Full Bayesian analysis of claims reserving uncertainty

Gareth W. Peters*   Rodrigo S. Targino*   Mario V. Wüthrich†‡

May 20, 2016

Abstract

We revisit the gamma-gamma Bayesian chain-ladder (BCL) model for claims reserving in non-life insurance. This claims reserving model is usually used in an empirical Bayesian way using plug-in estimates for variance parameters, because this empirical Bayesian framework allows us for closed form solutions. The main purpose of this paper is to develop the full Bayesian case also considering prior distributions for variance parameters, and to study the resulting sensitivities.

Keywords. Chain-ladder method; claims reserving uncertainty; claims development result; Mack's formula; Merz-Wüthrich's formula; conditional mean square error of prediction; run-off uncertainty; full Bayesian chain-ladder model.

1 Introduction

The chain-ladder (CL) algorithm is probably to most popular method to set the reserves for non-life insurance claims. Originally, the CL method was introduced in a purely algorithmic fashion and it was not based on a stochastic model. Stochastic models underpinning the CL algorithm with a statistical model were only developed much later. The two most commonly used stochastic representations are Mack's [5] distribution-free CL model and the over-dispersed Poisson (ODP) model of Renshaw and Verrall [6] and England and Verrall [1]. In this paper we study the gamma-gamma Bayesian chain-ladder (BCL) model which provides in its non-informative prior limit another stochastic representation for the CL method. This was first considered in the claims reserving context by Gisler [3] and Gisler and Wüthrich [4]. The typical application of the gamma-gamma BCL model was done under fixed (given) variance parameters where plug-in estimates were used, see Example 2.13 in Wüthrich and Merz [8] for such an empirical Bayesian analysis. Of course, this (partially) contradicts the Bayesian paradigm. In a full Bayesian approach one should also model these variance parameters with prior distributions. The aim of this paper is to study the influence of such a *full* Bayesian modeling approach and compare it to the *empirical* Bayesian modeling approach used in [8]. In particular, we aim at analyzing

---

*University College London, Department of Statistical Science, UK

†ETH Zurich, RiskLab, Department of Mathematics, Switzerland

‡Swiss Finance Institute SFI Professor

the sensitivity of the prediction uncertainty in the choice of these variance parameters. This is crucial in solvency consideration and improves the crude estimates usually used in practice.

Organization.
In the next section we introduce the gamma-gamma BCL model and show that we recover the classical CL reserves in its non-informative prior limit. In Section 3 we provide the prediction uncertainty formulas in the long-term view and the short-term view, respectively. In Section 4 we give several real data examples and analyze the resulting sensitivities in the choice of the prior distributions. All proofs are provided in Appendix A.

2 Gamma-gamma Bayesian chain-ladder model

We introduce the gamma-gamma BCL model in this section. In contrast to Model Assumptions 2.6 in [8] we also model the variance parameters in a Bayesian way. We then derive the claims predictors in the non-informative prior limit which turn out to be identical to the classical CL predictors, as seen in Theorem 2.2 below.

We denote accident years by $1 \le i \le I$ and development years by $0 \le j \le J$. Throughout, we assume $I > J$. The cumulative claim of accident year i after development year j is denoted by $C_{i,j}$, and $C_{i,J}$ denotes the ultimate claim of accident year i. For more background information on the claims reserving problem and the corresponding notation we refer to Chapter 1 in [8]. We make the following model assumptions.

Model Assumptions 2.1 (gamma-gamma BCL model)**.**

(a) *Conditionally given parameter vectors $\boldsymbol{\Theta} = (\Theta_0, \dots, \Theta_{J-1})$ and $\boldsymbol{\sigma} = (\sigma_0, \dots, \sigma_{J-1})$, the cumulative claims $(C_{i,j})_{0 \le j \le J}$ are independent (in accident year i), and Markov processes (in development year j) with conditional distributions*

$$C_{i,j+1} \mid_{\{C_{i,j}, \boldsymbol{\Theta}, \boldsymbol{\sigma}\}} \sim \Gamma\left(C_{i,j}\sigma_j^{-2}, \Theta_j \sigma_j^{-2} \right),$$

for all $1 \le i \le I$ and $0 \le j \le J - 1$.

(b) *The parameter vectors $\boldsymbol{\Theta}$ and $\boldsymbol{\sigma}$ are independent.*

(c) *The components Θ_j of $\boldsymbol{\Theta}$ are independent and $\Gamma(\gamma_j, f_j(\gamma_j - 1))$-distributed with prior parameters $f_j > 0$ and $\gamma_j > 1$ for $0 \le j \le J - 1$.*

(d) *The components σ_j of $\boldsymbol{\sigma}$ are independent and π_j-distributed having support in $(0, d_j)$ for given constants $0 < d_j < \infty$ for all $0 \le j \le J - 1$.*

(e) *$(\boldsymbol{\Theta}, \boldsymbol{\sigma})$ and $C_{1,0}, \dots, C_{I,0}$ are independent and $C_{i,0} > 0$, $\mathbb{P}$-a.s., for all $1 \le i \le I$.*

These model assumptions imply that we have the following CL properties

$$\mathbb{E}\left[C_{i,j+1} \mid C_{i,0}, \dots, C_{i,j}, \boldsymbol{\Theta}, \boldsymbol{\sigma} \right] = \Theta_j^{-1} C_{i,j}, \tag{2.1}$$

$$\mathrm{Var}\left(C_{i,j+1} \mid C_{i,0}, \dots, C_{i,j}, \boldsymbol{\Theta}, \boldsymbol{\sigma} \right) = \Theta_j^{-2} \sigma_j^2 C_{i,j}. \tag{2.2}$$

2

Thus, for given parameter vectors $\boldsymbol{\Theta}$ and $\boldsymbol{\sigma}$ we obtain a distributional example of Mack's [5] distribution-free CL model with CL factors Θ_j^{-1} and variance parameters $\Theta_j^{-2}\sigma_j^2$. Moreover, for $\pi_j(\cdot)$ being single point masses for all $0 \le j \le J-1$ we exactly obtain Model Assumptions 2.6 of [8] assuming given (known) variance parameters σ_j^2.

The main task in claims reserving is to predict the ultimate claims $C_{i,J}$, given observations

$$\mathcal{D}_t = \{C_{i,j}:\ i+j \le t,\ 1 \le i \le I,\ 0 \le j \le J\}, \qquad \text{at time } t \ge I.$$

In complete analogy to the derivations in Section 2.2.1 of [8], the application of Bayes' rule provides posterior π for the parameters $(\boldsymbol{\Theta}, \boldsymbol{\sigma})$, conditionally given observations $\mathcal{D}_t$, for $t \ge I$,

$$\pi\left(\boldsymbol{\theta}, \boldsymbol{\sigma}\,|\,\mathcal{D}_t\right)\ \propto\ \prod_{j=0}^{J-1}\left[\theta_j^{\gamma_j + \sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sigma_j^2} - 1} \exp\left\{-\theta_j\left[f_j(\gamma_j - 1) + \sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j+1}}{\sigma_j^2}\right]\right\}\right.$$
$$\left. \times\ \prod_{i=1}^{(t-j-1)\wedge I} \frac{\left(\frac{C_{i,j+1}}{\sigma_j^2}\right)^{\frac{C_{i,j}}{\sigma_j^2}}}{\Gamma\left(\frac{C_{i,j}}{\sigma_j^2}\right)}\ \pi_j(\sigma_j)\right]. \qquad (2.3)$$

From this we see that the posteriors of (Θ_j, σ_j) are independent for different development periods $0 \le j \le J-1$. A non-informative prior limit for Θ_j^{-1} corresponds to letting $\gamma_j \to 1$ (and the terms $f_j(\gamma_j - 1)$ will vanish in (2.3)). Therefore, we refer as the non-informative prior limit when the (component-wise) limits $\boldsymbol{\gamma} \to \mathbf{1}$ are taken, where we set $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_{J-1})$ and $\mathbf{1} = (1, \ldots, 1)$. We have the following theorem for the claims prediction under this non-informative prior limit (the proof is given in the appendix).

Theorem 2.2 (CL predictor)**.** *Under Model Assumptions 2.1 and for $t \ge I \ge i > t - n \ge t - J$, the Bayesian predictor for $C_{i,n}$ in its non-informative prior limit is given by*

$$\lim_{\boldsymbol{\gamma} \to \mathbf{1}} \mathbb{E}\left[C_{i,n}\,|\,\mathcal{D}_t\right] = C_{i,t-i} \prod_{j=t-i}^{n-1} \widehat{f}_j^{CL(t)} \stackrel{\text{def.}}{=} \widehat{C}_{i,n}^{CL(t)},$$

with CL factors $\widehat{f}_j^{CL(t)}$ defined by

$$\widehat{f}_j^{CL(t)} = \frac{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j+1}}{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j}}.$$

Theorem 2.2 states that in the non-informative prior limit $\boldsymbol{\gamma} \to \mathbf{1}$ we exactly obtain the classical CL predictor, see Mack [5]. Thus, we have found another stochastic representation that underpins the CL algorithm with a statistical model. Therefore, we (may) use this model in its non-informative prior limit to analyze the prediction uncertainty of the CL algorithm. In contrast to Remark 2.11 of [8] we now obtain this result in the full Bayesian framework, also considering prior distributions for the standard deviation parameters $\boldsymbol{\sigma}$.

3 Prediction uncertainty formulas

3.1 Long-term prediction uncertainty formula

The main purpose of this paper is to analyze the influence of the standard deviation parameters $\boldsymbol{\sigma}$ on the ultimate claim prediction uncertainty, where in contrast to Chapter 2 of [8], these standard deviation parameters $\boldsymbol{\sigma}$ are also modeled with prior distributions. We analyze the prediction uncertainty at time $t \geq I$ in terms of the conditional mean square error of prediction (MSEP) given by

$$
\begin{aligned}
\mathrm{msep}_{C_{i,J}|\mathcal{D}_t}\left(\mathbb{E}\left[C_{i,J}|\mathcal{D}_t\right]\right) &= \mathbb{E}\left[\left.(C_{i,J} - \mathbb{E}\left[C_{i,J}|\mathcal{D}_t\right])^2\right|\mathcal{D}_t\right] = \mathrm{Var}\left(C_{i,J}|\mathcal{D}_t\right) \\
&= \mathrm{Var}\left(\mathbb{E}\left[C_{i,J}|\mathcal{D}_t,\boldsymbol{\sigma}\right]|\mathcal{D}_t\right) + \mathbb{E}\left[\mathrm{Var}\left(C_{i,J}|\mathcal{D}_t,\boldsymbol{\sigma}\right)|\mathcal{D}_t\right].
\end{aligned} \quad (3.1)
$$

We aim at calculating this conditional MSEP in the gamma-gamma BCL model which provides, in its non-informative prior limit $\gamma \to 1$, an uncertainty estimate for the CL algorithm, that is, we aim at calculating the limit

$$
\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{D}_t}\left(\widehat{C}_{i,J}^{CL(t)}\right) \stackrel{\mathrm{def.}}{=} \lim_{\gamma \to 1} \mathrm{msep}_{C_{i,J}|\mathcal{D}_t}\left(\mathbb{E}\left[C_{i,J}|\mathcal{D}_t\right]\right).
$$

For $0 \leq j \leq J-1$ and $t \geq I$, we define on the intervals $(0, d_j)$ the functions

$$
h_j(\sigma_j|\mathcal{D}_t) = \frac{\Gamma\left(1 + \sum_{i=1}^{(t-j-1)\wedge I}\frac{C_{i,j}}{\sigma_j^2}\right)}{\left(\sum_{i=1}^{(t-j-1)\wedge I}\frac{C_{i,j+1}}{\sigma_j^2}\right)^{1+\sum_{i=1}^{(t-j-1)\wedge I}\frac{C_{i,j}}{\sigma_j^2}}} \prod_{i=1}^{(t-j-1)\wedge I} \frac{\left(\frac{C_{i,j+1}}{\sigma_j^2}\right)^{\frac{C_{i,j}}{\sigma_j^2}}}{\Gamma\left(\frac{C_{i,j}}{\sigma_j^2}\right)} \pi_j(\sigma_j). \quad (3.2)
$$

This function (3.2) is the un-normalized marginal posterior density of σ_j, given $\mathcal{D}_t$, in the non-informative prior limit (this we are going to see in detail below). The following lemma proves integrability of $h_j(\sigma_j|\mathcal{D}_t)$ on interval $(0, d_j)$ (the proof is provided in the appendix).

Lemma 3.1. *Choose $0 \leq j \leq J-1$ and $t \geq I$. Assume that either $(t-j-1)\wedge I = 1$ or that for at least one accident year $1 \leq i \leq (t-j-1)\wedge I$ we have*

$$
\frac{C_{i,j+1}}{C_{i,j}} \neq \widehat{f}_j^{CL(t)}. \quad (3.3)
$$

Then, we obtain integrability

$$
k_j(\mathcal{D}_t) = \int_0^{d_j} h_j(\sigma_j|\mathcal{D}_t)d\sigma_j < \infty.
$$

To obtain a finite conditional MSEP at time $t \geq I$ in the non-informative prior limit we require, for the data $\mathcal{D}_t$, that

$$
\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j} > d_j^2, \qquad \text{holds for all } t-I \leq j \leq J-1, \quad (3.4)
$$

see also proof of Theorem 3.2, below. We then define the variables

$$
\Psi_j^{(t)} = \frac{\sigma_j^2}{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j} - \sigma_j^2}.
$$

4

Under (3.4) we obtain, $\mathbb{P}$-a.s.,

$$0 \; < \; \Psi_j^{(t)} \; < \; \frac{d_j^2}{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j} - d_j^2}.$$

This implies that under (3.3) and (3.4) the following two expectations are well-defined

$$\widehat{\mathbb{E}}\left[\sigma_j^2\left(1+\Psi_j^{(t)}\right)\Big|\mathcal{D}_t\right] \;=\; k_j(\mathcal{D}_t)^{-1}\int_0^{d_j}\sigma_j^2\left(1+\Psi_j^{(t)}\right)h_j(\sigma_j|\mathcal{D}_t)\,d\sigma_j, \tag{3.5}$$

$$\widehat{\mathbb{E}}\left[\Psi_j^{(t)}\Big|\mathcal{D}_t\right] \;=\; k_j(\mathcal{D}_t)^{-1}\int_0^{d_j}\Psi_j^{(t)}\,h_j(\sigma_j|\mathcal{D}_t)\,d\sigma_j, \tag{3.6}$$

where $\widehat{\mathbb{P}}$ denotes the probability measure given by density $\widehat{\pi}_j(\sigma_j) = k_j(\mathcal{D}_t)^{-1}h_j(\sigma_j|\mathcal{D}_t)$ on $(0, d_j)$. We have the following estimate of the conditional MSEP for the CL algorithm (the proof is provided in the appendix).

Theorem 3.2. *Under Model Assumptions 2.1, (3.3) and (3.4), we have for $t \geq I \geq i > t - J$*

$$\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{D}_t}\left(\widehat{C}_{i,J}^{CL(t)}\right) = \left(\widehat{C}_{i,J}^{CL(t)}\right)^2 \sum_{j=t-i}^{J-1}\frac{1}{\widehat{C}_{i,j}^{CL(t)}}\,\widehat{\mathbb{E}}\left[\sigma_j^2\left(1+\Psi_j^{(t)}\right)\Big|\mathcal{D}_t\right]\prod_{n=j+1}^{J-1}\left(1+\widehat{\mathbb{E}}\left[\Psi_n^{(t)}\Big|\mathcal{D}_t\right]\right)$$

$$+ \left(\widehat{C}_{i,J}^{CL(t)}\right)^2\left(\prod_{j=t-i}^{J-1}\left(1+\widehat{\mathbb{E}}\left[\Psi_j^{(t)}\Big|\mathcal{D}_t\right]\right) - 1\right);$$

and the non-informative prior limit of the aggregated conditional MSEP is given by

$$\widehat{\mathrm{msep}}_{\sum_i C_{i,J}|\mathcal{D}_t}\left(\sum_i \widehat{C}_{i,J}^{CL(t)}\right) \;=\; \sum_i \widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{D}_t}\left(\widehat{C}_{i,J}^{CL(t)}\right)$$

$$+ 2\sum_{i<m}\widehat{C}_{i,J}^{CL(t)}\widehat{C}_{m,J}^{CL(t)}\left(\prod_{j=t-i}^{J-1}\left(1+\widehat{\mathbb{E}}\left[\Psi_j^{(t)}\Big|\mathcal{D}_t\right]\right) - 1\right),$$

where the summations run over $t - J + 1 \leq i \leq I$ and $t - J + 1 \leq i < m \leq I$, respectively.

To evaluate this prediction uncertainty formulas we still need to calculate the expected values given in (3.5) and (3.6). This can be done efficiently with importance sampling or using numerical integration (note that the corresponding density (3.2) is known up to the normalizing constant $k_j(\mathcal{D}_t)$), since one only needs to compute one-dimensional integrals.

3.2 Short-term prediction uncertainty formula

Recent solvency developments [2, 7] require a second, short-term uncertainty view. This second view studies the one-year update when new information becomes available (which has a direct influence on the actual earning statement of an insurance company). We therefore define the so-called claims development result. Choose $t \geq I \geq i > t - J$ and define the claims development result of accident year i at time $t + 1$ by

$$\mathrm{CDR}_i^{(t+1)} = \mathbb{E}\left[C_{i,J}|\mathcal{D}_t\right] - \mathbb{E}\left[C_{i,J}|\mathcal{D}_{t+1}\right].$$

Due to the martingale property of successive predictions (assuming integrability) we have

$$\mathbb{E}\left[\left.\mathrm{CDR}_i^{(t+1)}\right| \mathcal{D}_t\right] = 0.$$

For this reason the claims development result is predicted by 0. The conditional MSEP viewed from time t for this prediction is then defined by (subject to existence)

$$\mathrm{msep}_{\mathrm{CDR}_i^{(t+1)}|\mathcal{D}_t}(0) = \mathbb{E}\left[\left.\left(\mathrm{CDR}_i^{(t+1)} - 0\right)^2\right| \mathcal{D}_t\right] = \mathrm{Var}\left(\mathbb{E}\left[C_{i,J}|\mathcal{D}_{t+1}\right]|\mathcal{D}_t\right). \qquad (3.7)$$

The following lemma (which holds for any claims reserving method) relates the short-term view (3.7) to the long-term view (3.1) (the proof is provided in the appendix).

Lemma 3.3. *Assume* (3.1) *exists. Then, for* $t \geq I \geq i > t - J$ *we have*

$$\mathrm{msep}_{C_{i,J}|\mathcal{D}_t}\left(\mathbb{E}\left[C_{i,J}|\mathcal{D}_t\right]\right) \geq \mathrm{msep}_{\mathrm{CDR}_i^{(t+1)}|\mathcal{D}_t}(0).$$

The aim is to study (3.7) in the non-informative prior limit $\gamma \to 1$ of the gamma-gamma BCL model. We define the credibility weights for $t \geq I \geq t - j \geq 1$ and $0 \leq j \leq J - 1$ by

$$\alpha_j^{(t)} = \frac{C_{t-j,j}}{\sum_{\ell=1}^{t-j} C_{\ell,j}} \in (0, 1].$$

We have the following theorem (the proof is provided in the appendix).

Theorem 3.4. *Under Model Assumptions 2.1,* (3.3) *and* (3.4)*, we have for* $t \geq I \geq i > t - J$

$$\widehat{\mathrm{msep}}_{\mathrm{CDR}_i^{(t+1)}|\mathcal{D}_t}(0) \stackrel{\mathrm{def.}}{=} \lim_{\gamma \to 1} \mathrm{msep}_{\mathrm{CDR}_i^{(t+1)}|\mathcal{D}_t}(0)$$

$$= \left(\widehat{C}_{i,J}^{CL(t)}\right)^2 \left[\left(1 + \left(\alpha_{t-i}^{(t)}\right)^{-1} \widehat{\mathbb{E}}\left[\left.\Psi_{t-i}^{(t)}\right| \mathcal{D}_t\right]\right) \prod_{j=t-i+1}^{J-1} \left(1 + \alpha_j^{(t)} \widehat{\mathbb{E}}\left[\left.\Psi_j^{(t)}\right| \mathcal{D}_t\right]\right) - 1\right];$$

and the non-informative prior limit of the aggregated conditional MSEP is given by

$$\widehat{\mathrm{msep}}_{\sum_i \mathrm{CDR}_i^{(t+1)}|\mathcal{D}_t}(0) \stackrel{\mathrm{def.}}{=} \lim_{\gamma \to 1} \mathrm{msep}_{\sum_i \mathrm{CDR}_i^{(t+1)}|\mathcal{D}_t}(0)$$

$$= \sum_i \widehat{\mathrm{msep}}_{\mathrm{CDR}_i^{(t+1)}|\mathcal{D}_t}(0)$$

$$+ 2 \sum_{i<m} \widehat{C}_{i,J}^{CL(t)} \widehat{C}_{m,J}^{CL(t)} \left[\left(1 + \widehat{\mathbb{E}}\left[\left.\Psi_{t-i}^{(t)}\right| \mathcal{D}_t\right]\right) \prod_{j=t-i+1}^{J-1} \left(1 + \alpha_j^{(t)} \widehat{\mathbb{E}}\left[\left.\Psi_j^{(t)}\right| \mathcal{D}_t\right]\right) - 1\right],$$

where the summations run over $t - J + 1 \leq i \leq I$ *and* $t - J + 1 \leq i < m \leq I$*, respectively.*

4 Numerical sensitivities

In this section we study several examples in order to analyze the sensitivities of the prediction uncertainty to the choice of the prior distributions $\pi_j(\cdot)$ for σ_j. We start by describing the empirical Bayesian case considered in [8]. In the empirical Bayesian case we use plug-in estimates $\widehat{\sigma}_j$ for σ_j which is equivalent to choosing a degenerate prior distribution function $\pi_j(\sigma_j) =$

$\delta_{\{\sigma_j=\widehat{\sigma}_j\}}$ that has a point mass of size 1 in $\widehat{\sigma}_j$. Therefore, we first need to provide the empirical estimates $\widehat{\sigma}_j$. Observe that in the classical distribution-free CL model of Mack [5] one identifies $s_j^2 = \Theta_j^{-2}\sigma_j^2$, compare formula (3) in Mack [5] to formula (2.2) above. If we use Mack's [5] estimates for these variance parameters we obtain at time $t = I$, for $0 \le j \le (J-1) \wedge (I-3)$, the result

$$\widehat{s}_j^2 = \frac{1}{I-j-2} \sum_{\ell=1}^{I-j-1} C_{\ell,j} \left(\frac{C_{\ell,j+1}}{C_{\ell,j}} - \widehat{f}_j^{CL(I)} \right)^2,$$

and if $j = J - 1 = I - 2$, we set

$$\widehat{s}_{J-1}^2 = \min\left\{ \widehat{s}_{J-3}^2, \widehat{s}_{J-2}^2, \widehat{s}_{J-2}^4/\widehat{s}_{J-3}^2 \right\} = \min\left\{ \widehat{s}_{J-3}^2, \widehat{s}_{J-2}^4/\widehat{s}_{J-3}^2 \right\}.$$

Since Θ_j^{-1} plays the role of the CL factor, see (2.1), we set for the empirical standard deviation estimate of σ_j

$$\widehat{\sigma}_j = \frac{\sqrt{\widehat{s}_j^2}}{\widehat{f}_j^{CL(I)}}, \qquad \text{for } 0 \le j \le J - 1.$$

The model is now fully specified for obtaining the empirical Bayesian estimate [8] and Mack's distribution-free CL estimate [5] for the CL reserves and the corresponding conditional MSEPs. We calculate these for the data given in Table A.1 of [8], see also Table 2 in the appendix (where we additionally provide the parameter estimates $\widehat{f}_j^{CL(I)}$, $\widehat{s}_j$ and $\widehat{\sigma}_j$).

accident year i	CL reserves $\mathcal{R}_i^{CL(I)}$	emp. Bayes' $\widehat{\text{msep}}^{1/2}$	Mack's $\text{msep}^{1/2}$	emp. CDR $\widehat{\text{msep}}^{1/2}$
1	0			
2	15.126	0.267	0.267	0.267
3	26.257	0.914	0.914	0.884
4	34.538	3.058	3.058	2.948
5	85.302	7.628	7.628	7.018
6	156.494	33.341	33.341	32.470
7	286.121	73.467	73.467	66.178
8	449.167	85.399	85.398	50.296
9	1'043.242	134.338	134.337	104.311
10	3'950.815	410.850	410.817	385.773
total	6'047.061	462.990	462.960	420.220

Table 1: CL reserves $\mathcal{R}_i^{CL(I)}$ and rooted conditional long-term MSEPs in the non-informative prior limit of the gamma-gamma BCL model of Theorem 3.2 with priors $\pi_j(\sigma_j) = \delta_{\{\sigma_j=\widehat{\sigma}_j\}}$, Mack's formula [5] and the conditional short-term MSEP of Theorem 3.4 for the data of Table 2.

This then provides the results given in Table 1 where the CL reserves are defined as the prediction of the outstanding loss liabilities at time $t = I$, for $I \ge i > I - J$ given by

$$\mathcal{R}_i^{CL(I)} = \widehat{C}_{i,J}^{CL(I)} - C_{i,I-i}.$$

Note that these results coincide with Tables 2.4 and 2.5 of [8]. Moreover, we remark that

$$\sum_{\ell=1}^{I-j-1} C_{\ell,j} > \widehat{\sigma}_j^2 \qquad \text{for all } 0 \le j \le J - 1,$$

which, in particular, means that (3.4) holds for any $d_j \in (\widehat{\sigma}_j, \sum_{\ell=1}^{I-j-1} C_{\ell,j})$.

The goal now is to challenge these results by replacing the empirical estimates $\widehat{\sigma}_j$ by non-trivial prior densities $\pi_j(\cdot)$ for σ_j. To study the sensitivities we make different choices of the prior density $\pi_j(\cdot)$, namely, we choose integers $k \in \{2, \ldots, 20\}$ and select uniform priors for $\pi_j(\cdot)$ on the intervals

$$(0, d_j) = (0, k \cdot \widehat{\sigma}_j) \qquad \text{for } 0 \le j \le J-1. \tag{4.1}$$

These choices guarantee that the support $(0, d_j)$ of $\pi_j(\cdot)$ contains the empirical estimate $\widehat{\sigma}_j$ and for increasing k the prior distribution $\pi_j(\cdot)$ of σ_j gets less informative; for $k = 20$ there is a positive probability that σ_j is up to 20 times as large as the empirical estimate $\widehat{\sigma}_j$. Moreover, we note that in all our examples relation (3.4) is fulfilled for the maximal $d_j = 20 \cdot \widehat{\sigma}_j$.

For these choices $k \in \{2, \ldots, 20\}$ in (4.1) we calculate the rooted conditional long-term MSEP for aggregated accident years provided in Theorem 3.2 and the rooted conditional CDR MSEP (short-term view) for aggregated accident years provided in Theorem 3.4. The case $k = 1$ is identified with the empirical Bayesian estimate provided in Table 1 (note that this corresponds to the priors $\pi_j(\sigma_j) = \delta_{\{\sigma_j = \widehat{\sigma}_j\}}$). The only remaining difficulty for this analysis is the calculation of the one-dimensional integrals (3.5)-(3.6). This can be done either with importance sampling or with numerical integration. We have seen that in our examples it was sufficiently precise to divide the intervals $(0, d_j)$ into $M = 1'000$ equally spaced sub-intervals $(a_j^{(m)}, a_j^{(m+1)})$ with $a_j^{(1)} = 0$ and $a_j^{(M+1)} = d_j$ and then approximate the denominator by the discrete sum

$$k_j(\mathcal{D}_I) = \int_0^{d_j} h_j(\sigma_j | \mathcal{D}_I) \, d\sigma_j \approx \sum_{m=1}^{M} \frac{1}{a_j^{(m+1)} - a_j^{(m)}} \, h_j \left((a_j^{(m)} + a_j^{(m+1)})/2 \Big| \mathcal{D}_I \right),$$

and analogously for the remaining terms. In Figure 1 we provide these (un-normalized) posteriors $h_j(\sigma_j | \mathcal{D}_I)$ for $0 \le j \le J-1$ and for uniform priors $\pi_j(\cdot)$ having $k = 2$ (in red) and $k = 5$ (in blue), the dotted green line provides the empirical estimate $\widehat{\sigma}_j$. We see that the bigger development period j the more volatile is the posterior of σ_j. The extreme case $j = J-1 = I-2$ means that we have only one single observation $C_{1,J}$ from which we cannot estimate a variance parameter and, henceforth, the posterior of σ_{J-1} remains a uniform distribution, see Figure 1 (bottom rhs), this is also seen in derivation (A.3), below.

The conditional MSEP results are presented in Figure 2. The left-hand side (lhs) of Figure 2 gives the CL reserves $\sum_i \mathcal{R}_i^{CL(I)}$ aggregated over all accident years $2 \le i \le I$ and the corresponding confidence bounds of $\pm \, 2 \, \widehat{\text{msep}}_{\sum_i C_{i,J} | \mathcal{D}_I}^{1/2} (\sum_i \widehat{C}_{i,J}^{CL(I)})$, see Theorem 3.2, and $\pm \, 2 \, \widehat{\text{msep}}_{\sum_i \text{CDR}_i^{(I+1)} | \mathcal{D}_I}^{1/2} (0)$, see Theorem 3.4, for the different values of $k \in \{2, \ldots, 20\}$. As expected, we see that the confidence bounds increase with increasing prior uncertainty in $\pi_j(\cdot)$ (i.e. increasing k), however the graph seems to stabilize rather quickly (around $k = 5$) in this example. The right-hand side (rhs) of Figure 2 shows this increase relative to the empirical Bayesian case. We observe an overall increase of roughly 34% (rooted total MSEP) and 29% (rooted CDR MSEP) relative to the empirical Bayesian one. Thus, we observe an increase, but not a dramatic one.
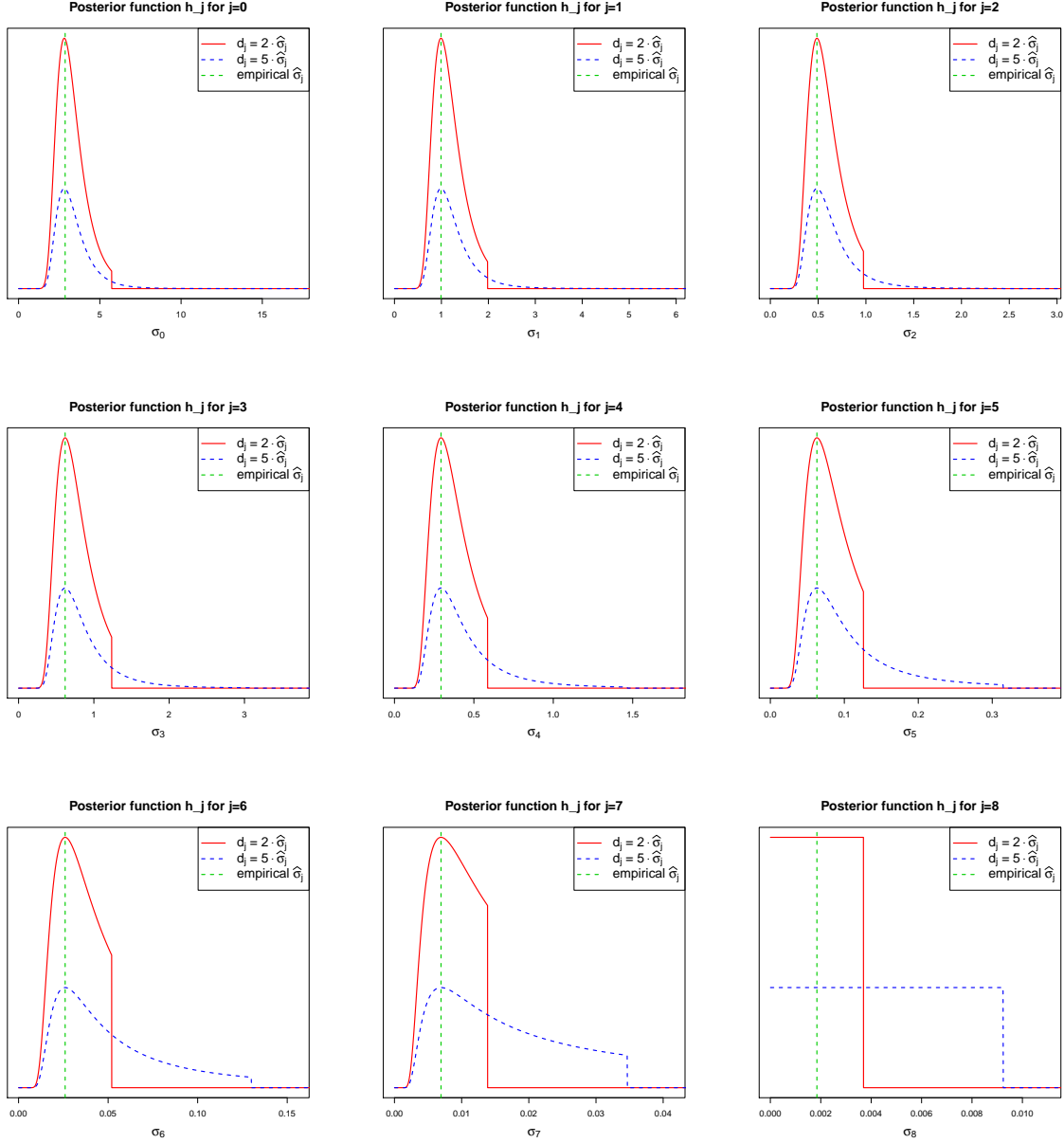
Figure 1: Un-normalized posteriors $h_j(\sigma_j|\mathcal{D}_I)$ for $0 \leq j \leq 8$ and uniform priors $\pi_j(\cdot)$ with $k = 2, 5$, the x-axis is truncated at $6 \cdot \widehat{\sigma}_j$.

In the remainder we present more examples of different lines of business. In Figure 3 we consider a motor third party liability (MTPL) insurance portfolio. For this portfolio we have $I = 22$ accident years and final development period $J = 15$. We see a similar picture as in the previous example: the increase of the rooted conditional MSEP stabilizes around $k = 5$ and the total increase relative to the empirical Bayesian estimate is roughly 18% and 16%, respectively. The reason for this fast convergence is that the posterior uncertainty in σ_j, given $\mathcal{D}_I$, is rather small for all development periods j regardless of the choice of the prior distribution. This comes from the fact that we have a large trapezoid (22×16) and even for the latest development period
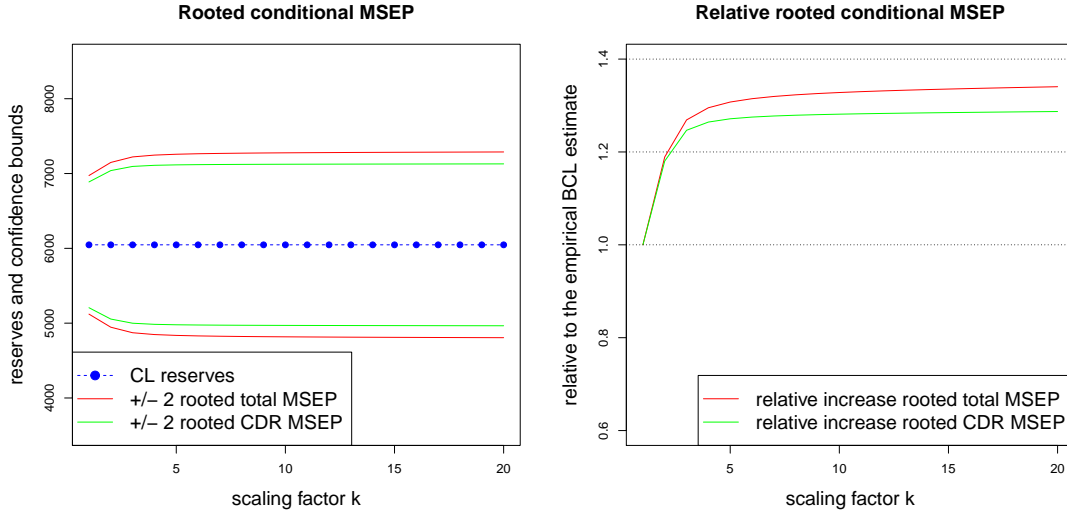
Figure 2: Example of Tables 1 and 2, the x-axis displays the different choices $k \in \{1, \ldots, 20\}$ in (4.1): (lhs) CL reserves $\sum_i \mathcal{R}_i^{CL(I)}$ aggregated over all accident years $2 \leq i \leq I$ and confidence bounds of ± 2 rooted conditional MSEPs in the long-term view (total MSEP over all development periods) and the short-term view (CDR MSEP); (rhs) rooted conditional MSEPs for $k \in \{2, \ldots, 20\}$ divided by the corresponding empirical rooted conditional MSEPs (which corresponds to $k = 1$).
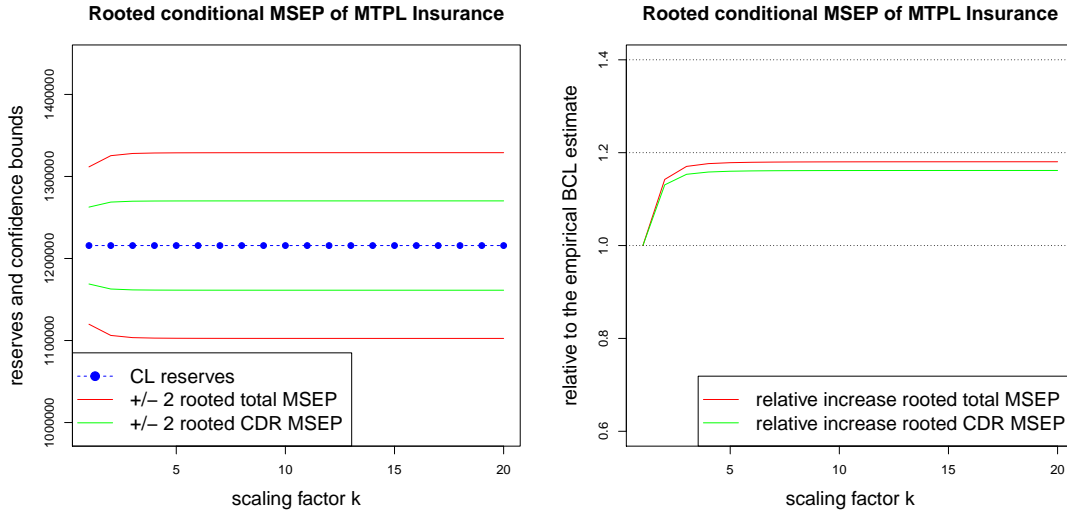

Figure 3: Motor third party liability (MTPL) insurance portfolio of a 22×16 trapezoid; relative increase is 18% (rooted total MSEP) and 16% (rooted CDR MSEP).

$J = 15$ we have 7 observations.

In Figure 4 we consider a general liability insurance portfolio of size $I = 21$ and $J = 13$, the relative increase amounts to roughly 14% and 12%, respectively, and the picture is very similar
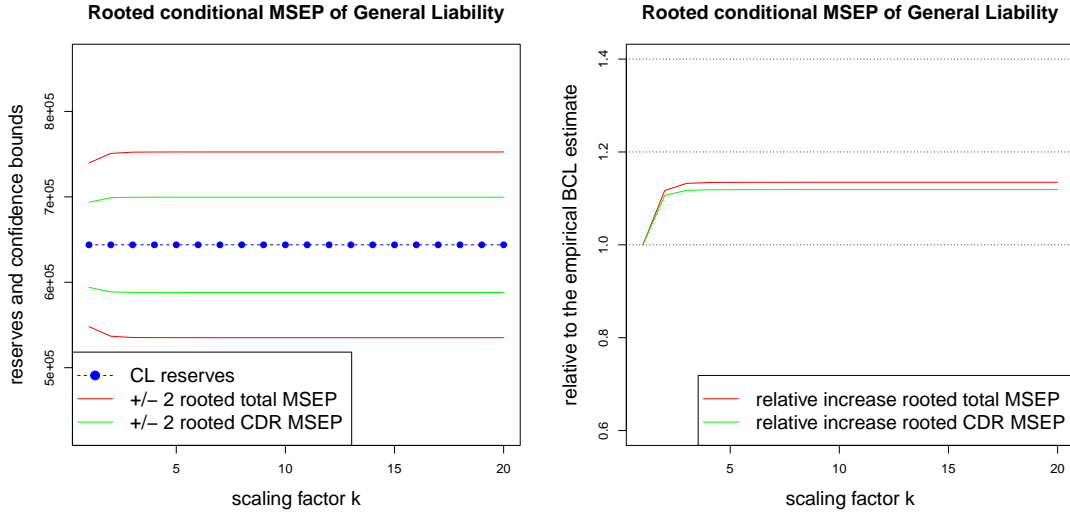
10

Figure 4: General liability insurance portfolio of a 21×14 trapezoid; relative increase is 14% (rooted total MSEP) and 12% (rooted CDR MSEP).
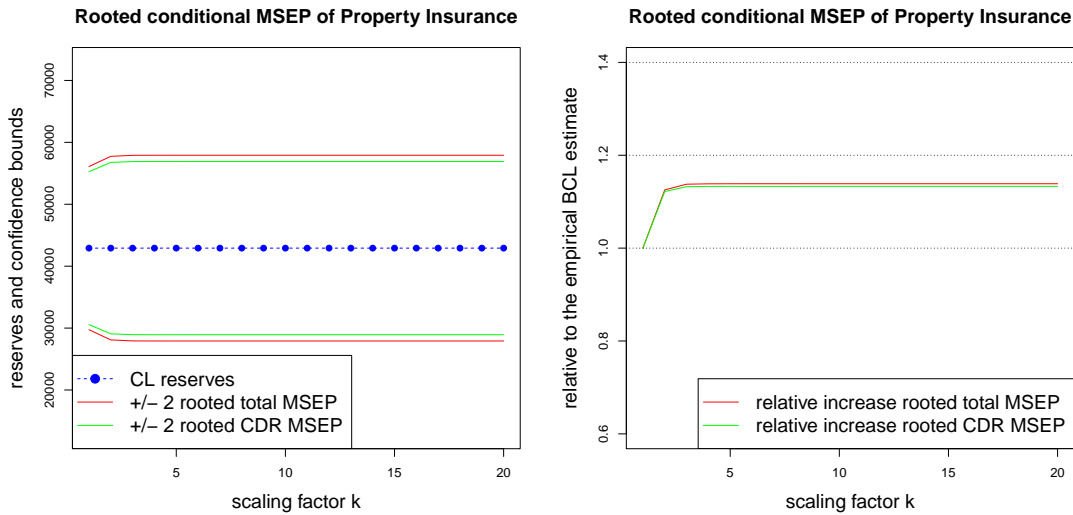

Figure 5: Industrial property insurance portfolio of a 15×7 trapezoid; relative increase is 14% (rooted total MSEP) and 13% (rooted CDR MSEP).

to MTPL insurance because posterior uncertainty in σ_j is again small due to the size of the trapezoid and the fact that we have 8 observations in the last column $J = 13$.

In Figure 5 we consider an industrial property insurance portfolio of size $I = 15$ and $J = 6$, the relative increase amounts to roughly 14% and 13%, respectively. Thus, again we obtain a rather similar picture to the previous examples. Note that here we have 9 observations in the last column $J = 6$.

As a rule of thumb we can say that if the right-end point of the posterior functions $h_j(\sigma_j | \mathcal{D}_I)$,

11

given in Figure 1, is sufficiently smooth then increasing the support of the uniform prior $\pi_j(\cdot)$ does not substantially increase the uncertainty estimates. This holds true as long as we do not get to close to the threshold (3.4) because at the threshold the conditional MSEP becomes infinite, see also Theorem 2.12 in [8].

5 Conclusions

We have considered the full Bayesian version of the gamma-gamma BCL model. This model provides the CL reserves in its non-informative prior limit. Therefore, it can be used to analyze the prediction uncertainty of the CL claims reserving method. The main purpose of this paper was to analyze the contribution of the variance parameters to the prediction uncertainty in the case where these variance parameters are also modeled in a Bayesian way. In our examples this additional source of uncertainty increases the rooted conditional MSEP between 10% and 40% (compared to the empirical Bayesian version presented in [8]). This increase is bigger for triangles $I = J - 1$ and becomes the smaller for trapezoids to bigger the difference $I - J$ is.

A Proofs

Proof of Theorem 2.2. For explicit details we refer to Section 2.2.1 in [8]. From (2.3) we see that the posteriors of (Θ_j, σ_j) are independent for different development periods $0 \leq j \leq J - 1$. Lemma 2.7 and Corollary 2.8 of [8] then imply for $t \geq I \geq i > t - n \geq t - J$

$$\mathbb{E}\left[C_{i,n} | \mathcal{D}_t, \boldsymbol{\sigma}\right] = C_{i,t-i} \prod_{j=t-i}^{n-1} \left(\omega_j^{(t)} \, \widehat{f}_j^{CL(t)} + (1 - \omega_j^{(t)})f_j\right), \tag{A.1}$$

with credibility weights

$$\omega_j^{(t)} = \frac{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j}}{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j} + \sigma_j^2 \, (\gamma_j - 1)} \in (0,1).$$

This implies using the tower property and posterior independence (2.3)

$$\mathbb{E}\left[C_{i,n} | \mathcal{D}_t\right] = \mathbb{E}\left[\mathbb{E}\left[C_{i,n} | \mathcal{D}_t, \boldsymbol{\sigma}\right] | \mathcal{D}_t\right] = C_{i,t-i} \prod_{j=t-i}^{n-1} \mathbb{E}\left[\left(\omega_j^{(t)} \, \widehat{f}_j^{CL(t)} + (1 - \omega_j^{(t)})f_j\right) \Big| \mathcal{D}_t\right]$$

$$= C_{i,t-i} \prod_{j=t-i}^{n-1} \left(\mathbb{E}\left[\omega_j^{(t)} \Big| \mathcal{D}_t\right] \widehat{f}_j^{CL(t)} + \left(1 - \mathbb{E}\left[\omega_j^{(t)} \Big| \mathcal{D}_t\right]\right) f_j\right).$$

The remains the consideration of the conditional expectations of the credibility weights. The bounded support assumption on σ_j implies, $\mathbb{P}$-a.s.,

$$\underline{\omega}_j^{(t)} \stackrel{\text{def.}}{=} \frac{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j}}{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j} + d_j^2 \, (\gamma_j - 1)} \leq \omega_j^{(t)} \leq 1. \tag{A.2}$$

Since the $\mathcal{D}_{t-1}$-measurable lower bound $\underline{\omega}_j^{(t)}$ converges to 1 as $\gamma_j \to 1$, the claim follows. $\qquad \square$

Proof of Lemma 3.1. The case $(t - j - 1) \wedge I = 1$ can easily be treated because in that case we have

$$k_j(\mathcal{D}_t) = \int_0^{d_j} \frac{C_{1,j}}{C_{1,j+1}} \, \pi_j(\sigma_j)d\sigma_j = \frac{C_{1,j}}{C_{1,j+1}} < \infty. \tag{A.3}$$

To prove the integrability on $(0, d_j)$ in the case $(t - j - 1) \wedge I > 1$ we need to check that $h_j(\sigma_j | \mathcal{D}_t)$ is well-behaved for $\sigma_j \to 0$. We consider the following logarithm

$$\log\left(h_j(\sigma_j | \mathcal{D}_t)\right) = \log\left(\sigma_j^2 \left(\sum_{i=1}^{(t-j-1)\wedge I} C_{i,j+1}\right)^{-\left(1 + \sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sigma_j^2}\right)} \prod_{i=1}^{(t-j-1)\wedge I} C_{i,j+1}^{\frac{C_{i,j}}{\sigma_j^2}}\right)$$

$$+ \log\Gamma\left(1 + \sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sigma_j^2}\right) - \sum_{i=1}^{(t-j-1)\wedge I} \log\Gamma\left(\frac{C_{i,j}}{\sigma_j^2}\right) + \log\left(\pi_j(\sigma_j)\right). \qquad \text{(A.4)}$$

The first line on the right-hand side behaves for $\sigma_j \to 0$ as

$$\log\left(\sigma_j^2\right) + \sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sigma_j^2} \log\left(\frac{C_{i,j+1}}{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j+1}}\right) + O(1).$$

The terms involving the gamma functions $\Gamma(\cdot)$ are treated by Stirling's formula

$$\log\Gamma(x) = \frac{1}{2}\log(2\pi) + \left(x - \frac{1}{2}\right)\log x - x + o(x), \qquad \text{as } x \to \infty.$$

The identity $\Gamma(1 + x) = x\Gamma(x)$ and Stirling's formula applied to the first two terms on the second line on the right-hand side of (A.4) provide for $\sigma_j \to 0$

$$\log\left(\sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sigma_j^2}\right) + \left(\sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sigma_j^2} - \frac{1}{2}\right) \log\left(\sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sigma_j^2}\right) - \sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sigma_j^2}$$

$$- \sum_{i=1}^{(t-j-1)\wedge I} \left[\left(\frac{C_{i,j}}{\sigma_j^2} - \frac{1}{2}\right) \log\left(\frac{C_{i,j}}{\sigma_j^2}\right) - \left(\frac{C_{i,j}}{\sigma_j^2}\right)\right] + O(1)$$

$$= -\log(\sigma_j^2) + \sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sigma_j^2} \log\left(\frac{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j}}{C_{i,j}}\right) + \frac{1 - (t-j-1)\wedge I}{2} \log\left(\sigma_j^2\right) + O(1).$$

If we merge these two terms we obtain for $\sigma_j \to 0$ behavior

$$\sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sigma_j^2} \log\left(\frac{C_{i,j+1}}{C_{i,j}} \frac{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j}}{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j+1}}\right) + \frac{1 - (t-j-1)\wedge I}{2}\log(\sigma_j^2) + O(1)$$

$$= \frac{1}{\sigma_j^2}\left[\sum_{i=1}^{(t-j-1)\wedge I} C_{i,j} \log\left(\frac{C_{i,j+1}}{C_{i,j}}\right) - \sum_{i=1}^{(t-j-1)\wedge I} C_{i,j} \log\left(\widehat{f}_j^{CL(t)}\right)\right] + \frac{1 - (t-j-1)\wedge I}{2}\log(\sigma_j^2) + O(1).$$

By assumption, for at least one accident year $1 \le i \le (t - j - 1) \wedge I$ we have $C_{i,j+1}/C_{i,j} \neq \widehat{f}_j^{CL(t)}$. Therefore, Jensen's inequality is strict in the following derivation

$$\sum_{i=1}^{(t-j-1)\wedge I} C_{i,j} \log\left(\frac{C_{i,j+1}}{C_{i,j}}\right) = \sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j} \sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j}} \log\left(\frac{C_{i,j+1}}{C_{i,j}}\right)$$

$$< \sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j} \log\left(\sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j}} \frac{C_{i,j+1}}{C_{i,j}}\right) = \sum_{i=1}^{(t-j-1)\wedge I} C_{i,j} \log\left(\widehat{f}_j^{CL(t)}\right).$$

Therefore, we can define the following strictly positive constant

$$\varepsilon = \sum_{i=1}^{(t-j-1)\wedge I} C_{i,j} \log\left(\widehat{f}_j^{CL(t)}\right) - \sum_{i=1}^{(t-j-1)\wedge I} C_{i,j} \log\left(\frac{C_{i,j+1}}{C_{i,j}}\right) > 0.$$

This implies for $\sigma_j \to 0$ and a strictly positive constant $\varepsilon > 0$

$$h_j(\sigma_j | \mathcal{D}_t) = \exp\left\{-\frac{\varepsilon}{\sigma_j^2} + \frac{1 - (t-j-1)\wedge I}{2}\log(\sigma_j^2) + O(1)\right\} \pi_j(\sigma_j).$$

This provides integrability in 0 and completes the proof. $\qquad \square$

Lemma A.1. *Under Model Assumptions 2.1, we have in the non-informative prior limit for* $t \geq I \geq i > t - J$ *(subject to existence)*

$$\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{D}_t}\left(\widehat{C}_{i,J}^{CL(t)}\right) = \lim_{\gamma \to 1} \mathrm{msep}_{C_{i,J}|\mathcal{D}_t}\left(\mathbb{E}\left[C_{i,J} \middle| \mathcal{D}_t\right]\right) = \lim_{\gamma \to 1} \mathbb{E}\left[\mathrm{Var}\left(C_{i,J} \middle| \mathcal{D}_t, \boldsymbol{\sigma}\right) \middle| \mathcal{D}_t\right].$$

Proof of Lemma A.1. We have for the first term in the conditional MSEP decomposition (3.1), using (A.1), see also Lemma 2.7 and Corollary 2.8 in [8], and using posterior independence (2.3)

$$\mathrm{Var}\left(\mathbb{E}\left[C_{i,J} \middle| \mathcal{D}_t, \boldsymbol{\sigma}\right] \middle| \mathcal{D}_t\right) = C_{i,t-i}^2 \mathrm{Var}\left(\prod_{j=t-i}^{J-1}\left(\omega_j^{(t)} \widehat{f}_j^{CL(t)} + (1-\omega_j^{(t)})f_j\right) \middle| \mathcal{D}_t\right)$$

$$= C_{i,t-i}^2 \left(\prod_{j=t-i}^{J-1} \mathbb{E}\left[\left(\omega_j^{(t)} \widehat{f}_j^{CL(t)} + (1-\omega_j^{(t)})f_j\right)^2 \middle| \mathcal{D}_t\right] - \prod_{j=t-i}^{J-1} \mathbb{E}\left[\omega_j^{(t)} \widehat{f}_j^{CL(t)} + (1-\omega_j^{(t)})f_j \middle| \mathcal{D}_t\right]^2\right).$$

Since $\omega_j^{(t)}$ is sandwiched, $\mathbb{P}$-a.s., between a $\mathcal{D}_{t-1}$-measurable lower bound $\underline{\omega}_j^{(t)}$ that converges to 1 as $\gamma_j \to 1$ and 1, see (A.2), we obtain that the right-hand side of the last identity converges to 0 as $\boldsymbol{\gamma} \to 1$. This proves the lemma. □

Formula (2.3) provides the posterior (marginal) distribution of σ_j, conditionally given $\mathcal{D}_t$. This posterior distribution depends on the explicit choice of the prior parameter $\gamma_j > 1$, to illustrate this we use the following notation $\pi_j^{(\gamma_j)}(\sigma_j|\mathcal{D}_t) = \pi_j(\sigma_j|\mathcal{D}_t)$.

Lemma A.2. *Choose* $0 \leq j \leq J - 1$ *and* $t \geq I$, *and assume that either* $(t - j - 1) \wedge I = 1$ *or for at least one accident year* $1 \leq i \leq (t - j - 1) \wedge I$ *we have* $C_{i,j+1}/C_{i,j} \neq \widehat{f}_j^{CL(t)}$. *Consider* $S_j^{(\gamma_j)} \sim \pi_j^{(\gamma_j)}(\sigma_j|\mathcal{D}_t)$. *We have the following convergence in distribution*

$$\lim_{\gamma_j \to 1} S_j^{(\gamma_j)} \overset{(\mathrm{d})}{=} S_j \sim \widehat{\pi}_j(\sigma_j) = k_j(\mathcal{D}_t)^{-1} h_j(\sigma_j|\mathcal{D}_t).$$

We use the notation $S_j^{(\gamma_j)}$ for that standard deviation parameter σ_j in the above lemma to more clearly indicate the dependence of the law of the corresponding random variable $S_j^{(\gamma_j)}$ from the choice of the prior parameter γ_j and to clearly distinguish random variables $S_j^{(\gamma_j)}$ from their potential realizations σ_j in the domain $(0, d_j)$.

Proof of Lemma A.2. Convergence in distribution is implied by the corresponding convergence of the moment generating functions, that is, for any $r \in \mathbb{R}$ we aim at proving

$$\lim_{\gamma_j \to 1} \mathbb{E}\left[\exp\{rS_j^{(\gamma_j)}\} \middle| \mathcal{D}_t\right] = \mathbb{E}\left[\exp\{rS_j\} \middle| \mathcal{D}_t\right].$$

From (2.3) we obtain posterior distribution, conditionally given $\mathcal{D}_t$,

$$\pi(\boldsymbol{\sigma}|\mathcal{D}_t) = \int \pi(\boldsymbol{\theta}, \boldsymbol{\sigma}|\mathcal{D}_t)d\boldsymbol{\theta} \propto \prod_{j=0}^{J-1}\left[\left(f_j(\gamma_j - 1) + \sum_{i=1}^{(t-j-1)\wedge I}\frac{C_{i,j+1}}{\sigma_j^2}\right)^{-\left(\gamma_j + \sum_{i=1}^{(t-j-1)\wedge I}\frac{C_{i,j}}{\sigma_j^2}\right)}\right.$$

$$\left.\times \Gamma\left(\gamma_j + \sum_{i=1}^{(t-j-1)\wedge I}\frac{C_{i,j}}{\sigma_j^2}\right)\prod_{i=1}^{(t-j-1)\wedge I}\frac{\left(\frac{C_{i,j+1}}{\sigma_j^2}\right)^{\frac{C_{i,j}}{\sigma_j^2}}}{\Gamma\left(\frac{C_{i,j}}{\sigma_j^2}\right)}\pi_j(\sigma_j)\right]. \quad (A.5)$$

14

This again provides posterior independence between the different development years $0 \leq j \leq J-1$. We define the corresponding marginal functions on the intervals $(0, d_j)$

$$h_j^{(\gamma_j)}(\sigma_j|\mathcal{D}_t) = \frac{\Gamma\left(\gamma_j + \sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sigma_j^2}\right)}{\left(f_j(\gamma_j - 1) + \sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j+1}}{\sigma_j^2}\right)^{\gamma_j + \sum_{i=1}^{(t-j-1)\wedge I} \frac{C_{i,j}}{\sigma_j^2}}} \prod_{i=1}^{(t-j-1)\wedge I} \frac{\left(\frac{C_{i,j+1}}{\sigma_j^2}\right)^{\frac{C_{i,j}}{\sigma_j^2}}}{\Gamma\left(\frac{C_{i,j}}{\sigma_j^2}\right)} \pi_j(\sigma_j),$$

thus the density of $S_j^{(\gamma_j)}$ is given by

$$\pi_j^{(\gamma_j)}(\sigma_j|\mathcal{D}_t) = \frac{h_j^{(\gamma_j)}(\sigma_j|\mathcal{D}_t)}{\int_0^{d_j} h_j^{(\gamma_j)}(\sigma_j|\mathcal{D}_t)d\sigma_j},$$

and we aim at proving the following convergence of moment generating functions

$$\lim_{\gamma_j \to 1} \mathbb{E}\left[\exp\{rS_j^{(\gamma_j)}\}\Big| \mathcal{D}_t\right] = \lim_{\gamma_j \to 1} \frac{\int_0^{d_j} e^{r\sigma_j} h_j^{(\gamma_j)}(\sigma_j|\mathcal{D}_t)d\sigma_j}{\int_0^{d_j} h_j^{(\gamma_j)}(\sigma_j|\mathcal{D}_t)d\sigma_j} = \frac{\int_0^{d_j} e^{r\sigma_j} h_j(\sigma_j|\mathcal{D}_t)d\sigma_j}{k_j(\mathcal{D}_t)} = \mathbb{E}\left[\exp\{rS_j\}| \mathcal{D}_t\right].$$

Firstly, observe that we the following point-wise (in σ_j) convergence on the interval $(0, d_j)$

$$\lim_{\gamma_j \to 1} h_j^{(\gamma_j)}(\sigma_j|\mathcal{D}_t) = h_j(\sigma_j|\mathcal{D}_t).$$

Thus, there remains to prove the existence of a uniform integrable upper bound such that one can apply Lebesgue's dominated convergence theorem which proves the claim.

To find an integrable upper bound we first we remark that for all $r \in \mathbb{R}$ we have uniform bound on $(0, d_j)$

$$0 \leq e^{r\sigma_j} \leq \max\{e^{rd_j}, 1\}.$$

There remains the consideration of $h_j^{(\gamma_j)}(\sigma_j|\mathcal{D}_t)$. We set $c_j = \sum_{i=1}^{(t-j-1)\wedge I} C_{i,j}$ and $c_{j+1} = \sum_{i=1}^{(t-j-1)\wedge I} C_{i,j+1}$ and $p(\sigma_j) = \prod_{i=1}^{(t-j-1)\wedge I} \left(\frac{C_{i,j+1}}{\sigma_j^2}\right)^{\frac{C_{i,j}}{\sigma_j^2}} \Gamma\left(\frac{C_{i,j}}{\sigma_j^2}\right)^{-1} \pi_j(\sigma_j)$. Then we have

$$\begin{aligned}
h_j^{(\gamma_j)}(\sigma_j|\mathcal{D}_t) &= \left(f_j(\gamma_j - 1) + c_{j+1}/\sigma_j^2\right)^{-\left(\gamma_j + c_j/\sigma_j^2\right)} \Gamma\left(\gamma_j + c_j/\sigma_j^2\right) p(\sigma_j) \\
&= p(\sigma_j) \int_0^\infty x^{\gamma_j + c_j/\sigma_j^2 - 1} \exp\{-(f_j(\gamma_j - 1) + c_{j+1}/\sigma_j^2)x\} \, dx \\
&= p(\sigma_j) \int_0^\infty \exp\{(\gamma_j - 1)(\log x - f_j x)\} \, x^{c_j/\sigma_j^2} \exp\{-xc_{j+1}/\sigma_j^2\} \, dx.
\end{aligned}$$

Define the function $\ell(x) = \log x - f_j x$. It satisfies

$$\lim_{x \to 0} \ell(x) = -\infty \qquad \text{and} \qquad \lim_{x \to \infty} \ell(x) = -\infty.$$

Since $\log x$ is concave there are only two cases:
(i) $\ell(x) \leq 0$ for all $x \in [0, \infty)$. In this case we have for $\gamma_j \geq 1$

$$\begin{aligned}
h_j^{(\gamma_j)}(\sigma_j|\mathcal{D}_t) &= p(\sigma_j) \int_0^\infty \exp\{(\gamma_j - 1)\ell(x)\} \, x^{c_j/\sigma_j^2} \exp\{-xc_{j+1}/\sigma_j^2\} \, dx \\
&\leq p(\sigma_j) \int_0^\infty x^{c_j/\sigma_j^2} \exp\{-xc_{j+1}/\sigma_j^2\} \, dx = p(\sigma_j) \frac{\Gamma(1 + c_j/\sigma_j^2)}{(c_{j+1}/\sigma_j^2)^{1+c_j/\sigma_j^2}} = h_j(\sigma_j|\mathcal{D}_t).
\end{aligned}$$

Lemma 3.1 then provides the integrable upper bound in the first case (i) and we may apply Lebesgue's dominated convergence theorem with proves the claim.
(ii) There exists $x \in (0, \infty)$ such that $\ell(x) > 0$. In this case concavity of $\log x$ implies that there exist $0 < x_1 < x_2 < \infty$ such that $\ell(x) \leq 0$ on $[0, x_1] \cup [x_2, \infty)$ and $\ell(x) > 0$ on $[x_1, x_2]$. We then have

$$\begin{aligned}
h_j^{(\gamma_j)}(\sigma_j|\mathcal{D}_t) &\leq p(\sigma_j) \int_0^\infty x^{c_j/\sigma_j^2} \exp\{-xc_{j+1}/\sigma_j^2\} \, dx \\
&\quad + p(\sigma_j) \int_{x_1}^{x_2} (\exp\{(\gamma_j - 1)\ell(x)\} - 1) \, x^{c_j/\sigma_j^2} \exp\{-xc_{j+1}/\sigma_j^2\} \, dx.
\end{aligned}$$

15

Since we consider the limit $\gamma_j \to 1$ we may (and will) w.l.o.g. assume that $\gamma_j \leq 2$. This implies

$$K = \sup_{1 \leq \gamma_j \leq 2, \; x_1 \leq x \leq x_2} \exp\left\{ (\gamma_j - 1)\, \ell(x) \right\} = (f_j e)^{-1} \in (1, \infty).$$

Thus, in this second case (ii) we have for $1 \leq \gamma_j \leq 2$

$$h_j^{(\gamma_j)}(\sigma_j | \mathcal{D}_t) \quad \leq \quad K\, h_j(\sigma_j | \mathcal{D}_t),$$

and we have an integrable upper bound also in this case. This finishes the proof of Lemma A.2. $\qquad\square$

Proof of Theorem 3.2. Due to Lemma A.1 it is sufficient to study the limit

$$\lim_{\gamma \to 1} \mathbb{E}\left[\operatorname{Var}\left(C_{i,J} | \mathcal{D}_t, \boldsymbol{\sigma} \right) | \mathcal{D}_t \right].$$

The inner conditional variance $\operatorname{Var}\left(C_{i,J} | \mathcal{D}_t, \boldsymbol{\sigma} \right)$ exactly corresponds to the situation of known standard deviation parameters $\boldsymbol{\sigma}$. We can therefore directly apply Theorem 2.12 of [8] to obtain this conditional variance. Assumption (3.4) implies that this conditional variance is finite, $\mathbb{P}$-a.s., in $\boldsymbol{\sigma}$; and posterior independence of parameters, see (A.5), implies that we can decouple the corresponding (outer) conditional expectations. This provides the following identity

$$
\begin{aligned}
\mathbb{E}\left[\operatorname{Var}\left(C_{i,J} | \mathcal{D}_t, \boldsymbol{\sigma} \right) | \mathcal{D}_t \right] \;=\; & C_{i,t-i} \sum_{j=t-i}^{J-1} \left\{ \prod_{m=t-i}^{j-1} \mathbb{E}\left[\widehat{f}_m^{\sigma(t)} \Big| \mathcal{D}_t \right] \mathbb{E}\left[\sigma_j^2 \left(\widehat{f}_j^{\sigma(t)} \right)^2 \left(1 + \Phi_j^{\gamma_j(t)} \right) \Big| \mathcal{D}_t \right] \right. \\
& \left. \times \prod_{n=j+1}^{J-1} \mathbb{E}\left[\left(\widehat{f}_n^{\sigma(t)} \right)^2 \left(1 + \Phi_n^{\gamma_n(t)} \right) \Big| \mathcal{D}_t \right] \right\} \\
& + C_{i,t-i}^2 \left(\prod_{j=t-i}^{J-1} \mathbb{E}\left[\left(\widehat{f}_j^{\sigma(t)} \right)^2 \left(1 + \Phi_j^{\gamma_j(t)} \right) \Big| \mathcal{D}_t \right] - \prod_{j=t-i}^{J-1} \mathbb{E}\left[\left(\widehat{f}_j^{\sigma(t)} \right)^2 \Big| \mathcal{D}_t \right] \right),
\end{aligned}
$$

with

$$\Phi_j^{\gamma_j(t)} \;=\; \frac{\sigma_j^2}{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j} + \sigma_j^2(\gamma_j - 2)},$$

and conditional BCL factor

$$\widehat{f}_j^{\sigma(t)} = \omega_j^{(t)}\, \widehat{f}_j^{CL(t)} + (1 - \omega_j^{(t)})f_j. \tag{A.6}$$

Since $\omega_j^{(t)}$ is sandwiched, $\mathbb{P}$-a.s., between a $\mathcal{D}_{t-1}$-measurable lower bound $\underline{\omega}_j^{(t)}$ that converges to 1 as $\gamma_j \to 1$ and 1, see (A.2), we obtain

$$\lim_{\gamma \to 1} \mathbb{E}\left[\widehat{f}_j^{\sigma(t)} \Big| \mathcal{D}_t \right] = \widehat{f}_j^{CL(t)} \qquad \text{and} \qquad \lim_{\gamma \to 1} \mathbb{E}\left[\left(\widehat{f}_j^{\sigma(t)} \right)^2 \Big| \mathcal{D}_t \right] = \left(\widehat{f}_j^{CL(t)} \right)^2,$$

and moreover

$$
\begin{aligned}
\lim_{\gamma \to 1} \mathbb{E}\left[\sigma_j^2 \left(\widehat{f}_j^{\sigma(t)} \right)^2 \left(1 + \Phi_j^{\gamma_j(t)} \right) \Big| \mathcal{D}_t \right] \;&=\; \left(\widehat{f}_j^{CL(t)} \right)^2 \lim_{\gamma \to 1} \mathbb{E}\left[\sigma_j^2 \left(1 + \Phi_j^{\gamma_j(t)} \right) \Big| \mathcal{D}_t \right], \\
\lim_{\gamma \to 1} \mathbb{E}\left[\left(\widehat{f}_j^{\sigma(t)} \right)^2 \left(1 + \Phi_j^{\gamma_j(t)} \right) \Big| \mathcal{D}_t \right] \;&=\; \left(\widehat{f}_j^{CL(t)} \right)^2 \lim_{\gamma \to 1} \mathbb{E}\left[1 + \Phi_j^{\gamma_j(t)} \Big| \mathcal{D}_t \right].
\end{aligned}
$$

Observe that we have point-wise (in $\boldsymbol{\sigma}$) convergence, $\mathbb{P}$-a.s.,

$$\lim_{\gamma_j \to 1} \Phi_j^{\gamma_j(t)} \;=\; \frac{\sigma_j^2}{\sum_{\ell=1}^{(t-j-1)\wedge I} C_{\ell,j} - \sigma_j^2} \;=\; \Psi_j^{(t)}.$$

Since for any $\gamma_j \in [1, 2]$ the random variables $\Phi_j^{\gamma_j(t)}$ are uniformly bounded in σ_j on the interval $(0, d_j)$, the claim of the theorem follows completely analogous to the one of Lemma A.2. The claim for the aggregated claim is also proved completely analogously. $\qquad\square$

Proof of Lemma 3.3. Observe that $(\mathbb{E}\left[C_{i,J}|\mathcal{D}_t\right])_{t=i,\dots,i+J}$ is a $\mathcal{D}_t$-martingale with $\mathbb{E}\left[C_{i,J}|\mathcal{D}_{i+J}\right]=C_{i,J}$. Using that the innovation process $(\mathbb{E}\left[C_{i,J}|\mathcal{D}_{t+1}\right]-\mathbb{E}\left[C_{i,J}|\mathcal{D}_t\right])_t=(-\mathrm{CDR}_i^{(t+1)})_t$ of a martingale has uncorrelated components we receive for $t\geq I\geq i>t-J$

$$
\begin{aligned}
\mathrm{msep}_{C_{i,J}|\mathcal{D}_t}\left(\mathbb{E}\left[C_{i,J}|\mathcal{D}_t\right]\right) &= \mathrm{Var}\left(C_{i,J}|\mathcal{D}_t\right) = \mathrm{Var}\left(\mathbb{E}\left[C_{i,J}|\mathcal{D}_{i+J}\right]|\mathcal{D}_t\right)\\
&= \mathrm{Var}\left(\mathbb{E}\left[C_{i,J}|\mathcal{D}_{t+1}\right]-(\mathbb{E}\left[C_{i,J}|\mathcal{D}_{t+1}\right]-\mathbb{E}\left[C_{i,J}|\mathcal{D}_{i+J}\right])|\mathcal{D}_t\right)\\
&= \mathrm{Var}\left(\mathbb{E}\left[C_{i,J}|\mathcal{D}_{t+1}\right]|\mathcal{D}_t\right)+\mathrm{Var}\left(\mathbb{E}\left[C_{i,J}|\mathcal{D}_{t+1}\right]-\mathbb{E}\left[C_{i,J}|\mathcal{D}_{i+J}\right]|\mathcal{D}_t\right)\\
&\geq \mathrm{Var}\left(\mathbb{E}\left[C_{i,J}|\mathcal{D}_{t+1}\right]|\mathcal{D}_t\right) = \mathrm{msep}_{\mathrm{CDR}_i^{(t+1)}|\mathcal{D}_t}(0).
\end{aligned}
$$

This proves the claim. $\qquad\square$

Lemma A.3. *We have the following identity for $t\geq I\geq t-j\geq 1$ and $0\leq j\leq J-1$*

$$
\left(1+\Psi_j^{(t)}\right)\left(\frac{\sigma_j^2}{C_{t-j,j}}+1\right)-1=\left(\alpha_j^{(t)}\right)^{-1}\Psi_j^{(t)}.
$$

Proof of Lemma A.3. We calculate the left-hand side

$$
\begin{aligned}
\left(1+\Psi_j^{(t)}\right)\left(\frac{\sigma_j^2}{C_{t-j,j}}+1\right)-1 &= \left(1+\frac{\sigma_j^2}{\sum_{\ell=1}^{(t-j-1)\wedge I}C_{\ell,j}-\sigma_j^2}\right)\left(\frac{\sigma_j^2}{C_{t-j,j}}+1\right)-1\\
&= \frac{\sigma_j^2}{C_{t-j,j}}+\frac{\sigma_j^2}{\sum_{\ell=1}^{(t-j-1)\wedge I}C_{\ell,j}-\sigma_j^2}+\frac{\sigma_j^2}{C_{t-j,j}}\frac{\sigma_j^2}{\sum_{\ell=1}^{(t-j-1)\wedge I}C_{\ell,j}-\sigma_j^2}\\
&= \sigma_j^2\left(\frac{\sum_{\ell=1}^{(t-j-1)\wedge I}C_{\ell,j}-\sigma_j^2+C_{t-j,j}+\sigma_j^2}{C_{t-j,j}\left(\sum_{\ell=1}^{(t-j-1)\wedge I}C_{\ell,j}-\sigma_j^2\right)}\right)\\
&= \sigma_j^2\frac{(\alpha_j^{(t)})^{-1}}{\sum_{\ell=1}^{(t-j-1)\wedge I}C_{\ell,j}-\sigma_j^2} = (\alpha_j^{(t)})^{-1}\Psi_j^{(t)}.
\end{aligned}
$$

This proves the claim of the lemma. $\qquad\square$

Proof Theorem 3.4. We have

$$
\mathrm{Var}\left(\mathbb{E}\left[C_{i,J}|\mathcal{D}_{t+1}\right]|\mathcal{D}_t\right) = \mathbb{E}\left[\mathbb{E}\left[C_{i,J}|\mathcal{D}_{t+1}\right]^2\Big|\mathcal{D}_t\right]-\mathbb{E}\left[C_{i,J}|\mathcal{D}_t\right]^2.
$$

Theorem 2.2 implies that the last term converges to $(\widehat{C}_{i,J}^{CL(t)})^2$, so there remains the consideration of the first term in the identity above. Recall the definition of the conditional BCL factor (A.6). Note that we have for $t\geq I\geq t-j\geq 1$ and $J-1\geq j\geq 0$

$$
\begin{aligned}
\widehat{f}_j^{\sigma(t+1)} &= \omega_j^{(t+1)}\,\widehat{f}_j^{CL(t+1)}+(1-\omega_j^{(t+1)})f_j\\
&= \frac{\sum_{\ell=1}^{(t-j)\wedge I}C_{\ell,j}}{\sum_{\ell=1}^{(t-j)\wedge I}C_{\ell,j}+\sigma_j^2\,(\gamma_j-1)}\frac{\sum_{\ell=1}^{(t-j)\wedge I}C_{\ell,j+1}}{\sum_{\ell=1}^{(t-j)\wedge I}C_{\ell,j}}+\left(1-\frac{\sum_{\ell=1}^{(t-j)\wedge I}C_{\ell,j}}{\sum_{\ell=1}^{(t-j)\wedge I}C_{\ell,j}+\sigma_j^2\,(\gamma_j-1)}\right)f_j\\
&= \nu_j^{(t)}\frac{C_{t-j,j+1}}{C_{t-j,j}}+\left(1-\nu_j^{(t)}\right)\left[\omega_j^{(t)}\,\widehat{f}_j^{CL(t)}+(1-\omega_j^{(t)})f_j\right] = \nu_j^{(t)}\frac{C_{t-j,j+1}}{C_{t-j,j}}+\left(1-\nu_j^{(t)}\right)\widehat{f}_j^{\sigma(t)},
\end{aligned}
$$

where we have defined

$$
\nu_j^{(t)}=\frac{C_{t-j,j}}{\sum_{\ell=1}^{(t-j)\wedge I}C_{\ell,j}+\sigma_j^2\,(\gamma_j-1)}\in(0,1).
$$

Using independence between different development periods j we obtain for $i+J\geq t+1$

$$
\mathbb{E}\left[C_{i,J}|\mathcal{D}_{t+1}\right]=C_{i,t-i+1}\prod_{j=t-i+1}^{J-1}\left(\mathbb{E}\left[\nu_j^{(t)}\Big|\mathcal{D}_{t+1}\right]\frac{C_{t-j,j+1}}{C_{t-j,j}}+\mathbb{E}\left[\left(1-\nu_j^{(t)}\right)\widehat{f}_j^{\sigma(t)}\Big|\mathcal{D}_{t+1}\right]\right),
$$

where these conditional expectations are taken w.r.t. σ_j, given $\mathcal{D}_{t+1}$. The above formula is analogous to the one found in the proof of Theorem 2.15 in [8]. We need to calculate the second moment of this expression. We are going to sandwich this random variable and then consider upper and lower bounds, respectively. First, note that we have $\mathbb{P}$-a.s., see (A.2),

$$\underline{\widehat{f}}_j^{(t)} \stackrel{\text{def.}}{=} \underline{\omega}_j^{(t)} \widehat{f}_j^{CL(t)} \;\;\leq\;\; \widehat{f}_j^{\sigma(t)} \;\;\leq\;\; \widehat{f}_j^{CL(t)} + \left(1 - \underline{\omega}_j^{(t)}\right) f_j \stackrel{\text{def.}}{=} \overline{\widehat{f}}_j^{(t)}.$$

Observe that $\underline{\widehat{f}}_j^{(t)}$ and $\overline{\widehat{f}}_j^{(t)}$ are $\mathcal{D}_t$-measurable. Moreover, we obtain $\mathcal{D}_t$-measurable bounds, $\mathbb{P}$-a.s.,

$$\underline{\nu}_j^{(t)} \stackrel{\text{def.}}{=} \frac{C_{t-j,j}}{\sum_{\ell=1}^{t-j} C_{\ell,j} + d_j^2(\gamma_j - 1)} \leq \nu_j^{(t)} \leq \frac{C_{t-j,j}}{\sum_{\ell=1}^{t-j} C_{\ell,j}} \;=\; \alpha_j^{(t)}.$$

Putting everything together we have (positive) lower and upper bounds, $\mathbb{P}$-a.s.,

$$C_{i,t-i+1} \prod_{j=t-i+1}^{J-1} \left(\underline{\nu}_j^{(t)} \frac{C_{t-j,j+1}}{C_{t-j,j}} + \left(1 - \alpha_j^{(t)}\right) \underline{\widehat{f}}_j^{(t)} \right)$$

$$\leq \;\; \mathbb{E}\left[C_{i,J} | \mathcal{D}_{t+1}\right] \;\; \leq \;\; C_{i,t-i+1} \prod_{j=t-i+1}^{J-1} \left(\alpha_j^{(t)} \frac{C_{t-j,j+1}}{C_{t-j,j}} + \left(1 - \underline{\nu}_j^{(t)}\right) \overline{\widehat{f}}_j^{(t)} \right).$$

Conditionally given $\mathcal{D}_t$, the only random terms in the lower and upper bounds are $C_{i,t-i+1}, \ldots, C_{t-J+1,J}$. These random terms all belong to different accident years and different development periods. Using conditional independence between different accident years, given $\boldsymbol{\Theta}$ and $\boldsymbol{\sigma}$, as well as posterior independence of parameters in different development periods j implies that we can exchange the conditional expectation and the product, which provides the following bounds

$$\mathbb{E}\left[C_{i,t-i+1}^2 \big| \mathcal{D}_t \right] \prod_{j=t-i+1}^{J-1} \mathbb{E}\left[\left(\underline{\nu}_j^{(t)} \frac{C_{t-j,j+1}}{C_{t-j,j}} + \left(1 - \alpha_j^{(t)}\right) \underline{\widehat{f}}_j^{(t)} \right)^2 \bigg| \mathcal{D}_t \right]$$

$$\leq \;\; \mathbb{E}\left[\mathbb{E}\left[C_{i,J} | \mathcal{D}_{t+1}\right]^2 \big| \mathcal{D}_t \right] \;\; \leq \;\; \mathbb{E}\left[C_{i,t-i+1}^2 \big| \mathcal{D}_t \right] \prod_{j=t-i+1}^{J-1} \mathbb{E}\left[\left(\alpha_j^{(t)} \frac{C_{t-j,j+1}}{C_{t-j,j}} + \left(1 - \underline{\nu}_j^{(t)}\right) \overline{\widehat{f}}_j^{(t)} \right)^2 \bigg| \mathcal{D}_t \right].$$

There remains the calculation of these terms. We start with the conditional variance of $C_{i,t-i+1}$, given $\mathcal{D}_t$,

$$\begin{aligned}
\text{Var}\left(C_{i,t-i+1} | \mathcal{D}_t\right) &= \mathbb{E}\left[\text{Var}\left(C_{i,t-i+1} | \mathcal{D}_t, \boldsymbol{\Theta}, \boldsymbol{\sigma}\right) | \mathcal{D}_t \right] + \text{Var}\left(\mathbb{E}\left[C_{i,t-i+1} | \mathcal{D}_t, \boldsymbol{\Theta}, \boldsymbol{\sigma}\right] | \mathcal{D}_t \right) \\
&= \mathbb{E}\left[\Theta_{t-i}^{-2} \sigma_{t-i}^2 C_{i,t-i} \big| \mathcal{D}_t \right] + \text{Var}\left(\Theta_{t-i}^{-1} C_{i,t-i} \big| \mathcal{D}_t \right) \\
&= \mathbb{E}\left[\Theta_{t-i}^{-2} \sigma_{t-i}^2 \big| \mathcal{D}_t \right] C_{i,t-i} + \mathbb{E}\left[\Theta_{t-i}^{-2} \big| \mathcal{D}_t \right] C_{i,t-i}^2 - \left(\mathbb{E}\left[\Theta_{t-i}^{-1} \big| \mathcal{D}_t \right] C_{i,t-i} \right)^2 \\
&= \mathbb{E}\left[\mathbb{E}\left[\Theta_{t-i}^{-2} \big| \mathcal{D}_t, \boldsymbol{\sigma} \right] \left(\sigma_{t-i}^2 C_{i,t-i} + C_{i,t-i}^2 \right) \big| \mathcal{D}_t \right] - \mathbb{E}\left[C_{i,t-i+1} | \mathcal{D}_t \right]^2 \\
&= \mathbb{E}\left[\left(\widehat{f}_{t-i}^{\sigma(t)} \right)^2 \left(1 + \Phi_{t-i}^{\gamma_{t-i}(t)} \right) \left(\sigma_{t-i}^2 C_{i,t-i} + C_{i,t-i}^2 \right) \bigg| \mathcal{D}_t \right] - \mathbb{E}\left[C_{i,t-i+1} | \mathcal{D}_t \right]^2,
\end{aligned}$$

this was also obtained in the empirical Bayesian model in Lemma C.1 of [8]. Using the definition of $\widehat{\mathbb{E}}[\cdot|\mathcal{D}_t]$ and the derivations in the proof of Theorem 3.2 in the second step as well as Lemma A.3 in the third step, we obtain

$$\begin{aligned}
\lim_{\gamma \to 1} \mathbb{E}\left[C_{i,t-i+1}^2 \big| \mathcal{D}_t \right] &= \lim_{\gamma \to 1} \mathbb{E}\left[\left(\widehat{f}_{t-i}^{\sigma(t)} \right)^2 \left(1 + \Phi_{t-i}^{\gamma_{t-i}(t)} \right) \left(\sigma_{t-i}^2 C_{i,t-i} + C_{i,t-i}^2 \right) \bigg| \mathcal{D}_t \right] \\
&= \left(\widehat{C}_{i,t-i+1}^{CL(t)} \right)^2 \widehat{\mathbb{E}}\left[\left(1 + \Psi_{t-i}^{(t)} \right) \left(\frac{\sigma_{t-i}^2}{C_{i,t-i}} + 1 \right) \bigg| \mathcal{D}_t \right] \\
&= \left(\widehat{C}_{i,t-i+1}^{CL(t)} \right)^2 \left(1 + \left(\alpha_{t-i}^{(t)} \right)^{-1} \widehat{\mathbb{E}}\left[\Psi_{t-i}^{(t)} \big| \mathcal{D}_t \right] \right).
\end{aligned}$$

18

Finally, we need to consider the terms under the product. We have for $j = t - i + 1, \ldots, J - 1$

$$
\mathbb{E}\left[\left(\underline{\nu}_j^{(t)} \, \frac{C_{t-j,j+1}}{C_{t-j,j}} + \left(1 - \alpha_j^{(t)}\right) \underline{\widehat{f}}_j^{(t)} \right)^2 \middle| \mathcal{D}_t \right]
$$

$$
= \; \mathrm{Var}\left(\underline{\nu}_j^{(t)} \, \frac{C_{t-j,j+1}}{C_{t-j,j}} + \left(1 - \alpha_j^{(t)}\right) \underline{\widehat{f}}_j^{(t)} \middle| \mathcal{D}_t \right) + \mathbb{E}\left[\underline{\nu}_j^{(t)} \, \frac{C_{t-j,j+1}}{C_{t-j,j}} + \left(1 - \alpha_j^{(t)}\right) \underline{\widehat{f}}_j^{(t)} \middle| \mathcal{D}_t \right]^2
$$

$$
= \; \left(\frac{\underline{\nu}_j^{(t)}}{C_{t-j,j}} \right)^2 \mathrm{Var}\left(C_{t-j,j+1} \middle| \mathcal{D}_t \right) + \left(\underline{\nu}_j^{(t)} \, \frac{\mathbb{E}\left[C_{t-j,j+1} \middle| \mathcal{D}_t \right]}{C_{t-j,j}} + \left(1 - \alpha_j^{(t)}\right) \underline{\widehat{f}}_j^{(t)} \right)^2 .
$$

Moreover, we have

$$
\lim_{\gamma \to 1} \underline{\nu}_j^{(t)} = \alpha_j^{(t)} \qquad \text{and} \qquad \widehat{f}_j^{CL(t)} = \lim_{\gamma \to 1} \underline{\widehat{f}}_j^{(t)} \leq \lim_{\gamma \to 1} \frac{\mathbb{E}\left[C_{t-j,j+1} \middle| \mathcal{D}_t \right]}{C_{t-j,j}} \leq \lim_{\gamma \to 1} \overline{\widehat{f}}_j^{(t)} = \widehat{f}_j^{CL(t)} .
$$

This implies

$$
\lim_{\gamma \to 1} \mathbb{E}\left[\left(\underline{\nu}_j^{(t)} \, \frac{C_{t-j,j+1}}{C_{t-j,j}} + \left(1 - \alpha_j^{(t)}\right) \underline{\widehat{f}}_j^{(t)} \right)^2 \middle| \mathcal{D}_t \right] = \left(\alpha_j^{(t)} \widehat{\mathbb{E}}\left[\Psi_j^{(t)} \middle| \mathcal{D}_t \right] + 1 \right) \left(\widehat{f}_j^{CL(t)} \right)^2 ,
$$

and completely analogously for the upper bound. The claim for the aggregated claim follows also completely analogously. This completes the proof of the theorem. □

References

[1] England, P.D., Verrall, R.J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal* **8/3**, 443-518.

[2] European Commission (2009). Framework Solvency II Directive (Directive 2009/138/EC).

[3] Gisler, A. (2006). The estimation error in the chain-ladder reserving method: a Bayesian approach. *ASTIN Bulletin* **36/2**, 554-565.

[4] Gisler, A., Wüthrich, M.V. (2008). Credibility for the chain ladder reserving method. *ASTIN Bulletin* **38/2**, 565-600.

[5] Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin* **23/2**, 213-225.

[6] Renshaw, A.E., Verrall, R.J. (1998). A stochastic model underlying the chain-ladder technique. *British Actuarial Journal* **4/4**, 903-923.

[7] Swiss Financial Market Supervisory Authority FINMA (2006). Swiss Solvency Test, Technisches Dokument, Version 2. October 2006.

[8] Wüthrich, M.V., Merz, M. (2015). Stochastic claims reserving manual: advances in dynamic modeling. *SSRN manuscript*, version of August 21, 2015, ID 2649057.

B   Data

accident year i	development year j									
	0	1	2	3	4	5	6	7	8	9
1	5'946.975	9'668.212	10'563.929	10'771.690	10'978.394	11'040.518	11'106.331	11'121.181	11'132.310	11'148.124
2	6'346.756	9'593.162	10'316.383	10'468.180	10'536.004	10'572.608	10'625.360	10'636.546	10'648.192	
3	6'269.090	9'245.313	10'092.366	10'355.134	10'507.837	10'573.282	10'626.827	10'635.751		
4	5'863.015	8'546.239	9'268.771	9'459.424	9'592.399	9'680.740	9'724.068			
5	5'778.885	8'524.114	9'178.009	9'451.404	9'681.692	9'786.916				
6	6'184.793	9'013.132	9'585.897	9'830.796	9'935.753					
7	5'600.184	8'493.391	9'056.505	9'282.022						
8	5'288.066	7'728.169	8'256.211							
9	5'290.793	7'648.729								
10	5'675.568									
$\widehat{f}_j^{CL(I)}$	1.4925	1.0778	1.0229	1.0148	1.0070	1.0051	1.0011	1.0010	1.0014	
$\widehat{s}_j$	4.2771	1.0689	0.4984	0.6276	0.2952	0.0633	0.0260	0.0069	0.0018	
$\widehat{\sigma}_j$	2.8656	0.9918	0.4872	0.6184	0.2932	0.0630	0.0260	0.0069	0.0018	

Table 2: Observed cumulative payments $C_{i,j}$ in $\mathcal{D}_I$ with $I = 10$ and $J = 9$, CL factor estimates $\widehat{f}_j^{CL(I)}$ and standard deviation parameter estimates $\widehat{s}_j$ and $\widehat{\sigma}_j$.