-estimator. In Section 9 the separate claim frequency and mean claim results are combined to risk premium results. Section 10 gives simulation results for the goodness of estimators of between-groups variance components. Section 11 treats the effect of regression to the mean and suggests an additional bias correction for this. Section 12 deals with model errors. Of these Section 12.1 treats iterations between GLM (Generalized Linear Model) factor estimation and credibility, while Section 12.2 deals with errors in the multiplicative design. In Section 13 we describe examples with real data, with all groups having small individual weights and with many groups have large individual weights, respectively. Appendix A gives proofs. Appendix B gives tables from simulations. and from real data with a few lines from the

Corresponding author. E-mail: stig.rosenlund@sverige.nu. Telephone: +46 730443491

examples.

2. Overview of some non-credibility multiplicative pricing models

In non-life insurance multiplicative pricing (rating) a Poisson model for claim numbers is often combined with a Gamma model with constant CV (coefficient of variation) for claim amounts. This combined model is here denoted S-GLM = Standard GLM model. The Tweedie model for claim cost with variance function $v(\mu) = \mu^p$ $(1 \leq p \leq 2)$ is also treated in the literature, although seldom used in practice by our experience.

Of previous work relaxing these variance assumptions we mention the following three papers. Smyth & Jørgensen (2002) analyzed a model with Poisson distributed claim numbers and Gamma-distributed claim amounts, such that both risk premiums and dispersion parameters vary over tariff cells and obey GLM models. The log link gives multiplicative risk premiums and dispersion parameters. This link is used for examples. Aggregated data giving claim numbers and total claim costs per tariff cell are used for inference.

Heller et al. (2007) treated models where claim numbers are Poisson, zero-inflated Poisson or negative binomial and claim amounts are normal, Gamma or inverse Gaussian. GLM models are assumed for claim frequency, mean claim, claim number dispersion and claim amount dispersion. Log links are used for the examples. Inference uses the vector of claim costs for individual customers over some time period.

Rosenlund (2014) studied the performance of MMT (Method of Marginal Totals), and a new non-parametric variance estimate method for MMT called MVW (MMT Variance estimates under Weak assumptions), under some moderate deviations from multiplicativity of claim frequency and mean claim and from the S-GLM assumption of constant claim amount CV. Overall the conclusion was that MMT with MVW is better than S-GLM and Tweedie for consumer insurance with many arguments and many claims. The MVW variance estimate method only presupposes that exposure is not extremely non-multiplicative. Dispersion is *not* assumed multiplicative, as in Smyth & Jørgensen (2002) and Heller et al. (2007). Such an assumption would be strong. Assume e.g. that a particular combination of argument values is prone to large and variable claims, in such a way that no multiplicative dispersion model can be reasonably fitted. We can think of young men in urban areas as being prone to such claims, while all other combinations - including young men in rural areas - have multiplicative dispersions. Such portfolios would not be uncommon. Then the S-GLM, Tweedie, Smyth & Jørgensen (2002) and Heller et al. (2007) variance assumptions certainly fail, while the MVW method gives approximately correct confidence intervals. Or assume that class x in argument 1 and/or class y in argument 2 are especially prone to large and variable claims, with no extra variability for the combination of x and y. The S-GLM, Tweedie, Smyth & Jørgensen (2002) and Heller et al. (2007) variance assumption would then assign a too high variance estimate for the combination of class x in argument 1 and class y in argument 2, while the MVW estimate works. Such an example was treated in Rosenlund (2014). Data for inference with MMT and MVW are aggregated claim numbers, claim amounts and claim amount squares per tariff cell.

There are good reasons for admitting arbitrarily distributed claim amounts. For example, large claims in the long tail do not always or even mostly admit parametric distributions as those used by Smyth & Jørgensen (2002) and Heller et al. (2007).

The non-Poisson claim number models of Heller et al. (2007) are unnecessarily complicated when dealing with insurance lines with many independent customers, each contributing a small percentage of the total claim number. The specific car insurance data used in Heller et al. (2007) are from such an insurance line. See Grigelionis (1963). Rosenlund (2010) gives perspectives on Poisson approximation in such cases and on macroscopic fluctuations affecting large parts of the portfolio in the same way. See Barbour et al. (1992) for a modern treatment.

An obvious drawback of the Smyth & Jørgensen (2002) inference method is that no information at all on observed claim amount variability within tariff cells is used, as can be realized by considering the example of a 2×2 tariff with four tariff cells. The Heller et al. (2007) inference method normally uses more such information. But suppose that there is only one single customer per tariff cell in the 2×2 example. Then no such information is used. By contrast, inclusion of the sums of claim amount squares in our method's data for inference makes full use of observed claim amount variability within tariff cells, given a non-parametric model.

3. Overview of some credibility models

We use here the term credibility argument for one where some classes have too few claims to admit basing the premium on the class alone. Credibility analysis must be used for this argument.

Bühlmann & Straub (1970) gave the Bühlmann-Straub estimator for a non-parametric credibility model with a credibility argument and no other rating factors. Campbell (1986) combined the credibility argument with a grouping of it by properties such as those we consider in the next section, 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. See below for a definition. For the Bühlmann-Straub model De Vylder (1996) III, Chapter 3, Section 3.4.7, gives improved estimators using pseudo-estimators credited to Bichsel & Straub. Ohlsson (2008) and Ohlsson & Johansson (2010), Chapter 4, treat a setting with multiplicative rating factors including the credibility argument and group properties for this. This is our setting, but these papers do not make the Compound Poisson assumption.

No parametric distributional assumptions for the claim cost of a combination of arguments, including the credibility one, are made in these studies. The data used are aggregated claim numbers, claim costs and exposures per combination of arguments and contract observations.

Cipra (1996), among others, treats very large claims (outliers) and missing observations. Bühlmann & Gisler (2005) give an overview of credibility. They do not treat pseudo-estimators. Garrido & Zhou (2009) study credibility of estimators obtained with GLM and GLMM (mixed GLM) in order to ascertain the needed number of claims for tariff calculation.

4. Summary of model and method

We deal with any number of ordinary non-credibility rating factors appearing in analyses described in Section 2, such as policy holder age, insured object age and year of construction. In addition we have a credibility argument. We assume that the claim cost of any combination of arguments, including the credibility one, has the Compound Poisson distribution, conditional on stochastic credibility parameters.

Given a number of credibility argument groups (e.g. geographical parishes) of policy holders, a risk premium is to be computed that weighs together the group's own risk premium with a multiplicatively computed risk premium, where the arguments are different group properties, e.g. median income and population density per parish. The properties might have been purchased from external sources and are supposed to be available as arguments in the data.

We use the word group instead of class to distinguish the credibility argument from ordinary arguments. Group properties are called Auxiliaries in Ohlsson & Johansson (2010).

With the Compound Poisson assumption, multiplicative risk premiums and in addition four natural assumptions **A1 - A4** given below, we will develop variance estimators yielding optimal credibility factors for the credibility argument, treating risk premium as one entity. Bias corrections and iterations between ordinary argument GLM factor estimation and credibility are part of the method. Also we give separate treatments of claim frequency and mean claim, with corresponding assumptions, which are combined to risk premium results.

The rating factors for ordinary arguments are preferably estimated in the same multiplicative analysis as the rating factors for the group properties. If that is impossible in practice, e.g. if data for ordinary arguments are to be sent to an external party which then performs multiclass analysis, then the rating factors for ordinary arguments can be estimated first, after which those factors are used for normalizing in a second run in order to predict the rating factors for the group properties. Then the iterative part of our method cannot be used, either. This was the case in Sweden for motor insurance before deregulation in the 1990s, when the Swedish Actuarial Research Board was the external party where the Campbell (1986) study was made.

The Compound Poisson assumption, suitable at least for mass consumer insurance, makes possible more precise estimators and predictions than previous credibility methods. Namely, we will rely on the variance formula $\mathrm{Var}[\sum_{i=1}^{N} Z_i] = \mathrm{E}[\sum_{i=1}^{N} Z_i^2]$ for N Poisson distributed and Z_i IID and independent of N. This frees us from the problem of defining a contract observation. Ohlsson & Johansson (2010) treat claim frequency and mean claim separately, but in the other models contracts with several claims in an observation period will enter the calculations with the sum of claim cost in the period. Thus information is lost. Setting a smaller period length separates claims more, but computations will degenerate with too small lengths. For the Ohlsson & Johansson (2010) model this drawback applies to claim frequency only.

The data for inference are the same as those described for the MVW method in Section 2. Namely aggregated claim numbers, claim amounts and claim amount squares per tariff cell, where a tariff cell now means a combination of arguments, including the credibility one. (The actual tariff might merge several credibility groups into one, making the number of combinations handled in insurance sales smaller than the number of tariff cells.) Our additional use of aggregated squares is sufficient under Compound Poisson and makes our method possible.

We assume that very large claims have been appropriately truncated. Also we assume that detailed data bases are available so that no observations are missing. Thus we can make do with simpler models than the one of e.g. Cipra (1996).

Like Dannenburg et al. (1996) and Ohlsson & Johansson (2010) we define risk parameters as real variables in variance component models, not as functions of abstract risk variables.

To the best of our knowledge our results for the Compound Poisson model integrated with GLM, together with the rather weak other conditions we impose in certain places, are new, unless otherwise stated, e.g. in the section on mean claim analysis.

Free program for credibility by this paper and by other methods, GLM for non-life insurance pricing, claim reserve algorithms, etc.: `www.stigrosenlund.se/rapp.htm`.

5. Notation for observables

Normalized duration for some time period and group j is

w_j = (some constant)$\times\sum$(estimated risk premium for ordinary arguments)$\times$duration.

(Risk premium constant estimate)$\times$(estimated factor product for group properties) is GLM factor estimated collective risk premium for a group and denoted $\hat{\mu}_j$ below. The factors are obtained with normalized duration as exposure. Normalized duration can be generalized to normalized sum insured under yearly risk. All formulas will be the same. See Rosenlund (2014) for an investigation into the suitability of different GLM methods.

An example of an ordinary argument is policy holder age. An example of a group property is population density for the credibility argument geographical parish.

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.

We define the following observables, where $J, \ldots, w_{jr}$ are considered deterministic.

$$J \quad = \quad \text{number of groups,} \tag{5.1}$$

$$R \quad = \quad \text{number of group properties,} \tag{5.2}$$

$$k_{jr} \quad = \quad \text{the class of group } j \text{ in group property } r,\ \ j \in \{1, \ldots, J\},\ \ r \in \{1, \ldots, R\}, \tag{5.3}$$

$$
\begin{aligned}
w_j &= \text{normalized duration group } j, && (5.4)\\
w_{jr} &= \sum_{i:k_{ir}=k_{jr}} w_i = \text{sum of normalized duration over group } j\text{'s value in group property } r, && (5.5)\\
X_j &= \text{claim cost group } j, && (5.6)\\
X_{jr} &= \sum_{i:k_{ir}=k_{jr}} X_i = \text{sum of claim cost } X_i \text{ over group } j\text{'s value in group property } r, && (5.7)\\
X &= \sum_{i=1}^{J} X_i = \text{total claim cost}, && (5.8)\\
Q_j &= \text{sum of squared claim amounts for individual claims group } j, && (5.9)\\
Q &= \sum_{i=1}^{J} Q_i = \text{total sum of squares}, && (5.10)\\
Y_j &= X_j/w_j = \text{observed risk premium group } j, && (5.11)\\
\hat{\mu}_j &= \text{factor estimated collective risk premium group } j \text{ for group properties.} && (5.12)
\end{aligned}
$$

We now define observables for claim frequency and mean claim. Parameters etc. are introduced in Sections 7 and 8. Subscript F and M denote Frequency and Mean, respectively.

$$
\begin{aligned}
N_j &= \text{number of claims group } j, && (5.13)\\
J_1 &= \sum_{j=1}^{J} \mathbf{1}_{\{N_j>0\}} = \text{number of groups with claims}, && (5.14)\\
N_0 &= \sum_{j=1}^{J} N_j = \text{total number of claims}, && (5.15)\\
N_{jr} &= \sum_{i:k_{ir}=k_{jr}} N_i = \text{sum of numbers } N_i \text{ over group } j\text{'s value in group property } r, && (5.16)\\
Z_{ji} &= \text{individual claim amounts in group } j \in \{1,\dots,J\},\ \ i \in \{1,\dots,N_j\}, && (5.17)\\
Y_{\mathsf{F}j} &= N_j/w_j = \text{observed claim frequency group } j, && (5.18)\\
Y_{\mathsf{M}j} &= X_j/N_j = \sum_{i=1}^{N_j} Z_{ji}/N_j = \text{observed mean claim group } j, && (5.19)\\
\hat{\mu}_{\mathsf{F}j} &= \text{factor estimated collective claim frequency group } j \text{ for group properties}, && (5.20)\\
\hat{\mu}_{\mathsf{M}j} &= \text{factor estimated collective mean claim group } j \text{ for group properties.} && (5.21)
\end{aligned}
$$

All j with the same k_{jr} have the same w_{jr}, X_{jr} and N_{jr}. For one group property, where $R = 1$, $\hat{\mu}_j$ is the univariate risk premium for group property class k_{j1}, namely X_{j1}/w_{j1}. If there are no group properties we set $R = 1$ and $k_{j1} \equiv 1$. Then we have $X_{j1} = X,\ \ w_{j1} = \sum_{i=1}^{J} w_i$ and $\hat{\mu}_j$ is the total observed risk premium.

6. Risk premium

We start with risk premium and formulate four assumptions, where **A4** is a variance function assumption that is often needed to decrease the variability of the variance estimators.

A1. Conditional on stochastic variables Θ_j ($j = 1,\dots,J$), with expectation $\mathrm{E}[\Theta_j] = 1$ and variance $\mathrm{Var}[\Theta_j] = \tau^2$, X_j has a Compound Poisson distribution with expectation $\mu_j \Theta_j w_j$, where μ_j is multiplicative in the group properties.

A2. $(\Theta_1, X_1, Q_1), \ \dots \ , (\Theta_J, X_J, Q_J)$ are independent.

A3. $\mathrm{E}[\hat{\mu}_j] = \mu_j$.

A4. $\mathrm{Var}[Y_j \mid \Theta_j] \ = \ \phi(\mu_j \Theta_j)^p / w_j$ for $\phi > 0$ and $1 \leq p \leq 2$, and $\mathrm{E}[\Theta_j^p]$ is independent of j.

Objective. To predict $\mu_j \Theta_j$ as well as possible.

We use the words predict and predictor, by the convention to reserve the word estimator for guesses on non-stochastic parameters (the Bayesian formulation set aside). Our predictors are guesses on stochastic variable outcomes that occurred in the past and can never be observed.

The distributional assumption of Compound Poisson in **A1** is stronger than the distribution-free assumptions in Ohlsson & Johansson (2010). For motivations, see Section 1 of this paper, Grigelionis (1963), Barbour et al. (1992), Rosenlund (2010) and Rosenlund (2014).

Assumption **A3** requires sufficiently many claims to be approximately correct for $R > 1$. For $R = 1$ it is always correct. The Ohlsson & Johansson (2010) assumptions mean that μ_j, which pertains to Auxiliaries in the language of these authors, is known and equal to $\hat{\mu}_j$. The estimator $\hat{\mu}_j$ is simply plugged into the formulas in lieu of μ_j. Here that would correspond to the assumption that $\mathrm{Var}[\hat{\mu}_j] = 0$. Now $\mathrm{Var}[\hat{\mu}_j]$ can often be assumed to be small insofar as it effects the credibility predictors. It can be shown to be of about the same magnitude as $\mathrm{Cov}(\hat{\mu}_j, Y_j)$, so if one is included the other one should also be. See equation (6.15). If $R = 1$ we can compute and estimate these quantities as well as $\mathrm{Cov}(\hat{\mu}_j, \Theta_j)$. We will use these estimators also for $R > 1$, believing that they anyway will entail an improvement of credibility predictors.

For ordinary arguments, however, such as policy holder age and year of construction, we use here the factor product estimates 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 $\hat{\mu}_j$. See Section 12 for more deliberations on this topic.

A4 is similar to (4.6) in Ohlsson & Johansson (2010). We can leave out this assumption with enough claims per group. That is shown below. Variance function assumptions concerning the outcomes for different argument classes are strong, and if every class has enough data for a variance estimate only from the claims of the class, then such an estimate is preferable.

The parameter p in assumption **A4** must be fixed initially. If the claim frequency ladder $\hat{\mu}_{\mathsf{F}j}$ is as steep as the mean claim ladder $\hat{\mu}_{\mathsf{M}j}$ (in ascending orders), loosely speaking, then $p = 1.5$ is suitable. If it is less steep p should be closer to 2, and if more steep closer to 1.

We now recapitulate parameters and functionals given above and define new ones.

$$\mu_j \quad = \quad \mathrm{E}[Y_j] \qquad \text{risk premium group } j, \tag{6.1}$$

$$\Lambda_j \quad = \quad \mu_j \Theta_j \qquad \text{risk premium group } j \text{ conditional on } \Theta_j, \tag{6.2}$$

$$\sigma^2 \quad = \quad \phi \mathrm{E}[\Theta_j^p] \qquad \text{assumed independent of } j, \tag{6.3}$$

$$\sigma_j^2 \quad = \quad \mathrm{E}[\mathrm{Var}[Y_j \mid \Theta_j]], \tag{6.4}$$

$$\tau^2 \quad = \quad \mathrm{Var}[\Theta_j] \qquad \text{assumed independent of } j, \tag{6.5}$$

$$\nu_j^2 \quad = \quad \frac{1}{{w_{j1}}^2} \sum_{i:k_{i1}=k_{j1}} w_i^2(\sigma_i^2 + \mu_j^2 \tau^2) \ \text{ equal to} \mathrm{Var}[\hat{\mu}_j] \text{ when } R = 1. \tag{6.6}$$

Our predictors and estimators are

$$\Lambda_j^* \quad = \quad z_j Y_j + (1 - z_j)\mu_j \quad \text{non-observable predictor of } \Lambda_j, \text{ see (6.15)}, \tag{6.7}$$

$$\widehat{\Lambda}_j \quad = \quad \widehat{z}_j Y_j + (1 - \widehat{z}_j)\hat{\mu}_j \quad \text{estimated predictor of } \Lambda_j, \text{ see (6.20)}, \tag{6.8}$$

$$\begin{aligned}
\widehat{\sigma}^2 &= Q\left(\sum_{j=1}^{J} w_j \hat{\mu}_j^p\right)^{-1}, && (6.9)\\
\overset{*}{\sigma}{}_j^2 &= Q_j/w_j^2, && (6.10)\\
\widehat{\sigma}_j^2 &= \widehat{\sigma}^2 \hat{\mu}_j^p / w_j, && (6.11)\\
\widetilde{\sigma}_j^2 &= q\,\overset{*}{\sigma}{}_j^2 + (1-q)\widehat{\sigma}_j^2,\quad q \in [0,1],\ \text{for a suitable } q, && (6.12)\\
\hat{\tau}^2 &= \text{pseudo-estimator of } \tau^2,\ \text{see (6.18) in Section 6.2.2}, && (6.13)\\
\hat{\nu}_j^2 &= \frac{1}{{w_{j1}}^2} \sum_{i:k_{i1}=k_{j1}} w_i^2(\widetilde{\sigma}_i^2 + \hat{\mu}_j^2 \hat{\tau}^2)\ \text{ estimator of } \mathrm{Var}[\hat{\mu}_j] \text{ when } R=1. && (6.14)
\end{aligned}$$

6.1. Best linear predictor of the credibility factor

For pricing we wish to use a predictor of $\mu_j\Theta_j$ as a risk premium factor for the value j of the credibility argument. It consists of the multiplicative risk premium μ_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 $\mu_j\Theta_j$ with a factor product for ordinary arguments, e.g. policy holder age, in order to obtain the risk premium for a specific object. The normalized duration w_j is a sum of such factor products, each multiplied with ordinary duration.

We seek the best linear predictor of $\mu_j\Theta_j$ in L^2-norm in the form (6.7), i.e. z_j is to be determined so that $\mathrm{E}[(\Lambda_j^* - \mu_j\Theta_j)^2]$ is minimized. Then Λ_j^* is the BLP, provided the BLP has the form (6.7). We show in Appendix A that this form holds.

Theorem 6.1. *For the BLP in L^2-norm of $\mu_j\Theta_j$ it holds*

$$z_j = \frac{\mathrm{Var}[\hat{\mu}_j] + \mu_j^2\tau^2 - \mathrm{Cov}(\hat{\mu}_j, Y_j) - \mu_j \mathrm{Cov}(\hat{\mu}_j, \Theta_j)}{\sigma_j^2 + \mathrm{Var}[\hat{\mu}_j] + \mu_j^2\tau^2 - 2\mathrm{Cov}(\hat{\mu}_j, Y_j)}. \tag{6.15}$$

Here σ_j^2 is a variance component for variation within group j and τ^2 a variance component for variation between the groups. And $\mathrm{Var}[\hat{\mu}_j]$ is a component that gives the estimation variance in the multiplicative model. In the following we shall give general estimators of the two first variance components which can be inserted in the expression (6.15).

If w_j/w_{jr} is small enough for all r, then $\hat{\mu}_j$ is approximately independent of (Θ_j, X_j, Q_j) and $\mathrm{Var}[\hat{\mu}_j]$, $\mathrm{Cov}(\hat{\mu}_j, Y_j)$ and $\mathrm{Cov}(\hat{\mu}_j, \Theta_j)$ will be small in absolute value. In applications it will not make sense to have few groups in some group property value. For instance, if the class Population density High consists only of a couple of groups (parishes), then the factor for the class will be too unstable.

Generally it is difficult to obtain computable estimators of $\mathrm{Var}[\hat{\mu}_j]$, $\mathrm{Cov}(\hat{\mu}_j, Y_j)$ and $\mathrm{Cov}(\hat{\mu}_j, \Theta_j)$, since GLM solutions have complicated structures unless unrealistic conditions are imposed. But for the special case one group property, estimates can be computed.

Corollary 6.1. *For the case $\hat{\mu}_j = X_{j1}/w_{j1}$, i.e. only one group property, we have*

$$z_j = \frac{\nu_j^2 + \mu_j^2\tau^2 - \dfrac{w_j}{w_{j1}}\sigma_j^2 - \dfrac{2w_j}{w_{j1}}\mu_j^2\tau^2}{\sigma_j^2 + \nu_j^2 + \mu_j^2\tau^2 - \dfrac{2w_j}{w_{j1}}\sigma_j^2 - \dfrac{2w_j}{w_{j1}}\mu_j^2\tau^2}. \tag{6.16}$$

6.2. Variance estimators

6.2.1. Estimators of variance components within a group

Lemma 6.1. *Given* **A1** - **A4** *and* $\widehat{\sigma}^2$ *by* (6.9), $\mathrm{E}[\widehat{\sigma}^2] \approx \sigma^2$.

Lemma 6.2. *Given* **A1** *and* $\overset{*}{\sigma}{}_j^2$ *by* (6.10), $\mathrm{E}[\overset{*}{\sigma}{}_j^2] = \sigma_j^2$.

Lemma 6.3. *Given* **A1** - **A4** *and* $\widehat{\sigma}_j^2$ *by* (6.11), $\mathrm{E}[\widehat{\sigma}_j^2] \approx \sigma_j^2$.

A weighting together of the two σ_j^2-estimators by expression (6.12) is also approximately unbiased, i.e. $\mathrm{E}[\widetilde{\sigma}_j^2] \approx \sigma_j^2$. Under the assumption **A4**, $\widehat{\sigma}_j^2$ shall be used since it has less variance, i.e. we set $q = 0$. If we desire to put a certain weight on the belief that **A4** is not true and the estimator $\overset{*}{\sigma}{}_j^2$ can be believed to have a moderately large variance, then we set $q > 0$, sometimes $q = 1$. This can be the case for insurance lines where claim amounts are not very variable in size, and for all j the number of claims N_j is not too small.

6.2.2. Pseudo-estimator of variance component between groups

Let

$$c_j = \frac{\hat{\mu}_j^2 \hat{\tau}^2}{\widetilde{\sigma}_j^2 + \hat{\mu}_j^2 \hat{\tau}^2} = \frac{\hat{\tau}^2}{\widetilde{\sigma}_j^2 / \hat{\mu}_j^2 + \hat{\tau}^2}, \tag{6.17}$$

$$\hat{\tau}^2 = \frac{1}{J} \sum_{j=1}^{J} c_j \left(\frac{Y_j}{\hat{\mu}_j} - 1 \right)^2. \tag{6.18}$$

Here $\hat{\tau}^2$ is a Bichsel-Straub type pseudo-estimator in the sense that (6.17) and (6.18) contain $\hat{\tau}^2$. They define an equation that must be solved numerically. We seek a positive $\hat{\tau}^2$ satisfying it, if possible. If not, then $\hat{\tau}^2 = 0$. De Vylder (1996) describes an iterative pseudo-estimator solution, but a simple binary search suffices. We solve the zero of the function of $\hat{\tau}^2$ that is given by the left side of (6.18) minus the right side. See Appendix A.12. Successive interval halvings quickly yield the solution. All pseudo-estimators in this paper are solved analogously.

We define excess $\mathrm{e}(\cdot)$ as 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

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

This gives

$$\mathrm{Var}[(Z-\mu)^2] = \mathrm{E}[(Z-\mu)^4] - \mathrm{E}[(Z-\mu)^2]^2 = [\mathrm{e}(Z)+2]\mathrm{E}[(Z-\mu)^2]^2 = [\mathrm{e}(Z)+2]\mathrm{Var}[Z]^2. \tag{6.19}$$

Lemma 6.4. *Assume* **A1** - **A3**. *If* $q < 1$ *assume also* **A4**. *With* $\widetilde{\sigma}_j^2$ *by* (6.12) *and* $\hat{\mu}_j$ *by* (5.12), *it holds* $\mathrm{E}[\hat{\tau}^2] \approx \tau^2$. *If* $\mathrm{e}(Y_j)$ *is constant, then* $\hat{\tau}^2$ *is approximately optimal.*

We write approximately, since true parameters are replaced by estimates in (6.17) and (6.18). Also, it can be shown that $\mathrm{e}(Y_j)$ will grow to infinity as w_j goes to 0, implying that the weights c_j will be too large for small w_j, although c_j will tend to 0 with w_j. Hence some optimality is lost, but $\hat{\tau}^2$ might be better than any practically feasible estimator when we treat risk premium as one entity. See Sections 7 and 8.3 for separate claim frequency and mean claim analyses. They are combined in Section 9.

Note that $\hat{\tau}^2$ is always non-negative, while the classical Bühlmann-Straub estimator of a variance component corresponding to τ^2 is not. In De Vylder (1996) $\mathrm{A}_{\mathrm{pseu}}$ is a pseudo-estimator and A is the classical estimator. Sections 3.3.7 and 3.4.5 have clarifying discussions of the bias of A related to the non-negativity of the variance component and the fact that A can furnish negative values.

6.3. *Credibility factor estimator for predictor and correction for bias*

We have now developed estimators that can be put into the right side of equation (6.15), with the exception of the variance $\mathrm{Var}[\hat{\mu}_j]$ and the covariances $\mathrm{Cov}(\hat{\mu}_j, Y_j)$ and $\mathrm{Cov}(\hat{\mu}_j, \Theta_j)$. For w_j/w_{jr} small enough for all r, these can be omitted, and c_j by (6.17) can serve as an estimator of z_j. Its form is recognized as a variation of the form of the classical Bühlmann-Straub estimator.

When $R = 1$ and $\hat{\mu}_j = X_{j1}/w_{j1}$ we have computable estimators of $\mathrm{Var}[\hat{\mu}_j]$, $\mathrm{Cov}(\hat{\mu}_j, Y_j)$ and $\mathrm{Cov}(\hat{\mu}_j, \Theta_j)$ by Corollary 6.1 and its proof in Appendix A.2. We shall use this expression also for the general case. (6.17) and (6.20) are approximately equal when w_j/w_{jr} is small enough for all r, and it is plausible that (6.20) is a better approximation than (6.17) also when $\hat{\mu}_j \not\equiv X_{j1}/w_{j1}$.

Theorem 6.2. *An estimated credibility factor using (6.16) is*

$$\widehat{z}_j = \frac{\hat{\nu}_j^2 + \hat{\mu}_j^2\left(1 - \dfrac{2w_j}{w_{j1}}\right)\hat{\tau}^2 - \dfrac{w_j}{w_{j1}}\widetilde{\sigma}_j^2}{\widetilde{\sigma}_j^2 + \hat{\nu}_j^2 + \hat{\mu}_j^2\left(1 - \dfrac{2w_j}{w_{j1}}\right)\hat{\tau}^2 - \dfrac{2w_j}{w_{j1}}\widetilde{\sigma}_j^2}. \tag{6.20}$$

We have thus determined the predictor $\widehat{\Lambda}_j = \widehat{z}_j Y_j + (1 - \widehat{z}_j)\hat{\mu}_j$ by (6.8) depending on variance estimates. For the non-observable predictors the following unbiasedness holds for the total.

$$\mathrm{E}[\sum_{j=1}^{J} \Lambda_j^* w_j] = \sum_{j=1}^{J}[z_j\mu_j + (1 - z_j)\mu_j]w_j = \sum_{j=1}^{J}\mu_j w_j = \mathrm{E}[\sum_{j=1}^{J} Y_j w_j] = \mathrm{E}[X]. \tag{6.21}$$

The replacement of z_j and μ_j with estimates $\widehat{z}_j$ and $\hat{\mu}_j$ can cause a bias. (In the examples of Section 13 the bias estimate is negligible. This is the case also for separate claim frequency and mean claim analyses, which are combined to risk premium.) In general we have $\sum_{j=1}^{J}\widehat{\Lambda}_j w_j \neq X$, although the total sum property holds for Y_j and – if the MMT method was used – $\hat{\mu}_j$, namely: $\sum_{j=1}^{J} Y_j w_j = \sum_{j=1}^{J}\hat{\mu}_j w_j = X$. A correction factor gives a new estimator

$$\widetilde{\Lambda}_j = \widehat{\Lambda}_j X\left(\sum_{i=1}^{J}\widehat{\Lambda}_i w_i\right)^{-1}. \tag{6.22}$$

7. Claim frequency

All derivations for risk premium are transferable, with X_j and Q_j replaced by N_j. We use the notation for risk premium, adding the subscript F for Frequency. E.g. $Y_{\mathsf{F}j} = N_j/w_j$. The assumptions **A1F - A4F** are the same as **A1 - A4**, after replacing X_j and Q_j with N_j and Compound Poisson with just Poisson.

Here ϕ_{F} and p_{F} are 1 and by (6.3) then $\sigma_{\mathsf{F}}^2 = 1$ and $\sigma_{\mathsf{F}j}^2 = \mu_{\mathsf{F}j}/w_j$. We should obtain $\hat{\mu}_{\mathsf{F}j}$ with GLM Poisson log link. Then by (6.9) $\widehat{\sigma}_{\mathsf{F}}^2 = 1$ and by (6.11) $\widehat{\sigma}_{\mathsf{F}j}^2 = \hat{\mu}_{\mathsf{F}j}/w_j$. Since **A4F** holds for claim numbers in the Compound Poisson model given **A1F**, **A2F** and **A3F**, we will use $\widetilde{\sigma}_{\mathsf{F}}^2 = \widehat{\sigma}_{\mathsf{F}}^2$, thus setting $q_{\mathsf{F}} = 0$. Then $\widetilde{\sigma}_{\mathsf{F}j}^2/\hat{\mu}_{\mathsf{F}j}^2 = 1/(\hat{\mu}_{\mathsf{F}j}w_j)$.

The assumption of constant excess fails for the mixed Poisson variable N_j, so that (6.18) will give too large weights for small w_j. To evade this problem we will give another pseudo-estimator of τ_{F}^2 under the reasonably weak conditions that the 3:rd central moment and the excess of $\Theta_{\mathsf{F}j}$ are 0. This is true if $\Theta_{\mathsf{F}j} \sim N(1, \tau_{\mathsf{F}}^2)$. Now $\Theta_{\mathsf{F}j}$ is non-negative, but the normal distribution can hold approximatively with a τ_{F}^2 giving a small probability to negative values in $N(1, \tau_{\mathsf{F}}^2)$.

See Appendix A.12.2 for the solution of equations (7.2) and (7.3). We have not been able to prove that there is exactly one solution $\hat{\tau}_{\mathsf{F}}^2$, only that there is at least one, which can be 0. For definiteness we thus define the solution as the largest solution of possibly more than one.

We give long formulas in order to be completely unambiguous. The factor $[\cdots]^{-1}$ of (7.2) serves to make $c_{\mathsf{F}1} + \ldots + c_{\mathsf{F}J}$ equal to 1.

Let

$$u(x,y) \;=\; \frac{1}{y^3} + \frac{7x+2}{y^2} + \frac{4x}{y} + 2x^2, \tag{7.1}$$

and

$$c_{\mathsf{F}j} \;=\; \left(\frac{1}{\hat\mu_{\mathsf{F}j}w_j} + \hat\tau_{\mathsf{F}}^2\right)^2 u(\hat\tau_{\mathsf{F}}^2, \hat\mu_{\mathsf{F}j}w_j)^{-1}\left[\sum_{i=1}^{J}\left(\frac{1}{\hat\mu_{\mathsf{F}i}w_i} + \hat\tau_{\mathsf{F}}^2\right)^2 u(\hat\tau_{\mathsf{F}}^2, \hat\mu_{\mathsf{F}i}w_i)^{-1}\right]^{-1}, \tag{7.2}$$

$$\hat\tau_{\mathsf{F}}^2 \;=\; \sum_{j=1}^{J} c_{\mathsf{F}j}\hat\tau_{\mathsf{F}}^2\left(\frac{1}{\hat\mu_{\mathsf{F}j}w_j} + \hat\tau_{\mathsf{F}}^2\right)^{-1}\left(\frac{Y_{\mathsf{F}j}}{\hat\mu_{\mathsf{F}j}} - 1\right)^2. \tag{7.3}$$

Lemma 7.1. *Assume* **A1F** - **A3F**. *Then also* **A4F** *holds. For the pseudo-estimator* $\hat\tau_{\mathsf{F}}^2$ *we have* $\mathrm{E}[\hat\tau_{\mathsf{F}}^2] \approx \tau_{\mathsf{F}}^2$. *If* $\mathrm{E}[(\Theta_{\mathsf{F}j}-1)^3] = 0$ *and* $\mathrm{e}(\Theta_{\mathsf{F}j}) = 0$, *then* $\mathrm{Var}[(Y_{\mathsf{F}j}/\mu_{\mathsf{F}j} - 1)^2] = u(\tau_{\mathsf{F}}^2, \mu_{\mathsf{F}j}w_j)$ *and* $\hat\tau_{\mathsf{F}}^2$ *is approximately optimal in the sense of having the smallest variance of estimators in the form (7.3) with* $c_{\mathsf{F}1} + \ldots + c_{\mathsf{F}J} = 1$.

8. Mean claim

For mean claim we condition on the claim numbers N_j. In assumption **A4M** and in (8.4) we condition on $N_j > 0$ as well. This conditioning will be implicit. It is not written out below. In the notation we add the subscript M for Mean claim.

To a large extent we follow Ohlsson & Johansson (2010), the differences being the covariance terms in Theorem 8.1 and the pseudo-estimator for the between-groups variance component given by (8.22) and (8.23). We will not deduce a pseudo-estimator for mean claim itself, since the properties of the multiplicative solutions intertwined with credibility can cause unrobustness.

For only one group property $\hat\mu_{\mathsf{M}j}$ is the simple univariate mean claim for group property class k_{j1}, namely X_{j1}/N_{j1}. Without group properties $\hat\mu_{\mathsf{M}j}$ is the total observed mean claim.

In Rosenlund (2014) we motivated why the MMT solution should be applied to risk premium in many applications. A consequence is that the mean claim factors should be the ratios between the corresponding factors for risk premium and claim frequency in those applications. Then $\hat\mu_{\mathsf{M}j}$ is obtained that way. In other applications the GLM gamma log link solution is preferable, and then $\hat\mu_{\mathsf{M}j}$ is that one. The specific solution will be indeterminate and will not affect our formulas.

We now formulate four assumptions like those for risk premium.

A1M. Conditional on stochastic variables $\Theta_{\mathsf{M}j}$ $(j = 1, \ldots, J)$, with expectation $\mathrm{E}[\Theta_{\mathsf{M}j}] = 1$ and variance $\mathrm{Var}[\Theta_{\mathsf{M}j}] = \tau_{\mathsf{M}}^2$, for any specific j the Z_{ji} are IID with expectation $\mu_{\mathsf{M}j}\Theta_{\mathsf{M}j}$, where the $\mu_{\mathsf{M}j}$ is multiplicative in the group properties.

A2M. $(\Theta_{\mathsf{M}1}, X_1, Q_1)$, $\ldots$, $(\Theta_{\mathsf{M}J}, X_J, Q_{\mathsf{M}J})$ are independent.

A3M. $\mathrm{E}[\hat\mu_{\mathsf{M}j}] = \mu_{\mathsf{M}j}$.

A4M. $\mathrm{Var}[Y_{\mathsf{M}j}|\Theta_{\mathsf{M}j}] = \phi_{\mathsf{M}}(\mu_{\mathsf{M}j}\Theta_{\mathsf{M}j})^{p_{\mathsf{M}}}/N_j$ for $\phi_{\mathsf{M}} > 0$ and $1 \le p_{\mathsf{M}} \le 2$, and $\mathrm{E}[\Theta_{\mathsf{M}j}^{p_{\mathsf{M}}}]$ is independent of j.

Objective. To predict $\mu_{\mathsf{M}j}\Theta_{\mathsf{M}j}$ as well as possible.

A4M is equivalent to $\mathrm{Var}[Z_{ji} \mid \Theta_{\mathsf{M}j}] = \phi_{\mathsf{M}}(\mu_{\mathsf{M}j}\Theta_{\mathsf{M}j})^{p_{\mathsf{M}}}$. Its formulation is for similarity with the corresponding risk premium assumptions. In Rosenlund (2014) we argued against such an assumption in ordinary non-credibility multiplicative tariff analysis. But here often many groups are supposed to have few claims, so that without the assumption we risk overparametrization.

The parameter p_{M} in assumption **A4M** is fixed initially. Ordinarily $p_{\mathsf{M}} = 2$ is the most suitable value. It is natural to assume a constant CV for claim severities, and this implies $p_{\mathsf{M}} = 2$.

Define

$$\mu_{\mathsf{M}j} = \mathrm{E}[Y_{\mathsf{M}j}] \quad \text{mean claim group } j, \tag{8.1}$$

$$\Lambda_{\mathsf{M}j} = \mu_{\mathsf{M}j}\Theta_{\mathsf{M}j} \quad \text{mean claim group } j \text{ conditional on } \Theta_{\mathsf{M}j}, \tag{8.2}$$

$$\sigma^2_{\mathsf{M}} = \phi_{\mathsf{M}}\mathrm{E}[\Theta^{p_{\mathsf{M}}}_{\mathsf{M}j}] \quad \text{assumed independent of } j, \tag{8.3}$$

$$\sigma^2_{\mathsf{M}j} = \mathrm{E}[\mathrm{Var}[Y_{\mathsf{M}j} \mid \Theta_{\mathsf{M}j}]], \tag{8.4}$$

$$\tau^2_{\mathsf{M}} = \mathrm{Var}[\Theta_{\mathsf{M}j}] \quad \text{assumed independent of } j, \tag{8.5}$$

$$\nu^2_{\mathsf{M}j} = \frac{1}{{N_{j1}}^2}\sum_{i:k_{i1}=k_{j1}} N_i^2(\sigma^2_{\mathsf{M}i} + \mu^2_{\mathsf{M}j}\tau^2_{\mathsf{M}}) \quad \text{equal to } \mathrm{Var}[\hat{\mu}_{\mathsf{M}j}] \text{ when } R=1. \tag{8.6}$$

The predictors and estimators are

$$\Lambda^*_{\mathsf{M}j} = z_{\mathsf{M}j}Y_{\mathsf{M}j} + (1-z_{\mathsf{M}j})\mu_{\mathsf{M}j} \quad \text{non-observable predictor of } \Lambda_{\mathsf{M}j}, \text{ see (8.17)}, \tag{8.7}$$

$$\widehat{\Lambda}_{\mathsf{M}j} = \widehat{z}_{\mathsf{M}j}Y_{\mathsf{M}j} + (1-\widehat{z}_{\mathsf{M}j})\hat{\mu}_{\mathsf{M}j} \quad \text{estimated predictor of } \Lambda_{\mathsf{M}j}, \text{ see (8.24)}, \tag{8.8}$$

$$\widehat{\sigma}^2_{\mathsf{M}} = \frac{\sum_{j=1}^{J} \hat{\mu}_{\mathsf{M}j}^{-p_{\mathsf{M}}} \sum_{i=1}^{N_j}(Z_{ji} - Y_{\mathsf{M}j})^2}{\sum_{j=1}^{J} \mathbf{1}_{\{N_j>0\}}(N_j-1)} = \frac{\sum_{j=1}^{J} \hat{\mu}_{\mathsf{M}j}^{-p_{\mathsf{M}}} \mathbf{1}_{\{N_j>0\}}(Q_j - X_j^2/N_j)}{\sum_{j=1}^{J} \mathbf{1}_{\{N_j>0\}}(N_j-1)}, \tag{8.9}$$

$$\overset{*}{\sigma}{}^2_{\mathsf{M}j} = \frac{\sum_{i=1}^{N_j}(Z_{ji} - Y_{\mathsf{M}j})^2}{N_j(N_j-1)} = \frac{Q_j - X_j^2/N_j}{N_j(N_j-1)}, \text{ if } N_j > 1, \tag{8.10}$$

$$\widehat{\sigma}^2_{\mathsf{M}j} = \widehat{\sigma}^2_{\mathsf{M}}\hat{\mu}^{p_{\mathsf{M}}}_{\mathsf{M}j}/N_j, \tag{8.11}$$

$$\widetilde{\sigma}^2_{\mathsf{M}j} = q_{\mathsf{M}}\, \overset{*}{\sigma}{}^2_{\mathsf{M}j} + (1-q_{\mathsf{M}})\widehat{\sigma}^2_{\mathsf{M}j}, \quad q_{\mathsf{M}} \in [0,1], \text{ for a suitable } q_{\mathsf{M}}, \tag{8.12}$$

$$\widetilde{Y} = \frac{1}{N_0}\sum_{j=1}^{J} X_j/\hat{\mu}_{\mathsf{M}j} = \frac{1}{N_0}\sum_{j=1}^{J} N_j Y_{\mathsf{M}j}/\hat{\mu}_{\mathsf{M}j}, \tag{8.13}$$

$$\overset{*}{\tau}{}^2_{\mathsf{M}} = \max\left(0, \frac{\sum_{j=1}^{J} N_j(Y_{\mathsf{M}j}/\hat{\mu}_{\mathsf{M}j} - \widetilde{Y})^2 - (J_1-1)\widehat{\sigma}^2_{\mathsf{M}}}{N_0 - \sum_{j=1}^{J} N_j^2/N_0}\right), \text{ estimator of } \tau^2_{\mathsf{M}}, \tag{8.14}$$

an adaptation of (4.27) in Ohlsson & Johansson (2010) for $p_{\mathsf{M}} = 2$.

$$\hat{\tau}^2_{\mathsf{M}} = \text{pseudo-estimator of } \tau^2_{\mathsf{M}}, \text{ see (8.23) in Section 8.3}, \tag{8.15}$$

$$\hat{\nu}^2_{\mathsf{M}j} = \frac{1}{{N_{j1}}^2}\sum_{i:k_{i1}=k_{j1}} N_i^2(\widetilde{\sigma}^2_{\mathsf{M}i} + \hat{\mu}^2_{\mathsf{M}j}\hat{\tau}^2_{\mathsf{M}}) \quad \text{estimator of } \mathrm{Var}[\hat{\mu}_{\mathsf{M}j}] \text{ when } R=1. \tag{8.16}$$

Replacing w_j with N_j, the derivations for risk premium can be transferred to mean claim to prove the (approximate) unbiasedness of these estimators, except those for $\sigma^2_{\mathsf{M}}, \sigma^2_{\mathsf{M}j}$ and τ^2_{M}. See Ohlsson & Johansson (2010) for the estimators of σ^2_{M} and $\sigma^2_{\mathsf{M}j}$ and for $\overset{*}{\tau}{}^2_{\mathsf{M}}$. The adaptation of the latter is explained in Section 8.3.1. We give a proof only for $\hat{\tau}^2_{\mathsf{M}}$.

8.1. Best linear predictor of the mean claim credibility factor

For pricing we wish to use a predictor of $\mu_{\mathsf{M}j}\Theta_{\mathsf{M}j}$ as a mean claim factor for the value j of the credibility argument. It consists of the multiplicative mean claim $\mu_{\mathsf{M}j}$, the same for all groups with the same values of the group properties, and the specific factor $\Theta_{\mathsf{M}j}$. We have to multiply the predictor of $\mu_{\mathsf{M}j}\Theta_{\mathsf{M}j}$ with a factor product for ordinary arguments, e.g. policy holder age, in order to obtain the mean claim for a specific object.

Theorem 8.1. *For the BLP in L^2-norm of $\mu_{\mathsf{M}j}\Theta_{\mathsf{M}j}$ it holds*

$$z_{\mathsf{M}j} = \frac{\mathrm{Var}[\hat{\mu}_{\mathsf{M}j}] + \mu_{\mathsf{M}j}^2\tau_{\mathsf{M}}^2 - \mathrm{Cov}(\hat{\mu}_{\mathsf{M}j}, Y_{\mathsf{M}j}) - \mu_{\mathsf{M}j}\mathrm{Cov}(\hat{\mu}_{\mathsf{M}j}, \Theta_{\mathsf{M}j})}{\sigma_{\mathsf{M}j}^2 + \mathrm{Var}[\hat{\mu}_{\mathsf{M}j}] + \mu_{\mathsf{M}j}^2\tau_{\mathsf{M}}^2 - 2\mathrm{Cov}(\hat{\mu}_{\mathsf{M}j}, Y_{\mathsf{M}j})}. \tag{8.17}$$

If N_j/N_{jr} is small enough for all r, then $\hat{\mu}_{\mathsf{M}j}$ is approximately independent of $(\Theta_{\mathsf{M}j}, X_j, Q_j)$ and $\mathrm{Var}[\hat{\mu}_{\mathsf{M}j}]$, $\mathrm{Cov}(\hat{\mu}_{\mathsf{M}j}, Y_{\mathsf{M}j})$ and $\mathrm{Cov}(\hat{\mu}_{\mathsf{M}j}, \Theta_{\mathsf{M}j})$ will be small in absolute value.

For the special case one group property, estimators can be computed.

Corollary 8.1. *For the case $\hat{\mu}_{\mathsf{M}j} = X_{j1}/N_{j1}$, i.e. only one group property, we have*

$$z_{\mathsf{M}j} = \frac{\nu_{\mathsf{M}j}^2 + \mu_{\mathsf{M}j}^2\tau_{\mathsf{M}}^2 - \dfrac{N_j}{N_{j1}}\sigma_{\mathsf{M}j}^2 - \dfrac{2N_j}{N_{j1}}\mu_{\mathsf{M}j}^2\tau_{\mathsf{M}}^2}{\sigma_{\mathsf{M}j}^2 + \nu_{\mathsf{M}j}^2 + \mu_{\mathsf{M}j}^2\tau_{\mathsf{M}}^2 - \dfrac{2N_j}{N_{j1}}\sigma_{\mathsf{M}j}^2 - \dfrac{2N_j}{N_{j1}}\mu_{\mathsf{M}j}^2\tau_{\mathsf{M}}^2}. \tag{8.18}$$

8.2. Estimators of mean claim variance components within a group

Lemma 8.1. *Given* **A1M** - **A4M** *and $\widehat{\sigma}_{\mathsf{M}}^2$ by (8.9),* $\mathrm{E}[\widehat{\sigma}_{\mathsf{M}}^2] \approx \sigma_{\mathsf{M}}^2$.

Lemma 8.2. *Given* **A1M** *and $\overset{*}{\sigma}{}_{\mathsf{M}j}^2$ by (8.10),* $\mathrm{E}[\overset{*}{\sigma}{}_{\mathsf{M}j}^2] = \sigma_{\mathsf{M}j}^2$.

Lemma 8.3. *Given* **A1M** - **A4M** *and $\widehat{\sigma}_{\mathsf{M}j}^2$ by (8.11),* $\mathrm{E}[\widehat{\sigma}_{\mathsf{M}j}^2] \approx \sigma_{\mathsf{M}j}^2$.

8.3. Estimators of mean claim variance component between groups

8.3.1. Non-pseudo-estimator

The adaptation of (4.27) in Ohlsson & Johansson (2010) into $\overset{*}{\tau}{}_{\mathsf{M}}^2$ by (8.14) is this. Firstly, we have to write $(J_1 - 1)$ instead of $(J - 1)$, since J is the total number of groups. The number J_1 with claims is relevant here. Secondly, we have specialized p on p. 82 to $p = 2$. This is our p_{M}. It is our experience that other values of p are seldom, if ever, used for mean claim in actuarial practice. Thirdly, we have divided the expression by the square of an estimate of the Ohlsson & Johansson (2010) base factor μ, defined on p. 82. This is necessary in our context, since we deal with risk premium estimates without specifying them into base factors and argument factors. Fourthly we truncate the estimator from below to 0.

As we shall see in tables of simulation results, sometimes $\overset{*}{\tau}{}_{\mathsf{M}}^2$ is best and sometimes $\hat{\tau}_{\mathsf{M}}^2$ by (8.15) is best. The results also give indications as to which estimator is preferable depending on the situation. Namely that the pseudo-estimator generally seems best for light-tailed conditional claim distributions, while the non-pseudo-estimator seems best for heavy-tailed ones.

8.3.2. Pseudo-estimator for conditionally gamma distributed claim amounts

We could adapt (6.18) straight-forwardly to a pseudo-estimator of τ^2_{M}, but when N_j differ substantially it is farfetched to assume constant excess for $Y_{\mathsf{M}j}$. Particularly so after mixing by $\Theta_{\mathsf{M}j}$, since this makes the CLT inapplicable.

So we use the same assumptions of zero third central moment and zero excess for $\Theta_{\mathsf{M}j}$ as for the claim frequency parameter $\Theta_{\mathsf{F}j}$. In addition we assume that $p_{\mathsf{M}} = 2$ and that the claim amounts Z_{ji} are Gamma distributed, conditional on $\Theta_{\mathsf{M}j}$. Then they have the same CV $\sqrt{\phi_{\mathsf{M}}}$ conditional on $\Theta_{\mathsf{M}j}$. In other words we assume the mean claim part of S-GLM. This enables a mathematically consistent optimization of the pseudo-estimator, which sometimes might perform well even if these assumptions are only partly true. The pseudo-estimator will anyway be approximatively unbiased.

The formulas are involved but present no difficulties to implement in programs. See Appendix A.12.3 for the solution of equations (8.22) and (8.23). The same definition of $\hat{\tau}^2_{\mathsf{M}}$ as the largest of possibly several solutions as the definition of $\hat{\tau}^2_{\mathsf{F}}$ applies.

With notation for manageable formulas let

$$a_4(x) = x^{-3}(x+3)(x+2)(x+1), \tag{8.19}$$

$$a_3(x) = x^{-2}(x+2)(x+1), \tag{8.20}$$

$$\mathrm{v}(x,y) = (3x^2+6x+1)a_4\left(\frac{x+1}{y}\right) - 4(3x+1)a_3\left(\frac{x+1}{y}\right) + 6(x+y) - (x+y)^2 + 3, \tag{8.21}$$

and

$$c_{\mathsf{M}j} = \mathbf{1}_{\{N_j>0\}}\left(\frac{\widehat{\sigma}^2_{\mathsf{M}j}}{\hat{\mu}^2_{\mathsf{M}j}} + \hat{\tau}^2_{\mathsf{M}}\right)^2 \mathrm{v}(\hat{\tau}^2_{\mathsf{M}}, \widehat{\sigma}^2_{\mathsf{M}}/N_j)^{-1}\left[\sum_{i=1}^{J}\mathbf{1}_{\{N_i>0\}}\left(\frac{\widehat{\sigma}^2_{\mathsf{M}i}}{\hat{\mu}^2_{\mathsf{M}i}} + \hat{\tau}^2_{\mathsf{M}}\right)^2 \mathrm{v}(\hat{\tau}^2_{\mathsf{M}}, \widehat{\sigma}_{\mathsf{M}}/N_i)^{-1}\right]^{-1}, \tag{8.22}$$

$$\hat{\tau}^2_{\mathsf{M}} = \sum_{j=1}^{J} c_{\mathsf{M}j}\hat{\tau}^2_{\mathsf{M}}\left(\frac{\widehat{\sigma}^2_{\mathsf{M}j}}{\hat{\mu}^2_{\mathsf{M}j}} + \hat{\tau}^2_{\mathsf{M}}\right)^{-1}\left(\frac{Y_{\mathsf{M}j}}{\hat{\mu}_{\mathsf{M}j}} - 1\right)^2. \tag{8.23}$$

Lemma 8.4. *Assume* **A1M** - **A4M** *and that* $p_{\mathsf{M}} = 2$ *and* Z_{ji} *are Gamma distributed, conditional on* $\Theta_{\mathsf{M}j}$*. For the pseudo-estimator* $\hat{\tau}^2_{\mathsf{M}}$ *we have* $\mathrm{E}[\hat{\tau}^2_{\mathsf{M}}] \approx \tau^2_{\mathsf{M}}$*. If* $\mathrm{E}[(\Theta_{\mathsf{M}j}-1)^3] = 0$ *and* $\mathrm{e}(\Theta_{\mathsf{M}j}) = 0$*, then* $\mathrm{Var}[(Y_{\mathsf{M}j}/\mu_{\mathsf{M}j}-1)^2] = \mathrm{v}(\tau^2_{\mathsf{M}}, \sigma^2_{\mathsf{M}}/N_j)$ *and* $\hat{\tau}^2_{\mathsf{M}}$ *is approximately optimal in the sense of having the smallest variance of estimators in the form (8.23) with* $c_{\mathsf{M}1} + \ldots + c_{\mathsf{M}J} = 1$.

8.3.3. Possible alternative distribution-free pseudo-estimator

The estimator (8.23) depends on the gamma distribution and will sometimes be unrobust against departures from that form. It is seen in Appendix A.10 that the optimality of a pseudo-estimator depends on the 3:rd and 4:th moments of the conditional claim amount distributions. These are deduced from the 1:st and 2:nd moments under the gamma assumption. Under some other assumption, such as log-normal distributions, this deduction is invalid.

Alternatively we might try to develop estimates of $\mathrm{E}[(Z_{ji}/(\Theta_{\mathsf{M}j}\mu_{\mathsf{M}j}))^r \mid \Theta_{\mathsf{M}j}]$, for $r = 3, 4$, by using $\sum_{j=1}^{J}\sum_{i=1}^{N_j}\hat{\mu}^{-r}_{\mathsf{M}j}Z^r_{ji}$. We devised a method for this, but simulations showed that the resulting estimates were too unstable. So we abandoned this approach.

8.4. Credibility factor estimator for mean claim predictor and correction for bias

For N_j/N_{jr} small enough for all r, the variance $\mathrm{Var}[\hat{\mu}_{\mathsf{M}j}]$ and the covariances $\mathrm{Cov}(\hat{\mu}_{\mathsf{M}j}, Y_{\mathsf{M}j})$ and $\mathrm{Cov}(\hat{\mu}_{\mathsf{M}j}, \Theta_{\mathsf{M}j})$ can be omitted, and an expression like (6.17) can be an estimator of $z_{\mathsf{M}j}$.

When $R = 1$ and $\hat{\mu}_{\mathsf{M}j} = X_{j1}/N_{j1}$ we have computable estimators of $\mathrm{Var}[\hat{\mu}_{\mathsf{M}j}]$, $\mathrm{Cov}(\hat{\mu}_{\mathsf{M}j}, Y_{\mathsf{M}j})$ and $\mathrm{Cov}(\hat{\mu}_{\mathsf{M}j}, \Theta_{\mathsf{M}j})$. These are as those for risk premium, with w_j replaced by N_j.

Theorem 8.2. *An estimated credibility factor using (8.18) is for $N_j > 0$*

$$\widehat{z}_{\mathsf{M}j} = \frac{\hat{\nu}^2_{\mathsf{M}j} + \hat{\mu}^2_{\mathsf{M}j}\left(1 - \dfrac{2N_j}{N_{j1}}\right)\hat{\tau}^2_{\mathsf{M}} - \dfrac{N_j}{N_{j1}}\widetilde{\sigma}^2_{\mathsf{M}j}}{\widetilde{\sigma}^2_{\mathsf{M}j} + \hat{\nu}^2_{\mathsf{M}j} + \hat{\mu}^2_{\mathsf{M}j}\left(1 - \dfrac{2N_j}{N_{j1}}\right)\hat{\tau}^2_{\mathsf{M}} - \dfrac{2N_j}{N_{j1}}\widetilde{\sigma}^2_{\mathsf{M}j}}. \tag{8.24}$$

We could use $\overset{*}{\tau}{}^2_{\mathsf{M}}$ instead of $\hat{\tau}^2_{\mathsf{M}}$. For $N_j = 0$ we set $\widehat{z}_{\mathsf{M}j} = 0$.

We have thus determined the predictor $\widehat{\Lambda}_{\mathsf{M}j} = \widehat{z}_{\mathsf{M}j} Y_{\mathsf{M}j} + (1 - \widehat{z}_{\mathsf{M}j})\hat{\mu}_{\mathsf{M}j}$ by (8.8) depending on variance estimates. For the non-observable predictors the following unbiasedness holds for the total.

$$\mathrm{E}[\sum_{j=1}^{J} \Lambda^*_{\mathsf{M}j} N_j] = \sum_{j=1}^{J} [z_{\mathsf{M}j}\mu_{\mathsf{M}j} + (1 - z_{\mathsf{M}j})\mu_{\mathsf{M}j}] N_j = \sum_{j=1}^{J} \mu_{\mathsf{M}j} N_j = \mathrm{E}[\sum_{j=1}^{J} Y_{\mathsf{M}j} N_j] = \mathrm{E}[X]. \tag{8.25}$$

Inserting estimates can cause a bias. A correction factor gives a new estimator

$$\widetilde{\Lambda}_{\mathsf{M}j} = \widehat{\Lambda}_{\mathsf{M}j} X \left(\sum_{i=1}^{J} \widehat{\Lambda}_{\mathsf{M}i} N_i\right)^{-1}. \tag{8.26}$$

9. Combining claim frequency and mean claim

The claim frequency results of Section 7 are combined with the mean claim results of Section 8, if we wish to use these results for risk premium. We then define

$$\widehat{\Lambda}_{\mathsf{FM}j} = \widehat{\Lambda}_{\mathsf{F}j} \widehat{\Lambda}_{\mathsf{M}j} \tag{9.1}$$

and use it instead of (6.8). Then we correct it to $\widetilde{\Lambda}_{\mathsf{FM}j}$ as in (6.22). It serves as final rating factor for risk premium. That we were not able to obtain optimal weights c_j in Lemma 6.4 due to overestimation for small w_j, with risk premium as one entity, speaks for this combination. So does also the uncertainty of the proper value of p in assumption **A4**, whereas $p_{\mathsf{F}} = 1$ and $p_{\mathsf{M}} = 2$ are natural values.

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, which motivates (9.1).

The contract definition problem mentioned in Section 4 can be avoided completely also with this combination.

Remark 1. *Corrections to $\widehat{\Lambda}_{\mathsf{F}j}$ or $\widehat{\Lambda}_{\mathsf{M}j}$ with a constant factor will have no effect on risk premium. We might define $\widehat{z}_{\mathsf{FM}j}$ so that a counterpart to (6.8) holds by $\widehat{z}_{\mathsf{FM}j} = (\widehat{\Lambda}_{\mathsf{F}j}\widehat{\Lambda}_{\mathsf{M}j} - \hat{\mu}_j)/(Y_j - \hat{\mu}_j)$, but this number will not always be in the interval* [0,1].

10. Simulation results for between-groups variance component estimators

10.1. Setup of comparison

To obtain guidelines for choice of estimator, depending on the situation, we have compared our pseudo-estimators with the classical type ones that are rendered in Ohlsson & Johansson (2010).

For claim frequency the classical type one on p. 82 of Ohlsson & Johansson (2010) was modified by assuming $\sigma^2 = \mu$ and specialized to $p = 1$. That is, we used the Poisson assumption. This will give an estimator at least as good than the classical type one, although we could not detect any differences when comparing them. The modified estimator allows us to avoid the contract definition problem that arises, if we use the formulas (4.25) and (4.26) in Ohlsson & Johansson (2010) for within-groups variance component estimators. For mean claim we used $\overset{*}{\tau}{}^2_{\mathsf{M}}$ by (8.14) as the classical type estimator.

Four J-values 200, 2000, 8000 and 90000 were studied. For each value a fictitious insurance file was made with very varied exposure sizes per group j. No ordinary arguments were constructed. A single group property, or Auxiliary, with five classes was assigned to each j. Thus the model errors that will be described in Section 12 were avoided.

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

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

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

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

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.

Claim frequency and mean claim were analyzed separately. The basic model assumptions were obeyed, but we did not let the distributions of $\Theta_{\mathsf{F}j}$ and $\Theta_{\mathsf{M}j}$ have zero excess or even mostly third central moment zero, as our pseudo-estimator theorems presuppose. From a practical viewpoint these are unrealistic assumptions, but ones that admit relatively simple and mathematically consistent estimators. The estimators have to be reasonably robust against departures from the assumptions in order to be useful, though. Neither do we include here conditionally gamma distributed claim amounts. Simulations with such claim amounts showed our mean claim pseudo-estimator to be generally best, as could be expected.

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 thousands of their square roots for easier perception and appropriate scale, are tabulated in Table 2 and 3. The comparison determining the column Best was slightly involved, as far as the occurrences of question marks are concerned. Namely, we observed differences of goodness estimates per simulation. For claim frequency, let $\delta_1 = (\hat{\tau}^2_{\mathsf{F}} - \tau^2_{\mathsf{F}})^2$ be the observed mean square deviation of $\hat{\tau}^2_{\mathsf{F}}$ and let δ_2 be the corresponding deviation for the classical type estimator. We are interested in the sign of $\mathrm{E}[\delta_1] - \mathrm{E}[\delta_2]$. If -1, then $\hat{\tau}^2_{\mathsf{F}}$ is best and vice versa. It is intuitively clear that δ_1 and δ_2 are positively correlated. Hence $\delta_0 = \delta_1 - \delta_2$ has a smaller variance than $\mathrm{Var}[\delta_1] + \mathrm{Var}[\delta_1]$. A confidence interval for $\mathrm{E}[\delta_0]$ will be smaller if the sequence of δ_0-outcomes is analyzed, rather than if the sequences of δ_1-outcomes and δ_2-outcomes are treated as independent of each other. The sign of the estimate of $\mathrm{E}[\delta_0]$ determines which estimator is denoted as Best. If the confidence interval, with level 99 %, contains 0, then a question mark is added. Analogously for mean claim. However, when δ_0 has a very large variance, such as on some lines of Table 3 for lognormal conditional claim distribution, then the CLT normal approximation for the sample mean is not so good.

10.2. Distributions and results

The results are given in Table 2 and Table 3 in Appendix B. Our pseudo-estimators are denoted by SR, the classical type ones by OJ (Ohlsson & Johansson).

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_r^0 have the gamma distribution of (r, r).

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.

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

Let X be F or M for claim frequency or mean claim, respectively. For $\tau_{\mathsf{X}}^2 = 0$ the confidence intervals (confidence level 95 %) are for $10^5\times$parameter. This is marked with a †. Otherwise confidence intervals (95 %) are for the relative biases in percent of the SR and OJ τ^2-estimates, i.e. for 100(estimate - truevalue)/truevalue.

Conditional claim amount distributions and Θ-distributions are listed by ascending CV order.

Distribution of $\Theta_{\mathsf{X}j}$	Meaning	τ_{X}^2
D1	1 always	0.00000000
D2	U(0.875,1.125)	0.00520833
D3	$0.25\Theta_4^0 + 0.75$	0.01562500
D4	$0.25\Theta_2^0 + 0.75$	0.03125000
D5	$0.25\Theta_1^0 + 0.75$	0.06250000
D6	U(0.500,1.500)	0.08333333
D7	Θ_4^0	0.25000000
D8	Θ_2^0	0.50000000
D9	Θ_1^0	1.00000000

10.3. Conclusions from simulations

Almost all estimators have negative bias, except of course for zero τ_{F}^2 and τ_{M}^2, respectively.

10.3.1. Claim frequency

In Table 2 it is seen that the pseudo-estimator $\hat{\tau}_{\mathsf{F}}^2$ is the best one, except for some cases with small τ_{F}^2. For the smaller number of groups $J = 200$ it is possibly worse than the classical type one also for large τ_{F}^2 in Case 9. The result for small τ_{F}^2, especially in Case 1, is somewhat counter-intuitive, since the pseudo-estimator is always non-negative while the classical type one can attain negative values if not truncated to 0 from below. An explanation of this is likely that the function $g(x)$ of Section A.12.2 , whose zero is computed by binary interval halving, is unstable in the vicinity of 0 for Case 1. A guideline would be to recommend that $\hat{\tau}_{\mathsf{F}}^2$ is used, unless J is small and the suspected τ_{F}^2 is also small.

10.3.2. Mean claim

From Table 3 we can conclude that the pseudo-estimator $\hat{\tau}_{\mathsf{M}}^2$ is generally best for the uniform conditional claim distributions, while the classical type one is generally best for the heavy-tailed lognormal ones unless J is large. The latter seems clear despite the question marks in cases 10–18 and 28–36. The stability of the function $g(x)$ of Section A.12.3 is likely severely affected in heavy-tailed cases and moderate J, to a larger extent than the stability of the classical type estimator is. The pseudo-estimator is also best for gamma conditional claim distributions, although that is not listed here. One might generalize this to a proposition that the pseudo-estimator is best for light-tailed conditional claim distributions, but not for heavy-tailed ones unless J is large.. This is somewhat vague and should be corroborated by more simulations on various types of light-tailed and heavy-tailed conditional claim distributions, while making the proposition more precise.

We investigated also an adaptation of (6.18) to mean claim, to see if it performed better than $\hat{\tau}^2_{\mathsf{M}}$ in some cases, since the assumptions for $\hat{\tau}^2_{\mathsf{M}}$ are not met. We found that it did not.

11. 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 $1, \dots, 9$ with the distribution of exposures on Λ_j and discrete predictions $\widetilde{\Lambda}_j$ according to Table 1. 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 1. 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	$\sum w_j$	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

$$\widetilde{\widetilde{\Lambda}}_j = k_0 \widetilde{\Lambda}_j^{\delta} \quad \text{or, if claim frequency and mean claim are combined,} \quad \widetilde{\widetilde{\Lambda}}_{\mathsf{FM}j} = k_0 \widetilde{\Lambda}_{\mathsf{FM}j}^{\delta}. \tag{11.1}$$

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

12. Model errors

12.1. 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. We stated in Section 6 that factor product estimates for ordinary arguments, such as policy holder age, are treated as deterministic constants, i.e. as having zero variance. This is a model assumption for risk premium, claim frequency and mean claim.

In practice, however, the ordinary factor product estimates often have non-zero covariances with the Θ_j (and $\Theta_{\mathsf{F}j}, \Theta_{\mathsf{M}j}$), 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 **A2** holds. 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 exposures (not the normalized ones) with

$$\widehat{\Theta}_j = \widetilde{\Lambda}_j/\hat{\mu}_j \quad \text{or, if claim frequency and mean claim are combined,} \quad \widehat{\Theta}_{\mathsf{FM}j} = \widetilde{\Lambda}_{\mathsf{FM}j}/\hat{\mu}_j \qquad (12.1)$$

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 exposure and with all arguments, ordinary and group properties, that were part of the initial factor estimation. That gives new w_j and $\hat{\mu}_j$ for credibility analysis. 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 it does 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.

12.2. Errors in the multiplicative model

The multiplicative hypothesis in **A1** 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. Hence Assumption **A3** also fails. That means that the weight $\widehat{z}_j$ *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 predictor of $\mu_j\Theta_j$ 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 $\hat{\mu}_j$ will be too unstable. With $R = 1$, it holds $\mathrm{E}[\hat{\mu}_j] = \mu_j$ exactly and Corollary 6.1 and expression (6.20) are applicable without reservations concerning w_j/w_{jr}.

This model error is not observable.

13. Examples with real data

In Tables 4 and 5 we give results from combining claim frequency and mean claim analyses. We set $q_{\mathsf{M}} = 0$, since several groups have no or few claims, making the individual within-group variance estimators unsuitable. The exponents were $p_{\mathsf{F}} = 1$ and $p_{\mathsf{M}} = 2$. Five iterations were performed, i.e. six GLM and credibility analyses including the first one. MMT / Poisson log link was used for risk premium and claim frequency, after which mean claim $\hat{\mu}_{\mathsf{M}j}$ was obtained as the ratio $\hat{\mu}_j/\hat{\mu}_{\mathsf{F}j}$. In Example 1 with one group property the solutions $(\hat{\mu}_j, \hat{\mu}_{\mathsf{F}j}, \hat{\mu}_{\mathsf{M}j})$ by MMT, S-GLM and Tweedie are equal.

The differences between these results and the results from treating risk premium as one entity were substantial for some groups. We give $\widehat{z}_{\mathsf{FM}j}$ as defined in Remark 1 of Section 9. When either $\{Y_{\mathsf{F}j} < \mu_{\mathsf{F}j}\} \bigcap \{Y_{\mathsf{M}j} > \mu_{\mathsf{M}j}\}$ or $\{Y_{\mathsf{F}j} > \mu_{\mathsf{F}j}\} \bigcap \{Y_{\mathsf{M}j} < \mu_{\mathsf{M}j}\}$ we might obtain $\widehat{z}_{\mathsf{FM}j} \notin [0, 1]$. See group 247 in Example 2, Table 5, for an extreme example.

13.1. Example 1

In Table 4 in Appendix B we show a few selected lines from a report with 6183 groups. We show $\overset{*}{\sigma}_j = \sqrt{\overset{*}{\sigma}{}_j^2}$ in percent as an example of variance estimators per j. For this case with many groups most individual weights $\widehat{z}_{\mathsf{FM}j}$ are small. One group property (auxiliary variable) with 22 values was used. The claim frequency version of Corollary 6.1 and Corollary 8.1 for mean claim are applicable without reservations.

Parameters and estimates

$\hat{\tau}^2_{\mathsf{F}}$	$\hat{\tau}^2_{\mathsf{M}}$	$\widetilde{\Lambda}_{\mathsf{FM}j}/\widehat{\Lambda}_{\mathsf{FM}j}$	k_0	δ
0.0708	0.0223	0.9908	2.592	0.8

13.2. Example 2

In Table 5 in Appendix B a few lines are given from a report with 248 groups. Two group properties with 13 and 4 values, respectively, were used.

Parameters and estimates

$\hat{\tau}^2_{\mathsf{F}}$	$\hat{\tau}^2_{\mathsf{M}}$	$\widetilde{\Lambda}_{\mathsf{FM}j}/\widehat{\Lambda}_{\mathsf{FM}j}$	k_0	δ
0.0480	0.0142	1.0025	6.716	0.6

14. Conclusion

We have developed a model with reasonably weak assumptions for credibility analysis, that lets us combine multiplicative collective properties with individual ones of groups, such that a good approximation of the best linear predictor is obtained. Using the Compound Poisson model, which is solidly based on general limit results for Poisson approximation, enables more precise estimators than previous methods and the avoidance of the contract definition problem. We give a method for risk premium as one entity, and also separate methods for claim frequency and mean claim which are to be combined to risk premium. Optimal or near optimal pseudo-estimators for the between-groups variance component are given under rather reasonable conditions. We have pointed out the regression effect as a difficulty in applying credibility and devised a pragmatic method to handle it. Also we have pointed out the problem of deviations from the multiplicative hypothesis with many group properties.

Acknowledgement

Thanks are due to Dr. Adhitya Ronnie Effendie, Universitas Gadjah Mada, for valuable suggestions.

References

Barbour, A. D., Holst, L. & Janson, S. (1992). *Poisson Approximation.* Oxford University Press.

Bichsel, F. & Straub, E. Erfahrungstarifierung in der Kollektif-Krankenversicherung. *Internal note of the Swiss-Re.*

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.

Cipra, T. (1996). Dynamic credibility with outliers and missing observations. *Applications of Mathematics* **41** (2), 149–159.

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

Dannenburg, D. R., Kaas, R, Goovaerts, M. J. (1996). *Practical Actuarial Credibility Models.* Institute of Actuarial Science and Econometrics, Amsterdam.

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

Garrido, J. & Zhou, J. (2009). Full Credibility with Generalized Linear and Mixed Models. *ASTIN Bulletin* **39** (1), 61–80.

Grigelionis, B. (1963). On the convergence of sums of random step processes to a Poisson process. *Theory of Probability and its Applications.* **8** (2), 177–182.

Heller, G. Z., Stasinopoulos, M., Rigby R. A. & De Jong, P. (2007). Mean and dispersion modelling for policy claim costs. *Scandinavian Actuarial Journal*, **2007:4**, 281–292.

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.

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.

Smyth, G. K. & Jørgensen, B. (2002). Fitting Tweedie's Compound Poisson Model to Insurance Claims Data: Dispersion Modelling. *ASTIN Bulletin* **32(1)**, 143–157.

Appendix A: Proofs

A.1. Proof of Theorem 6.1

Optimality of BLP form

We use Ohlsson & Johansson (2010), Section 4.2, which can be specialized to our model. The cell index i on p. 82 has in our model only one value per group. These authors consider both ordinary arguments and group properties (Auxiliaries). We deal only with the latter, since ordinary arguments have been used for normalizing and are present only in w_j. Thus we write the cell index i as i_j. The authors' observation index t is 1 in our model. The Ohlsson & Johansson (2010) formulas on p. 82-83, applied to risk premium, translate as follows.

$U_j = \Theta_j$,

$\mu\gamma_i = \mu\gamma_{i_j} = \mu_j$, where $\mu = \mu_k$ for a base combination of Auxiliaries, e.g. (1,1, ..., 1),

$Y_{i_j j1} = Y_j$,

$\overline{\overline{\widetilde{Y}}}_{.j.} = \widetilde{w}_{i_j j1}\widetilde{Y}_{i_j j1}/\widetilde{w}_{i_j j1} = \widetilde{Y}_{i_j j1} = Y_{i_j j1}/\gamma_{i_j} = \mu Y_j/\mu_j$,

$\widehat{U}_j = \widetilde{z}_j \frac{\overline{\overline{\widetilde{Y}}}_{.j.}}{\mu} + (1-\widetilde{z}_j) = \widetilde{z}_j Y_j/\mu_j + (1-\widetilde{z}_j)$,

and the final rating factor for group j is

$\mu\gamma_{i_j}\widehat{U}_j = \mu_j\widehat{U}_j = \widetilde{z}_j Y_j + (1-\widetilde{z}_j)\mu_j$,

i.e. in the form (6.7). Since the Ohlsson & Johansson (2010) credibility estimator is $\widehat{U}_j$, we have shown the optimality of the form. Our z_j and its estimator corresponding to $\widetilde{z}_j$ is derived by quite different methods than in Ohlsson & Johansson (2010).

Remark 2. *The Ohlsson & Johansson (2010) Tweedie type variance assumption (4.19) corresponds to our assumption* **A4***, but in the latter the exponent p does not apply to ordinary arguments. Thus* **A4** *is weaker than the former, which can be seen as an advantage. Furthermore, if we would make the Ohlsson & Johansson (2010) assumption, then Y_j would be a linear combination of several Compound Poisson variables such that there is no c for which $c \times Y_j$ is Compound Poisson, conditional on Θ_j. Thus our computations based on this distribution would not be possible.*

We now compute z_j. We use that $\mathrm{E}[Y_j \mid \Theta_j] = \mu_j\Theta_j$, but not the Compound Poisson assumption. We have

$$\mathrm{E}[(\Lambda_j^* - \mu_j\Theta_j)^2] = \mathrm{E}[(z_jY_j + (1-z_j)\hat{\mu}_j - \mu_j\Theta_j)^2] = \mathrm{E}[(z_j(Y_j - \mu_j\Theta_j) + (1-z_j)(\hat{\mu}_j - \mu_j\Theta_j))^2]$$

$$= z_j^2\mathrm{E}[(Y_j - \mu_j\Theta_j)^2] + 2(z_j - z_j^2)\mathrm{E}[(Y_j - \mu_j\Theta_j)(\hat{\mu}_j - \mu_j\Theta_j)] + (1-z_j)^2\mathrm{E}[(\hat{\mu}_j - \mu_j\Theta_j)^2].$$

Set 1/2 of the derivative of this expression with respect to z_j equal to 0. It is clear that if this equation has only one solution, then the minimum has been found.

$$z_j\mathrm{E}[(Y_j - \mu_j\Theta_j)^2] + (1-2z_j)\mathrm{E}[(Y_j - \mu_j\Theta_j)(\hat{\mu}_j - \mu_j\Theta_j)] - (1-z_j)\mathrm{E}[(\hat{\mu}_j - \mu_j\Theta_j)^2] = 0. \quad (A.1)$$

We proceed to simplify the expressions in (A.1).

First term

From assumption **A1** we get

$$\begin{aligned} \mathrm{E}[Y_j \mid \Theta_j] &= \mu_j\Theta_j, \\ \mathrm{E}[Y_j] &= \mu_j. \end{aligned}$$

For any stochastic variable Z and σ-algebra $\mathcal{F}$ we have $\mathrm{Var}[Z] = \mathrm{E}[\mathrm{Var}[Z \mid \mathcal{F}]] + \mathrm{Var}[\mathrm{E}[Z \mid \mathcal{F}]]$. That gives in this context

$$\mathrm{E}[(Y_j - \mu_j\Theta_j)^2] = \mathrm{Var}[Y_j - \mu_j\Theta_j] = \mathrm{E}[\mathrm{Var}[Y_j - \mu_j\Theta_j \mid \Theta_j]] + \mathrm{Var}[\mathrm{E}[Y_j - \mu_j\Theta_j \mid \Theta_j]]$$

$$= \mathrm{E}[\mathrm{Var}[Y_j - \mu_j\Theta_j \mid \Theta_j]] = \mathrm{E}[\mathrm{Var}[Y_j \mid \Theta_j]] = \sigma_j^2.$$

Second term

By using $\mathrm{E}[Y_j] = \mathrm{E}[\hat{\mu}_j] = \mu_j$ we obtain

$$\mathrm{E}[(Y_j - \mu_j\Theta_j)(\hat{\mu}_j - \mu_j\Theta_j)] = \mathrm{E}[(Y_j - \mu_j + \mu_j - \mu_j\Theta_j)(\hat{\mu}_j - \mu_j + \mu_j - \mu_j\Theta_j)]$$

$$= \mathrm{E}[(Y_j - \mu_j)(\hat{\mu}_j - \mu_j)] + \mathrm{E}[(Y_j - \mu_j)(\mu_j - \mu_j\Theta_j)] + \mathrm{E}[(\mu_j - \mu_j\Theta_j)(\hat{\mu}_j - \mu_j)] + \mathrm{E}[(\mu_j - \mu_j\Theta_j)^2]$$

$$= \mathrm{Cov}(\hat{\mu}_j, Y_j) - \mu_j\mathrm{Cov}(Y_j, \Theta_j) - \mu_j\mathrm{Cov}(\hat{\mu}_j, \Theta_j) + \mu_j^2\mathrm{Var}[\Theta_j]$$

$$= \mathrm{Cov}(\hat{\mu}_j, Y_j) - \mu_j\mathrm{Cov}(Y_j, \Theta_j) - \mu_j\mathrm{Cov}(\hat{\mu}_j, \Theta_j) + \mu_j^2\tau^2.$$

We have

$$\mathrm{Cov}(Y_j, \Theta_j) = \mathrm{E}[\mathrm{E}[(Y_j - \mu_j)(\Theta_j - 1) \mid \Theta_j] = \mathrm{E}[(\Theta_j - 1)(\mu_j\Theta_j - \mu_j)] = \mu_j\tau^2.$$

Hence

$$\mathrm{E}[(Y_j - \mu_j\Theta_j)(\hat{\mu}_j - \mu_j\Theta_j)] = \mathrm{Cov}(\hat{\mu}_j, Y_j) - \mu_j\mathrm{Cov}(\hat{\mu}_j, \Theta_j).$$

Third term

We have

$$\mathrm{E}[(\hat{\mu}_j - \mu_j\Theta_j)^2] = \mathrm{E}[(\hat{\mu}_j - \mu_j + \mu_j - \mu_j\Theta_j)^2]$$

$$= \mathrm{E}[(\hat{\mu}_j - \mu_j)^2] - 2\mu_j\mathrm{E}[(\hat{\mu}_j - \mu_j)(\Theta_j - 1)] + \mu_j^2\mathrm{E}[(\Theta_j - 1)^2]$$

$$= \mathrm{Var}[\hat{\mu}_j] - 2\mu_j\mathrm{Cov}(\hat{\mu}_j, \Theta_j) + \mu_j^2\mathrm{Var}[\Theta_j] = \mathrm{Var}[\hat{\mu}_j] - 2\mu_j\mathrm{Cov}(\hat{\mu}_j, \Theta_j) + \mu_j^2\tau^2.$$

Thus (A.1) reduces to

$$z_j\sigma_j^2 + (1-2z_j)[\mathrm{Cov}(\hat{\mu}_j, Y_j) - \mu_j\mathrm{Cov}(\hat{\mu}_j, \Theta_j)] - (1-z_j)[\mathrm{Var}[\hat{\mu}_j] - 2\mu_j\mathrm{Cov}(\hat{\mu}_j, \Theta_j) + \mu_j^2\tau^2] = 0,$$

i.e.

$$z_j\left[\sigma_j^2 - 2\mathrm{Cov}(\hat{\mu}_j, Y_j) + 2\mu_j\mathrm{Cov}(\hat{\mu}_j, \Theta_j) + \mathrm{Var}[\hat{\mu}_j] - 2\mu_j\mathrm{Cov}(\hat{\mu}_j, \Theta_j) + \mu_j^2\tau^2\right]$$

$$= -\mathrm{Cov}(\hat{\mu}_j, Y_j) + \mu_j\mathrm{Cov}(\hat{\mu}_j, \Theta_j) + \mathrm{Var}[\hat{\mu}_j] - 2\mu_j\mathrm{Cov}(\hat{\mu}_j, \Theta_j) + \mu_j^2\tau^2,$$

i.e.

$$z_j[\sigma_j^2 + \mathrm{Var}[\hat{\mu}_j] + \mu_j^2\tau^2 - 2\mathrm{Cov}(\hat{\mu}_j, Y_j)] = \mathrm{Var}[\hat{\mu}_j] + \mu_j^2\tau^2 - \mathrm{Cov}(\hat{\mu}_j, Y_j) - \mu_j\mathrm{Cov}(\hat{\mu}_j, \Theta_j).$$

This gives the expression (6.15). □

A.2. Proof of Corollary 6.1

When $R = 1$ and $\hat{\mu}_j = X_{j1}/w_{j1}$, we will show that

$$\mathrm{Var}[\hat{\mu}_j] = \frac{1}{{w_{j1}}^2}\sum_{i:k_{i1}=k_{j1}} w_i^2(\sigma_i^2 + \mu_j^2\tau^2) = \nu_j^2, \tag{A.2}$$

$$\mathrm{Cov}(\hat{\mu}_j, Y_j) = \frac{w_j}{w_{j1}}\sigma_j^2 + \frac{w_j}{w_{j1}}\mu_j^2\tau^2, \tag{A.3}$$

$$\mathrm{Cov}(\hat{\mu}_j, \Theta_j) = \frac{w_j}{w_{j1}}\mu_j\tau^2. \tag{A.4}$$

By assumption **A1** we have

$$\mathrm{Var}[Y_j] = \mathrm{E}[\mathrm{Var}[Y_j \mid \Theta_j]] + \mathrm{Var}[\mathrm{E}[Y_j \mid \Theta_j]] = \sigma_j^2 + \mathrm{Var}[\mu_j \Theta_j] = \sigma_j^2 + \mu_j^2 \tau^2.$$

Hence by the independence assumption **A2**

$$\mathrm{Var}[\hat{\mu}_j] = \mathrm{Var}[X_{j1}/w_{j1}] = \frac{1}{{w_{j1}}^2}\mathrm{Var}\left[\sum_{i:k_{i1}=k_{j1}} w_i Y_i\right] = \frac{1}{{w_{j1}}^2}\sum_{i:k_{i1}=k_{j1}} w_i^2 \mathrm{Var}[Y_i]$$

$$= \frac{1}{{w_{j1}}^2}\sum_{i:k_{i1}=k_{j1}} w_i^2(\sigma_i^2 + \mu_i^2\tau^2) = \nu_j^2.$$

Since $\mu_i = \mu_j$ for $\{i : k_{i1} = k_{j1}\}$ we have proved (A.2).

By the independence assumption **A2**, for $\mathrm{Cov}(\hat{\mu}_j, Y_j)$ we can eliminate all terms in $\hat{\mu}_j$ not containing Y_j. Hence we show (A.3) by

$$\mathrm{Cov}(\hat{\mu}_j, Y_j) = \mathrm{Cov}(\frac{w_j Y_j}{w_{j1}}, Y_j) = \frac{w_j}{w_{j1}}\mathrm{Var}[Y_j] = \frac{w_j}{w_{j1}}\sigma_j^2 + \frac{w_j}{w_{j1}}\mu_j^2\tau^2.$$

For $\mathrm{Cov}(\hat{\mu}_j, \Theta_j)$, by Appendix A.1 under Second term,

$$\mathrm{Cov}(\hat{\mu}_j, \Theta_j) = \mathrm{Cov}(\frac{w_j Y_j}{w_{j1}}, \Theta_j) = \frac{w_j}{w_{j1}}\mu_j\tau^2.$$

Insertion of (A.2), (A.3) and (A.4) in (6.15) gives (6.16). □

A.3. Proof of Lemma 6.1

By the variance formula for Compound Poisson it holds

$$\mathrm{E}[Q \mid \Theta_1 \ldots \Theta_J] = \mathrm{Var}[X \mid \Theta_1 \ldots \Theta_J]$$

and furthermore it holds

$$\mathrm{Var}[X \mid \Theta_1 \ldots \Theta_J] = \mathrm{Var}\left[\sum_{j=1}^{J} w_j Y_j \mid \Theta_1 \ldots \Theta_J\right] = \sum_{j=1}^{J} w_j^2 \mu_j^p \phi \Theta_j^p / w_j = \sum_{j=1}^{J} w_j \mu_j^p \phi \Theta_j^p$$

and thus

$$\mathrm{E}[Q] = \mathrm{E}[\mathrm{E}[Q \mid \Theta_1 \ldots \Theta_J]] = \mathrm{E}[\mathrm{Var}[X \mid \Theta_1 \ldots \Theta_J]] = \sum_{j=1}^{J} w_j \mu_j^p \phi \mathrm{E}[\Theta_j^p] = \sum_{j=1}^{J} w_j \mu_j^p \sigma^2.$$

So we have

$$\mathrm{E}\left[\frac{Q}{\sum_{j=1}^{J} w_j \mu_j^p}\right] = \sigma^2.$$

If we assume that $\mathrm{Var}[\hat{\mu}_j] = 0$, so that $\mathrm{P}(\hat{\mu}_j = \mu_j) = 1$, then Lemma 6.1 follows with $\approx$ replaced by $=$. In any case it can be assumed that the CV of $\sum_{j=1}^{J} w_j \hat{\mu}_j^p$ is small. □

A.4. Proof of Lemma 6.2

Again by the variance formula for Compound Poisson it holds

$$\mathrm{E}[Q_j \mid \Theta_j] = \mathrm{Var}[X_j \mid \Theta_j] = w_j^2 \mathrm{Var}[Y_j \mid \Theta_j]$$

and hence

$$\mathrm{E}[\overset{*}{\sigma}{}_j^2] = w_j^{-2}\mathrm{E}[Q_j] = w_j^{-2}\mathrm{E}[\mathrm{E}[Q_j \,|\, \Theta_j]] = \mathrm{E}[\mathrm{Var}[Y_j \,|\, \Theta_j]] = \sigma_j^2.$$

A.5. Proof of Lemma 6.3

By assumption **A4** and the expressions (6.4), (6.3) it holds

$$\sigma_j^2 = \mathrm{E}[\mathrm{Var}[Y_j \mid \Theta_j]] = \mu_j^p \phi \mathrm{E}[\Theta_j^p]/w_j = \sigma^2 \mu_j^p / w_j.$$

If, as in Ohlsson & Johansson (2010), we replace σ^2 with the estimator $\widehat{\sigma}^2$ and μ_j^p with the estimator $\hat{\mu}_j^p$, then Lemma 6.3 follows. □

A.6. Proof of Lemma 6.4

We start by referring to the Bichsel-Straub estimator given in De Vylder (1996) III, Chapter 3, Section 3.4.7, Theorem 18. It treats the original Bühlmann-Straub model with several contracts, weighted observations, and no contract excesses. We obtain it by specializing our model to $\mu_j \equiv \mu$. We translate the De Vylder (1996) notation, but adjust it by setting estimators where De Vylder (1996) has known parameters. The left sides are the notation of De Vylder (1996).

$$\begin{aligned}
\mathrm{a} &= \mu^2\tau^2, \\
z_j &= \frac{\mathrm{A_{pseu}}}{\widehat{\sigma}_j^2 + \mathrm{A_{pseu}}}, \\
k_0 &= J, \\
\mathrm{X_{zw}} &= \sum_{j=1}^{J} \frac{z_j}{z_1 + \ldots + z_J} Y_j, \\
\mathrm{A_{pseu}} &= \frac{1}{J-1} \sum_{j=1}^{J} z_j (Y_j - \mathrm{X_{zw}})^2.
\end{aligned}$$

Then $\mathrm{A_{pseu}}$ is the optimal Bichsel-Straub estimator of a.

In our model with several μ_j we cannot use the pseudo-estimator $\mathrm{X_{zw}}$. We have to make do with the standardized variable Y_j/μ_j and its expectation 1 in (6.18), where we replace μ_j with its estimator $\hat{\mu}_j$.

We now give a self-contained account, where we weaken the condition of zero excess to constant excess. We can see in Appendix A.2 that

$$\mathrm{E}\left[\frac{Y_j}{\mu_j}\right] = 1, \qquad \mathrm{Var}\left[\frac{Y_j}{\mu_j}\right] = \frac{\sigma_j^2}{\mu_j^2} + \tau^2.$$

Define the random variables

$$U_j = \tau^2 \left(\frac{\sigma_j^2}{\mu_j^2} + \tau^2\right)^{-1} \left(\frac{Y_j}{\mu_j} - 1\right)^2 \text{ with expectation } \mathrm{E}[U_j] = \tau^2.$$

Given that the optimal estimator of τ^2 is a linear combination

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

the minimum variance standard solution is $a_j = \text{const}/\text{Var}[U_j]$.

Assume that Y_j has constant excess γ_2. Then

$$\text{Var}\left[\left(\frac{Y_j}{\mu_j} - 1\right)^2\right] = (2+\gamma_2)\text{Var}\left[\frac{Y_j}{\mu_j}\right]^2 = (2+\gamma_2)\left(\frac{\sigma_j^2}{\mu_j^2} + \tau^2\right)^2.$$

Hence

$$\text{Var}[U_j] = \tau^4\left(\frac{\sigma_j^2}{\mu_j^2} + \tau^2\right)^{-2} \text{Var}\left[\left(\frac{Y_j}{\mu_j} - 1\right)^2\right] = \tau^4(2+\gamma_2).$$

So a_j are all equal to $1/J$. But we have to replace μ_j with $\hat{\mu}_j$, σ_j^2 with $\widetilde{\sigma}_j^2$ and τ^2 with $\hat{\tau}^2$ in a_j to obtain c_j. This entails a bias, but due to the complicated nature of $\hat{\mu}_j$ there is no obvious way to correct it, by a factor like $1/(J-1)$ instead of $1/J$ or otherwise. So we keep $1/J$. If J is moderately large and the $\hat{\mu}_j$ have reasonably small variances the bias will be moderate. □

A.7. Proof of Lemma 7.1

First we show that $\text{Var}[(Y_{\mathsf{F}j}/\mu_{\mathsf{F}j} - 1)^2] = u(\tau_{\mathsf{F}}^2, \mu_{\mathsf{F}j} w_j)$.

Suppressing j for shortness, we let

$$N = N_j,$$

$$m = \mu_{\mathsf{F}j} w_j,$$

$$\Delta = m\Theta_{\mathsf{F}j},$$

$$\rho = m^2 \tau_{\mathsf{F}}^2.$$

Then $N \mid \Delta$ is distributed Poisson(Δ). We can use semiinvariants, described in Cramér (1946), Chapter 15, Section 15.10, and the fact that all Poisson semiinvariants are equal to the mean. With the help of equation (15.10.4) in Cramér (1946) we obtain

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

Assume that Θ_j, and hence also Δ, has 3:rd central moment 0 and excess 0. Then the 3:rd and 4:th semiinvariants of Δ are 0. Again from Cramér (1946), equation (15.10.4), we deduce that

$$\begin{aligned}
\text{E}[\Delta] &= m, \\
\text{E}[\Delta^2] &= \rho + m^2, \\
\text{E}[\Delta^3] &= 3\rho m + m^3, \\
\text{E}[\Delta^4] &= 3\rho^2 + 6\rho m^2 + m^4.
\end{aligned} \tag{A.5}$$

Since $\mathrm{E}[N^r] = \mathrm{E}[\mathrm{E}[N^r \mid \Delta]]$ we get

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

For the 2:nd and 4:th central moments of N we have, regardless of the distribution of Δ,

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

We are interested in

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

After somewhat lengthy calculations we arrive at the comparatively simple expression

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

Hence we get, replacing ρ with $m^2\tau_{\mathsf{F}}^2$,

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

Returning completely to the original notation, we have $Y_{\mathsf{F}j}/\mu_{\mathsf{F}j} = N/m$ and $m = \mu_{\mathsf{F}j}w_j$, which gives

$$\mathrm{Var}\left[\left(\frac{Y_{\mathsf{F}j}}{\mu_{\mathsf{F}j}} - 1\right)^2\right] = \frac{1}{\mu_{\mathsf{F}j}^3 w_j^3} + \frac{7\tau_{\mathsf{F}}^2 + 2}{\mu_{\mathsf{F}j}^2 w_j^2} + \frac{4\tau_{\mathsf{F}}^2}{\mu_{\mathsf{F}j} w_j} + 2\tau_{\mathsf{F}}^4.$$

Now for the optimal pseudo-estimator, we note that

$$\mathrm{E}\left[\frac{Y_{\mathsf{F}j}}{\mu_{\mathsf{F}j}}\right] = 1, \qquad \mathrm{Var}\left[\frac{Y_{\mathsf{F}j}}{\mu_{\mathsf{F}j}}\right] = \frac{\sigma_{\mathsf{F}j}^2}{\mu_{\mathsf{F}j}^2} + \tau_{\mathsf{F}}^2 = \frac{1}{\mu_{\mathsf{F}j}w_j} + \tau_{\mathsf{F}}^2.$$

Define the random variables

$$T_j = \tau_{\mathsf{F}}^2\left(\frac{1}{\mu_{\mathsf{F}j}w_j} + \tau_{\mathsf{F}}^2\right)^{-1}\left(\frac{Y_{\mathsf{F}j}}{\mu_{\mathsf{F}j}} - 1\right)^2 \text{ with expectation } \mathrm{E}[T_j] = \tau_{\mathsf{F}}^2.$$

Given that the optimal estimator of τ_{F}^2 is a linear combination

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

the minimum variance standard solution is $a_j = \mathrm{const}/\mathrm{Var}[T_j]$.

From the above we have, with $u(,)$ as in the statement of Lemma 7.1,

$$\mathrm{Var}[T_j] = \tau_{\mathsf{F}}^4\left(\frac{1}{\mu_{\mathsf{F}j}w_j} + \tau_{\mathsf{F}}^2\right)^{-2}\mathrm{Var}\left[\left(\frac{Y_{\mathsf{F}j}}{\mu_{\mathsf{F}j}} - 1\right)^2\right]$$

$$= \tau_{\mathsf{F}}^4\left(\frac{1}{\mu_{\mathsf{F}j}w_j} + \tau_{\mathsf{F}}^2\right)^{-2}\left(\frac{1}{\mu_{\mathsf{F}j}^3 w_j^3} + \frac{7\tau_{\mathsf{F}}^2 + 2}{\mu_{\mathsf{F}j}^2 w_j^2} + \frac{4\tau_{\mathsf{F}}^2}{\mu_{\mathsf{F}j} w_j} + 2\tau_{\mathsf{F}}^4\right) = \tau_{\mathsf{F}}^4\left(\frac{1}{\mu_{\mathsf{F}j}w_j} + \tau_{\mathsf{F}}^2\right)^{-2} u(\tau_{\mathsf{F}}^2, \mu_{\mathsf{F}j}w_j).$$

In the expression for the minimizing a_j we can cancel out τ_{F}^4 and obtain

$$a_j = \left(\frac{1}{\mu_{\mathsf{F}j} w_j} + \tau_{\mathsf{F}}^2\right)^2 u(\tau_{\mathsf{F}}^2, \mu_{\mathsf{F}j} w_j)^{-1} \left[\sum_{i=1}^{J} \left(\frac{1}{\mu_{\mathsf{F}i} w_i} + \tau_{\mathsf{F}}^2\right)^2 u(\tau_{\mathsf{F}}^2, \mu_{\mathsf{F}i} w_i)^{-1}\right]^{-1}. \tag{A.6}$$

The optimal estimator of τ_{F}^2 using unknown true parameters is then

$$\overset{*}{\tau}{}_{\mathsf{F}}^2 = \sum_{j=1}^{J} a_j T_j = \sum_{j=1}^{J} a_j \tau_{\mathsf{F}}^2 \left(\frac{1}{\mu_{\mathsf{F}j} w_j} + \tau_{\mathsf{F}}^2\right)^{-1} \left(\frac{Y_{\mathsf{F}j}}{\mu_{\mathsf{F}j}} - 1\right)^2. \tag{A.7}$$

It is unbiased. Substituting estimators for true values in the right sides we obtain $c_{\mathsf{F}j}$ and $\hat{\tau}_{\mathsf{F}}^2$. The latter will not be unbiased.

Remark 3. *It can be shown that*

$$\operatorname{Var}\left[\left(\frac{Y_{\mathsf{F}j}}{\mu_{\mathsf{F}j}} - 1\right)^2\right] \sim \frac{1}{\mu_{\mathsf{F}j}^3 w_j^3} \text{ as } w_j \to 0,$$

regardless of the 3:rd and 4:th moments of $\Theta_{\mathsf{F}j}$. □

A.8. Proof of Theorem 8.1

The same as for Theorem 6.1 with w_j replaced by N_j. □

A.9. Proof of Corollary 8.1

The same as for Corollary 6.1 with the same replacement as above. □

A.10. Proof of Lemma 8.4

First we show that $\operatorname{Var}[(Y_{\mathsf{M}j}/\mu_{\mathsf{M}j} - 1)^2] = v(\tau_{\mathsf{M}}^2, \sigma_{\mathsf{M}}^2/N_j)$.

The conditions of the Lemma are that

$$\operatorname{Var}[Z_{ji} \mid \Theta_{\mathsf{M}j}] = \phi_{\mathsf{M}} \operatorname{E}[Z_{ji} \mid \Theta_{\mathsf{M}j}]^2 = \phi_{\mathsf{M}} \mu_{\mathsf{M}j}^2 \Theta_{\mathsf{M}j}^2,$$

where $Z_{ji} \mid \Theta_{\mathsf{M}j}$ is Gamma distributed. Then the CV conditional on $\Theta_{\mathsf{M}j}$ is $\sqrt{\phi_{\mathsf{M}}}$ for all j and i. By (8.3) we have

$$\phi_{\mathsf{M}} = \sigma_{\mathsf{M}}^2 / \operatorname{E}[\Theta_{\mathsf{M}j}^2] = \sigma_{\mathsf{M}}^2/(\tau_{\mathsf{M}}^2 + 1).$$

Let (α, β) be the usual Gamma distribution parameter, such that the mean is α/β and the variance is α/β^2. Then it is easily shown that, conditional on $\Theta_{\mathsf{M}j}$, the parameters are

Variable	α	β
Z_{ji}	$(\tau_{\mathsf{M}}^2 + 1)/\sigma_{\mathsf{M}}^2$	$(\tau_{\mathsf{M}}^2 + 1)/(\sigma_{\mathsf{M}}^2 \Theta_{\mathsf{M}j} \mu_{\mathsf{M}j})$
$Y_{\mathsf{M}j}/\mu_{\mathsf{M}j}$	$N_j(\tau_{\mathsf{M}}^2 + 1)/\sigma_{\mathsf{M}}^2$	$N_j(\tau_{\mathsf{M}}^2 + 1)/(\sigma_{\mathsf{M}}^2 \Theta_{\mathsf{M}j})$

We are interested in $\operatorname{Var}[(Y_{\mathsf{M}j}/\mu_{\mathsf{M}j} - 1)^2]$.

The moments of the Gamma distribution are, for $Z \sim \Gamma(\alpha, \beta)$,

$$\operatorname{E}[Z^r] = \beta^{-r} \prod_{i=0}^{r-1} (\alpha + i).$$

For shortness we suppress indices etc. in the notation below. With $a_4()$ and $a_3()$ by (8.19) and (8.20), let

$$\Theta = \Theta_{\mathsf{M}j},$$

$$Y = Y_{\mathsf{M}j}/\mu_{\mathsf{M}j},$$

$$A_4 = a_4\left(\frac{N_j(\tau_{\mathsf{M}}^2+1)}{\sigma_{\mathsf{M}}^2}\right),$$

$$A_3 = a_3\left(\frac{N_j(\tau_{\mathsf{M}}^2+1)}{\sigma_{\mathsf{M}}^2}\right),$$

$$V = \frac{\sigma_{\mathsf{M}}^2}{N_j} + \tau_{\mathsf{M}}^2 = \frac{\sigma_{\mathsf{M}j}^2}{\mu_{\mathsf{M}j}^2} + \tau_{\mathsf{M}}^2 = \mathrm{Var}[Y] \quad \text{(see below)}.$$

From $\mathrm{E}[Z^r]$ above we obtain $\mathrm{E}[Y^r \mid \Theta] = A_r\Theta^r$, and so $\mathrm{E}[Y^r] = A_r\mathrm{E}[\Theta^r]$ for $r = 3, 4$. We want

$$\mathrm{Var}[(Y-1)^2] = \mathrm{E}[(Y-1)^4] - \mathrm{E}[(Y-1)^2]^2 = \mathrm{E}[(Y-1)^4] - V^2 = \mathrm{E}[Y^4 - 4Y^3 + 6Y^2 - 4Y + 1] - V^2,$$

which gives

$$\mathrm{Var}[(Y-1)^2] = A_4\mathrm{E}[\Theta^4] - 4A_3\mathrm{E}[\Theta^3] + 6(V+1) - 3 - V^2.$$

As in equation (A.5) we have from the moment conditions on Θ

$$\mathrm{E}[\Theta^4] = 3\tau_{\mathsf{M}}^4 + 6\tau_{\mathsf{M}}^2 + 1, \qquad \mathrm{E}[\Theta^3] = 3\tau_{\mathsf{M}}^2 + 1.$$

This gives

$$\mathrm{Var}[(Y-1)^2] = (3\tau_{\mathsf{M}}^4 + 6\tau_{\mathsf{M}}^2 + 1)A_4 - 4(3\tau_{\mathsf{M}}^2 + 1)A_3 + 6\left(\tau_{\mathsf{M}}^2 + \frac{\sigma_{\mathsf{M}}^2}{N_j}\right) - \left(\tau_{\mathsf{M}}^2 + \frac{\sigma_{\mathsf{M}}^2}{N_j}\right)^2 + 3.$$

Returning to the original notation we get $\mathrm{Var}[(Y_{\mathsf{M}j}/\mu_{\mathsf{M}j} - 1)^2] = v(\tau_{\mathsf{M}}^2, \sigma_{\mathsf{M}}^2/N_j)$.

The rest of the proof is as the one for Lemma 7.1 with the obvious modifications. As before we have

$$\mathrm{E}\left[\frac{Y_{\mathsf{M}j}}{\mu_{\mathsf{M}j}}\right] = 1, \qquad \mathrm{Var}\left[\frac{Y_{\mathsf{M}j}}{\mu_{\mathsf{M}j}}\right] = \frac{\sigma_{\mathsf{M}j}^2}{\mu_{\mathsf{M}j}^2} + \tau_{\mathsf{M}}^2.$$

Define the random variables

$$V_j = \tau_{\mathsf{M}}^2\left(\frac{\sigma_{\mathsf{M}j}^2}{\mu_{\mathsf{M}j}^2} + \tau_{\mathsf{M}}^2\right)^{-1}\left(\frac{Y_{\mathsf{M}j}}{\mu_{\mathsf{M}j}} - 1\right)^2 \text{ with expectation } \mathrm{E}[V_j] = \tau_{\mathsf{M}}^2.$$

Given that the optimal estimator of τ_{M}^2 is a linear combination

$$\sum_{\substack{j=1\\N_j>0}}^{J} a_jV_j, \text{ where } \sum_{\substack{j=1\\N_j>0}}^{J} a_j = 1,$$

the minimum variance standard solution is $a_j = \mathrm{const}/\mathrm{Var}[V_j]$.

From the above we have

$$\mathrm{Var}[V_j] = \tau_{\mathsf{M}}^4\left(\frac{\sigma_{\mathsf{M}j}^2}{\mu_{\mathsf{M}j}^2} + \tau_{\mathsf{M}}^2\right)^{-2}\mathrm{Var}\left[\left(\frac{Y_{\mathsf{M}j}}{\mu_{\mathsf{M}j}} - 1\right)^2\right] = \tau_{\mathsf{M}}^4\left(\frac{\sigma_{\mathsf{M}j}^2}{\mu_{\mathsf{M}j}^2} + \tau_{\mathsf{M}}^2\right)^{-2} v(\tau_{\mathsf{M}}^2, \sigma_{\mathsf{M}}^2/N_j).$$

We then proceed as in Appendix A.7 by developing an unbiased estimator, analogous to (A.7), using true values and a biased one with plugged-in estimates. This gives $c_{\mathsf{M}j}$ and $\hat{\tau}_{\mathsf{M}}^2$ by (8.22) and (8.23). □

A.11. Proof of Theorem 8.2

The same as for Theorem 6.2 with w_j replaced by N_j. □

A.12. Solutions of pseudo-estimator equations

A.12.1. Risk premium

For (6.17) and (6.18), let $A_j = \widetilde{\sigma}_j^2/\hat{\mu}_j^2$ and $B_j = (Y_j/\hat{\mu}_j - 1)^2/J$. Let x stand for $\hat{\tau}^2$. Then we can write these equations as

$$c_j(x) = (A_j x^{-1} + 1)^{-1},$$
$$f(x) = x - \sum_{j=1}^{J} c_j(x) B_j = 0.$$

Since $f(0) = 0$ a non-negative solution of $f(x) = 0$ exists. Since $c_j(x) \le 1$ we have $f(x) \ge x - \sum_{j=1}^{J} B_j = x - R$, say. Hence $f(R) \ge 0$ and $f(x) > 0$ for $x > R$. This can be taken as the right endpoint of the interval where the solution is. The left endpoint is 0. The solution is in the closed interval [0,R].

Convexity of $f(x)$ follows from the second derivative.

$$f'(x) = 1 - \sum_{j=1}^{J} B_j \left[(-1)(A_j x^{-1}+1)^{-2}(-1)A_j x^{-2}\right] = 1 - \sum_{j=1}^{J} B_j \left[(A_j x^{-1}+1)^{-2} A_j x^{-2}\right],$$

$$f''(x) = -\sum_{j=1}^{J} B_j \left[(A_j x^{-1}+1)^{-2}(-2)A_j x^{-3} + (-2)(A_j x^{-1}+1)^{-3}(-1)A_j x^{-2} A_j x^{-2}\right]$$

$$= \sum_{j=1}^{J} 2B_j A_j x^{-3}(A_j x^{-1}+1)^{-2}\left[1 - A_j x^{-1}(A_j x^{-1}+1)^{-1}\right]$$

$$= \sum_{j=1}^{J} 2B_j A_j x^{-3}(A_j x^{-1}+1)^{-2}\left(1 - \frac{A_j}{A_j + x}\right) > 0 \text{ for } x > 0.$$

So if $f(\hat{\tau}^2) = 0$ for $\hat{\tau}^2 > 0$, then it is the unique positive solution. Otherwise the solution is 0.

A.12.2. Claim frequency

For (7.2) and (7.3) we let $A_j = \widetilde{\sigma}_{\mathsf{F}j}^2/\hat{\mu}_{\mathsf{F}j}^2 = 1/(\hat{\mu}_{\mathsf{F}j} w_j)$ and $B_j = (Y_{\mathsf{F}j}/\hat{\mu}_{\mathsf{F}j} - 1)^2$. With $x = \hat{\tau}_{\mathsf{F}}^2$ we write $c_{\mathsf{F}j}(x) = c_{\mathsf{F}j}$. We have to solve

$$f(x) = x - \sum_{j=1}^{J} c_{\mathsf{F}j}(x) \frac{x}{x + A_j} B_j = 0.$$

Let $A_{\min} = \min\{A_1, \ldots, A_J\}$ and $B_{\max} = \max\{B_1, \ldots, B_J\}$. Then

$$f(x) \ge x - \sum_{j=1}^{J} c_{\mathsf{F}j} \frac{x}{x + A_{\min}} B_{\max} = x\left(1 - \frac{B_{\max}}{x + A_{\min}}\right).$$

Let $R = B_{\max} - A_{\min}$. If $R > 0$ it is the positive solution of $\left(1 - \dfrac{B_{\max}}{x + A_{\min}}\right) = 0$, and since this function is increasing it holds $f(R) \geq 0$ and $f(x) > 0$ for $x > R$. This can be taken as the right endpoint of the interval where the solution is. The left endpoint is 0. The solution is in the closed interval [0,R]. But if $R \leq 0$ we have $f(x) > 0$ for $x > 0$ and no positive solution exists. Otherwise, consider $g(x) = f(x)/x$, namely

$$g(x) = 1 - \sum_{j=1}^{J} c_{\mathsf{F}j}(x) \frac{1}{x + A_j} B_j.$$

If we can show that $g'(x) > 0$ for $x > 0$, then $g(x) = 0$ has at most one positive solution. This is not obvious. We have $\lim_{x \to \infty} g(x) = 1$. From the definition (7.2) and some calculations we get

$$\lim_{x \downarrow 0} g(x) = 1 - \left[\sum_{j=1}^{J} B_j A_j^{-1} (A_j + 2)^{-1}\right] \left[\sum_{i=1}^{J} (A_i + 2)^{-1}\right]^{-1}.$$

So if this expression is negative a positive solution exists. It remains to show that it is unique, and that there is no positive solution if the expression is non-negative.

A.12.3. Mean claim

We treat (8.22) and (8.23) in the same way as (7.2) and (7.3). Here $A_j = \widehat{\sigma}^2_{\mathsf{M}j}/\hat{\mu}^2_{\mathsf{M}j} = \widehat{\sigma}^2_{\mathsf{M}}/N_j$ and $B_j = (Y_{\mathsf{M}j}/\hat{\mu}_{\mathsf{M}j} - 1)^2$. With $x = \hat{\tau}^2_{\mathsf{M}}$ the functions $f(x)$ and $g(x)$ are as in Appendix A.12.2 with $c_{\mathsf{F}j}(x)$ replaced by $c_{\mathsf{M}j}(x)$. A right endpoint R is obtained as before. Here we get

$$\lim_{x \downarrow 0} g(x) = 1 - \left[\sum_{j=1}^{J} B_j A_j^{-1} (3A_j + 1)^{-1}\right] \left[\sum_{i=1}^{J} (3A_i + 1)^{-1}\right]^{-1}.$$

It remains to show that this determines the existence or not of a unique positive solution.

Appendix B: Tables from simulations and from real data

B.1. Simulation results

The tables are explained in Section 10.

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

Case	J	$\Theta_{\mathsf{F}j}$ distr	Best	SR sMseK	SR Lo95	SR Point	SR Up95	OJ sMseK	OJ Lo95	OJ Point	OJ Up95
1	200	D1	OJ?	0.43	19.8†	20.2†	20.7†	0.43	20.5†	21.0†	21.4†
2	200	D2	OJ	1.32	-6.0	-5.7	-5.4	1.18	-5.3	-5.1	-4.8
3	200	D3	SR?	2.94	-4.3	-4.1	-3.9	2.96	-4.0	-3.8	-3.6
4	200	D4	SR	5.58	-4.1	-3.9	-3.7	6.09	-4.0	-3.8	-3.6
5	200	D5	SR	12.07	-4.6	-4.4	-4.2	13.96	-4.7	-4.4	-4.1
6	200	D6	SR	8.49	-1.6	-1.5	-1.4	8.90	-2.4	-2.3	-2.2
7	200	D7	SR	32.09	-2.2	-2.1	-1.9	35.86	-3.5	-3.4	-3.2
8	200	D8	SR	71.16	-2.0	-1.9	-1.7	76.27	-4.3	-4.1	-3.9
9	200	D9	OJ?	171.45	-2.0	-1.8	-1.6	169.57	-5.7	-5.5	-5.3
10	2000	D1	SR	0.14	7.5†	7.6†	7.7†	0.15	7.8†	7.9†	8.0†
11	2000	D2	OJ	0.41	-0.6	-0.6	-0.5	0.36	-0.5	-0.5	-0.4
12	2000	D3	SR	0.93	-0.5	-0.4	-0.4	0.94	-0.5	-0.4	-0.4
13	2000	D4	SR	1.77	-0.4	-0.4	-0.3	1.96	-0.4	-0.3	-0.3
14	2000	D5	SR	3.90	-0.5	-0.4	-0.4	4.59	-0.5	-0.4	-0.4
15	2000	D6	SR	2.64	-0.2	-0.1	-0.1	2.76	-0.3	-0.2	-0.2
16	2000	D7	SR	9.98	-0.2	-0.2	-0.1	11.48	-0.4	-0.3	-0.3
17	2000	D8	SR	22.30	-0.3	-0.2	-0.2	25.10	-0.5	-0.4	-0.4
18	2000	D9	SR	52.87	-0.2	-0.2	-0.1	57.24	-0.6	-0.6	-0.5
19	8000	D1	SR	0.07	4.0†	4.1†	4.2†	0.08	4.1†	4.2†	4.3†
20	8000	D2	OJ	0.21	-0.2	-0.2	-0.1	0.18	-0.2	-0.1	-0.1
21	8000	D3	SR	0.46	-0.2	-0.1	-0.1	0.47	-0.2	-0.1	-0.1
22	8000	D4	SR	0.89	-0.1	-0.1	-0.1	0.99	-0.1	-0.1	-0.0
23	8000	D5	SR	1.94	-0.2	-0.1	-0.1	2.30	-0.2	-0.1	-0.1
24	8000	D6	SR	1.31	-0.1	-0.0	-0.0	1.39	-0.1	-0.1	-0.1
25	8000	D7	SR	5.04	-0.1	-0.1	-0.0	5.81	-0.1	-0.1	-0.1
26	8000	D8	SR	10.95	-0.1	-0.1	-0.1	12.43	-0.2	-0.1	-0.1
27	8000	D9	SR	26.46	-0.1	-0.1	-0.0	28.84	-0.2	-0.1	-0.1
28	90000	D1	SR	0.02	1.2†	1.2†	1.3†	0.02	1.2†	1.3†	1.3†
29	90000	D2	OJ	0.07	-0.1	0.1	0.4	0.06	-0.1	0.1	0.3
30	90000	D3	SR?	0.14	-0.0	0.0	0.0	0.14	-0.0	0.0	0.0
31	90000	D4	SR	0.27	-0.0	0.0	0.0	0.30	-0.0	0.0	0.0
32	90000	D5	SR	0.59	-0.0	-0.0	0.0	0.68	-0.0	-0.0	0.0
33	90000	D6	SR	0.40	-0.0	-0.0	0.0	0.42	-0.0	-0.0	-0.0
34	90000	D7	SR	1.50	-0.1	-0.0	0.1	1.74	-0.2	-0.1	0.0
35	90000	D8	SR	3.00	-0.1	0.0	0.1	3.70	-0.1	-0.0	0.1
36	90000	D9	SR	7.72	-0.0	0.1	0.1	8.61	-0.0	0.1	0.2

Table 3. Mean claim comparison of τ^2-estimates.

Case	J	Cond cldst	Θ_{Mj} distr	Best	SR sMseK	SR Lo95	SR Point	SR Up95	OJ sMseK	OJ Lo95	OJ Point	OJ Up95
1	200	Unif	D1	OJ	0.14	6.2†	6.4†	6.6†	0.13	6.3†	6.5†	6.7†
2	200	Unif	D2	OJ	0.76	-3.7	-3.5	-3.3	0.69	-3.5	-3.3	-3.1
3	200	Unif	D3	SR	2.37	-3.3	-3.0	-2.8	2.64	-3.8	-3.5	-3.2
4	200	Unif	D4	SR	4.98	-3.5	-3.3	-3.0	5.83	-4.5	-4.2	-3.9
5	200	Unif	D5	SR	11.52	-4.2	-3.9	-3.6	13.79	-5.9	-5.5	-5.2
6	200	Unif	D6	SR	7.61	-1.0	-0.9	-0.7	8.70	-2.1	-1.9	-1.7
7	200	Unif	D7	SR	33.99	-0.8	-0.6	-0.4	38.63	-4.2	-4.0	-3.7
8	200	Unif	D8	OJ?	88.28	1.5	1.8	2.0	86.88	-5.8	-5.5	-5.2
9	200	Unif	D9	OJ	308.36	9.5	10.0	10.5	208.59	-8.1	-7.8	-7.4
10	200	Logn	D1	OJ?	620.96	0.0†	829.8†	1659.6†	13.03	244.7†	265.8†	286.8†
11	200	Logn	D2	OJ?	1243.29	-100.0	200.0	499.9	16.59	-0.3	4.9	10.2
12	200	Logn	D3	OJ?	232.10	-21.2	3.3	27.8	13.88	-10.7	-9.2	-7.8
13	200	Logn	D4	OJ?	268.55	-15.5	-1.4	12.8	19.81	-8.6	-7.6	-6.5
14	200	Logn	D5	OJ?	247.86	-10.6	-4.0	2.5	23.30	-8.3	-7.7	-7.1
15	200	Logn	D6	OJ?	636.82	-10.3	2.4	15.0	25.91	-4.0	-3.5	-3.0
16	200	Logn	D7	OJ?	1893.23	-11.6	0.9	13.4	56.59	-5.1	-4.8	-4.4
17	200	Logn	D8	OJ?	2317.52	-4.1	3.6	11.2	102.96	-6.2	-5.9	-5.5
18	200	Logn	D9	OJ?	7359.93	0.7	12.8	25.0	233.28	-8.2	-7.8	-7.4
19	2000	Unif	D1	OJ	0.05	2.4†	2.4†	2.5†	0.04	2.4†	2.5†	2.5†
20	2000	Unif	D2	OJ	0.23	-0.4	-0.3	-0.3	0.21	-0.4	-0.3	-0.3
21	2000	Unif	D3	SR	0.74	-0.3	-0.3	-0.2	0.84	-0.4	-0.3	-0.2
22	2000	Unif	D4	SR	1.59	-0.4	-0.3	-0.2	1.91	-0.5	-0.4	-0.3
23	2000	Unif	D5	SR	3.74	-0.5	-0.4	-0.3	4.62	-0.6	-0.5	-0.4
24	2000	Unif	D6	SR	2.36	-0.1	-0.1	-0.1	2.72	-0.2	-0.2	-0.2
25	2000	Unif	D7	SR	10.40	-0.1	-0.1	-0.0	12.74	-0.5	-0.5	-0.4
26	2000	Unif	D8	SR	25.68	0.1	0.1	0.2	29.75	-0.7	-0.7	-0.6
27	2000	Unif	D9	SR?	76.97	1.0	1.1	1.2	77.11	-1.2	-1.1	-0.9
28	2000	Logn	D1	OJ?	6.76	90.2†	99.5†	108.7†	5.42	122.9†	130.2†	137.5†
29	2000	Logn	D2	OJ?	6.81	1.4	3.2	5.1	6.28	-1.4	0.3	2.0
30	2000	Logn	D3	SR?	5.89	-0.6	-0.1	0.5	6.70	-1.6	-1.0	-0.4
31	2000	Logn	D4	OJ?	9.62	-0.6	-0.2	0.2	7.72	-1.1	-0.8	-0.4
32	2000	Logn	D5	OJ?	8.94	-0.7	-0.5	-0.3	8.31	-0.9	-0.7	-0.5
33	2000	Logn	D6	OJ	8.24	-0.4	-0.2	-0.1	6.82	-0.6	-0.4	-0.3
34	2000	Logn	D7	OJ?	335.82	-0.8	1.1	3.0	29.11	-0.6	-0.4	-0.2
35	2000	Logn	D8	OJ?	121.19	-0.0	0.3	0.6	39.89	-0.7	-0.6	-0.5
36	2000	Logn	D9	OJ?	2147.41	0.2	3.2	6.1	87.70	-1.3	-1.1	-1.0
37	8000	Logn	D1	SR	1.72	51.1†	53.4†	55.7†	2.50	76.3†	79.6†	82.9†
38	8000	Logn	D2	SR	2.13	-0.3	0.3	0.9	2.77	-2.0	-1.3	-0.5
39	8000	Logn	D3	SR	2.46	-0.3	-0.1	0.2	2.89	-0.7	-0.5	-0.2
40	8000	Logn	D4	SR?	3.40	-0.2	-0.0	0.1	3.66	-0.3	-0.2	0.0
41	8000	Logn	D5	OJ?	4.81	-0.1	-0.0	0.1	4.56	-0.2	-0.1	-0.0
42	8000	Logn	D6	OJ?	4.67	-0.1	-0.0	0.0	4.08	-0.2	-0.1	-0.0
43	8000	Logn	D7	OJ	11.47	-0.1	-0.1	-0.0	9.52	-0.2	-0.2	-0.1
44	8000	Logn	D8	OJ?	25.98	-0.0	0.0	0.1	18.78	-0.2	-0.2	-0.1
45	8000	Logn	D9	OJ?	115.88	0.2	0.3	0.5	46.31	-0.4	-0.3	-0.2
46	90000	Logn	D1	SR?	0.50	15.3†	16.7†	18.2†	1.10	25.9†	29.2†	32.4†
47	90000	Logn	D2	SR	0.41	-4.1	-2.6	-1.2	0.64	-7.0	-4.8	-2.6
48	90000	Logn	D3	SR	0.73	-0.2	-0.1	0.1	0.95	-0.4	-0.2	-0.0
49	90000	Logn	D4	SR	0.99	0.0	0.1	0.3	1.20	0.0	0.2	0.3
50	90000	Logn	D5	SR?	1.42	-0.1	-0.0	0.0	1.52	-0.1	-0.0	0.1
51	90000	Logn	D6	SR?	1.38	-0.1	-0.0	0.0	1.40	-0.1	-0.0	0.0
52	90000	Logn	D7	OJ	3.05	-0.2	-0.0	0.1	2.48	-0.2	-0.1	0.1
53	90000	Logn	D8	OJ	6.53	-0.3	-0.1	0.1	5.38	-0.2	-0.1	0.1
54	90000	Logn	D9	OJ	19.98	-0.1	0.1	0.3	14.55	-0.2	0.0	0.2

B.2. Applications to real data

Table 4. Example 1 with $J = 6183$ groups.

Group j	Normalized duration w_j	N_j	$\widehat{z}_{\mathsf{FM}j}$	Indiv. riskpr/ normdur Y_j	100 $\overset{*}{\sigma}_j$	Factorest risk premium $\hat{\mu}_{\mathsf{F}j}\hat{\mu}_{\mathsf{M}j}$	Weighted risk premium $\widetilde{\Lambda}_{\mathsf{FM}j}$	Regression effect corr. risk premium $\widetilde{\widetilde{\Lambda}}_{\mathsf{FM}j}$
1	24.55	0	0.252	0.000	0.0	67.355	49.934	59.212
2	0.00	0	0.000	0.000	0.0	67.355	66.733	74.673
3	72.65	12	0.702	61.204	64.6	67.355	62.455	70.818
4	1.19	0	0.016	0.000	0.0	67.355	65.653	73.705
5	53.20	17	0.160	245.849	18.8	67.355	95.120	99.154
...								
41	210.81	51	0.100	69.992	45.7	103.298	99.042	102.411
42	25.11	15	0.699	166.451	55.7	103.298	146.061	139.739
43	35.62	13	0.220	201.821	30.9	76.936	103.465	106.054
44	282.34	72	1.111	106.497	23.5	90.609	107.267	109.160
45	178.72	37	0.224	168.962	18.2	87.617	104.869	107.204
...								
6179	4.15	0	0.050	0.000	0.0	76.936	72.454	79.752
6180	0.13	0	0.002	0.000	0.0	67.355	66.616	74.569
6181	0.59	0	0.008	0.000	0.0	67.355	66.197	74.193
6182	0.06	0	0.001	0.000	0.0	67.355	66.680	74.625
6183	0.07	0	0.001	0.000	0.0	67.355	66.676	74.622

Table 5. Example 2 with $J = 248$ groups.

Group j	Normalized duration w_j	N_j	$\widehat{z}_{\mathsf{FM}j}$	Indiv. riskpr/ normdur Y_j	100 $\overset{*}{\sigma}_j$	Factorest risk premium $\hat{\mu}_{\mathsf{F}j}\hat{\mu}_{\mathsf{M}j}$	Weighted risk premium $\widetilde{\Lambda}_{\mathsf{FM}j}$	Regression effect corr. risk premium $\widetilde{\widetilde{\Lambda}}_{\mathsf{FM}j}$
1	4271.50	1088	0.572	141.671	6.1	137.132	140.083	130.291
2	176.76	52	0.610	122.555	23.0	79.408	105.992	110.217
3	316.34	55	0.560	71.252	39.6	116.847	91.529	100.930
4	0.01	0	0.000	0.000	0.0	46.637	46.752	67.447
5	0.02	0	0.000	0.000	0.0	46.637	46.746	67.443
...								
41	479.43	134	0.581	155.123	11.7	85.350	126.198	122.381
42	379.83	118	0.647	182.390	13.2	106.893	156.132	139.052
43	16006.58	5144	0.993	156.120	2.5	117.378	156.233	139.106
44	53.07	23	0.412	186.637	32.6	99.005	135.496	127.714
45	315.03	58	-2.837	91.699	25.4	90.161	86.017	97.237
...								
244	2735.37	574	0.827	81.828	8.0	70.449	80.064	93.142
245	0.17	0	0.002	0.000	0.0	36.418	36.438	58.079
246	0.12	0	0.003	0.000	0.0	137.132	137.084	128.610
247	133.08	14	-8.526	98.328	37.9	95.174	68.453	84.785
248	3969.62	678	0.877	66.957	8.5	75.226	68.150	84.559