An incremental loss ratio method using prior information on calendar year effects

Ulrich Riegel

Received: 29 December 2021 / Revised: 29 March 2022 / Accepted: 11 April 2022 / Published online: 20 May 2022 © EAJ Association 2022

Abstract
In a run-off triangle external factors can have a similar influence on all incremental losses of the same calendar year. This can distort the triangle such that reserving methods like chain ladder or the loss ratio method do not work properly. A very recent example of such an external factor is the Covid-19 pandemic. In many countries, the insurance industry is in the process of establishing market knowledge about the impact of the pandemic on premiums and losses. We extend the additive claims reserving model to allow for calendar year effects and develop a variant of the incremental loss ratio method (also known as the additive method) that can make use of such market knowledge. We derive formulas for the mean squared error of prediction and provide a detailed numerical example.

Keywords Additive claims reserving model · Loss development · Calendar year effects modeling · EM algorithm

1 Introduction
Pricing and reserving of long-tailed Casualty business are often based on run-off triangles where incremental claims are grouped according to their accident and development year. The increments on a diagonal of such a triangle share the same calendar year and can therefore be affected by the same external factors. For instance, inflation may have similar effects for all incremental claims on the same diagonal (see, e.g., Clark [6]). Such calendar year effects are problematic, since they violate an assumption that is made in many reserving models: the independency of the accident years (see, e.g., Hertig [10] or Mack [13]).

There is quite some literature on calendar year effects in claims reserving. In the 1970s Verbeek [21] and Taylor [19] introduced a separation technique for claims...
inflation. Björkwall et al. [2] provide a bootstrap procedure for assessing the prediction error of this separation method. Calendar year effects in chain ladder-type models are studied, for example, in Barnett and Zenwirth [3], Brehm [4], de Jong [8], Kirschner et al. [11] and Kuang et al. [12]. These papers use either maximum-likelihood methods or bootstrap simulation methods. Other contributions study calendar year dependence in a Bayesian inference framework, see, e.g., Wüthrich [23], Shi et al. [18] or Wüthrich and Happ [24]. Bühlmann and Moriconi [5] provide a credibility model that extends the Buhlmann–Straub claims reserving model to consider calendar year effects. A recent paper by Gigante et al. [9] tackles calendar year effects using hierarchical generalized linear models.

The topic is presently very relevant due to Covid-19. The pandemic leaves its traces in nearly all relevant premium and claims data. Apart from accident year effects there are substantial calendar year effects. The impacts are manifold and depend on country and line of business. For instance, there can be increased or reduced claims activity, delays in court decisions or inflation shocks. These effects will complicate calculations not only for 1 or 2 years, but also in the medium term, as the distortions in the run-off triangles will persist for many years. Due to the relevance of the topic, the impact of the pandemic is being closely monitored by the insurance industry and to a certain extent market knowledge is available that can help to quantify these effects. For instance, it may be estimated from market data that payments are x% below or above average in a certain calendar year. The author is not aware of a reserving model that can easily make use of this kind of information. In the present paper we introduce a stochastic reserving method that can factor in such prior market knowledge on calendar year effects in a systematic way. For this purpose, we extend the incremental loss ratio method (also known as the additive method) and the underlying stochastic model (the additive claims reserving model or simply additive model). The incremental loss ratio method is one of the most popular reserving methods worldwide. According to the ASTIN report on non-life reserving [1] only chain ladder and Bornhuetter–Ferguson are more widely used.

In our stochastic model the incremental losses are generated by a two level stochastic mechanism. The calendar year effects are modeled on the first level. The incremental losses, which are modeled on the second level, are assumed to be conditionally independent, given the calendar year effects. In that respect, our model is similar to the conditionally independent loss increments (CILI) model with additive diagonal risk (ADR) defined by Bühlmann and Moriconi [5]. Apart from this similarity, the two models (and the proposed reserving methods) are, however, quite different. For example, in contrast to the CILI model with ADR, we do not make use of benchmark information like prior estimates for expected ultimates or development patterns.

In Sect. 2 we define the additive model with calendar year effects, which is an extension of the classical additive model. In Sect. 3 we use the expectation maximization algorithm (EM algorithm) to derive an iteration for parameter estimation in this model. Using the resulting estimators to predict future claims leads to the incremental loss ratio method with calendar year effects. The conditional mean square error of prediction (MSEP) of this reserving method is calculated in Sect. 4.
A shortcoming of the method is, that it uses many parameters. In particular, the estimation of individual variance parameters for each development year can lead to instabilities. Therefore, we do not recommend to use the method without robustification. In Sect. 5 we describe a possibility to make the approach more robust by using a variance function like in a generalized linear model. Eventually, this robustified version of the method is illustrated with an extensive numerical example in Sect. 6. Some technical calculations have been moved to the appendix.

An implementation of the (robustified) incremental loss ratio method with calendar year effects in an Excel tool is available as an online appendix.

2 An additive model with calendar year effects

For \( i \in \{1, \ldots, m\} \) and \( k \in \{1, \ldots, n\} \) let \( S_{i,k} \) denote the incremental claims of accident year \( i \) in the development year \( k \) (payments or incurred amounts). Then \( S_{i,k} \) is the increment of the accident year \( i \) in the calendar year \( i + k - 1 \). We assume that the random variables \( S_{i,k} \) are observable up to the calendar year \( p \) (i.e., for \( i + k \leq p + 1 \)) for a \( p \) with \( \max(m, n) \leq p < m + n - 1 \). Let \( \mathcal{A} := \{1, \ldots, m\} \times \{1, \ldots, n\} \) denote the set of all pairs of accident and development years and let \( \mathcal{O} := \{(i, k) \in \mathcal{A} | i + k \leq p + 1\} \) be the subset containing the indices of the observable increments. The goal of a reserving method is to predict the increments \( S_{i,k} \) with \((i,k) \in \mathcal{P} := \mathcal{A} \setminus \mathcal{O} \). We define

\[
I(k) := \min\{p-k+1, m\} \quad \text{and} \quad K(i) := \min\{p-i+1, n\}
\]

such that \( S_{1,k}, \ldots, S_{I(k),k} \) are the observable incremental losses from development year \( k \) and \( S_{i,1}, \ldots, S_{i,K(i)} \) are the observable increments from accident year \( i \).

The incremental loss ratio method or additive method is motivated by a stochastic model, the additive model, see, for instance, Mack [14] or Schmidt and Zocher [17]. In this model it is assumed that volume measures \( \nu_1, \ldots, \nu_n > 0 \) are available. The additive model is defined by the following three model assumptions:

AM1: The increments \( S_{i,k}, (i,k) \in \mathcal{A} \), are independent.

AM2: There are parameters \( \mu_1, \ldots, \mu_n \) such that

\[
\mathbb{E}(S_{i,k}) = \nu_i \mu_k
\]

for \((i,k) \in \mathcal{A}\).

AM3: There are parameters \( \sigma_1, \ldots, \sigma_n > 0 \) such that

\[
\text{Var}(S_{i,k}) = \nu_i \sigma_k^2
\]

for \((i,k) \in \mathcal{A}\).

It is easy to show that
\[ \hat{\mu}_k := \frac{\sum_{i=1}^{I(k)} S_{i,k}}{\sum_{i=1}^{I(k)} v_i} \]

and
\[ \hat{\sigma}_k^2 := \frac{1}{I(k) - 1} \sum_{i=1}^{I(k)} v_i \left( \frac{S_{i,k}}{v_i} - \hat{\mu}_k \right)^2 \]

for \( I(k) > 1 \)

are unbiased estimators for \( \mu_k \) and \( \sigma_k^2 \), see Mack [14]. If \( p = n \), then \( I(n) = 1 \) and the parameter \( \sigma_n^2 \) is often estimated with the extrapolation
\[ \hat{\sigma}_n := \min(\hat{\sigma}_{n-1}, \hat{\sigma}_{n-2}^2). \]

In this paper we will use maximum likelihood (ML) techniques for parameter estimation. The additive model, as defined by the assumptions AM1 to AM3, is distribution-free. If we want to use ML estimators instead of moment estimators we need a distribution assumption. The moment estimators \( \hat{\mu}_k \) can be interpreted as weighted least squares estimators. It is well known that the weighted least square estimators are also the ML estimators if the errors are normally distributed. Thus, it is quite natural to add the following model assumption.

**AM4:** The increments \( S_{i,k} \) are normally distributed.

The normal distribution is not always a good choice for aggregated claims, because it is symmetric and can yield negative values. In the context of the additive model, however, the normal distribution is not a bad assumption, since it is used to model the incremental claims \( S_{i,k} \) which may well be negative, particularly if we consider a triangle containing reported claims.

Using the model assumptions AM1 to AM4 it is straightforward to calculate the ML estimators
\[ \hat{\mu}_k^{\text{ML}} := \frac{\sum_{i=1}^{I(k)} S_{i,k}}{\sum_{i=1}^{I(k)} v_i} \]

and
\[ (\hat{\sigma}_k^{\text{ML}})^2 := \frac{1}{I(k)} \sum_{i=1}^{I(k)} v_i \left( \frac{S_{i,k}}{v_i} - \hat{\mu}_k^{\text{ML}} \right)^2 \]

for \( I(k) > 1 \).

We have \( \hat{\sigma}_k^{\text{ML}} = \hat{\sigma}_k \) for all \( k \). The ML estimators \( (\hat{\sigma}_k^{\text{ML}})^2 \) differ from the moment estimators \( \hat{\sigma}_k^2 \) and have a negative bias.

We will now modify the model assumptions AM1 to AM4 of the additive model to include simple calendar year effects. For \( q \in \{1, \ldots, m+n-1\} \) let
\[ \mathcal{C}(q) := \{(i,k) \in \mathcal{A} \mid i + k - 1 = q\} \]
be the set of indices of the increments in calendar year \( q \). We assume that there is a factor \( F_q \) for each calendar year which scales the means \( E(S_{i,k}) = v_{i,k} \mu_k \) from AM2 for all \((i,k) \in C(q)\) simultaneously. We model these calendar year factors \( F_q \) as independent random variables and assume that the first two moments of the \( F_q \) are provided as prior information, i.e., that \( E(F_q) = \mu_q \) and \( \text{Var}(F_q) = \tau^2_q > 0 \) are known for all \( q \). The intention is to have \( \mu_q = 1 \) for ‘normal’ calendar years, \( \mu_q < 1 \) for calendar years where we assume below average increments and \( \mu_q > 1 \) for calendar years where we assume above average increments. Straightforward generalization of AM1 to AM4 leads to the additive model with simple calendar year effects:

AMSCY1: The calendar years are independent, i.e., the sigma algebras
\[
\sigma\{F_q, S_{i,k} \mid (i,k) \in C(q)\}, \quad q = 1, \ldots, m + n - 1
\]
are independent. Conditionally, given \( F_1, \ldots, F_{m+n-1} \), the random variables \( S_{i,k}, (i,k) \in A \), are independent.

AMSCY2: There are parameters \( \mu_1, \ldots, \mu_n \) such that
\[
E(S_{i,k} \mid F_{i+k-1}) = F_{i+k-1} v_{i,k} \mu_k
\]
for \((i,k) \in A\).

AMSCY3: There are parameters \( \sigma_1, \ldots, \sigma_n > 0 \) such that
\[
\text{Var}(S_{i,k} \mid F_{i+k-1}) = v_{i,k} \sigma_k^2
\]
for \((i,k) \in A\).

AMSCY4: The calendar year factors \( F_1, \ldots, F_{m+n-1} \) are normally distributed. Conditionally, given \( F_1, \ldots, F_{m+n-1} \), the random variables \( S_{i,k}, (i,k) \in A \), are normally distributed.

Remark 1 In AMSCY3 we assume that the conditional variances \( \text{Var}(S_{i,k} \mid F_{i+k-1}) \) do not depend on the calendar year effect \( F_{i+k-1} \). Alternatively, it would also be plausible to assume that the conditional variances scale with the calendar year factors, i.e.,
\[
\text{Var}(S_{i,k} \mid F_{i+k-1}) = F_{i+k-1} v_{i,k} \sigma_k^2.
\]
We do not use this alternative assumption, since it is more difficult to handle in the EM algorithm. Note that the two assumptions are similar if the calendar year factors \( F_{i+k-1} \) are close to 1.

We will now make this model slightly more flexible. Instead of the prior information \( \mu_q \) and \( \tau^2_q \) we allow to provide the first two moments \( E(F_{i,k}) = F_{i,k} \) and \( \text{Var}(F_{i,k}) = \tau^2_{i,k} \geq 0 \) of the scaling factors \( F_{i,k} \) individually for each \((i,k) \in A\). Since we still want to model calendar year effects, we assume that the \( F_{i,k} \) from the same calendar year are perfectly positively dependent.

This modification enables the use of very detailed prior information per individual increment. But this is not the main motivation for the additional flexibility.
Actually, it is questionable whether the ability to include more detailed prior information would justify adding complexity to the model. But the flexibility can be very beneficial in certain situations. For instance, setting all $\tau_{i,k} := 0$ for a development year $k$ means that the calendar year effects do not apply to this development year $k$ (apart from a deterministic scaling with $F_{i,k} = f_{i,k}$). This can be used to handle development years that contain negative increments, see Remark 6.

Let $\mathcal{F}_q$ denote the sigma algebra generated by $\{F_{i,k} \mid (i, k) \in \mathcal{C}(q)\}$. Straightforward generalization of the model assumptions AMSCY1 to AMSCY4 leads to the additive model with calendar year effects:

**AMCY1:** The calendar years are independent, i.e., the sigma algebras

$$\sigma\{F_{i,k}, S_{i,k} \mid (i, k) \in \mathcal{C}(q)\}, \quad q = 1, \ldots, m + n - 1$$

are independent. For every $q$ the vector $(F_{i,k})_{(i,k)\in\mathcal{C}(q)}$ is comonotonic. Conditionally, given $\{F_{i,k} \mid (i, k) \in \mathcal{A}\}$, the random variables $S_{i,k}$, $(i, k) \in \mathcal{A}$, are independent.

**AMCY2:** There are parameters $\mu_1, \ldots, \mu_n$ such that

$$\mathbb{E}(S_{i,k} \mid \mathcal{F}_{i+k-1}) = F_{i,k} v_i \mu_k$$

for $(i, k) \in \mathcal{A}$.

**AMCY3:** There are parameters $\sigma_1, \ldots, \sigma_n > 0$ such that

$$\text{Var}(S_{i,k} \mid \mathcal{F}_{i+k-1}) = v_i \sigma_k^2$$

for $(i, k) \in \mathcal{A}$.

**AMCY4:** The calendar year factors $F_{i,k}$, $(i, k) \in \mathcal{A}$, are either normally distributed (if $\tau_{i,k} > 0$) or constant (if $\tau_{i,k} = 0$). Conditionally, given $\{F_{i,k} \mid (i, k) \in \mathcal{A}\}$, the random variables $S_{i,k}$, $(i, k) \in \mathcal{A}$, are normally distributed.

For the remainder of this paper we work in this additive model with calendar year effects.

**Remark 2** As a consequence of the model assumptions AMCY1 and AMCY4 there exist random variables $X_q \sim \text{Norm}(0, 1)$, $q \in \{1, \ldots, m + n - 1\}$, such that

$$F_{i,k} = f_{i,k} + \tau_{i,k} X_{i+k-1}$$

almost surely for all $(i, k) \in \mathcal{A}$. We refer to the $X_q$ as residual calendar year effects. Note that $\mathcal{F}_q$ is the sigma algebra generated by $X_q$.

**Remark 3** If $f_{i,k} = f_q$ and $\tau_{i,k}^2 = \tau_q^2 > 0$ for all $(i, k) \in \mathcal{C}(q)$, $q = 1, \ldots, m + n - 1$, then we obtain the model assumptions AMSCY1 to AMSCY4 of the additive model with simple calendar year effects.
If \( f_{i,k} = 1 \) and \( \tau_{i,k} = 0 \) for all \((i, k) \in \mathcal{A}\), then we obtain the model assumptions AM1 to AM4 of the additive model. Therefore, the additive model with calendar year effects is a generalization of the classical additive model.

3 An incremental loss ratio method with calendar year effects

The goal of this section is to calculate the ML estimators for the parameters \( \mu_k \) and \( \sigma_k \) of the additive model with calendar year effects. For this purpose, we apply the Expectation Maximization algorithm (EM algorithm). The EM algorithm has been described by Dempster et al. [7] and provides a generic framework to deal with log-likelihood functions that depend on so-called latent variables, i.e., random variables that are not observable. In our case, the latent variables are the residual calendar year effects \( X_q \) which are not observable. We assume \( p > n \) to avoid problems with the estimation of \( \sigma_n \). A pragmatic solution for the case \( p = n \) is provided in Remark 4 below.

The resulting estimators for \( \mu_k \) and \( \sigma_k \) can be used to predict the future incremental claims, leading to the incremental loss ratio method with calendar year effects. This method suffers from the problem that many parameters have to be estimated. In particular, the estimation of individual variance parameters \( \sigma_k \) for each development year \( k \) can lead to instabilities and unexpected results. In Sect. 5 we address this issue and introduce a more robust variant which uses less parameters. For application in practice, we recommend to use this robustified method.

Let \( S := (S_{i,k})_{(i,k) \in \mathcal{O}} \) denote the vector consisting of the observable incremental losses and \( X := (X_1, \ldots, X_p) \) the vector containing the residual calendar year effects of the observable calendar years. Note that \( X \) is not observable. Let \( \Theta := (\mu, \sigma) := (\mu_1, \ldots, \mu_n, \sigma_1, \ldots, \sigma_n) \). The density \( p_{(X,S)}(\cdot\,;\theta) \) of \((X, S)\) has a very simple form:

\[
p_{(X,S)}(x,s;\theta) = \prod_{q=1}^{p} \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{x_q^2}{2} \right) \cdot \prod_{(i,k) \in \mathcal{O}} \frac{1}{\sqrt{2\pi \sigma_k^2}} \exp\left( -\frac{[s_{i,k} - (\tau_{i,k} x_{i,k-1} + f_{i,k}) \cdot \nu_i \mu_k]^2}{2 \nu_i \sigma_k^2} \right).
\]

In order to calculate the ML estimators, we have to maximize the log-likelihood of the observable data \( S \)

\[
l_S(\theta) = \log p_S(S;\theta) = \log \int_{\mathbb{R}^p} p_{(X,S)}(x,S;\theta) \, dx.
\]

Unfortunately, \( l_S \) is not as simple as the log-likelihood \( l_{(X,S)}(\theta) = \log p_{(X,S)}(X,S;\theta) \) of the complete data \((X, S)\). The EM algorithm is an iterative method to find (local) maxima of the observable log-likelihood in such a situation.

The EM iteration alternates between performing an expectation step (E step), which calculates the expected log-likelihood function using the current estimate for the parameters, and a maximization step (M step), which computes new parameters.
maximizing the expected log-likelihood found in the E step. These parameter estimates are then used to determine the distribution of the latent variables and the expected log-likelihood function in the next E step. In our case we define

$$Q(\Theta|\Theta') := E_{X|S;\Theta'} \left( \log \left( p(X,S;\Theta) \right) \right)$$

and start with an initial parameter vector $\Theta^{(0)}$. Given $\hat{\Theta}^{(v)}$ we calculate $Q(\Theta|\hat{\Theta}^{(v)})$ in the E step and in the M step we choose $\hat{\Theta}^{(v+1)}$ to be any value which maximizes $Q(\cdot|\hat{\Theta}^{(v)})$.

The following proposition provides the conditional distribution of the latent variables $X$, which is needed in the calculation of the expected log-likelihood in the E step. We define the functions $a_q : \mathbb{R}^n \times (0, \infty)^n \rightarrow \mathbb{R}$ and $b_q : \mathbb{R}^n \times (0, \infty)^n \rightarrow \mathbb{R}$, $q = 1, \ldots, p$, by

$$a_q(\mu, \sigma) := \left( 1 + \sum_{(i,k) \in C(q)} \frac{v_i \mu^2_k \tau_{i,k}^2}{\sigma_k^2} \right)^{-1}$$

and

$$b_q(\mu, \sigma) := \left( 1 + \sum_{(i,k) \in C(q)} \frac{v_i \mu^2_k \tau_{i,k}^2}{\sigma_k^2} \right)^{-1} \cdot \left( \sum_{(i,k) \in C(q)} \frac{S_{i,k} \mu_k - v_i \mu_k^2 \tau_{i,k} f_{i,k}^2}{\sigma_k^2} \tau_{i,k} \right).$$

**Proposition 1** Conditionally, given $S$ and a parameter vector $\Theta$, the vector $X$ is normally distributed with mean $b(\Theta)$ and covariance matrix $\Sigma(\Theta)$, where

$$b(\Theta) := \begin{pmatrix} b_1(\Theta) \\ \vdots \\ b_p(\Theta) \end{pmatrix} \quad \text{and} \quad \Sigma(\Theta) := \begin{pmatrix} a_1(\Theta) & 0 \\ \vdots & \ddots \\ 0 & \ddots & a_p(\Theta) \end{pmatrix}.$$

**Proof** For $q = 1, \ldots, p$ we have

\[
\begin{align*}
\frac{x^2_q}{2} + \frac{1}{2} \sum_{(i,k) \in C(q)} \left( S_{i,k} - (\tau_{i,k} x_{i+k} - f_{i,k}) \cdot v_i \mu_k \right)^2 \\
= \frac{x^2_q}{2} + \frac{1}{2} \sum_{(i,k) \in C(q)} \frac{v_i \mu^2_k \tau^2_{i,k}}{\sigma_k^2} \left( \tau_{i,k} x_q + f_{i,k} \right)^2 - 2 S_{i,k} (\tau_{i,k} x_q + f_{i,k}) v_i \mu_k + S_{i,k}^2 \\
= \frac{1}{2} \left( 1 + \sum_{(i,k) \in C(q)} \frac{v_i \mu^2_k \tau^2_{i,k}}{\sigma_k^2} \right) x^2_q - \left( \sum_{(i,k) \in C(q)} \frac{S_{i,k} \mu_k \tau_{i,k} - v_i \mu_k^2 \tau_{i,k} f_{i,k} \tau_{i,k}}{\sigma_k^2} \right) x_q + \text{const} \\
= \frac{1}{2} a_q(\Theta)^{-1} x^2_q - a_q(\Theta)^{-1} b_q(\Theta) x_q + \text{const} \\
= \frac{1}{2} a_q(\Theta)^{-1} (x_q - b_q(\Theta))^2 + \text{const}.
\end{align*}
\]
In this calculation we have used ‘const’ for expressions that do not depend on \((x_1, \ldots, x_p)\). Consequently, we have

\[
p_{X|S;\Theta}(x) = \text{const} \cdot \prod_{q=1}^{p} \exp \left( -\frac{1}{2} a_q(\Theta)^{-1} (x_q - b_q(\Theta))^2 \right).
\]

We define

\[
c_{i,k}(\Theta) := \tau_{i,k}^2 \cdot a_{i+k-1}(\Theta), \quad d_{i,k}(\Theta) := f_{i,k} + \tau_{i,k} b_{i+k-1}(\Theta).
\]

Then \(E(F_{i,k} | S;\Theta) = d_{i,k}(\Theta)\) and \(\text{Var}(F_{i,k} | S;\Theta) = c_{i,k}(\Theta)\). Given an estimate \(\hat{\Theta}^{(v)}\), we can use Proposition 1 to perform the E step:

\[
Q(\Theta|\hat{\Theta}^{(v)}) = E_{X|S;\hat{\Theta}^{(v)}} \left( \log \left( p(X,S)(X,S;\Theta) \right) \right)
\]

\[
= E_{X|S;\hat{\Theta}^{(v)}} \left( \sum_{q=1}^{p} \left( \frac{X_q^2}{2} \right) \right) + E_{X|S;\hat{\Theta}^{(v)}} \left( \sum_{(i,k) \in O} \left( -\log(\sigma_k) - \frac{(S_{i,k} - F_{i,k} \cdot v_i \mu_k)^2}{2v_i \sigma_k^2} \right) \right) + \text{const}
\]

\[
= \sum_{q=1}^{p} \left( \frac{a_q(\hat{\Theta}^{(v)}) + b_q(\hat{\Theta}^{(v)})^2}{2} \right) + \sum_{(i,k) \in O} \left( -\log(\sigma_k) - \frac{v_i^2 \mu_k^2 c_{i,k}(\hat{\Theta}^{(v)}) + (S_{i,k} - d_{i,k}(\hat{\Theta}^{(v)}) v_i \mu_k)^2}{2v_i \sigma_k^2} \right) + \text{const}.
\]

In the M step we have to solve the following equations for \(\Theta\) to obtain \(\hat{\Theta}^{(v+1)}\):

\[
0 = \frac{\partial}{\partial \mu_k} Q(\Theta|\hat{\Theta}^{(v)}) = \sum_{i=1}^{I(k)} d_{i,k}(\hat{\Theta}^{(v)}) S_{i,k} - (c_{i,k}(\hat{\Theta}^{(v)}) + d_{i,k}(\hat{\Theta}^{(v)})^2) v_i \mu_k,
\]

\[
0 = \frac{\partial}{\partial \sigma_k} Q(\Theta|\hat{\Theta}^{(v)}) = \sum_{i=1}^{I(k)} \left( \frac{1}{\sigma_k} + \frac{(S_{i,k} - d_{i,k}(\hat{\Theta}^{(v)}) v_i \mu_k)^2 + v_i^2 \mu_k^2 c_{i,k}(\hat{\Theta}^{(v)})}{v_i \sigma_k^2} \right).
\]

We obtain the following iteration.

**Iteration for parameter estimation** We start with the ML estimators in the normal additive model:

\[
\hat{\mu}^{(0)} := \frac{\sum_{i=1}^{I(k)} S_{i,k}}{\sum_{i=1}^{I(k)} v_i}, \quad \text{and} \quad \left(\hat{\sigma}^{(0)}\right)^2 := \frac{1}{I(k)} \sum_{i=1}^{I(k)} v_i \left( \frac{S_{i,k}}{v_i} - \hat{\mu}^{(0)} \right)^2.
\]

We set \(\hat{\Theta}^{(0)} := (\hat{\mu}^{(0)}, \ldots, \hat{\sigma}^{(0)}, \ldots, \hat{\sigma}^{(0)}). \) Given \(\hat{\Theta}^{(v)} := (\hat{\mu}^{(v)}, \ldots, \hat{\sigma}^{(v)}, \ldots, \hat{\sigma}^{(v)})\) we calculate \(\hat{\Theta}^{(v+1)} := (\hat{\mu}^{(v+1)}, \ldots, \hat{\mu}^{(v+1)}, \hat{\sigma}^{(v+1)}, \ldots, \hat{\sigma}^{(v+1)})\) by

 UserModel. Springer
\[ \hat{\mu}^{(v+1)}_k := \frac{\sum_{i=1}^{l(k)} d_{i,k}(\hat{\Theta}^{(v)}) S_{i,k}}{\sum_{i=1}^{l(k)} (c_{i,k}(\hat{\Theta}^{(v)}) + d_{i,k}(\hat{\Theta}^{(v)})^2) v_i} \]

and

\[ (\hat{\sigma}^{(v+1)}_k)^2 := \frac{1}{I(k)} \sum_{i=1}^{l(k)} \left[ v_i \cdot \left( \frac{S_{i,k}}{v_i} - d_{i,k}(\hat{\Theta}^{(v)}) \cdot \hat{\mu}^{(v+1)}_k \right)^2 + v_i \cdot (\hat{\sigma}^{(v+1)}_k)^2 \cdot c_{i,k}(\hat{\Theta}^{(v)}) \right]. \]

If the algorithm converges, we define

\[ \hat{\mu}^{(\infty)}_k := \lim_{v \to \infty} \hat{\mu}^{(v)}_k, \quad \hat{\sigma}^{(\infty)}_k := \lim_{v \to \infty} \hat{\sigma}^{(v)}_k \quad \text{and} \quad \hat{\Theta}^{(\infty)} := \lim_{v \to \infty} \hat{\Theta}^{(v)}. \]

Moreover, we call \( \hat{F}_{i,k}^{(\infty)} := d_{i,k}(\hat{\Theta}^{(\infty)}) = E(F_{i,k} | S; \hat{\Theta}^{(\infty)}) \) for \( (i, k) \in O \) the posterior calendar year factors.

**Remark** It has been shown by Dempster et al. [7] that the EM algorithm increases the likelihood of the observable data in every step. In general, it is known that the EM algorithm converges to the ML estimator in many cases, although convergence is often slow. The convergence properties of the EM algorithm have been studied by Wu [22] and Vaida [20]. In our situation, unfortunately, we cannot guarantee convergence to the ML estimators for arbitrary loss triangles and arbitrary prior information.

Using the estimators from the EM algorithm we obtain the following reserving method.

**Incremental loss ratio method with calendar year effects** For \( (i, k) \in P \)

\[ \hat{S}_{i,k}^{(\infty)} := v_i \cdot f_{i,k} \cdot \hat{\mu}^{(\infty)}_k \]

are called the predictions of the incremental loss ratio method with calendar year effects.

**Remark 4** For the application of the EM algorithm we have assumed \( p > n \). If \( p = n \), we use the formulas provided in the iteration for \( \hat{\mu}_1^{(v+1)}, \ldots, \hat{\mu}_n^{(v+1)} \) and for \( \hat{\sigma}_1^{(v+1)}, \ldots, \hat{\sigma}_n^{(v+1)} \). For \( \hat{\sigma}_n^{(v+1)} \) we use the extrapolation

\[ \hat{\sigma}_n^{(v+1)} := \min(\hat{\sigma}_{n-1}^{(v+1)}, (\hat{\sigma}_{n-1}^{(v+1)})^2 / \hat{\sigma}_{n-2}^{(v+1)}). \]

**Remark 5** If \( f_{i,k} = 1 \) and \( r_{i,k} = 0 \) for all \( (i, k) \in A \), then the loss ratio method with calendar year effects leads to the same predictions as the classical loss ratio method. In this case, the estimators for the standard errors that are derived in the next section equal the estimators from the additive model with ML estimators for the \( \sigma_k^2 \).

Since the ML estimators for \( \sigma_k^2 \) have a negative bias, one may want to be more consistent with the standard errors of the additive model with moment estimators. Replacing the factor \( I(k)^{-1} \) by \( (I(k) - 1)^{-1} \) in \( (\hat{\sigma}_k^{(0)})^2 \) and in the first summand of \( (\hat{\sigma}_k^{(v+1)})^2 \) in the iteration yields
An incremental loss ratio method using prior information on…

\[(\hat{\sigma}_k^{(0)})^2 := \frac{1}{I(k) - 1} \sum_{i=1}^{I(k)} v_i \left( \frac{S_{i,k}}{v_i} - \hat{\mu}_k^{(0)} \right)^2 \]

and

\[(\hat{\sigma}_k^{(v+1)})^2 := \frac{1}{I(k) - 1} \sum_{i=1}^{I(k)} v_i \left( \frac{S_{i,k}}{v_i} - d_{i,k}(\hat{\Theta}^{(v)}) \cdot \hat{\mu}_k^{(v+1)} \right)^2 + \frac{1}{I(k)} \sum_{i=1}^{I(k)} v_i \cdot (\hat{\mu}_k^{(v+1)})^2 \cdot c_{i,k}(\hat{\Theta}^{(v)}) \]

for \(k\) with \(I(k) > 1\). With this adjustment we obtain the moment estimators for \(\sigma_k^2\) from the iteration if \(f_{i,k} \to 1\) and \(r_{i,k} \to 0\) for all \((i, k) \in A\). Such a modification could also be considered plausible because the means \(d_{i,k}(\hat{\Theta}^{(v)}) \cdot \hat{\mu}_k^{(v+1)}\) in the sum

\[\sum_{i=1}^{I(k)} v_i \cdot \left( \frac{S_{i,k}}{v_i} - d_{i,k}(\hat{\Theta}^{(v)}) \cdot \hat{\mu}_k^{(v+1)} \right)^2 \]

are not known but estimated, which reduces the degrees of freedom from \(I(k)\) to \(I(k) - 1\). It is, however, not obvious whether the resulting estimators have any desirable properties apart from being more consistent with the moment estimators in the additive model.

If we use this modified iteration, we speak of the \textit{moment adjusted variant} of the method. For the estimation of the MSEP, we use the formulas from Sect. 4 as in the case without adjustment. Note that the adjustment does not only influence the estimated standard errors, but also the predictions of the future losses.

\textbf{Remark 6} It is questionable whether our multiplicative calendar year effects are plausible for development years with negative increments. If all increments of a triangle are positive, then a calendar year factor greater than 1 means that the loss increments of this calendar year are above average. For negative increments, however, such a factor greater than 1 means that the calendar year is \textit{better} than average. In certain cases it can be realistic to apply the same calendar year factor to positive and negative increments, for instance, if the calendar year effect is caused by a delay in claims handling, which may have a similar multiplicative impact on positive and negative developments. But in many cases this behavior is not desirable. In these cases, we suggest to use \(r_{i,k} := 0, i = 1, \ldots, m\), for development years \(k\) where at least one negative increment has been observed. This means that the calendar year effects do not apply to such development years \(k\) (apart from a deterministic scaling with \(f_{i,k}\)).

\section{Prediction error}

In this section, we derive a formula for the mean squared error of prediction (MSEP) of the incremental loss ratio method with calendar year effects. As mentioned above, a more robust version of the method will be introduced in Sect. 5. There,
we will also derive formulas for the MSEP of this robustified method from the formulas that are presented in the current section. For \((i, k) \in \mathcal{P}\) let \(w_{i,k} \in [0, 1]\). Let 
\[
\hat{R}^{(\infty)}_w := \sum_{(i,k) \in \mathcal{P}} w_{i,k} \hat{s}^{(\infty)}_{i,k} \quad \text{and} \quad R_w := \sum_{(i,k) \in \mathcal{P}} w_{i,k} S_{i,k}.
\]

If \(w_{i,k} = 1\) for all \((i, k) \in \mathcal{P}\) then \(R_w\) is the total reserve for all accident years. In this case, we write \(\hat{R}^{(\infty)}_{\text{total}} := \hat{R}^{(\infty)}_w\) and \(R_{\text{total}} := R_w\). If 
\[
w_{j,k} = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i, \end{cases}
\]
then \(R_w\) is the reserve for the accident year \(i\) and we write \(\hat{R}^{(\infty)}_i := \hat{R}^{(\infty)}_w\) and \(R_i := R_w\). Let 
\[
\mathcal{D} := \sigma\{S_{i,k} \mid (i, k) \in \mathcal{O}\}
\]
be the sigma algebra generated by the observable loss data. Actually, we are interested in the conditional MSEP 
\[
\text{mse}_{R_w \mid \mathcal{D}}(\hat{R}^{(\infty)}_w) := E\left(\left(\hat{R}_w - \hat{R}^{(\infty)}_w\right)^2 \mid \mathcal{D}\right)
\]
\[
= \text{Var}(R_w \mid \mathcal{D}) + \left(\hat{R}^{(\infty)}_w - E(\hat{R}_w \mid \mathcal{D})\right)^2
\]
\[
= \text{Var}(R_w) + \left(\hat{R}^{(\infty)}_w - E(R_w)\right)^2.
\]

Since it is not possible to calculate \((\hat{R}^{(\infty)}_w - E(R_w))^2\), we use the approximation 
\[
(\hat{R}^{(\infty)}_w - E(R_w))^2 \approx (\hat{R}^{(\infty)}_w - E(\hat{R}^{(\infty)}_w))^2 \approx \text{Var}(\hat{R}^{(\infty)}_w)
\]
which leads to 
\[
\text{mse}_{R_w \mid \mathcal{D}}(\hat{R}^{(\infty)}_w) \approx \text{Var}(R_w) + \text{Var}(\hat{R}^{(\infty)}_w)
\]
with the process variance \(\text{Var}(R_w)\) and the squared estimation error \(\text{Var}(\hat{R}^{(\infty)}_w)\).

**Theorem 1** For the process variance we have 
\[
\text{Var}(R_w) = \sum_{(i,k) \in \mathcal{P}} \sum_{(j,l) \in \mathcal{P}} w_{i,k} w_{j,l} \left(\delta_{i,j} \delta_{k,l} \sigma_k^2 + \delta_{i+k+l,j+l} \mu_k \mu_l \tau_{i,k} \tau_{j,l}\right),
\]
where 
\[
\delta_{i,j} := \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j, \end{cases}
\]
denotes the Kronecker symbol. For the squared estimation error we have
\[ \text{Var}(\hat{R}_w^{(\infty)}) = \sum_{(i,k) \in \mathcal{P}} \sum_{(j,l) \in \mathcal{P}} w_{i,k} w_{j,l} v_{i,j} \hat{\mu}_k \hat{\mu}_l \text{Cov}(\hat{\mu}_k^{(\infty)}, \hat{\mu}_l^{(\infty)}). \]

**Proof** For \((i, k) \in \mathcal{P}\) we have

\[
\text{Var}(S_{i,k}) = \text{E}(\text{Var}(S_{i,k} | X_{i+k-1})) + \text{Var}(\text{E}(S_{i,k} | X_{i+k-1})) \\
= \text{E}(v_i \sigma_k^2) + \text{Var}((f_{i,k} + \tau_{i,k} X_{i+k-1}) v_i \mu_k) \\
= v_i \sigma_k^2 + v_i^2 \mu_k^2 \tau_k^2.
\]

For \((i, k), (j, l) \in \mathcal{C}(q)\) with \((i, k) \neq (j, l)\) we obtain

\[
\text{Cov}(S_{i,k}, S_{j,l}) = \text{E}(\text{Cov}(S_{i,k} S_{j,l} | X_{i+k-1}, X_{j+l-1})) + \text{Cov}(\text{E}(S_{i,k} | X_{i+k-1}), \text{E}(S_{j,l} | X_{j+l-1})) \\
= 0 + \text{Cov}((f_{i,k} + \tau_{i,k} X_{i+k-1}) v_i \mu_k, (f_{j,l} + \tau_{j,l} X_{j+l-1}) v_j \mu_l) \\
= v_i v_j \mu_k \mu_l \tau_{i,k} \tau_{j,l} \text{Var}(X_q) = v_i v_j \mu_k \mu_l \tau_{i,k} \tau_{j,l}.
\]

Eventually, we have for \((i, k), (j, l) \in \mathcal{P}\) from different calendar years (i.e. \(i + k \neq j + l\))

\[
\text{Cov}(S_{i,k}, S_{j,l}) = \text{E}(\text{Cov}(S_{i,k} S_{j,l} | X_{i+k-1}, X_{j+l-1})) + \text{Cov}(\text{E}(S_{i,k} | X_{i+k-1}), \text{E}(S_{j,l} | X_{j+l-1})) \\
= 0 + \text{Cov}((f_{i,k} + \tau_{i,k} X_{i+k-1}) v_i \mu_k, (f_{j,l} + \tau_{j,l} X_{j+l-1}) v_j \mu_l) \\
= v_i v_j \mu_k \mu_l \tau_{i,k} \tau_{j,l} \text{Cov}(X_{i+k-1}, X_{j+l-1}) = 0.
\]

Therefore

\[
\text{Var}(R_w) = \text{Var}\left( \sum_{(i,k) \in \mathcal{P}} w_{i,k} S_{i,k} \right) \\
= \sum_{(i,k) \in \mathcal{P}} \sum_{(j,l) \in \mathcal{P}} w_{i,k} w_{j,l} \text{Cov}(S_{i,k}, S_{j,l}) \\
= \sum_{(i,k) \in \mathcal{P}} \sum_{(j,l) \in \mathcal{P}} w_{i,k} w_{j,l} \left( \delta_{i,j} \delta_{k,l} v_i \sigma_k^2 + \delta_{i+k,j+l} v_i v_j \mu_k \mu_l \tau_{i,k} \tau_{j,l} \right).
\]

Moreover, we have

\[
\text{Var}(\hat{R}_w^{(\infty)}) = \text{Var}\left( \sum_{(i,k) \in \mathcal{P}} w_{i,k} \hat{S}_{i,k}^{(\infty)} \right) \\
= \sum_{(i,k) \in \mathcal{P}} \sum_{(j,l) \in \mathcal{P}} w_{i,k} w_{j,l} \text{Cov}(\hat{S}_{i,k}^{(\infty)}, \hat{S}_{j,l}^{(\infty)}) \\
= \sum_{(i,k) \in \mathcal{P}} \sum_{(j,l) \in \mathcal{P}} w_{i,k} w_{j,l} \text{Cov}(v_i f_{i,k} \hat{\mu}_k^{(\infty)}, v_j f_{j,l} \hat{\mu}_l^{(\infty)}) \\
= \sum_{(i,k) \in \mathcal{P}} \sum_{(j,l) \in \mathcal{P}} w_{i,k} w_{j,l} v_i v_j f_{i,k} f_{j,l} \text{Cov}(\hat{\mu}_k^{(\infty)}, \hat{\mu}_l^{(\infty)}).
\]
Let \( \hat{\alpha}_q := a_q(\hat{\Theta}^{(\infty)}) \), \( \hat{\beta}_q := b_q(\hat{\Theta}^{(\infty)}) \), \( \hat{\gamma}_q := c_q(\hat{\Theta}^{(\infty)}) \) and \( \hat{\delta}_q := d_q(\hat{\Theta}^{(\infty)}) \). Let
\[
\pi_2 : \mathbb{N} \times \mathbb{N} \to \mathbb{N}, \quad (i, k) \mapsto k
\]
denote the projection on the second coordinate. We define
\[
\mathcal{K}(q) := \pi_2(\mathcal{C}(q)) = \{ k \in \mathbb{N} \mid q - k + 1 \leq m, k \leq \min(q, n) \}.
\]

In order to estimate the covariances \( \text{Cov}(\hat{\mu}_k^{(\infty)}, \hat{\mu}_l^{(\infty)}) \) we define for \( q \in \{1, \ldots, p\} \) and \( l \in \{1, \ldots, n\} \) with \( l \in \mathcal{K}(q) \)
\[
\hat{\alpha}_{q,l}^{(\mu)} := -\hat{\alpha}_q^2 \cdot \frac{2v_{q-l+1} \hat{\mu}_l^{(\infty)} \tau_{q-l+1,l}}{(\hat{\sigma}_l^{(\infty)})^2},
\]
\[
\hat{\alpha}_{q,l}^{(\sigma)} := \hat{\alpha}_q^2 \cdot \frac{2v_{q-l+1} (\hat{\mu}_l^{(\infty)})^2 \tau_{q-l+1,l}^2}{(\hat{\sigma}_l^{(\infty)})^3},
\]
\[
\hat{\rho}_{q,l}^{(\mu)} := \hat{\alpha}_{q,l}^{(\mu)} \sum_{(i,k) \in \mathcal{C}(q)} \frac{S_{i,k} \hat{\mu}_k^{(\infty)} - v_i (\hat{\mu}_k^{(\infty)})^2 f_{i,k}}{(\hat{\sigma}_k^{(\infty)})^2} + \hat{\alpha}_q \cdot \frac{S_{q-l+1,l} - 2v_{q-l+1} \hat{\mu}_l^{(\infty)} f_{q-l+1,l}}{(\hat{\sigma}_l^{(\infty)})^2} \tau_{q-l+1,l},
\]
\[
\hat{\rho}_{q,l}^{(\sigma)} := \hat{\alpha}_{q,l}^{(\sigma)} \sum_{(i,k) \in \mathcal{C}(q)} \frac{S_{i,k} \hat{\mu}_k^{(\infty)} - v_i (\hat{\mu}_k^{(\infty)})^2 f_{i,k}}{(\hat{\sigma}_k^{(\infty)})^2} - 2\hat{\alpha}_q \cdot \frac{S_{q-l+1,l} \hat{\mu}_l^{(\infty)} - 2v_{q-l+1} (\hat{\mu}_l^{(\infty)})^2 f_{q-l+1,l}}{(\hat{\sigma}_l^{(\infty)})^3} \tau_{q-l+1,l}
\]
and for \( l \notin \mathcal{K}(q) \)
\[
\hat{\alpha}_{q,l}^{(\mu)} := \hat{\alpha}_{q,l}^{(\sigma)} := \hat{\rho}_{q,l}^{(\mu)} := \hat{\rho}_{q,l}^{(\sigma)} := 0.
\]

Moreover, let
\[
\hat{\gamma}_{i,k,l}^{(\mu)} := \tau_{i,k} \hat{\alpha}_{i+k-l,1}^{(\mu)}, \quad \hat{\gamma}_{i,k,l}^{(\sigma)} := \tau_{i,k} \hat{\alpha}_{i+k-l,1}^{(\sigma)}, \quad \hat{\delta}_{i,k,l}^{(\mu)} := \tau_{i,k} \hat{\rho}_{i+k-1,l}^{(\mu)}, \quad \hat{\delta}_{i,k,l}^{(\sigma)} := \tau_{i,k} \hat{\rho}_{i+k-1,l}^{(\sigma)}.
\]

We define \( \hat{\mathcal{I}}^{(\infty)} := (\hat{I}_{i,k}^{(\infty)})_{i,k} \in \mathbb{R}^{2n \times 2n} \) by
Theorem 1 by as estimator for i.e., we use an approximation.

\[ \hat{\gamma}_{k,l}^{(\infty)} := \delta_{k,l} \cdot \left( \sum_{i=1}^{n(k)} \frac{(\hat{c}_{i,k} + \hat{d}_{i,k}^2)v_i}{(\hat{\sigma}_k^{(\infty)})^2} \right) \]

\[ - \sum_{i=1}^{n(k)} \frac{(\hat{\delta}_{i,k,l}^{(\mu)})S_{i,k} - (\hat{\gamma}_{i,k,l}^{(\mu)} + 2\hat{d}_{i,k}\hat{\delta}_{i,k,l}^{(\mu)})v_i\hat{\mu}_k^{(\infty)}}{(\hat{\sigma}_k^{(\infty)})^2}, \]

\[ \hat{\gamma}_{k,n+1}^{(\infty)} := \hat{\gamma}_{n+1,k}^{(\infty)} := 2\delta_{k,l} \cdot \left( \sum_{i=1}^{n(k)} \frac{\hat{d}_{i,k}S_{i,k} - (\hat{c}_{i,k} + \hat{d}_{i,k}^2)v_i\hat{\mu}_k^{(\infty)}}{(\hat{\sigma}_k^{(\infty)})^3} \right) \]

\[ - \sum_{i=1}^{n(k)} \frac{(\hat{\delta}_{i,k,l}^{(\sigma)})S_{i,k} - (\hat{\gamma}_{i,k,l}^{(\sigma)} + 2\hat{d}_{i,k}\hat{\delta}_{i,k,l}^{(\sigma)})v_i\hat{\mu}_k^{(\infty)}}{(\hat{\sigma}_k^{(\infty)})^2}, \]

\[ \hat{\gamma}_{n+1,n+1}^{(\infty)} := \delta_{k,l} \cdot \left( \sum_{i=1}^{n(k)} - \frac{1}{(\hat{\sigma}_k^{(\infty)})^2} + 3 \frac{(S_{i,k} - \hat{d}_{i,k}v_i\hat{\mu}_k^{(\infty)})^2 + v_i^2(\hat{\mu}_k^{(\infty)})^2\hat{c}_{i,k}}{(\hat{\sigma}_k^{(\infty)})^4} \right) \]

\[ + \sum_{i=1}^{n(k)} 2(S_{i,k} - \hat{d}_{i,k}v_i\hat{\mu}_k^{(\infty)})\hat{\delta}_{i,k,l}^{(\sigma)}\hat{\gamma}_{i,k,l}^{(\sigma)} \frac{v_i(\hat{\mu}_k^{(\infty)})^2\hat{c}_{i,k}}{(\hat{\sigma}_k^{(\infty)})^3}, \]

for \( k, l = 1, \ldots, n. \)

**Theorem 2** \( \hat{\gamma}^{(\infty)} \) is the observed information matrix at \( \hat{\Theta}^{(\infty)}. \)

The proof of Theorem 2 is based on a result from Oakes [15] and is found in Appendix A.

If \( \hat{\gamma}^{(\infty)} \) is invertible, we use the approximation

\[ \text{Cov}((\hat{\mu}_1^{(\infty)}, \ldots, \hat{\mu}_n^{(\infty)}, \hat{\sigma}_1^{(\infty)}, \ldots, \hat{\sigma}_n^{(\infty)})) \approx (\hat{\gamma}^{(\infty)})^{-1}, \]

i.e., we use

\[ \hat{\text{Cov}}(\hat{\mu}_i^{(\infty)}, \hat{\mu}_k^{(\infty)}) := [(\hat{\gamma}^{(\infty)})^{-1}]_{i,k} \]

as estimator for \( \text{Cov}(\hat{\mu}_i^{(\infty)}, \hat{\mu}_k^{(\infty)}) \) for \( i, k = 1, \ldots, n. \) Note that the inverse Fisher information \((\hat{\gamma}^{(\infty)})^{-1}\) is only the limiting covariance matrix of the ML estimators when the sample size goes to infinity. In our case, we often deal with small sample sizes (as in the numerics section), which means that the inverse Fisher information is only an approximation.

We can now use Theorem 1 to obtain an estimator for the mean squared error of prediction for the method.

**Estimator 1** Replacing the parameters \( \mu_k, \sigma_i \) and the covariances \( \text{Cov}(\hat{\mu}_k^{(\infty)}, \hat{\mu}_k^{(\infty)}) \) in Theorem 1 by \( \hat{\mu}_k^{(\infty)}, \hat{\sigma}_k^{(\infty)} \) and \( \hat{\text{Cov}}(\hat{\mu}_k^{(\infty)}, \hat{\mu}_k^{(\infty)}) \), respectively, we obtain the estimator
\[ \hat{\text{Var}}(R_w) := \sum_{(i,k) \in P} \sum_{(j,l) \in P} w_{i,k} w_{j,l} \left( \delta_{i,j} \delta_{k,l} v_i (\hat{\sigma}^2_k) + \delta_{i+k,j+l} v_i v_j \hat{\mu}_k (\hat{\mu}_l) \tau_{i,k} \tau_{j,l} \right) \]

for the process variance and the estimator
\[ \hat{\text{Var}}(\hat{R}_w^{(\infty)}) := \sum_{(i,k) \in P} \sum_{(j,l) \in P} w_{i,k} w_{j,l} v_i v_j f_{i,k} f_{j,l} \hat{\text{Cov}}(\hat{\mu}_k (\hat{\mu}_l)) \]

for the squared estimation error. The conditional MSEP \( \text{mse}_{R_i | \mathcal{S}}(\hat{R}_w^{(\infty)}) \) can then be estimated by
\[ \hat{\text{mse}}_{R_i | \mathcal{S}}(\hat{R}_w^{(\infty)}) := \hat{\text{Var}}(R_w) + \hat{\text{Var}}(\hat{R}_w^{(\infty)}). \]

In particular, we have for the accident year \( i \)
\[ \hat{\text{mse}}_{R_i | \mathcal{S}}(\hat{R}_i^{(\infty)}) = \hat{\text{Var}}(R_i) + \hat{\text{Var}}(\hat{R}_i^{(\infty)}) \]

with
\[ \hat{\text{Var}}(R_i) = \sum_{k=K(i)+1}^{n} \left( v_i (\hat{\sigma}^2_k) + v_i^2 (\hat{\mu}_k) \tau_{i,k}^2 \right) \]

and
\[ \hat{\text{Var}}(\hat{R}_i^{(\infty)}) = \sum_{k=K(i)+1}^{n} \sum_{l=K(i)+1}^{n} v_i^2 f_{i,k} f_{j,l} \hat{\text{Cov}}(\hat{\mu}_k (\hat{\mu}_l)). \]

For the total reserve we obtain
\[ \hat{\text{mse}}_{R_{\text{total}} | \mathcal{S}}(\hat{R}_{\text{total}}^{(\infty)}) = \hat{\text{Var}}(R_{\text{total}}) + \hat{\text{Var}}(\hat{R}_{\text{total}}^{(\infty)}), \]

where
\[ \hat{\text{Var}}(R_{\text{total}}) = \sum_{(i,k) \in P} v_i (\hat{\sigma}^2_k) + \sum_{q=p+1}^{m+n-1} \sum_{(k,l) \in C(q)} v_i v_j \hat{\mu}_k (\hat{\mu}_l) \tau_{i,k} \tau_{j,l} \]

and
\[ \hat{\text{Var}}(\hat{R}_{\text{total}}^{(\infty)}) = \sum_{(i,k) \in P} \sum_{(j,l) \in P} v_i v_j f_{i,k} f_{j,l} \hat{\text{Cov}}(\hat{\mu}_k (\hat{\mu}_l)). \]

5 Robustification of the method

As mentioned in Sect. 3, the incremental loss ratio method with calendar year effects can be instable since the variance parameters \( \sigma_1, \ldots, \sigma_n \) are being estimated for all development years individually. In particular, outliers can lead to unexpected effects.
Therefore, it is desirable to stabilize the estimation of the \( \sigma_k \) and to reduce the number of parameters. In this section we describe how this can be achieved by using a variance function. Moreover, we derive an estimator for the MSEP of the resulting robustified method from the results of Sect. 4. Of course, also other approaches to stabilize the estimation of the variance parameter are conceivable.

5.1 A robustified incremental loss ratio method with calendar year effects

We select a continuously differentiable variance function \( V : \mathbb{R} \to (0, \infty), \mu \mapsto V(\mu) \) and assume that \( \sigma_k^2 = V(\mu_k)\sigma_0^2 \) for all \( k \). Note that this assumption implies

\[
\text{Var} \left( \frac{S_{i,k}}{v_i} \bigg| F_{i+k-1} \right) = \frac{\sigma_k^2}{v_i} = \frac{\sigma_0^2}{v_i} \cdot V(\mu_k),
\]

which is consistent with the meaning of the variance function of an exponential dispersion family with weights \( v_i \). Natural candidates for such variance functions are continuously differentiable functions \( V \) with

\[
V(\mu) = |\mu|^p \quad \text{for } |\mu| > \varepsilon
\]

for a small \( \varepsilon > 0 \). For \( \mu > \varepsilon \) this means \( V(\mu) = \mu^p \) which is a common variance function for GLMs. In particular, for \( 1 < p < 2 \) we obtain the variance structure of the Tweedie family, which is a popular choice for ratemaking with GLMs since it can be interpreted as a collective risk model with Poisson frequency and Gamma severity. For \( p > 1 \) one could alternatively use

\[
V(\mu) := V_0 + |\mu|^p
\]

with a \( V_0 > 0 \), which is continuously differentiable and similar to \( V(\mu) = |\mu|^p \) for \( |\mu|^p \gg V_0 \).

**Remark 7** We require that \( V \) is differentiable, since we use the derivative in the calculation of the MSEP below. However, for applications in practice, it is also acceptable to use

\[
V(\mu) = \max(|\mu|^p, V_0)
\]

with an \( V_0 > 0 \). A pragmatic way to define \( V_0 \) is to select a percentage \( c \) (for instance \( c = 1\% \)) and to set \( V_0 := c^2 \cdot \max\{|\tilde{\mu}_1|^p, \ldots, |\tilde{\mu}_n|^p\} \), where \( \tilde{\mu}_1, \ldots, \tilde{\mu}_n \) are the estimators from the classical loss ratio method. For a fixed accident year this means that the standard deviations of all development years are at least \( c \) times the standard deviation of the most volatile development year.

**Remark 8** If a variance function \( V \) with \( V(\mu) = |\mu|^p \) for \( |\mu| > \varepsilon \) is used, then we have

\[
\log(\sigma_k^2) = p \log(|\mu_k|) + \log(\sigma_0^2),
\]

i.e., the exponent \( p \) can be estimated using log-linear regression.
Consider the mapping
\[ \iota : \mathbb{R}^n \times (0, \infty) \to \mathbb{R}^n \times (0, \infty)^n, \]
\[ (\mu, \sigma_0) \mapsto (\mu, \sqrt{V(\mu_1)}\sigma_0, \ldots, \sqrt{V(\mu_n)}\sigma_0). \]

Instead of \( \Theta = (\mu_1, \ldots, \mu_n, \sigma_1, \ldots, \sigma_n) \) we now want to estimate the parameter vector \( \Psi = (\mu_1, \ldots, \mu_n, \sigma_0) \) using a variation of the EM algorithm. Since there is no risk of misinterpretation, we drop the \( \iota \) if we apply the functions from Sect. 3 to \( \iota(\Psi) \). For instance we write \( a_q(\Psi) \) instead of \( a_q(\iota(\Psi)) \). Given an estimate \( \mathcal{Q}(\nu) = (\mu_1^{(\nu)}, \ldots, \mu_n^{(\nu)}, \sigma_0^{(\nu)}) \) we then obtain the following expected log-likelihood in the E step:

\[
Q(\Psi | \mathcal{Q}(\nu)) = \sum_{q=1}^{p} \left( -\frac{a_q(\mathcal{Q}(\nu)) + b_q(\mathcal{Q}(\nu))^2}{2} \right) + \sum_{(i,k) \in O} \left( -\log(\sqrt{V(\mu_k)}\sigma_0) \right) - \frac{\nu^2 \mu_k^2 c_{i,k}(\mathcal{Q}(\nu)) + (S_{i,k} - d_{i,k}(\mathcal{Q}(\nu))v_i\mu_k)^2}{2v_i V(\mu_k)\sigma_0^2} + \text{const.}
\]

Since this function is more difficult to maximize than the corresponding function in Sect. 3, we replace it by the following function to perform the M step:

\[
\mathcal{Q}(\Psi | \mathcal{Q}(\nu)) = \sum_{q=1}^{p} \left( -\frac{a_q(\mathcal{Q}(\nu)) + b_q(\mathcal{Q}(\nu))^2}{2} \right) + \sum_{(i,k) \in O} \left( -\log(\sqrt{V(\mu_k)}\sigma_0) \right) - \frac{\nu^2 \mu_k^2 c_{i,k}(\mathcal{Q}(\nu)) + (S_{i,k} - d_{i,k}(\mathcal{Q}(\nu))v_i\mu_k)^2}{2v_i V(\mu_k)\sigma_0^2} + \text{const.}
\]

That is, we calculate \( \mathcal{Q}(\nu+1) \) such that it maximizes \( \mathcal{Q}(\cdot | \mathcal{Q}(\nu)) \). Of course, this does not mean that \( \mathcal{Q}(\nu+1) \) also maximizes \( Q(\cdot | \mathcal{Q}(\nu)) \). However, there is a good chance that at least

\[
Q(\mathcal{Q}(\nu+1) | \mathcal{Q}(\nu)) \geq Q(\mathcal{Q}(\nu) | \mathcal{Q}(\nu)).
\]

If this condition is violated for a \( \nu \), then we do not use the \( \mathcal{Q}(\nu+1) \) that maximizes \( \mathcal{Q}(\cdot | \mathcal{Q}(\nu)) \) but employ the gradient ascent approach described in Appendix B to calculate an alternative \( \mathcal{Q}(\nu+1) \) that fulfills

\[
Q(\mathcal{Q}(\nu+1) | \mathcal{Q}(\nu)) > Q(\mathcal{Q}(\nu) | \mathcal{Q}(\nu))
\]

(unless \( \mathcal{Q}(\nu) \) already maximizes \( Q(\cdot | \mathcal{Q}(\nu)) \) locally, which means that we can stop the iteration). The resulting iteration is a generalized EM algorithm, as defined in Dempster et al. [7]. In particular, this means that we obtain a monotonically increasing likelihood of the observable data.
In order to maximize $\widetilde{Q}(\cdot | \widetilde{\Psi}^{(v)})$ in the modified M step we have to solve the following equations for $\Psi$ to obtain $\widetilde{\Psi}^{(v+1)}$:

$$0 = \frac{\partial}{\partial \mu_k} \widetilde{Q}(\Psi | \widetilde{\Psi}^{(v)}) = \frac{\sum_{i=1}^{l(k)} d_{ik}(\widetilde{\Psi}^{(v)}) S_{i,k} - (c_{i,k}(\widetilde{\Psi}^{(v)}) + d_{ik}(\widetilde{\Psi}^{(v)})^2)v_i \mu_k}{V(\widetilde{\mu}^{(v)}) \sigma_0^2},$$

$$0 = \frac{\partial}{\partial \sigma_0} \widetilde{Q}(\Psi | \widetilde{\Psi}^{(v)}) = \frac{\sum_{(i,k) \in O} - \frac{1}{\sigma_0} + \frac{(S_{i,k} - d_{ik}(\widetilde{\Psi}^{(v)})v_i \mu_k)^2 + \frac{v_i^2 \mu_k^2 c_{i,k}(\widetilde{\Psi}^{(v)})}{v_i V(\widetilde{\mu}^{(v)}) \sigma_0^3}}{V(\widetilde{\mu}^{(v)}) \sigma_0^3}.}{\sum_{(i,k) \in O}}$$

The observed triangle can contain columns that should not be used in the estimation of $\sigma_0$. Examples are columns that contain only zeros or the last column in the case $p = n$. In order to handle such columns we allow to select a subset $\emptyset \neq \mathcal{U} \subset \{1, \ldots, n\}$ of columns that are used to estimate $\sigma_0$. We then define

$$\mathcal{O}^* := \{(i, k) \in \mathcal{O} | k \in \mathcal{U}\}$$

and maximize only the summands of $\widetilde{Q}$ with indices in $\mathcal{O}^*$. This leads to the following

**Iteration for parameter estimation with variance function $V$** We start with

$$\widetilde{\mu}^{(0)} = \frac{\sum_{i=1}^{l(k)} S_{i,k}}{\sum_{i=1}^{l(k)} v_i} \quad \text{and} \quad (\widetilde{\sigma}^{(0)})^2 = \frac{1}{|\mathcal{O}^*|} \sum_{(i,k) \in \mathcal{O}^*} \frac{v_i}{V(\widetilde{\mu}^{(0)})} \left( \frac{S_{i,k}}{v_i} - \widetilde{\mu}^{(0)} \right)^2.$$

We set $\widetilde{\Psi}^{(0)} := (\widetilde{\mu}^{(0)}, \ldots, \widetilde{\sigma}^{(0)})$. Given the estimator $\widetilde{\Psi}^{(v)} = (\widetilde{\mu}^{(v)}, \ldots, \widetilde{\mu}^{(v)}, \widetilde{\sigma}^{(v)})$ we calculate $\widetilde{\Psi}^{(v+1)} = (\widetilde{\mu}^{(v+1)}, \ldots, \widetilde{\mu}^{(v+1)}, \widetilde{\sigma}^{(v+1)})$ by

$$\widetilde{\mu}^{(v+1)} = \frac{\sum_{i=1}^{l(k)} d_{ik}(\widetilde{\Psi}^{(v)}) S_{i,k}}{\sum_{i=1}^{l(k)} c_{i,k}(\widetilde{\Psi}^{(v)}) + d_{ik}(\widetilde{\Psi}^{(v)})^2 v_i} \quad \text{and}$$

$$(\widetilde{\sigma}^{(v+1)})^2 = \frac{1}{|\mathcal{O}^*|} \sum_{(i,k) \in \mathcal{O}^*} \frac{v_i}{V(\widetilde{\mu}^{(v)})} \left[ \left( \frac{S_{i,k}}{v_i} - d_{ik}(\widetilde{\Psi}^{(v)}) \cdot \widetilde{\mu}^{(v+1)} \right)^2 + (\widetilde{\mu}^{(v+1)})^2 \cdot c_{i,k}(\widetilde{\Psi}^{(v)}) \right].$$

If the condition $Q(\widetilde{\Psi}^{(v+1)} | \widetilde{\Psi}^{(v)}) \geq Q(\widetilde{\Psi}^{(v)} | \widetilde{\Psi}^{(v)})$ is violated for a $v$, then we do not use the above formulas for this $v$, but calculate an alternative $\widetilde{\Psi}^{(v+1)}$ using the gradient ascent approach from Appendix B. If the algorithm converges, we define

$$\widetilde{\mu}^{(\infty)} := \lim_{v \to \infty} \widetilde{\mu}^{(v)}, \quad \widetilde{\sigma}^{(\infty)} := \lim_{v \to \infty} \widetilde{\sigma}^{(v)} \quad \text{and} \quad \widetilde{\Psi}^{(\infty)} := \lim_{v \to \infty} \widetilde{\Psi}^{(v)}.$$

For $k = 1, \ldots, n$ we set

$$\widetilde{\sigma}^{(\infty)} = \sqrt{V(\mu^{(\infty)}) \cdot \sigma_0^{(\infty)}}.$$

 Springer
Moreover, we define the posterior calendar year factors analogously to Sect. 3 by
\[
\tilde{F}_{i,k}^{(\infty)} := d_{i,k}(\tilde{\Psi}^{(\infty)}) = \mathbb{E}(F_{i,k} | S; \tilde{\Psi}^{(\infty)})
\]
for \((i, k) \in \mathcal{O}\).

**Robustified incremental loss ratio method with calendar year effects** For \((i, k) \in \mathcal{P}\)
\[
\tilde{S}_{i,k}^{(\infty)} := v_i \cdot f_{i,k} \cdot \tilde{\mu}_{k}^{(\infty)}
\]
are called the predictions of the robustified incremental loss ratio method with calendar year effects.

**Remark 9** Sample calculations indicate that for the application of the method in practice, it might be acceptable to implement the iteration for parameter estimation without using the gradient ascending approach from Appendix B.

Alternatively, it is of course also possible to use the gradient ascent approach in every step, even if \(Q(\tilde{\Psi}^{(v+1)}|\tilde{\Psi}^{(v)}) \geq Q(\tilde{\Psi}^{(v)}|\tilde{\Psi}^{(v)})\) is fulfilled for the \(\tilde{\Psi}^{(v+1)}\) that maximizes \(\tilde{Q}(\cdot | \tilde{\Psi}^{(v)})\).

**Remark 10** In order to obtain a moment adjusted variant of the robustified method analogously to Remark 5, we can replace the formula for \(\tilde{\sigma}_{0}^{(v+1)}\) in the iteration by
\[
\tilde{\sigma}_{0}^{(v+1)} := \frac{1}{|\mathcal{O}^*| - |\mathcal{U}|} \sum_{(i,k) \in \mathcal{O}^*} \frac{v_i}{V(\tilde{\mu}_{k}^{(v)})} \cdot \left[ \left( \frac{S_{i,k}}{v_i} - d_{i,k}(\tilde{\Psi}^{(v)}) \cdot \tilde{\mu}_{k}^{(v+1)} \right)^2 \right] + \frac{1}{|\mathcal{O}^*|} \sum_{(i,k) \in \mathcal{O}^*} \frac{v_i}{V(\tilde{\mu}_{k}^{(v)})} \cdot \left[ (\tilde{\mu}_{k}^{(v+1)})^2 \cdot c_{i,k}(\tilde{\Psi}^{(v)}) \right].
\]

Note that such an adjustment only makes sense for the simplified iteration without gradient ascent steps as described in Remark 9.

**5.2 Calculation of the mean squared error of prediction**

Using the results of Sect. 4 we can easily derive a formula for the MSEP of the robustified version of the method. Let \(J(\Psi)\) denote the Jacobian matrix of \(i\) at \(\Psi\), i.e.
We calculate $\tilde{I}^{(\infty)}$ using the definition of $\tilde{I}^{(\infty)}$ from Sect. 4, but replacing $\tilde{\Theta}^{(\infty)}$ by $\tilde{\Psi}^{(\infty)}$. Then $\tilde{I}^{(\infty)}$ is the observed information matrix at $\tilde{\Psi}^{(\infty)}$ (see Theorem 2). The observed information matrix $\tilde{J}^{(\infty)}$ at $\tilde{\Psi}^{(\infty)}$ can be calculated as

$$\tilde{J}^{(\infty)} = J(\tilde{\Psi}^{(\infty)}) \tilde{I}^{(\infty)} J(\tilde{\Psi}^{(\infty)}).$$

If $\tilde{J}^{(\infty)}$ is invertible, we use

$$\hat{\text{Cov}}(\hat{\mu}^{(\infty)}_i, \hat{\mu}^{(\infty)}_k) := \left( (\tilde{J}^{(\infty)})^{-1} \right)_{i,k}$$

as estimator for $\text{Cov}(\hat{\mu}^{(\infty)}_i, \hat{\mu}^{(\infty)}_k)$ for $i, k = 1, \ldots, n$.

**Estimator 2** If we replace $\tilde{\Theta}^{(\infty)} = (\hat{\mu}^{(\infty)}_1, \ldots, \hat{\mu}^{(\infty)}_n, \hat{\sigma}^{(\infty)}_1, \ldots, \hat{\sigma}^{(\infty)}_n)$ in Estimator 1 by

$$\tilde{\Psi}^{(\infty)} = (\hat{\mu}^{(\infty)}_1, \ldots, \hat{\mu}^{(\infty)}_n, \sqrt{V(\hat{\mu}^{(\infty)}_1)\hat{\sigma}^{(\infty)}_1}, \ldots, \sqrt{V(\hat{\mu}^{(\infty)}_n)\hat{\sigma}^{(\infty)}_n})$$

and $\hat{\text{Cov}}(\hat{\mu}^{(\infty)}_i, \hat{\mu}^{(\infty)}_k)$ by $\hat{\text{Cov}}(\hat{\mu}^{(\infty)}_i, \hat{\mu}^{(\infty)}_k)$, then we obtain a formula for the MSEP of the robustified incremental loss ratio method with calendar year effects.

### 6 Numerical example

We consider a run-off triangle containing incremental payments of $m = 6$ accident and $n = 6$ development years. The triangle contains simulated data. Nevertheless, the example should be quite realistic since the parameters used to simulate the data were estimated from the payment triangle for the accident years 2014–2019 of a large and relatively long-tailed European Motor Liability portfolio. The expected increments of the fifth calendar year were slightly reduced in the simulation to obtain a calendar year effect. Table 1 contains the volumes and the incremental payments.
We want to apply the robustified incremental loss ratio method with calendar year effects to this triangle. For this purpose, we select a variance function of the type $V(\mu) = \max(|\mu|^p, V_0)$. In order to get a feeling for a reasonable exponent $p$ we apply log-linear regression to the moment estimators $\hat{\mu}_1, \ldots, \hat{\mu}_5$ and $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_5^2$ of the classical loss ratio method, see Fig. 1.

Of course, such a regression does not always yield reasonable results, but in our case we obtain a slope of 1.84 which leads to a plausible variance structure (Tweedie). Therefore, we use $p := 1.84$ in this numerical example. Moreover, we select

$$V_0 := 0.01^2 \cdot \max\{|\hat{\mu}_1|^p, \ldots, |\hat{\mu}_5|^p\}.$$ 

This lower bound for the variance function is, however, not relevant in this case, since none of the estimated incremental loss ratios $\hat{\mu}_1, \ldots, \hat{\mu}_5$ is close to zero.

To illustrate what kind of prior knowledge can be used and how the method works, we define six scenarios. For each scenario we describe the prior information as follows:

| $i$ | $v_i$ | $S_{i,1}$ | $S_{i,2}$ | $S_{i,3}$ | $S_{i,4}$ | $S_{i,5}$ | $S_{i,6}$ |
|-----|-------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1   | 6,240,413 | 703,670   | 842,367   | 801,442   | 484,742   | 345,426   | 346,250   |
| 2   | 6,365,221 | 745,821   | 919,642   | 809,694   | 469,578   | 390,161   |
| 3   | 6,428,873 | 698,393   | 945,805   | 872,165   | 555,202   |
| 4   | 6,750,317 | 730,997   | 872,849   | 1,042,499 |
| 5   | 7,020,330 | 744,750   | 922,575   |
| 6   | 7,160,736 | 731,366   |

**Table 1** Volumes $v_i$ and incremental run-off triangle $(S_{i,k})_{i,k\in\mathbb{N}}$

![Fig. 1](image-url) Log-linear regression to estimate the exponent $p$ of the variance function $V(\mu) = \max(|\mu|^p, V_0)$
and apply the robustified incremental loss ratio method with calendar year effects. For scenario \( v \) we denote the predicted ultimates by \( \widetilde{U}_i^{[v]} := \sum_{k=1}^n \widetilde{S}_{i,k}^{(\infty)} \), where \( \widetilde{S}_{i,k}^{(\infty)} := S_{i,k} \) for \((i, k) \in \mathcal{O}\), and the reserves by \( \widetilde{R}_i^{[v]} := \sum_{k=K(i)+1}^n \widetilde{S}_{i,k}^{(\infty)} \). The predicted ultimate losses are found in Table 7, the reserves in Table 8. The estimated standard errors

\[
\text{s.e.}(\widetilde{R}_i^{[v]}) := \sqrt{\text{mse}_{\mathcal{R}_i^{[v]}}(\widetilde{R}_i^{[v]})}
\]

are shown in Table 9.

**Scenario 1:** We assume that there have been no calendar year effects in the observation period and that we do not expect any calendar year effects in the future. Thus, we use the prior information \( f_{i,k} := 1 \) and \( \tau_{i,k} := 0 \) for all \((i, k) \in \mathcal{A}\). Then the predictions are equal to the predictions of the classical loss ratio method.

**Scenario 2:** We assume that there have been no calendar year effects in the observation period, but we see the potential for calendar year effects in the future. Thus we use the prior information \( f_{i,k} := 1 \) for \((i, k) \in \mathcal{A}\), \( \tau_{i,k} := 0 \) for \((i, k) \in \mathcal{O}\) and \( \tau_{i,k} := 0.05 \) for \((i, k) \in \mathcal{P}\). Since only the \( f_{i,k} \) and \( \tau_{i,k} \) with \((i, k) \in \mathcal{O}\) influence the parameter estimation and since \( f_{i,k} = 1 \) for \((i, k) \in \mathcal{P}\) we get the same predictions as in Scenario 1. However, the estimated standard errors increase due to the potential calendar year effects in the future (see Table 9).

**Scenario 3:** We start with the prior information from Scenario 2 but we suspect that there was an exceptional effect in the 5th calendar year. Since we do not know whether the 5th calendar year was better or worse than average, we do not change \( f_{i,k} = 1 \) but use \( \tau_{i,k} := 0.05 \) for \((i, k) \in \mathcal{C}(5)\). Table 2 shows the estimated posterior factors \( \widetilde{F}_{i,k}^{[3]} := \mathbb{E}(F_{i,k} | S; \widetilde{\Psi}^{(\infty)}) \) for \((i, k) \in \mathcal{O}\) for these assumptions. In Table 7 we see that the predicted ultimates are larger than for the first two scenarios, since the observed loss triangle contains a diagonal with untypically low claims.

**Scenario 4:** We start with the prior information from Scenario 3 but we now want to take into account that the increments in the 5th calendar year were on average about 15% below normal for the entire market. For \((i, k) \in \mathcal{C}(5)\) we therefore replace the factors \( f_{i,k} \) from Scenario 3 by \( f_{i,k} := 0.85 \) (we do not change \( \tau_{i,k} = 0.05 \) for these cells). Table 3 shows the estimated posterior factors \( \widetilde{F}_{i,k}^{[4]} := \mathbb{E}(F_{i,k} | S; \widetilde{\Psi}^{(\infty)}) \) for \((i, k) \in \mathcal{O}\) for these assumptions. It is not surprising that the posterior factors for

| \( i \) | \( \widetilde{F}_{i,1}^{[3]} \) (%) | \( \widetilde{F}_{i,2}^{[3]} \) (%) | \( \widetilde{F}_{i,3}^{[3]} \) (%) | \( \widetilde{F}_{i,4}^{[3]} \) (%) | \( \widetilde{F}_{i,5}^{[3]} \) (%) | \( \widetilde{F}_{i,6}^{[3]} \) (%) |
|---|---|---|---|---|---|---|
| 1 | 100.0 | 100.0 | 100.0 | 100.0 | 95.7 | 100.0 |
| 2 | 100.0 | 100.0 | 100.0 | 95.7 | 100.0 | 100.0 |
| 3 | 100.0 | 100.0 | 95.7 | 100.0 | 100.0 |
| 4 | 100.0 | 95.7 | 100.0 | 100.0 |
| 5 | 95.7 | 100.0 | 100.0 |
| 6 | 100.0 |
calendar year 5 are lower than in Scenario 3. This leads to an increase of the predicted ultimates compared to Scenario 3 (see Table 7).

**Scenario 5:** For comparison with Scenario 3 we now use $f_{i,k} := 1$ and $\tau_{i,k} := 0.05$ for all $(i,k) \in \mathcal{A}$. This means that we allow for calendar year effects but we do not provide the prior information, that calendar year 5 is not typical. Table 4 shows the estimated posterior factors $\hat{F}_{i,k}^{[5]} := \mathbb{E}(F_{i,k} | S; \hat{\Psi}^{(\infty)})$ for $(i,k) \in \mathcal{O}$ for these assumptions. We see that the factors for the calendar year 5 are slightly larger than in Scenario 3. The method still detects that calendar year 5 has lower increments than the other calendar years, but in contrast to Scenario 3 the method infers that some of the other years are slightly above average. Compared to Scenario 3 this leads to lower predictions of the ultimate losses (see Table 7).

**Scenario 6:** So far, we have only used ‘simple’ calendar year effects. In order to show the flexibility of the approach, we finally consider a scenario with more complex prior information. For the calendar years 1–4 and 6 we use $f_{i,k} := 1$ and $\tau_{i,k} := 0.02$, i.e., we are not aware of any exceptional effects in these calendar years, but we allow for small effects if they can be detected from the data. Assume that, according to market statistics, the incremental losses of calendar year 5 were not typical in the respective segment. Increments in the first development year were 15% below average, increments in the second development year 10% and increments in the third development year 5% below average. Moreover, we think there could also be calendar year effects in the future. Eventually, we expect an exceptional inflation shock of 5% in calendar year 7 (which is not observable yet). This situation can be captured with the prior information from Table 5.

Table 6 shows the estimated posterior factors $\tilde{F}_{i,k}^{[6]} := \mathbb{E}(F_{i,k} | S; \tilde{\Psi}^{(\infty)})$ for $(i,k) \in \mathcal{O}$ for these assumptions.
Table 5: Prior Information for Scenario 6

| i  | $f_{i,1}$ (%) | $f_{i,2}$ (%) | $f_{i,3}$ (%) | $f_{i,4}$ (%) | $f_{i,5}$ (%) | $f_{i,6}$ (%) |
|----|--------------|--------------|--------------|--------------|--------------|--------------|
| 1  | 100          | 100          | 100          | 100          | 100          | 100          |
| 2  | 100          | 100          | 100          | 100          | 100          | 100          |
| 3  | 100          | 100          | 95           | 100          | 100          | 100          |
| 4  | 100          | 100          | 90           | 100          | 100          | 100          |
| 5  | 100          | 100          | 85           | 100          | 105          | 105          |
| 6  | 100          | 100          | 85           | 100          | 105          | 105          |
In Table 7 we see that Scenario 6 leads to the greatest predicted ultimates, which is mainly due to the assumed inflation shock of 5% in calendar year 7 in this scenario.

The six scenarios use prior information with different levels of complexity. Unfortunately, it is not easy to provide a criterion for accepting more complex models. However, residuals can be used to obtain a feeling whether including prior information results in an improvement. Figure 2 contains plots of the residuals

\[ S_{i,k} = v_i \tilde{F}_{i,k}^{(\infty)} - \mu_k \]

\[ \sqrt{v_i \cdot \tilde{\sigma}_{i,k}^{(\infty)}} \]

| Table 6 | Posterior factors
| | $F_{i,k}^{(6)} := \text{E}(F_{i,k} \mid \mathcal{S}; F^{(\infty)})$ for $(i, k) \in \mathcal{C}$ under the hypothesis of Scenario 6 |
| | $i$ | $F_{i,1}^{(6)}$ (%) | $F_{i,2}^{(6)}$ (%) | $F_{i,3}^{(6)}$ (%) | $F_{i,4}^{(6)}$ (%) | $F_{i,5}^{(6)}$ (%) | $F_{i,6}^{(6)}$ (%) |
| | 1 | 100.2 | 100.3 | 99.7 | 99.7 | 101.2 | 101.0 |
| | 2 | 100.3 | 99.7 | 99.7 | 101.2 | 101.0 |
| | 3 | 99.7 | 99.7 | 98.1 | 101.0 |
| | 4 | 99.7 | 94.9 | 101.0 |
| | 5 | 91.1 | 101.0 |
| | 6 | 101.0 |

| Table 7 | Predicted ultimate losses for the six scenarios |
| | $i$ | $U_{i}^{(1)}$ | $U_{i}^{(2)}$ | $U_{i}^{(3)}$ | $U_{i}^{(4)}$ | $U_{i}^{(5)}$ | $U_{i}^{(6)}$ |
| | 1 | 3,523,897 | 3,523,897 | 3,523,897 | 3,523,897 | 3,523,897 | 3,523,897 |
| | 2 | 3,688,071 | 3,688,071 | 3,686,960 | 3,686,964 | 3,679,559 | 3,700,641 |
| | 3 | 3,803,422 | 3,803,422 | 3,809,810 | 3,816,567 | 3,796,453 | 3,830,017 |
| | 4 | 3,950,125 | 3,950,125 | 3,964,782 | 3,978,160 | 3,946,066 | 4,001,470 |
| | 5 | 3,983,211 | 3,983,211 | 4,011,979 | 4,034,705 | 3,986,413 | 4,090,615 |
| | 6 | 4,076,540 | 4,076,540 | 4,114,075 | 4,144,315 | 4,080,401 | 4,243,248 |
| Total | 23,025,266 | 23,025,266 | 23,111,504 | 23,184,609 | 23,012,788 | 23,389,889 |

| Table 8 | Reserves for the six scenarios |
| | $i$ | $R_{i}^{(1)}$ | $R_{i}^{(2)}$ | $R_{i}^{(3)}$ | $R_{i}^{(4)}$ | $R_{i}^{(5)}$ | $R_{i}^{(6)}$ |
| | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 2 | 353,175 | 353,175 | 352,063 | 352,068 | 344,663 | 365,744 |
| | 3 | 731,856 | 731,856 | 738,244 | 745,002 | 724,887 | 758,451 |
| | 4 | 1,303,779 | 1,303,779 | 1,318,436 | 1,331,815 | 1,299,720 | 1,355,125 |
| | 5 | 2,315,886 | 2,315,886 | 2,344,655 | 2,367,380 | 2,319,088 | 2,423,290 |
| | 6 | 3,345,174 | 3,345,174 | 3,382,710 | 3,412,949 | 3,349,035 | 3,511,883 |
| Total | 8,049,870 | 8,049,870 | 8,136,108 | 8,209,214 | 8,037,393 | 8,414,493 |
Table 9: Estimated standard errors for the six scenarios

|    | $\hat{\sigma}_i(R_{i1}^{[1]})$ | $\hat{\sigma}_i(R_{i2}^{[2]})$ | $\hat{\sigma}_i(R_{i3}^{[3]})$ | $\hat{\sigma}_i(R_{i4}^{[4]})$ | $\hat{\sigma}_i(R_{i5}^{[5]})$ | $\hat{\sigma}_i(R_{i6}^{[6]})$ |
|----|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| 1  | 0                               | 0                               | 0                               | 0                               | 0                               | 0                               |
| 2  | 31,588                          | 36,189                          | 33,930                          | 33,877                          | 34,407                          | 37,045                          |
| 3  | 42,861                          | 50,070                          | 47,814                          | 48,096                          | 48,604                          | 51,223                          |
| 4  | 57,749                          | 69,210                          | 66,554                          | 67,193                          | 67,929                          | 70,940                          |
| 5  | 84,588                          | 105,037                         | 101,510                         | 102,773                         | 103,994                         | 108,457                         |
| 6  | 104,508                         | 131,802                         | 127,341                         | 128,995                         | 130,492                         | 136,545                         |
| Total | 207,888                         | 297,637                         | 290,712                         | 294,507                         | 293,114                         | 306,071                         |

Fig. 2: Residual plots for the six scenarios
versus the calendar year $i + k - 1$ for the six scenarios. For each calendar year these residuals should be randomly distributed around zero. The plots for Scenarios 1 and 2 are identical since we do not allow for calendar year effects in the observed data in both cases. In these two scenarios all the residuals of calendar year 5 are below zero, which means that the increments in this calendar year tend to be below average and therefore indicate a calendar year effect. The method seems to correct this calendar year effect quite well in Scenarios 3, 5 and 6. In Scenario 4, where we have used $f_{i,k} = 0.85$ for $(i,k) \in \mathcal{C}(5)$, the method seems to over-correct the calendar year effect.

7 Conclusion

We propose an extension of the classical additive claims reserving model that allows for prior information on calendar year effects to be taken into account. The model is able to process quite detailed prior information and may therefore look over-parameterized. The philosophy behind the approach is, however, to use as much market information on calendar years as is available. If no prior market knowledge is available, then the model reduces to the classical additive model.

The extension of the additive model is done by using scaling factors that are comonotonic within the same calendar year and independent if they belong to different calendar years. Since these scaling factors are not observable, we use the EM algorithm to estimate the parameters. For the considered model, the EM algorithm leads to a simple and explicit iteration that can be easily implemented in a pricing or a reserving tool. The mean squared error of prediction is calculated using the Cramér-Rao bound provided by the inverse Fisher information.

The predictions of the method do not only depend on the estimators for the expected incremental loss ratios but also on the estimated variance parameters. Therefore, it makes sense to reduce the number of parameters. To tackle this issue, we describe a more robust version of the model, making use of variance functions to reduce the number of parameters.

Of course, the proposed model will not solve all the problems that actuaries will face in pricing and reserving due to the traces left by the pandemic in run-off triangles. It is just a starting point that should be further developed, depending on what market data on Covid-19 effects actually become available in the years to come.

Appendix A: Proof of Theorem 2

According to Oakes [15], the observed information matrix at $\Theta$ can be calculated as

$$
I(\Theta) := \left\{ \frac{-\partial^2 Q(\Theta, \Theta')}{(\partial \Theta)^2} - \frac{\partial^2 Q(\Theta, \Theta')}{\partial \Theta \partial \Theta'} \right\}_{\Theta' = \Theta}.
$$

For $l \in K(q)$ we have
\[
\frac{\partial}{\partial \mu_l} a_q(\Theta) = -a_q(\Theta)^2 \cdot \frac{2v_{q-l+1} \mu_l \tau_{q-l+1,l}^2}{\sigma_l^2},
\]
\[
\frac{\partial}{\partial \sigma_l} a_q(\Theta) = a_q(\Theta)^2 \cdot \frac{2v_{q-l+1} \mu_l^2 \tau_{q-l+1,l}^2}{\sigma_l^3},
\]
\[
\frac{\partial}{\partial \mu_l} b_q(\Theta) = \left[ \frac{\partial}{\partial \mu_l} a_q(\Theta) \right] \sum_{(i,k) \in \mathcal{C}(q)} S_{i,k} \mu_k - v_i \mu_k^2 f_i,k \tau_{i,k} \sigma_k^2 + a_q(\Theta) \frac{S_{q-l+1,l} - 2v_{q-l+1} \mu_l f_{q-l+1,l}^2}{\sigma_l^2} \tau_{q-l+1,l},
\]
\[
\frac{\partial}{\partial \sigma_l} b_q(\Theta) = \left[ \frac{\partial}{\partial \sigma_l} a_q(\Theta) \right] \sum_{(i,k) \in \mathcal{C}(q)} S_{i,k} \mu_k - v_i \mu_k^2 f_i,k \tau_{i,k} \sigma_k^2 - 2a_q(\Theta) \cdot \frac{S_{q-l+1,l} - v_{q-l+1} \mu_l f_{q-l+1,l}^2}{\sigma_l^3} \tau_{q-l+1,l}.
\]

For \( l \notin \mathcal{K}(q) \) we have

\[
\frac{\partial}{\partial \mu_l} a_q(\Theta) = \frac{\partial}{\partial \sigma_l} a_q(\Theta) = \frac{\partial}{\partial \mu_l} a_q(\Theta) = \frac{\partial}{\partial \sigma_l} a_q(\Theta) = 0.
\]

Moreover, we have

\[
\frac{\partial}{\partial \mu_l} c_{i,k}(\Theta) = \tau_{i,k}^2 \frac{\partial}{\partial \mu_l} a_{i+k-1}(\Theta), \quad \frac{\partial}{\partial \sigma_l} c_{i,k}(\Theta) = \tau_{i,k}^2 \frac{\partial}{\partial \sigma_l} a_{i+k-1}(\Theta),
\]
\[
\frac{\partial}{\partial \mu_l} d_{i,k}(\Theta) = \tau_{i,k} \frac{\partial}{\partial \mu_l} b_{i+k-1}(\Theta), \quad \frac{\partial}{\partial \sigma_l} d_{i,k}(\Theta) = \tau_{i,k} \frac{\partial}{\partial \sigma_l} b_{i+k-1}(\Theta).
\]

Using the notation

\[
y_{i,k,l}^{(\mu)}(\Theta) := \frac{\partial}{\partial \mu_l} c_{i,k}(\Theta), \quad y_{i,k,l}^{(\sigma)}(\Theta) := \frac{\partial}{\partial \sigma_l} c_{i,k}(\Theta)
\]
\[
\delta_{i,k,l}^{(\mu)}(\Theta) := \frac{\partial}{\partial \mu_l} d_{i,k}(\Theta), \quad \delta_{i,k,l}^{(\sigma)}(\Theta) := \frac{\partial}{\partial \sigma_l} d_{i,k}(\Theta)
\]

we can calculate the second partial derivatives of

\[
Q(\Theta | \Theta') = \sum_{q=1}^{p} \left( -a_q(\Theta') + b_q(\Theta')^2 \right)
+ \sum_{(i,k) \in \mathcal{C}} \left( -\log(\sigma_k) - \frac{v_i^2 \mu_k^2 c_{i,k}(\Theta') + (S_{i,k} - d_{i,k}(\Theta') v_i \mu_k)^2}{2v_i \sigma_k^2} \right) + \text{const.}
\]

We have

\( \odot \) Springer
\[
\frac{\partial^2}{\partial \mu_k^2} Q(\Theta|\Theta') = \frac{\partial}{\partial \mu_k} \sum_{i=1}^{L(k)} [d_{ik}(\hat{\Theta}^{(v)}) S_{i,k} - (c_{ik}(\hat{\Theta}^{(v)}) + d_{ik}(\hat{\Theta}^{(v)})^2) v_i \mu_k] \frac{1}{\sigma_k^2} \\
= - \sum_{i=1}^{L(k)} \frac{(c_{ik}(\hat{\Theta}^{(v)}) + d_{ik}(\hat{\Theta}^{(v)})^2) v_i}{\sigma_k^2},
\]

\[
\frac{\partial^2}{\partial \sigma_k^2} Q(\Theta|\Theta') = \frac{\partial}{\partial \sigma_k} \sum_{i=1}^{L(k)} \left[ \frac{1}{\sigma_k} + \frac{(S_{i,k} - d_{ik}(\hat{\Theta}^{(v)}) v_i \mu_k)^2 + v_i^2 \mu_k^2 c_{ik}(\hat{\Theta}^{(v)})}{\sigma_k^3} \right] v_i, \\
= \sum_{i=1}^{L(k)} \frac{1}{\sigma_k^2} \left[ \frac{(S_{i,k} - d_{ik}(\hat{\Theta}^{(v)}) v_i \mu_k)^2 + v_i^2 \mu_k^2 c_{ik}(\hat{\Theta}^{(v)})}{\sigma_k^3} \right] v_i.
\]

\[
\frac{\partial^2}{\partial \mu_k \partial \sigma_i} Q(\Theta|\Theta') = \frac{\partial}{\partial \sigma_i} \sum_{i=1}^{L(k)} \left[ \frac{1}{\sigma_i} + \frac{(S_{i,k} - d_{ik}(\hat{\Theta}^{(v)}) v_i \mu_k)^2 + v_i^2 \mu_k^2 c_{ik}(\hat{\Theta}^{(v)})}{\sigma_i^3} \right] v_i, \\
= \sum_{i=1}^{L(k)} \frac{1}{\sigma_i^2} \left[ \frac{(S_{i,k} - d_{ik}(\hat{\Theta}^{(v)}) v_i \mu_k)^2 + v_i^2 \mu_k^2 c_{ik}(\hat{\Theta}^{(v)})}{\sigma_i^3} \right] v_i.
\]

\[
\frac{\partial^2}{\partial \mu_k \partial \sigma_i} Q(\Theta|\Theta') = -2 \sum_{i=1}^{L(k)} \left[ \frac{d_{ik}(\hat{\Theta}^{(v)}) S_{i,k} - (c_{ik}(\hat{\Theta}^{(v)}) + d_{ik}(\hat{\Theta}^{(v)})^2) v_i \mu_k}{\sigma_k^2} \right] v_i.
\]

Moreover, we have

\[
\frac{\partial^2}{\partial \mu_k \partial \mu_l} Q(\Theta|\Theta') = \frac{\partial}{\partial \mu_l} \sum_{i=1}^{L(k)} \left[ \frac{d_{ik}(\hat{\Theta}^{(v)}) S_{i,k} - (c_{ik}(\hat{\Theta}^{(v)}) + d_{ik}(\hat{\Theta}^{(v)})^2) v_i \mu_k}{\sigma_k^2} \right] v_i, \\
= \sum_{i=1}^{L(k)} \frac{d_{ik}(\hat{\Theta}^{(v)})^2 S_{i,k} - (c_{ik}(\hat{\Theta}^{(v)}) + d_{ik}(\hat{\Theta}^{(v)})^2)^2 v_i \mu_k}{\sigma_k^4}.
\]

and for \( k \neq l \)

\[
\frac{\partial^2}{\partial \mu_k \partial \sigma_l} Q(\Theta|\Theta') = \frac{\partial}{\partial \sigma_l} \sum_{i=1}^{L(k)} \left[ \frac{(S_{i,k} - d_{ik}(\hat{\Theta}^{(v)}) v_i \mu_k)^2 + v_i^2 \mu_k^2 c_{ik}(\hat{\Theta}^{(v)})}{\sigma_k^3} \right] v_i, \\
= \sum_{i=1}^{L(k)} \frac{(S_{i,k} - d_{ik}(\hat{\Theta}^{(v)}) v_i \mu_k)^2 + v_i^2 \mu_k^2 c_{ik}(\hat{\Theta}^{(v)})}{\sigma_k^3} v_i.
\]
Replacing $\Theta$ and $\Theta'$ in the formulas above by $\hat{\Theta}^{(\infty)}$ and taking into account that

$$\gamma^{(\mu)}_{i,k,l} = \gamma^{(\mu)}_{i,k,l}(\hat{\Theta}^{(\infty)}), \quad \gamma^{(\sigma)}_{i,k,l} = \gamma^{(\sigma)}_{i,k,l}(\hat{\Theta}^{(\infty)}), \quad \delta^{(\mu)}_{i,k,l} = \delta^{(\mu)}_{i,k,l}(\hat{\Theta}^{(\infty)}) \quad \text{and} \quad \delta^{(\sigma)}_{i,k,l} = \delta^{(\sigma)}_{i,k,l}(\hat{\Theta}^{(\infty)})$$

we see that

$$\tilde{\gamma}^{(\infty)} = \left\{ -\frac{\partial^2 Q(\Theta, \Theta')}{(\partial \Theta)^2} - \frac{\partial^2 Q(\Theta, \Theta')}{\partial \Theta \partial \Theta'} \right\}_{\Theta' = \hat{\Theta}^{(\infty)}, \Theta = \hat{\Theta}^{(\infty)}} = I(\hat{\Theta}^{(\infty)})$$

and we obtain Theorem 2.

**Appendix B: Gradient ascent approach for calculating $\tilde{\Psi}^{(\nu+1)}$**

The gradient

$$g(\Psi) := \text{grad}_{Q(\cdot | \tilde{\Psi}^{(\nu)})}(\Psi) = \begin{pmatrix} \frac{\partial}{\partial \mu_1} Q(\Psi | \tilde{\Psi}^{(\nu)}) \\ \vdots \\ \frac{\partial}{\partial \mu_n} Q(\Psi | \tilde{\Psi}^{(\nu)}) \\ \frac{\partial}{\partial \sigma_0} Q(\Psi | \tilde{\Psi}^{(\nu)}) \end{pmatrix}$$

of $Q(\cdot | \tilde{\Psi}^{(\nu)})$ can easily be calculated. We have

$$\frac{\partial}{\partial \mu_k} Q(\Psi | \tilde{\Psi}^{(\nu)}) = -\frac{I(k)}{2} \frac{V'(\mu_k)}{V(\mu_k)} + \sum_{i=1}^{I(k)} \frac{d_{i,k}(\tilde{\Psi}^{(\nu)}) S_{i,k} - (c_{i,k}(\tilde{\Psi}^{(\nu)}) + d_{i,k}(\tilde{\Psi}^{(\nu)})^2) v_i \mu_k}{V(\mu_k) \sigma_0^2}$$

$$+ \sum_{i=1}^{I(k)} \left[ \frac{v^2_i \mu_k^2 c_{i,k}(\tilde{\Psi}^{(\nu)}) + (S_{i,k} - d_{i,k}(\tilde{\Psi}^{(\nu)}) v_i \mu_k)^2}{2 v_i V(\mu_k) \sigma_0^2} \right] \cdot V'(\mu_k)$$

and

$$\frac{\partial}{\partial \sigma_0} Q(\Psi | \tilde{\Psi}^{(\nu)}) = -\frac{|\Omega|}{\sigma_0} + \sum_{(i,k) \in \Omega} \frac{v^2_i \mu_k^2 c_{i,k}(\tilde{\Psi}^{(\nu)}) + (S_{i,k} - d_{i,k}(\tilde{\Psi}^{(\nu)}) v_i \mu_k)^2}{v_i V(\mu_k) \sigma_0^2}.$$
where $\min\{c, +\infty\} := c$ for real numbers $c$. Standard methods (like the golden section search, cf. Press et al. [16]) can be used to find a (local) minimum $\psi^{(v+1)}$ of the function

$$q^{(v)} : [0, T^{(v)}] \to \mathbb{R}, \quad \psi \mapsto Q(\tilde{\psi}^{(v)} + \psi \cdot g(\tilde{\psi}^{(v)}))$$

such that $q^{(v)}(\psi^{(v+1)}) > q^{(v)}(0)$. Then $\tilde{\psi}^{(v+1)} := \tilde{\psi}^{(v)} + \psi^{(v+1)} \cdot g(\tilde{\psi}^{(v)})$ satisfies the condition

$$Q(\tilde{\psi}^{(v+1)}|\tilde{\psi}^{(v)}) > Q(\tilde{\psi}^{(v)}|\tilde{\psi}^{(v)}).$$

If not all columns are used for the estimation of $(\tilde{\sigma}^{(v+1)}_{\nu} )^2$ in the iteration of Sect. 5.1 (i.e., if $O^* \neq O$), then we simply replace $\frac{\partial}{\partial \theta_0} Q(\Psi|\Psi^{(v)})$ in the definition of $g$ by

$$-\frac{|O^*|}{\sigma_0} + \sum_{(i,j) \in O^*} \frac{v_i^2 \mu_k^2 c_{i,k}(\tilde{\psi}^{(v)}) + (S_{i,k} - d_{i,k}(\tilde{\psi}^{(v)})v_i \mu_k)^2}{v_i V(\mu_k)\sigma_0^3}.$$

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s13385-022-00315-3.

**Acknowledgements** I am grateful to Dave Clark for his valuable comments on an earlier version of the manuscript. Moreover, I would also like to thank the anonymous reviewers for their constructive suggestions that helped improve the paper.

**References**

1. ASTIN Working Party on Non-Life Reserving (2016) Non-life reserving practices report. http://www.actuaries.org/ASTIN/Documents/ASTIN_WP_NL_Reserving_Report1.0_2016-06-15.pdf. Accessed 11 Apr 2022
2. Björkwall S, Hössjer O, Ohlsson E (2010) Bootstrapping the separation method in claims reserving. ASTIN Bull 40(2):845–869
3. Barnett G, Zenwirth B (2000) Best estimates for reserves. Proceedings CAS 87:245–321
4. Brehm PJ (2002) Correlation and the aggregation of unpaid loss distributions. CAS Forum 2:1–23
5. Bühlmann H, Moriconi F (2015) Credibility claims reserving with stochastic diagonal effects. ASTIN Bull 45(2):309–353
6. Clark DR (2006) Variance and covariance due to inflation. CAS Forum (Fall) 61:61–95
7. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). J R Stat Soc Ser B 39:1–38
8. de Jong P (2006) Forecasting runoff triangles. N Am Actuar J 10(2):28–38
9. Gigante P, Picech L, Sigalotti L (2019) Calendar year effect modeling for claims reserving in HGLM. ASTIN Bull 49(3):763–786
10. Hertig J (1985) A statistical approach to the IBNR-reserves in marine reinsurance. ASTIN Bull 15(2):171–183
11. Kirschner GS, Kerley C, Isaacs B (2008) Two approaches to calculating correlated reserves indications across multiple lines of business. Variance 2(1):15–38
12. Kuang D, Nielsen B, Nielsen JP (2008) Forecasting with the age-period-cohort model and the extended chain-ladder model. Biometrika 95:987–991
13. Mack Th (1993) Distribution-free calculation of the standard error of chain ladder reserve estimates. ASTIN Bull 23(2):213–225
14. Mack Th (2002) Schadenversicherungsmathematik. VVW, Karlsruhe
15. Oakes D (1999) Direct calculation of the information matrix via the EM algorithm. J R Stat Soc Ser B 61(2):479–482
16. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical recipes: the art of scientific computing. 3rd edn. Cambridge University Press, New York
17. Schmidt KD, Zocher M (2016) Additive method. In: Radtke M, Schmidt KD, Schnaus A (eds) Handbook on loss reserving. Springer International Publishing, Cham
18. Shi P, Basu S, Meyers GG (2012) A Bayesian log-normal model for multivariate loss reserving. N Am Actuar J 16(1):29–51
19. Taylor GC (1977) Separation of inflation and other effects from the distribution of non-life insurance claim delays. ASTIN Bull 9:217–230
20. Vaida F (2005) Parameter convergence for EM and MM algorithms. Stat Sin 15:831–840
21. Verbeek HG (1972) An approach to the analysis of claims experience in motor liability excess-of-loss reinsurance. ASTIN Bull 6:195–203
22. Wu CFJ (1983) On the convergence properties of the EM algorithm. Ann Stat 11(1):95–103
23. Wüthrich MV (2010) Accounting year effects modeling in the stochastic chain ladder reserving method. N Am Actuar J 14(2):235–255
24. Wüthrich MV, Happ S (2013) Paid-incurred chain reserving method with dependence modeling. ASTIN Bull 43(1):1–20

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.