Verrall, R. J. & Wüthrich, M. V. (2015). Parameter Reduction in Log-normal Chain-ladder Models. European Actuarial Journal, 5(2), pp. 355-380. doi: 10.1007/s13385-015-0114-7

European Actuarial Journal

Parameter Reduction in Log-Normal Chain-Ladder Models

--Manuscript Draft--

Manuscript Number:	EUAJ-D-13-00036R2
Full Title:	Parameter Reduction in Log-Normal Chain-Ladder Models
Article Type:	Original Research Paper
Keywords:	claims reserving; loss reserving; CL method; CL factors; non-life insurance; risk; general insurance; parameter reduction; tail factor
Corresponding Author:	Richard Verrall UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	
Corresponding Author's Secondary Institution:	
First Author:	Richard Verrall
First Author Secondary Information:	
Order of Authors:	Richard Verrall
	Mario Wüthrich
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Multiplicative chain-ladder (CL) models are characterized by CL factors that explain the development of claims from one period to the next. In classical CL models every development period has its own CL factor. In the present paper we give a method describing how some of these CL factors can be modeled by a joint functional dependence. This joint functional form reduces the number of model parameters needed.
Response to Reviewers:	We have implemented all proposed changes, in particular, we have changed the interpretation that was not fully appropriate.

Parameter Reduction in Log-Normal Chain-Ladder Models

Richard J. Verrall* Mario V. Wüthrich†‡

July 24, 2015

Abstract

Multiplicative chain-ladder (CL) models are characterized by CL factors that explain the development of claims from one period to the next. In classical CL models every development period has its own CL factor. In the present paper we give a method describing how some of these CL factors can be modeled by a joint functional dependence. This joint functional form reduces the number of model parameters needed.

*Cass Business School, City University London, 106 Bunhill Row, London EC1Y 8T2, UK
†ETH Zurich, RiskLab Switzerland, Department of Mathematics, 8092 Zurich, Switzerland
‡Swiss Finance Institute Professor

Parameter Reduction in Log-Normal Chain-Ladder Models

July 24, 2015

Abstract

Multiplicative chain-ladder (CL) models are characterized by CL factors that explain the development of claims from one period to the next. In classical CL models every development period has its own CL factor. In the present paper we give a method describing how some of these CL factors can be modeled by a joint functional dependence. This joint functional form reduces the number of model parameters needed.

1 Introduction

Verrall-Wüthrich [14] considered the practical issue of parameter reduction with applications to claims reserving models. The issue of over-parametrization is often raised in relation to the chain-ladder (CL) technique because there is an extra parameter (CL factor) for each development period. This also means that it is not possible to extrapolate individual CL factors to create "tail factors" without making further assumptions. In this paper and in Verrall-Wüthrich [14] we tackle this problem by formulating what is often done practically in an ad hoc manner into a statistical method of model selection. The drawback of Verrall-Wüthrich [14] was that it was necessary to solve the problem using numerical methods since closed-form solutions do not exist for the general case. For this reason, reversible jump Markov chain Monte Carlo (RJMCMC) methods were used. These methods are not straightforward to implement and they can be unstable and time-consuming. In this paper, we restrict attention to the log-normal distribution with conjugate priors. This leads to elegant analytical results that allow to do model selection in a direct manner.

An important feature of the approach, both in this paper and in Verrall-Wüthrich [14], is that a functional dependence allows for extrapolation of the claims development beyond the latest observed delay period in the data, and for creation of tail factors in a very natural way. Thus, these two papers show how the important issue of parameter reduction in claims reserving can be addressed. The choice is between the more general distributional assumptions and more complex implementation in Verrall-Wüthrich [14] or the more restricted log-normal case here with closed form solutions.

The paper is set out as follows: in the next section we define the Bayesian log-normal CL model and we provide first properties. In Section 3 we discuss parameter estimation, followed by Section 4 where we describe model selection techniques. In Section 5 we explain claims

prediction, uncertainty analysis and extrapolation for tail factor estimation. Finally, in Section 6 we revisit the liability example of Verrall-Wüthrich [14].

2 Bayesian log-normal CL model with conjugate priors

2.1 Model assumptions

To study the problem of parameter reduction in CL models we embed Hertig's [7] log-normal CL model into a Bayesian modeling framework. Based on this Bayesian approach we aim to determine the set of CL factors that can be explained by a common functional dependence. For this purpose we introduce a *truncation index* $k \in \{1, \ldots, J-1\}$. Before the truncation index each CL factor is characterized by an individual development parameter and beyond the truncation index all CL factors are described by a functional dependence of common parameters. This way we obtain a *whole family* of models $\mathcal{M}^{(k)}$, $k \in \{1, \ldots, J-1\}$, each model $\mathcal{M}^{(k)}$ having a different number of model parameters. Model selection will then be done on this family of models. The models are set out formally in Model Assumptions 2.1, below.

We introduce the following notation, for detailed background information on claims reserving we refer to Chapter 9 in Wüthrich [16]. We denote accident years by $i \in \{1, \ldots, I\}$ and development years by $j \in \{0, \ldots, J\}$. I is the last accident year considered and J is the maximal possible development delay. Throughout we assume $I > J$. The cumulative claim of accident year i after development year j is denoted by $C_{i,j}$, and $C_{i,J}$ is called ultimate claim or total claim amount of accident year i. At time I we have observations

$$\mathcal{D}_I = \{C_{i,j};\ i+j \leq I,\ 1 \leq i \leq I,\ 0 \leq j \leq J\},$$

and our aim is to predict the inexperienced part of the claims given by

$$\mathcal{D}_I^c = \{C_{i,j};\ i+j > I,\ 1 \leq i \leq I,\ 0 \leq j \leq J\}.$$

In order to achieve this task we make the following model assumptions for the claims $(C_{i,j})_{i,j}$.

Model Assumptions 2.1 (Bayesian log-normal CL models) *Choose a fixed truncation index $k \in \{1, \ldots, J-1\}$. Model $\mathcal{M}^{(k)}$ is given by the following assumptions. There are given standard deviation parameters $\boldsymbol{\sigma} = (\sigma_0, \ldots, \sigma_J)' \in \mathbb{R}_+^{J+1}$.*

- *Given parameter $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_J)'$, the sequences $(C_{i,j})_{j=0,\ldots,J}$ are independent (in i) Markov processes (in j) with log-link ratios*

$$\xi_{i,j} \ = \ \log\left(\frac{C_{i,j}}{C_{i,j-1}} - 1\right)\Big|_{\{\boldsymbol{\theta}, C_{i,j-1}\}} \ \sim \ \mathcal{N}(\theta_j, \sigma_j^2) \qquad \text{for } j = 0, \ldots, J,$$

  where we set $C_{i,-1} = \nu_i$ for $i \in \{1, \ldots, I\}$ with given constants $\nu_i > 0$.

- *Assume that the parameter $\boldsymbol{\Theta}^{(k)} = (\theta_0, \ldots, \theta_{k-1}, \alpha, \beta)'$ has a multivariate Gaussian distribution*

$$\boldsymbol{\Theta}^{(k)} = (\theta_0, \ldots, \theta_{k-1}, \alpha, \beta)' \ \sim \ \mathcal{N}\left(\boldsymbol{\mu}^{(k)}, T^{(k)}\right),$$

with given prior mean $\boldsymbol{\mu}^{(k)} = (\mu_0, \ldots, \mu_{k-1}, \mu_\alpha, \mu_\beta)' \in \mathbb{R}^{k+2}$ *and positive definite prior covariance matrix* $T^{(k)} = \operatorname{diag}(\tau_0^2, \ldots, \tau_{k-1}^2, \tau_\alpha^2, \tau_\beta^2) \in \mathbb{R}^{(k+2)\times(k+2)}$. *For* $j \in \{k, \ldots, J\}$ *we set*

$$\theta_j = \alpha - j\beta.$$

□

The interpretation of Model Assumptions 2.1 will be divided into three parts, see Remarks 2.2, 2.4 and 2.6 below.

Remarks 2.2 (to Model Assumptions 2.1, part 1/3)

- Cumulative claims satisfy the multiplicative structure

$$C_{i,j} = C_{i,j-1} \left(\exp\{\xi_{i,j}\} + 1 \right). \tag{2.1}$$

  Thus, we have a multiplicative CL structure described by a shifted log-normal distribution, and the excess claim $C_{i,j} - C_{i,j-1}$ has a multiplicative random structure described by a log-normal distribution. This structure is in particular appealing for inflation modeling on payments, see Shi et al. [12] and Wüthrich [15].

  Property (2.1) might be criticized because it requires non-negative excess claims. If this is an undesired model property one could also study the model $C_{i,j} = C_{i,j-1} \exp\{\xi_{i,j}\}$. The mathematical techniques would be exactly the same because we will only work on the log-link ratios $\xi_{i,j}$, see Lemma 2.5 below, however the modeling of the tail behavior would become more sophisticated. Therefore, we refrain from considering the latter model in this work.

- For fixed truncation index k parameter $\boldsymbol{\theta}$ in model $\mathcal{M}^{(k)}$ takes the following form

$$\boldsymbol{\theta} = (\theta_0, \ldots, \theta_{k-1}, \alpha - k\beta, \ldots, \alpha - J\beta)'. \tag{2.2}$$

  The first k components of $\boldsymbol{\theta}$ are modeled by individual parameters θ_j for $j < k$ and the remaining components are characterized by the two common parameters α and β and a linear functional dependence, that is, $\theta_{j+1} = \theta_j - \beta$ for $j \geq k$. The aim will be to find the optimal truncation index k and the optimal model $\mathcal{M}^{(k)}$, respectively, for a given data set $\mathcal{D}_I$.

- The distribution $\boldsymbol{\Theta}^{(k)} \sim \mathcal{N}\left(\boldsymbol{\mu}^{(k)}, T^{(k)}\right)$ reflects the prior knowledge about parameters. This can come from expert opinion, from market information or from a regulatory viewpoint. If there is only little information available or if we have heterogeneous beliefs we choose a covariance matrix $T^{(k)}$ with big variances, which reflects heterogeneity and/or uncertainty.

- Note that we have assumed prior independence $T^{(k)} = \operatorname{diag}(\tau_0^2, \ldots, \tau_{k-1}^2, \tau_\alpha^2, \tau_\beta^2)$ between the components of parameter $\boldsymbol{\Theta}^{(k)}$. In view of the following derivations this seems to be

an unnecessary restriction because the whole theory holds true for any positive definite covariance matrix $T^{(k)}$. We do this choice to more clearly separate the effects coming from different model characteristics. In practical applications this choice should be revised since dependence in $T^{(k)}$ may also help to incorporate shape constraints in tail factors.

- If there are known differences $\nu_i > 0$ between the accident years $i \in \{1, \ldots, I\}$ we can implement these differences by initializing $C_{i,-1} = \nu_i$. If there is no prior information available about these differences we set $\nu_i = 1$ for all i. These choices will not influence the prediction of $\mathcal{D}_I^c$, given the observations $\mathcal{D}_I$ and the parameters $\boldsymbol{\Theta}^{(k)}$, under our independence assumptions, because $C_{i,0} \in \mathcal{D}_I$ for all $i \leq I$. This is demonstrated in the next corollary. However, these differences will become important if we choose more general correlation structures in $T^{(k)}$ and between the components of $(\xi_{i,j})_{i,j}$, see (2.5) below, for the latter we also refer to Shi et al. [12] and Wüthrich [15].

An easy consequence of the model assumptions is the following corollary (the proof is completely similar to the one of Lemma 5.2 in Wüthrich-Merz [17]).

Corollary 2.3 *Choose model $\mathcal{M}^{(k)}$. Under Model Assumptions 2.1 we have for $i > I - J$*

$$\begin{aligned}\mathbb{E}\left[C_{i,J}\,\middle|\,\mathcal{D}_I, \boldsymbol{\Theta}^{(k)}\right] &= C_{i,I-i} \prod_{j=I-i+1}^{J} \left(\exp\left\{\theta_j + \sigma_j^2/2\right\} + 1\right) \\ &= C_{i,I-i} \prod_{j=I-i+1}^{k-1} \left(\exp\left\{\theta_j + \sigma_j^2/2\right\} + 1\right) \prod_{j=(I-i+1)\vee k}^{J} \left(\exp\left\{\alpha - j\beta + \sigma_j^2/2\right\} + 1\right),\end{aligned}$$

where $x \vee y = \max\{x, y\}$ and an empty product is defined to be equal to 1.

Remarks 2.4 (to Model Assumptions 2.1, part 2/3)

- Corollary 2.3 provides the multiplicative CL structure for given parameters $\boldsymbol{\Theta}^{(k)}$ in model $\mathcal{M}^{(k)}$ with CL factors defined by $f_j = (\exp\{\theta_j + \sigma_j^2/2\} + 1)$. It also shows that the choice of ν_i has no influence on the prediction under our independence assumptions.

- Before the truncation index k every CL factor is modeled individually by θ_j, $j < k$, after the truncation index k the CL factors are modeled by the two common parameters α and β using an exponential decay with rate β for $j \geq k$, that is,

$$f_j = \exp\left\{\alpha - j\beta + \sigma_j^2/2\right\} + 1,$$

see also (2.2). Thus, we use a curve fitting method by specifying an exponentially decaying function. From a purely theoretical point of view the fitted curve could have any other functional form and our theory would still work. Choice (2.2) has the advantage of simplicity and tractability whereas many other functional forms will in general require simulation based solutions similar to Verrall-Wüthrich [14]. The interested reader is referred to De Jong-Zehnwirth [4], Section 4 in England-Verrall [5], Verrall [13] and, in particular, Section 5 of Boor [2] for other functional forms. In our numerical example, the exponential decay seems quite reasonable, see Figure 5 below, an other possible choice is provided in (6.3) below.

2.2 Properties of Bayesian log-normal CL models

We introduce some notation that simplifies the outline. The cardinality of the set of indexes $\mathcal{I} = \{1, \ldots, I\} \times \{0, \ldots, J\}$ is denoted by $d = I(J+1)$ and we define the vector of log-link ratios

$$\boldsymbol{\xi} = (\xi_{i,j})'_{i=1,\ldots,I;j=0,\ldots,J} = (\xi_{1,0}, \ldots, \xi_{1,J}, \ldots, \xi_{I,0}, \ldots, \xi_{I,J})' \in \mathbb{R}^d.$$

The joint density in model $\mathcal{M}^{(k)}$ of $\boldsymbol{\xi}$ and parameters $\boldsymbol{\Theta}^{(k)}$ at position $(\boldsymbol{\xi}, \boldsymbol{\theta}^{(k)})$ is given by

$$f^{(k)}\left(\boldsymbol{\xi}, \boldsymbol{\theta}^{(k)}\right) = f^{(k)}\left(\boldsymbol{\xi} \,\middle|\, \boldsymbol{\theta}^{(k)}\right)\; p^{(k)}\left(\boldsymbol{\theta}^{(k)}\right), \tag{2.3}$$

where we have prior density

$$p^{(k)}\left(\boldsymbol{\theta}^{(k)}\right) = \frac{1}{(2\pi)^{k/2+1}\,\det(T^{(k)})^{1/2}}\; \exp\left\{-\frac{1}{2}\,(\boldsymbol{\theta}^{(k)} - \boldsymbol{\mu}^{(k)})'\,(T^{(k)})^{-1}\,(\boldsymbol{\theta}^{(k)} - \boldsymbol{\mu}^{(k)})\right\},$$

and likelihood function of the log-link ratios $\boldsymbol{\xi}$, given parameters $\boldsymbol{\theta}^{(k)}$,

$$f^{(k)}\left(\boldsymbol{\xi} \,\middle|\, \boldsymbol{\theta}^{(k)}\right) = \frac{1}{(2\pi)^{d/2}\,\det(\Sigma)^{1/2}}\; \exp\left\{-\frac{1}{2}\,(\boldsymbol{\xi} - A^{(k)}\boldsymbol{\theta}^{(k)})'\,\Sigma^{-1}\,(\boldsymbol{\xi} - A^{(k)}\boldsymbol{\theta}^{(k)})\right\}, \tag{2.4}$$

where we denote the diagonal covariance matrix of the log-link ratios by

$$\Sigma = \operatorname{diag}(\sigma_0^2, \ldots, \sigma_J^2, \ldots, \sigma_0^2, \ldots, \sigma_J^2) \in \mathbb{R}^{d\times d}, \tag{2.5}$$

and we define the matrix $A^{(k)} = (B'_k, \ldots, B'_k)' \in \mathbb{R}^{d\times(k+2)}$ such that matrix $B_k \in \mathbb{R}^{(J+1)\times(k+2)}$ describes the parameters for a single accident year i and is given by

$$B'_k = \left(\begin{array}{ccc|ccc} & & & 0 & \ldots & 0 \\ & \mathbb{1} & & \vdots & & \vdots \\ & & & 0 & \ldots & 0 \\ \hline 0 & \cdots & 0 & 1 & \ldots & 1 \\ 0 & \cdots & 0 & -k & \ldots & -J \end{array}\right),$$

with $\mathbb{1} \in \mathbb{R}^{k\times k}$ being the identity matrix. This choice implies, see also (2.2),

$$\boldsymbol{\theta} = B_k\boldsymbol{\Theta}^{(k)} \in \mathbb{R}^{J+1} \qquad \text{and} \qquad \mathbb{E}\left[\boldsymbol{\xi} \,\middle|\, \boldsymbol{\Theta}^{(k)}\right] = A^{(k)}\boldsymbol{\Theta}^{(k)} \in \mathbb{R}^d.$$

From this we see that $A^{(k)}$ allows the conditional expectation of the log-link ratios $\boldsymbol{\xi}$ to be expressed in terms of the parameters $\boldsymbol{\Theta}^{(k)}$ in model $\mathcal{M}^{(k)}$. An easy consequence of Model Assumptions 2.1 is the following lemma (we leave the proof to the reader).

Lemma 2.5 *Set Model Assumptions 2.1 for model $\mathcal{M}^{(k)}$. The joint density $f^{(k)}(\boldsymbol{\xi}, \boldsymbol{\theta}^{(k)})$ of $\boldsymbol{\xi}$ and $\boldsymbol{\Theta}^{(k)}$ describes a multivariate Gaussian distribution with*

$$\begin{pmatrix} \boldsymbol{\xi} \\ \boldsymbol{\Theta}^{(k)} \end{pmatrix} \;\sim\; \mathcal{N}\left(\begin{pmatrix} A^{(k)}\boldsymbol{\mu}^{(k)} \\ \boldsymbol{\mu}^{(k)} \end{pmatrix}, \begin{pmatrix} \Sigma + A^{(k)}T^{(k)}(A^{(k)})' & A^{(k)}T^{(k)} \\ T^{(k)}(A^{(k)})' & T^{(k)} \end{pmatrix}\right).$$

The random vector $\boldsymbol{\xi}$ has in model $\mathcal{M}^{(k)}$ a multivariate Gaussian distribution with mean $\boldsymbol{m}^{(k)} = A^{(k)}\boldsymbol{\mu}^{(k)}$ and covariance matrix $\Sigma^{(k)} = \Sigma + A^{(k)}T^{(k)}(A^{(k)})'$.

Remarks 2.6 (to Model Assumptions 2.1, part 3/3)

- We have assumed that all log-link ratios $\xi_{i,j}$ are conditionally independent. This independence assumption could be replaced by any multivariate Gaussian distribution and we would still get closed form solutions, see Merz et al. [11], Shi et al. [12] and Wüthrich [15]. At the current stage we refrain from doing so because we do not want to mix dependence with tail factor estimation, but it is worth analyzing this extension numerically in future work. Basically, it means that the covariance matrix Σ, defined in (2.5), needs to be replaced by any symmetric positive definite matrix and then the whole theory, as presented in this article, still runs through.

- One might criticize the Gaussian distribution assumption of the log-link ratios $\xi_{i,j}$ and the corresponding Gaussian priors $\boldsymbol{\Theta}^{(k)}$. Our model belongs to the Bayesian models with conjugate priors which have the advantage of staying in the same family of distributions for the posteriors, see Section 2.5 in Bühlmann-Gisler [3] and Section 8.1 in Wüthrich [16]. Thanks to the multivariate Gaussian assumptions we obtain closed form solutions as seen in Lemma 2.5 and Corollaries 3.1 and 3.2. Other distributional assumptions, in general, only allow for simulation results, similar to Verrall-Wüthrich [14], where (complicated) reversible jump Markov chain Monte Carlo (RJMCMC) simulations are used. The purpose of this paper is to present a model which has a closed form solution, this facilitates sensitivity analysis. The explicit choice of the distributional assumption will become especially important for the calculation of tail-sensitive risk measures such as Tail-Value-at-Risk. Our risk measure choice (5.1) is less tail-sensitive and, therefore, the log-normal model is usually sufficient.

- If we relax the distributional assumptions we may consider distributions from the exponential dispersion family with conjugate priors, see Lee-Nelder [9, 10], Gigante et al. [6], Section 2.5 in Bühlmann-Gisler [3] and Section 8.1 in Wüthrich [16]. In our situation this would provide the following distributional form

$$\begin{aligned}\log f^{(k)}_{\text{HGLM}}\left(\boldsymbol{\xi}, \boldsymbol{\theta}^{(k)}\right) &= \sum_{j=0}^{J}\sum_{i=1}^{I} \frac{w_{i,j}}{\phi}\left[\xi_{i,j}\theta_j - b(\theta_j)\right] + \log c(\xi_{i,j}, w_{i,j}/\phi) \\ &\quad + \sum_{j=0}^{k-1} \frac{1}{\psi_j}\left[\mu_j\theta_j - b(\theta_j)\right] + \sum_{a\in\{\alpha,\beta\}} \frac{1}{\psi_a}\left[\mu_a a - b(a)\right] + \text{const.}\end{aligned} \tag{2.6}$$

  The first term describes the log-likelihood function of the log-link ratios $\boldsymbol{\xi}$, given the parameters $\boldsymbol{\theta} = B_k\boldsymbol{\theta}^{(k)}$. The second term describes the parameters $\boldsymbol{\theta}^{(k)}$ that model the first k of the CL factors individually and the remaining CL factors are modeled by common parameters α and β (using the linear functional dependence). Our Gaussian model uses the generic choice $b(\theta) = \theta^2/2$. The general form (2.6) is less tractable than the Gaussian one because marginals do not have explicit forms. Formula (2.6) is very close to hierarchical generalized likelihood models (HGLMs) and calibration can also be done using maximum

likelihood estimation (MLE) methods, see Gigante et al. [6]. Since this framework is less tractable, we have decided to stay within Assumptions 2.1 for the present work.

3 Model calibration in a fixed model $\mathcal{M}^{(k)}$

We fix a truncation index k and a model $\mathcal{M}^{(k)}$ and assume that we have observations $\mathcal{D}_I$ with $I > J$. There are the parameters $\Sigma = \text{diag}(\sigma_0^2, \ldots, \sigma_J^2, \ldots, \sigma_0^2, \ldots, \sigma_J^2)$, $\boldsymbol{\mu}^{(k)}$ and $T^{(k)}$ that need to be specified. As explained in Remarks 2.2, the parameters $\boldsymbol{\mu}^{(k)}$ and $T^{(k)}$ correspond to prior knowledge or a market view and, therefore, cannot be calibrated from individual data $\mathcal{D}_I$. Thus, we either have this prior knowledge and then $\boldsymbol{\mu}^{(k)}$ and $T^{(k)}$ describe this information or there is no prior knowledge in which case we choose large variances for $T^{(k)}$ which make the influence of the prior distribution negligible (non-informative) for the prediction of $\mathcal{D}_I^c$.
There remains the calibration of Σ. This is described in Subsection 3.1 below. In view of the HGLM approach (2.6) we could also get to a slightly different model interpretation (more in the light of a frequentist's approach). This leads to another way of model calibration which we would briefly like to present in Subsection 3.2.

3.1 Model calibration using an empirical Bayesian approach

By an abuse of notation we set (note that the corresponding σ-fields generated by $C_{i,j}$ and $\xi_{i,j}$, $i + j \leq I$, are the same)

$$\mathcal{D}_I = \{\xi_{i,j};\ i + j \leq I,\ 1 \leq i \leq I,\ 0 \leq j \leq J\}\,.$$

Let $c = |\mathcal{D}_I| < d$ denote the cardinality of $\mathcal{D}_I$. Then, we define the projections $P_1 : \mathbb{R}^d \to \mathbb{R}^c$ and $P_2 : \mathbb{R}^d \to \mathbb{R}^{d-c}$ such that we obtain a bijective decomposition

$$\boldsymbol{\xi} \ \mapsto\ \left(\boldsymbol{\xi}_{\mathcal{D}_I}, \boldsymbol{\xi}_{\mathcal{D}_I^c}\right) = (P_1\boldsymbol{\xi}, P_2\boldsymbol{\xi})\,,$$

with $\boldsymbol{\xi}_{\mathcal{D}_I} = P_1\boldsymbol{\xi}$ containing exactly the components of $\boldsymbol{\xi}$ which are in $\mathcal{D}_I$, i.e. are observed at time I, and $\boldsymbol{\xi}_{\mathcal{D}_I^c} = P_2\boldsymbol{\xi}$ are the remaining components of $\boldsymbol{\xi}$. A direct consequence of Lemma 2.5 is that the parameters $\boldsymbol{\theta}^{(k)}$ can be integrated out in the following sense.

Corollary 3.1 (marginal likelihood functions) *Set Model Assumptions 2.1 for model $\mathcal{M}^{(k)}$. The random vector $(\boldsymbol{\xi}_{\mathcal{D}_I}, \boldsymbol{\xi}_{\mathcal{D}_I^c})$ has a multivariate Gaussian distribution with the first two moments given by*

$$\begin{aligned}
\boldsymbol{\mu}_I^{(k)} &= \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{D}_I}\right] = P_1 A^{(k)} \boldsymbol{\mu}^{(k)} \qquad \text{and} \qquad \Sigma_I^{(k)} = \text{Cov}\left(\boldsymbol{\xi}_{\mathcal{D}_I}\right) = P_1 \Sigma^{(k)} P_1', \\
\boldsymbol{\mu}_{I^c}^{(k)} &= \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{D}_I^c}\right] = P_2 A^{(k)} \boldsymbol{\mu}^{(k)} \qquad \text{and} \qquad \Sigma_{I^c}^{(k)} = \text{Cov}\left(\boldsymbol{\xi}_{\mathcal{D}_I^c}\right) = P_2 \Sigma^{(k)} P_2'.
\end{aligned}$$

The covariance matrix between the components $\boldsymbol{\xi}_{\mathcal{D}_I}$ and $\boldsymbol{\xi}_{\mathcal{D}_I^c}$ is given by

$$(\Sigma_{I^c,I}^{(k)})' = \Sigma_{I,I^c}^{(k)} = \text{Cov}\left(\boldsymbol{\xi}_{\mathcal{D}_I}, \boldsymbol{\xi}_{\mathcal{D}_I^c}\right) = P_1 \Sigma^{(k)} P_2'.$$

This corollary is an easy consequence of Lemma 2.5 because it only describes a permutation (relabeling) of the components of $\boldsymbol{\xi}$. However, it is useful for parameter calibration, prediction and model selection as we will see below.

Corollary 3.2 (predictive distribution) *Set Model Assumptions 2.1 for model $\mathcal{M}^{(k)}$. The conditional distribution of $\boldsymbol{\xi}_{\mathcal{D}_I^c}$, given $\boldsymbol{\xi}_{\mathcal{D}_I}$, is a multivariate Gaussian distribution with conditional mean given by*

$$\boldsymbol{\mu}_{I^c}^{\text{post}(k)} = \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{D}_I^c}\middle|\,\boldsymbol{\xi}_{\mathcal{D}_I}\right] = \boldsymbol{\mu}_{I^c}^{(k)} + \Sigma_{I^c,I}^{(k)}\left(\Sigma_I^{(k)}\right)^{-1}\left(\boldsymbol{\xi}_{\mathcal{D}_I} - \boldsymbol{\mu}_I^{(k)}\right),$$

and conditional covariance matrix given by

$$\Sigma_{I^c}^{\text{post}(k)} = \text{Cov}\left(\boldsymbol{\xi}_{\mathcal{D}_I^c}\middle|\,\boldsymbol{\xi}_{\mathcal{D}_I}\right) = \Sigma_{I^c}^{(k)} - \Sigma_{I^c,I}^{(k)}\left(\Sigma_I^{(k)}\right)^{-1}\Sigma_{I,I^c}^{(k)}.$$

Proof of Corollary 3.2. The proof is a standard consequence of Corollary 3.1, see for instance Result 4.6 in Johnson-Wichern [8]. □

In view of Corollary 3.2 we only need to calibrate the standard deviation parameters $\boldsymbol{\sigma}$ which enter Σ and $\Sigma^{(k)}$, respectively, and then we can predict the lower triangle in model $\mathcal{M}^{(k)}$. In a full Bayesian approach we choose a prior distribution for these standard deviation parameters. However, then we lose analytical tractability. Therefore, we turn here to an empirical Bayesian viewpoint, estimating these parameters with MLE methods. The marginal likelihood function of the observations $\boldsymbol{\xi}_{\mathcal{D}_I}$ for the standard deviation parameters $\boldsymbol{\sigma}$ is given by, see Corollary 3.1,

$$\begin{aligned}\mathcal{L}^{(k)}\left(\boldsymbol{\sigma}|\,\boldsymbol{\xi}_{\mathcal{D}_I}\right) &= f^{(k)}(\boldsymbol{\xi}_{\mathcal{D}_I}) \qquad (3.1)\\ &= \frac{1}{(2\pi)^{c/2}\,\det(\Sigma_I^{(k)})^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{\xi}_{\mathcal{D}_I} - \boldsymbol{\mu}_I^{(k)})'(\Sigma_I^{(k)})^{-1}(\boldsymbol{\xi}_{\mathcal{D}_I} - \boldsymbol{\mu}_I^{(k)})\right\}.\end{aligned}$$

Maximization of this marginal likelihood function provides the MLE $\widehat{\boldsymbol{\sigma}}^{(k)}$ for $\boldsymbol{\sigma}$ in model $\mathcal{M}^{(k)}$. If we replace the standard deviation parameters $\boldsymbol{\sigma}$ by their MLEs $\widehat{\boldsymbol{\sigma}}^{(k)}$ then Corollary 3.2 provides the full predictive distribution of the lower triangle $\mathcal{D}_I^c = \{\xi_{i,j};\ i+j > I\}$, conditionally given the observations $\mathcal{D}_I$, that is,

$$\widehat{\boldsymbol{\xi}}_{\mathcal{D}_I^c}^{(k)}\Big|_{\{\mathcal{D}_I\}} \ \sim\ \mathcal{N}\left(\widehat{\boldsymbol{\mu}}_{I^c}^{\text{post}(k)}, \widehat{\Sigma}_{I^c}^{\text{post}(k)}\right), \qquad (3.2)$$

where $\widehat{\boldsymbol{\mu}}_{I^c}^{\text{post}(k)}$ and $\widehat{\Sigma}_{I^c}^{\text{post}(k)}$ correspond to $\boldsymbol{\mu}_{I^c}^{\text{post}(k)}$ and $\Sigma_{I^c}^{\text{post}(k)}$ with $\boldsymbol{\sigma}^{(k)}$ replaced by $\widehat{\boldsymbol{\sigma}}^{(k)}$.

3.2 Hierarchical maximum likelihood estimation

The model calibration in the previous section was done using the interpretation of having a Bayesian model. However, we could also interpret this model as a HGLM, see Lee-Nelder [9, 10] and Gigante et al. [6]. We explain this in more detail next. For HGLM we assume a hierarchical model in the sense that there is a first level of effects

$$\boldsymbol{\theta}^{(k)} \sim \mathcal{N}(\boldsymbol{\mu}^{(k)}, T^{(k)}).$$

This may reflect the regulatory viewpoint where $\boldsymbol{\mu}^{(k)}$ describes the insurance market in average and $\boldsymbol{\theta}^{(k)}$ the company specific features. Based on this first level, we then have responses $\boldsymbol{\xi}|_{\{\boldsymbol{\theta}^{(k)}\}} \sim f^{(k)}(\boldsymbol{\xi}|\boldsymbol{\theta}^{(k)})$ according to (2.4). These use the linear predictor given by $A^{(k)}\boldsymbol{\theta}^{(k)}$. Lee-Nelder [9] introduced the h-likelihood of the data $\mathcal{D}_I$ and the effects $\boldsymbol{\theta}^{(k)}$ given by, see also formula (8) in Gigante et al. [6],

$$\begin{aligned} h^{(k)}\left(\boldsymbol{\xi}_{\mathcal{D}_I}, \boldsymbol{\theta}^{(k)}\right) \;\propto\; & \sum_{j=0}^{k-1}\left[\sum_{i=1}^{I-j}\left(-\frac{1}{2\sigma_j^2}\left(\xi_{i,j}-\theta_j\right)^2-\log\sigma_j\right)-\frac{1}{2\tau_j^2}(\theta_j-\mu_j)^2\right] \\ & + \sum_{j=k}^{J}\sum_{i=1}^{I-j}\left(-\frac{1}{2\sigma_j^2}\left(\xi_{i,j}-(\alpha-j\beta)\right)^2-\log\sigma_j\right)+\sum_{a\in\{\alpha,\beta\}}-\frac{1}{2\tau_a^2}(a-\mu_a)^2, \end{aligned} \tag{3.3}$$

where all remaining normalizing constants are put into the proportionality sign $\propto$. For conjugate HGLMs, the h-likelihood (3.3) can be viewed as an *augmented* GLM with data $\mathcal{D}_I$ and *pseudo-data* $\boldsymbol{\mu}^{(k)}$. Therefore, for given $T^{(k)}$, the effects $\boldsymbol{\theta}^{(k)}$ and the standard deviation parameters $\boldsymbol{\sigma}$ can be estimated by MLE providing the following system equations

$$\frac{\partial h^{(k)}\left(\boldsymbol{\xi}_{\mathcal{D}_I}, \boldsymbol{\theta}^{(k)}\right)}{\partial \theta_j}=0 \quad \text{and} \quad \frac{\partial h^{(k)}\left(\boldsymbol{\xi}_{\mathcal{D}_I}, \boldsymbol{\theta}^{(k)}\right)}{\partial \sigma_j}=0 \qquad \text{for } j=0,\ldots,k-1, \tag{3.4}$$

$$\frac{\partial h^{(k)}\left(\boldsymbol{\xi}_{\mathcal{D}_I}, \boldsymbol{\theta}^{(k)}\right)}{\partial (\alpha,\beta)}=\mathbf{0} \quad \text{and} \quad \frac{\partial h^{(k)}\left(\boldsymbol{\xi}_{\mathcal{D}_I}, \boldsymbol{\theta}^{(k)}\right)}{\partial \sigma_j}=0 \qquad \text{for } j=k,\ldots,J. \tag{3.5}$$

The solution of (3.4)-(3.5) provides the MLEs $\widehat{\widehat{\boldsymbol{\theta}}}^{(k)}$ and $\widehat{\widehat{\boldsymbol{\sigma}}}^{(k)}$ in model $\mathcal{M}^{(k)}$. Observe that for $j \le k-1$ we can do an optimization for each development year individually, whereas for $j \ge k$ we do an optimization over all development years simultaneously.

If there is no market view about $\boldsymbol{\mu}^{(k)}$, we set $\tau_1 = \ldots = \tau_{k-1} = \tau_\alpha = \tau_\beta = \infty$ on the right-hand side of (3.3). Then the (non-existent) market knowledge $\boldsymbol{\mu}^{(k)}$ disappears and we are back in the classical GLM context, optimizing the right-hand side of (3.3) neglecting the terms containing any additional knowledge (no augmentation).

The random variables $\xi_{i,j} \in \mathcal{D}_I^c$ in the lower triangle are then approximated by independent (for $i+j>I$) log-link ratios

$$\widehat{\widehat{\xi}}_{i,j}^{(k)}\Big|_{\{\mathcal{D}_I\}} \;\sim\; \mathcal{N}\left(\widehat{\widehat{\theta}}_j^{(k)}, \widehat{\widehat{\sigma}}_j^{(k)}\right). \tag{3.6}$$

Note that this second estimation approach also works in the more general HGLM situation given in (2.6). The disadvantage of (3.6) is that it only considers the point estimator $\widehat{\widehat{\boldsymbol{\theta}}}^{(k)}$ and then simulates conditionally on this estimator according to (3.6). But unlike (3.2), it does not consider uncertainty in this estimator and the quantification of this uncertainty is only obtained by rather involved approximations, using asymptotic MLE results (see for instance Section 6.4.3 in Wüthrich-Merz [17] and Gigante et al. [6]) or bootstrap simulations. Therefore, we prefer the empirical Bayesian approach of Subsection 3.1.

4 Model selection and parameter reduction

In the previous sections we have defined a whole family of models $\mathcal{M}^{(k)}$, $k \in \{1, \ldots, J-1\}$. We now try to find the model that fits the data $\boldsymbol{\xi}_{\mathcal{D}_I}$ best.

4.1 Model selection: Bayesian approach

The marginal distribution (3.1) has a very appealing form that allows for model selection. The general difficulty in such model selection problems is that the dimension of the parameter space may be different in each model $\mathcal{M}^{(k)}$. Therefore, a simulation approach for model selection needs a sophisticated design because these simulations need to experience parameter spaces having (potentially) different dimensions. To overcome this problem state-of-the-art simulation uses RJMCMC methods as demonstrated in Verrall-Wüthrich [14]. The general design of this RJMCMC simulation is very involved. The beauty of our model lies in the fact that we can completely avoid simulations because (3.1) has a sufficiently nice closed form. This we explain next.

We choose a prior distribution on the models $\mathcal{M}^{(k)}$, $k \in \{1, \ldots, J-1\}$, themselves, i.e.

$$p\left(\mathcal{M}^{(k)}\right) > 0 \qquad \text{with} \qquad \sum_{k=1}^{J-1} p\left(\mathcal{M}^{(k)}\right) = 1.$$

The posterior distribution on the model space, given observation $\boldsymbol{\xi}_{\mathcal{D}_I}$, is given by

$$p\left(\mathcal{M}^{(k)}\Big|\, \boldsymbol{\xi}_{\mathcal{D}_I}\right) \;=\; \frac{f^{(k)}(\boldsymbol{\xi}_{\mathcal{D}_I})\; p\left(\mathcal{M}^{(k)}\right)}{\sum_{l=1}^{J-1} f^{(l)}(\boldsymbol{\xi}_{\mathcal{D}_I})\; p\left(\mathcal{M}^{(l)}\right)} \;\propto\; f^{(k)}(\boldsymbol{\xi}_{\mathcal{D}_I})\; p\left(\mathcal{M}^{(k)}\right), \tag{4.1}$$

where the marginal $f^{(k)}(\boldsymbol{\xi}_{\mathcal{D}_I})$ is explicitly given in (3.1). Thus, model selection tells that we should choose the model with the maximal posterior probability weight $p(\mathcal{M}^{(k)}|\boldsymbol{\xi}_{\mathcal{D}_I})$ in (4.1). If there is no dominant model we may also choose model averaging using these posterior model probabilities. Averaging then also includes a component for model uncertainty within this family of models $\mathcal{M}^{(k)}$, $k \in \{1, \ldots, J-1\}$.

The latter may also raise the question whether we should speak about different models $\mathcal{M}^{(k)}$, $k \in \{1, \ldots, J-1\}$, or whether we have simply an overall Bayesian model. Following RJMCMC methods we prefer the first terminology because it emphasizes that parameter spaces may have different dimensions which need to be experienced. This always needs a rather careful treatment, in particular, if simulation methods are applied.

4.2 Model selection: HGLM approach

In the previous subsection, we have done model selection in a Bayesian approach. If we use the HGLM model interpretation of Section 3.2, we could also use other statistical measures for model selection. In order to find the parameters in the HGLM approach we consider the h-likelihood $h^{(k)}$ in each model $\mathcal{M}^{(k)}$, see (3.3). That is, in model $\mathcal{M}^{(k)}$, $k = 1, \ldots, J-1$, and with augmented observations $(\boldsymbol{\xi}_{\mathcal{D}_I}, \boldsymbol{\mu}^{(k)})$ we maximize the log-likelihood function

$$\log \mathcal{L}^{(k)}\left(\boldsymbol{\theta}^{(k)}, \boldsymbol{\sigma}\Big|\, \boldsymbol{\xi}_{\mathcal{D}_I}, \boldsymbol{\mu}^{(k)}\right) = h^{(k)}\left(\boldsymbol{\xi}_{\mathcal{D}_I}, \boldsymbol{\theta}^{(k)}\right).$$

This provides MLEs $\widehat{\widehat{\boldsymbol{\theta}}}^{(k)}$ and $\widehat{\widehat{\boldsymbol{\sigma}}}^{(k)}$, see (3.4)-(3.5). Akaike's information criterion (AIC) [1] of model $\mathcal{M}^{(k)}$ is given by

$$\mathrm{AIC}(k) = -2 \log \mathcal{L}^{(k)} \left(\widehat{\widehat{\boldsymbol{\theta}}}^{(k)}, \widehat{\widehat{\boldsymbol{\sigma}}}^{(k)} \middle| \boldsymbol{\xi}_{\mathcal{D}_I}, \boldsymbol{\mu}^{(k)} \right) + 2(k+2) + 2(J+1). \tag{4.2}$$

Statistical theory says that the model with the smallest AIC should be preferred. The second term in (4.2) accounts for the $k+2$ parameters in $\boldsymbol{\Theta}^{(k)}$ and the last term for the $J+1$ standard deviation parameters $\boldsymbol{\sigma}$. Since all models have the same number of standard deviation parameters $\boldsymbol{\sigma}$, the term $J+1$ is irrelevant for model selection and can be dropped for our analysis. Note that the second term in (4.2) punishes more complex models, thus prefers joint functional dependence in the CL factors which gives parameter reduction in $\boldsymbol{\Theta}^{(k)}$.

5 Claims prediction and tail factors

5.1 Claims prediction and uncertainty analysis

Having selected a model $\mathcal{M}^{(k)}$, we calculate the predictive distribution in the lower triangle $\mathcal{D}_I^c$, conditionally given the upper triangle $\mathcal{D}_I$, see Corollary 3.2 and formula (3.2). In practical applications this is done numerically by constructing the empirical distribution of the outstanding loss liabilities as follows:

1. Simulate $\widehat{\boldsymbol{\xi}}^{(k)}_{\mathcal{D}_I^c}$ according to (3.2).

2. Calculate for each accident year $i = I-J+1, \ldots, I$ the ultimate claim $C_{i,J}$ and the corresponding outstanding loss liabilities given by

$$R_i = C_{i,J} - C_{i,I-i} = C_{i,I-i} \left[\prod_{j=I=i+1}^{J} \left(\exp\left\{ \left(\widehat{\boldsymbol{\xi}}^{(k)}_{\mathcal{D}_I^c} \right)_{i,j} \right\} + 1 \right) - 1 \right],$$

   where $\left(\widehat{\boldsymbol{\xi}}^{(k)}_{\mathcal{D}_I^c} \right)_{i,j}$ denotes the component of $\widehat{\boldsymbol{\xi}}^{(k)}_{\mathcal{D}_I^c}$ that corresponds to cell (i,j) in the lower triangle $\mathcal{D}_I^c$.

3. Calculate the total outstanding loss liabilities in the lower triangle $R = \sum_{i=I-J+1}^{I} R_i$.

4. Repeat steps 1.-3. and obtain the empirical distribution of the outstanding loss liabilities R in the lower triangle. Its conditional mean is denoted by $\widehat{R} = \mathbb{E}\left[R \middle| \mathcal{D}_I \right]$ and the corresponding conditional mean square error of prediction (MSEP) by

$$\mathrm{msep}_{R|\mathcal{D}_I}(\widehat{R}) = \mathbb{E}\left[\left(R - \widehat{R} \right)^2 \middle| \mathcal{D}_I \right] = \mathrm{Var}\left(R \middle| \mathcal{D}_I \right). \tag{5.1}$$

$\widehat{R}$ is called (best-estimate) claims reserves and is used as predictor for the outstanding loss liabilities R at time I. The conditional MSEP (5.1) is a measure for quantifying prediction uncertainty of $\widehat{R}$. For a detailed explanation and discussion of best-estimate claims reserves and conditional MSEP we refer to Section 9.3 in Wüthrich [16].

Note that in the above simulation algorithm we have chosen a fixed model $\mathcal{M}^{(k)}$. If one wants to include model uncertainty within the considered family of models, then we should also integrate the model selection step with posterior model probabilities given by (4.1) into the simulation algorithm.

5.2 First order approximation

In the previous section we have estimated the claims reserves using simulations, in this section we derive an analytical approximation using a first order expansion. Fix accident year $i \in \{I-J+1,\ldots,I\}$. The best-estimate claims reserves $\widehat{R}_i = \mathbb{E}[R_i|\mathcal{D}_I] = \mathbb{E}[C_{i,J}|\mathcal{D}_I] - C_{i,I-i}$ are in model $\mathcal{M}^{(k)}$ given by, see also Corollary 2.3,

$$\begin{aligned}\widehat{R}_i &= \mathbb{E}\left[\mathbb{E}\left[C_{i,J}\Big|\mathcal{D}_I,\boldsymbol{\Theta}^{(k)}\right]\Big|\mathcal{D}_I\right] - C_{i,I-i} \\ &= C_{i,I-i}\mathbb{E}\left[\prod_{j=I-i+1}^{k-1}\left(e^{\theta_j+\sigma_j^2/2}+1\right)\prod_{j=(I-i+1)\vee k}^{J}\left(e^{\alpha-j\beta+\sigma_j^2/2}+1\right)\Big|\mathcal{D}_I\right] - C_{i,I-i}.\end{aligned}$$

For truncation index k the development periods $j<k$ behave mutually independently, given $\mathcal{D}_I$, and they are also independent of all development periods $j'\geq k$. This provides decomposition

$$\widehat{R}_i = C_{i,I-i}\left(\prod_{j=I-i+1}^{k-1}\mathbb{E}\left[e^{\theta_j+\sigma_j^2/2}+1\Big|\mathcal{D}_I\right]\mathbb{E}\left[\prod_{j=(I-i+1)\vee k}^{J}\left(e^{\alpha-j\beta+\sigma_j^2/2}+1\right)\Big|\mathcal{D}_I\right]-1\right). \quad (5.2)$$

The second product in (5.2) cannot easily be decoupled because all terms depend on the same random variables α and β. We have the following equality

$$\mathbb{E}\left[\prod_{j=(I-i+1)\vee k}^{J}\left(e^{\alpha-j\beta+\sigma_j^2/2}+1\right)\Big|\mathcal{D}_I\right] = 1+\sum_{n=1}^{J-((I-i+1)\vee k)+1}\sum_{j_1<\cdots<j_n}\mathbb{E}\left[\prod_{m=1}^{n}e^{\alpha-j_m\beta+\sigma_{j_m}^2/2}\Big|\mathcal{D}_I\right].$$

Thus, we need to calculate the last expected values. They are given by

$$\mathbb{E}\left[\prod_{m=1}^{n}e^{\alpha-j_m\beta+\sigma_{j_m}^2/2}\Big|\mathcal{D}_I\right] = \prod_{m=1}^{n}\mathbb{E}\left[e^{\alpha-j_m\beta+\sigma_{j_m}^2/2}\Big|\mathcal{D}_I\right]e^{\sum_{m<m'}\mathrm{Cov}(\alpha-j_m\beta,\alpha-j_{m'}\beta|\mathcal{D}_I)}.$$

If the last (co-)variance terms are comparably small compared to the posterior means we can set the last terms equal to one. This will be the case in our numerical example below (see narrow confidence bounds in Figure 5). This then justifies the first order approximation

$$\mathbb{E}\left[\prod_{m=1}^{n}e^{\alpha-j_m\beta+\sigma_{j_m}^2/2}\Big|\mathcal{D}_I\right] \approx \prod_{m=1}^{n}\mathbb{E}\left[e^{\alpha-j_m\beta+\sigma_{j_m}^2/2}\Big|\mathcal{D}_I\right], \quad (5.3)$$

which provides approximation (and lower bound)

$$\mathbb{E}\left[\prod_{j=(I-i+1)\vee k}^{J}\left(e^{\alpha-j\beta+\sigma_j^2/2}+1\right)\Big|\mathcal{D}_I\right] \approx \prod_{j=(I-i+1)\vee k}^{J}\mathbb{E}\left[e^{\alpha-j\beta+\sigma_j^2/2}+1\Big|\mathcal{D}_I\right],$$

and, moreover, we obtain approximation $\widehat{R}_i \approx \widehat{R}_i^{\text{approx}}$ with

$$\widehat{R}_i^{\text{approx}} = C_{i,I-i} \left(\prod_{j=I-i+1}^{k-1} \mathbb{E}\left[e^{\theta_j + \sigma_j^2/2} + 1 \middle| \mathcal{D}_I \right] \prod_{j=(I-i+1)\vee k}^{J} \mathbb{E}\left[e^{\alpha - j\beta + \sigma_j^2/2} + 1 \middle| \mathcal{D}_I \right] - 1 \right). \quad (5.4)$$

Note that all terms on the right-hand side of (5.4) can be calculated explicitly and no simulations are needed.

5.3 Tail factors

In view of (5.2) we can also model a tail factor expansion for the claims development beyond the last observed development period J. This modeling needs some care in the sense that we choose a final development period $J_\infty \in \mathbb{N}$ that needs to be finite. This then allows to expand the model to

$$\widehat{R}_i^{\text{ult}} = C_{i,I-i} \left(\prod_{j=I-i+1}^{k-1} \mathbb{E}\left[e^{\theta_j + \sigma_j^2/2} + 1 \middle| \mathcal{D}_I \right] \mathbb{E}\left[\prod_{j=(I-i+1)\vee k}^{J_\infty} \left(e^{\alpha - j\beta + \sigma_j^2/2} + 1 \right) \middle| \mathcal{D}_I \right] - 1 \right),$$

where we set $\sigma_j^2 = \sigma_J^2$ for $j > J$. The reason for choosing J_∞ finite is that β can become negative with positive probability which would provide an infinite mean in the last term for an infinite product. In our example below $J_\infty = 50$ is sufficient for capturing the tail, this can be seen by expanding/approximating the tail as in the first order approximation (5.4).

6 Liability insurance example

We consider the liability insurance run-off data from Verrall-Wüthrich [14]. The data is provided in Table 4, below. The data consists of claims payments for $I = 22$ accident years and $J+1 = 22$ development years. We would like to calibrate models $\mathcal{M}^{(k)}$, $k = 1, \ldots, 20$, to this data set. Thus, we have the choice between 20 different truncation indexes k.

6.1 HGLM model selection without market knowledge

We start the analysis of the data by using classical MLE ignoring any market knowledge $\boldsymbol{\mu}^{(k)}$ in the h-likelihood (3.3). As described in Subsection 3.2 we therefore set $\tau_1 = \ldots = \tau_{k-1} = \tau_\alpha = \tau_\beta = \infty$ on the right-hand side of (3.3) and calculate the corresponding MLEs dropping the prior knowledge part.

Observe that for our data we have $I = J + 1$, and therefore we have only one observation $C_{1,21}$ for the last development year $J = 21$. This implies that we cannot estimate variance parameter σ_{21}^2 from the data, therefore we (simply) set in all derivations $\sigma_{21}^2 = \sigma_{20}^2$.

We start with the MLEs $\widehat{\widehat{\boldsymbol{\theta}}}^{(k)}$ and $\widehat{\widehat{\boldsymbol{\sigma}}}^{(k)}$ for $k = 20$, thus all development periods are estimated individually, see (3.4). This provides the estimates given in Figure 1. We see a negative slope in the estimates $\widehat{\widehat{\theta}}_j^{(20)}$ as a function of j and one is tempted to fit a straight line either at truncation index $k = 7$ or at truncation index $k = 10$. This we are going to analyze in the sequel.

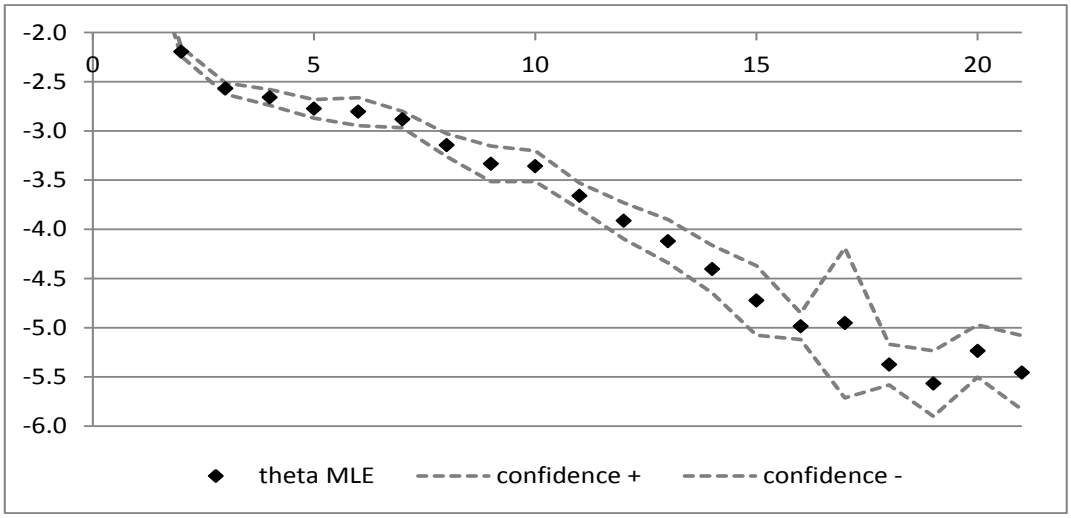

Figure 1: MLEs $\widehat{\widehat{\boldsymbol{\theta}}}^{(k)}$ for $k = 20$ supported by confidence bounds of two standard deviations (we have excluded $\widehat{\widehat{\theta}}_j^{(20)}$ for $j = 0, 1$ in the picture because they are much bigger than the remaining values).

For $k \in \{1, \ldots, 20\}$ we estimate $\boldsymbol{\theta}^{(k)}$ and $\boldsymbol{\sigma}^{(k)}$ using the MLE system (3.4)-(3.5) and then we calculate AIC given in formula (4.2) for each model $\mathcal{M}^{(k)}$ (we drop all terms coming from the market view because at the moment we do not assume market knowledge, i.e. $\tau_j = \infty$). The AICs (4.2) are provided in Figure 2. The picture confirms that the optimal truncation index for our data set is $k = 10$, closely followed by $k = 9, 7, 11$. Thus, the analysis explains that we should model individually the development parameters for $j = 1, \ldots, 9$ and for $j \geq 10$ we model them by the common parameters α and β. This choice provides estimates for the intercept $\widehat{\widehat{\alpha}}^{(10)} = -0.9060$ and for the slope $\widehat{\widehat{\beta}}^{(10)} = 0.2497$ in model $\mathcal{M}^{(10)}$. In Figure 3 we compare the model where we estimate each development year j individually to the model with truncation index $k = 10$. We see that the straight line fitted for $j \geq 10$ is close to the individual MLEs of θ_j, with only the last two periods $j = 20, 21$ differing considerably. These differences should not be over-stated because the individual estimates are based on 2 observations for $j = 20$ and 1 observation for $j = 21$, only. Moreover, the confidence bounds for $j = 20, 21$ may also be questioned because they are based on 2 observations only, recall that we set $\sigma_{21}^2 = \sigma_{20}^2$ because σ_{21}^2 cannot be estimated from 1 observation. The latter suggests that also the variance parameters σ_j^2 may be modeled by a functional form after some truncation index. For the time-being we refrain from doing so.

6.2 Bayesian model selection with different prior knowledge

We turn to the Bayesian case where we directly work on the marginals $\boldsymbol{\xi}_{\mathcal{D}_I}$ given by Corollary 3.1. Then, we only need to estimate the standard deviation parameters $\boldsymbol{\sigma}$ for the different models $\mathcal{M}^{(k)}$ using likelihood (3.1). We first specify the prior parameters $\boldsymbol{\mu}^{(k)}$ and $T^{(k)}$. Since for the

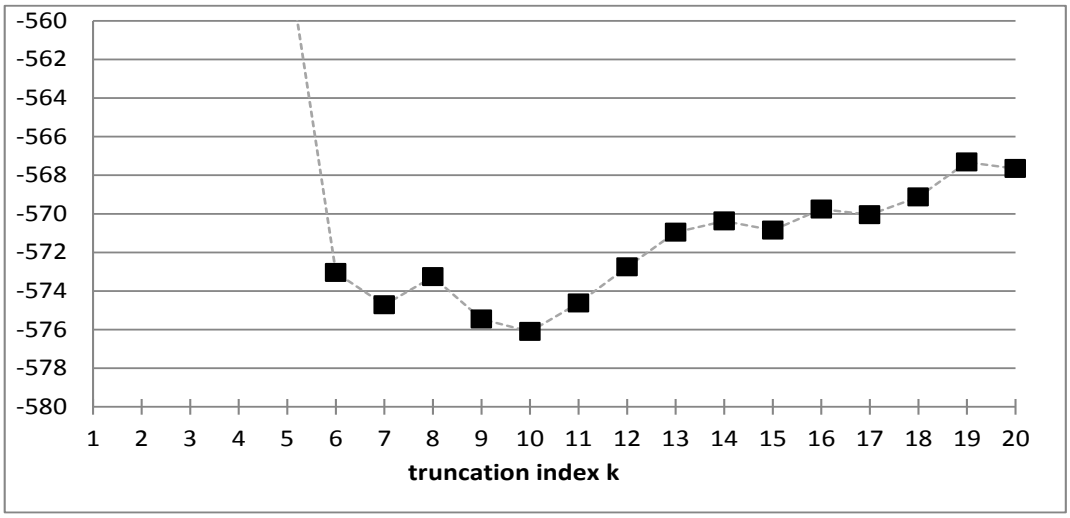

Figure 2: AIC(k) for $k = 1, \ldots, 20$, see formula (4.2).

present example we do not have this information we make an ad-hoc choice that allows to study the sensitivities in the reliability of the prior knowledge. We choose the MLEs for $\mu_0, \ldots, \mu_{k-1}$ and we choose prior tail parameters $\mu_\alpha = -1$ and $\mu_\beta = 0.25$, the latter two choices are motivated by the findings in the previous subsection. For $T^{(k)}$ we make three different choices: (1) Prior Model 1: all coefficients of variation are set equal to 1; (2) Prior Model 2: coefficients of variation of $\theta_0, \ldots, \theta_{k-1}$ are set equal to 0.1 and coefficients of variation of tail parameters α and β are set equal to 1; (3) Prior Model 3: all coefficients of variation are set equal to 0.1. Prior Model 1 corresponds to vague prior knowledge. Prior Models 2 and 3 put more emphasis on the prior knowledge, in Prior Model 2 we have informative prior knowledge for the individual parameters θ_j and vague prior knowledge for the tail parameters of α and β, and in Prior Model 3 we have informative prior knowledge for all parameters.

Finally, we assume that all models $\mathcal{M}^{(k)}$ are equally likely a priori, resulting in the choices $p(\mathcal{M}^{(k)}) = 1/(J-1)$ for $k = 1, \ldots, J-1$. Under these assumptions we calculate the posterior model probabilities (4.1) for the three prior information choices Prior Models 1-3, the results are presented in Figure 4. In Prior Model 2 (informative prior knowledge for θ_j's and vague prior knowledge for α and β) there is not a clear preference, truncation index $k = 7$ has the biggest posterior model probability of about 30% and truncation index $k = 10$ receives posterior model probability of about 20%. In Prior Model 3 (informative prior knowledge for θ_j's, α and β) truncation indexes $k = 9, 10$ are clearly favored with a posterior model probability of almost 40% each. The reason for the differences between Prior Models 2 and 3 is that in Prior Model 3 we have a pre-specified mean of $\mu_\beta = 0.25$ (with coefficient of variation 0.1) which fits to models $\mathcal{M}^{(10)}$ and $\mathcal{M}^{(9)}$. Therefore, these models obtain more posterior probability weight in Prior Model 3 compared to Prior Model 2 where the information about the prior slope μ_β has only little credibility (and may easily be changed by observations).

For vague prior information (Prior Model 1) we clearly favor truncation index $k = 6$. This might

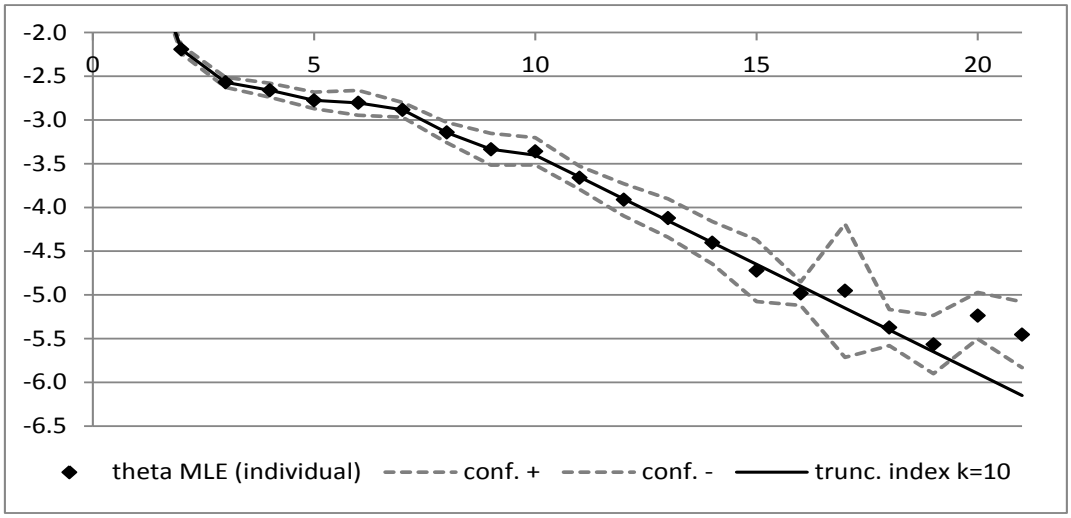

Figure 3: Individual MLEs $\widehat{\widehat{\boldsymbol{\theta}}}^{(k)}$ for $k = 20$ compared to the model with truncation index $k = 10$.

be a small surprise because intuitively we would prefer $k = 7$. The choice $k = 6$ reflects the fact that in case of little information about individual parameters we target for models with only few parameters and hence rather go for a smaller truncation index k due to parameter uncertainty. These considerations lead to the following conclusions. If we have informative prior knowledge we choose truncation index $k = 9$. If we have vague prior knowledge we try to reduce the number of parameters (to reduce parameter uncertainty) and we go for truncation index $k = 6$.
In Figure 5 we present the posterior estimates of the parameters given by

$$\boldsymbol{\theta}^{\text{post}(k)} = \mathbb{E}\left[\boldsymbol{\theta}^{(k)} \middle| \boldsymbol{\xi}_{\mathcal{D}_I}\right] = \boldsymbol{\mu}^{(k)} + T^{(k)}(A^{(k)})' P_1' \left(\Sigma_I^{(k)}\right)^{-1} \left(\boldsymbol{\xi}_{\mathcal{D}_I} - \boldsymbol{\mu}_I^{(k)}\right), \tag{6.1}$$

which are surrounded by intervals of two posterior standard deviations obtained from the conditional covariance matrix

$$T^{\text{post}(k)} = \text{Var}\left(\boldsymbol{\theta}^{(k)} \middle| \boldsymbol{\xi}_{\mathcal{D}_I}\right) = T^{(k)} - T^{(k)}(A^{(k)})' P_1' \left(\Sigma_I^{(k)}\right)^{-1} P_1 A^{(k)} T^{(k)}. \tag{6.2}$$

On the left-hand side (a) of Figure 5 we plot truncation index $k = 9$ for Prior Model 3 and on the right-hand side (b) truncation index $k = 6$ for Prior Model 1. We observe rather narrow intervals in both situations, which says that posterior parameter distributions are very concentrated. For Prior Model 1 they are slightly larger because we have more uncertainty concerning the prior knowledge. This is partly compensated by the fact that we use more observations for $k = 6$ to estimate the parameters of α and β compared to $k = 9$. These intervals for parameters mean that there is only little tail parameter uncertainty, if we believe into the truncation index model, and the dispersion in the MLEs in Figures 3 and 5 comes from process uncertainty.

6.3 Claims prediction and uncertainty analysis

Using the previous model selection analysis we calculate the predictive distribution in the lower triangle $\mathcal{D}_I^c$, conditionally given the upper triangle $\mathcal{D}_I$, see Corollary 3.2, in the selected model(s).

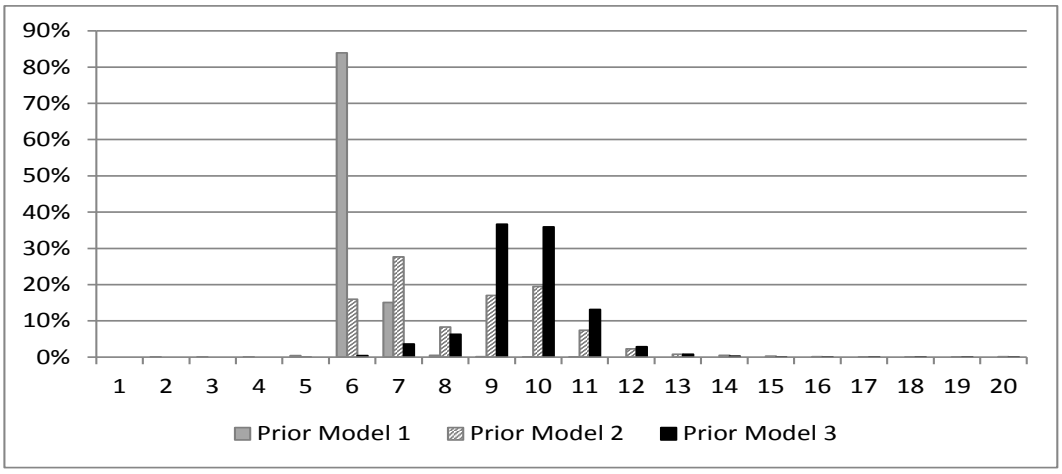

Figure 4: Posterior probability weights (4.1) for models $\mathcal{M}^{(k)}$, $k = 1, \ldots, J-1$, for the three different Prior Models 1-3.

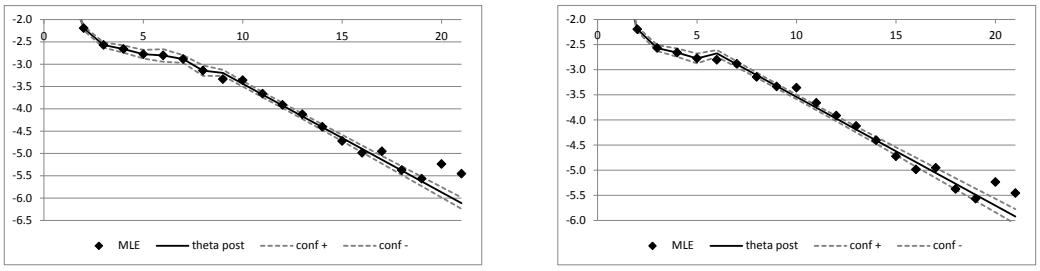

Figure 5: Resulting estimates $\boldsymbol{\theta}^{\text{post}(k)}$ surrounded by intervals of two posterior standard deviations, see (6.1)-(6.2). (a) lhs: Prior Model 3 for truncation index $k = 9$; (b) rhs: Prior Model 1 for truncation index $k = 6$.

From this posterior distribution we can then calculate the best-estimate reserves and the corresponding conditional MSEP. Since, in general, the closed form solution is too complicated we use the simulation algorithm presented in Section 5.1. In Table 1 we present the empirical results which are based on 300'000 simulations. We first consider lines (a) and (b) of Table 1. We see that more prior information reduces prediction uncertainty (conditional MSEP of Prior Model 3 with $k = 9$ versus conditional MSEP of Prior Model 1 with $k = 6$). The resulting reserves are very similar, they are slightly higher in Prior Model 1 because its slope β of the tail parameter is slightly smaller in Prior Model 1.

Lines (c) and (d) in Table 1 correspond to the case were each development period is treated individually. We see a strong increase in the conditional MSEP. Thus, estimating each development period individually strongly increases uncertainty which is in line with the statements of over-parametrization. Moreover, we can see that prior information strongly helps to decrease

	claims reserves $\widehat{R}$	$\text{msep}_{R\vert\mathcal{D}_I}(\widehat{R})^{1/2}$
(a) Prior Model 3, truncation index $k = 9$	1'438'947	52'090
(b) Prior Model 1, truncation index $k = 6$	1'440'738	53'321
(c) Prior Model 3, individual development periods	1'485'416	59'816
(d) Prior Model 1, individual development periods	1'492'284	68'240
(e) Bayesian ODP model of Verrall-Wüthrich [14]	1'476'301	54'073

Table 1: Resulting best-estimate claims reserves $\widehat{R}$ and corresponding conditional MSEP for (a) Prior Model 3 with truncation index $k = 9$ and (b) Prior Model 1 with truncation index $k = 6$. These are compared to the best-estimate claims reserves and the conditional MSEP if we model each development period individually in (c) Prior Model 3 and (d) Prior Model 1, respectively, and to (e) the Bayesian ODP model of Verrall-Wüthrich [14], Table 4.

uncertainty, line 3 versus line 4 in Table 1. We also observe that the resulting best-estimate claims reserves are higher in the individual development period modeling approach compared to the truncation index model. This comes from the fact that the truncation index model judges the tail decay more favorably for our data set, see Figure 5.
If we compare these results, given in Table 1, to the Bayesian over-dispersed Poisson (ODP) model of Verrall-Wüthrich [14], Table 4, we see that they are quite similar. With comparable prior uncertainty as in Prior Model 3, we choose truncation index $k = 7$ in the Bayesian ODP model and the resulting reserves and uncertainties are rather similar to our model (though the model assumptions are very different). Again, our model has the advantage over the Bayesian ODP model that we do not need involved RJMCMC simulations, see Section 3 of Verrall-Wüthrich [14], but obtain model selection and posterior parameters analytically.

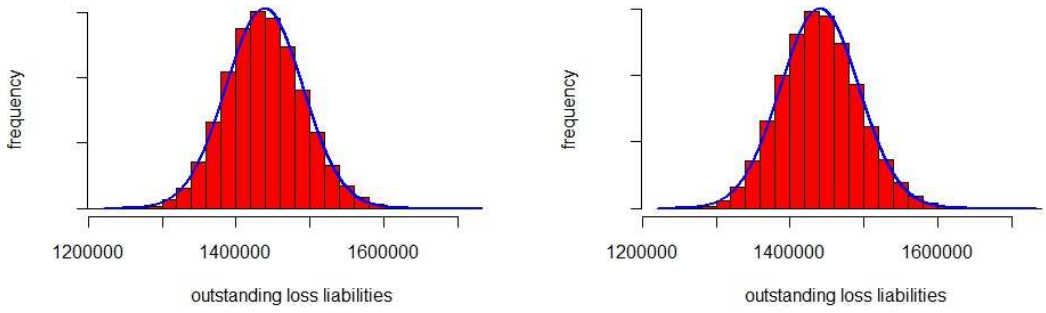

Figure 6: Resulting empirical density (red histogram) and Gaussian approximation (blue line). (a) lhs: Prior Model 3 for truncation index $k = 9$; (b) rhs: Prior Model 1 for truncation index $k = 6$.

In Figure 6 we plot the resulting empirical density and the corresponding Gaussian approximation using $\widehat{R}$ and $\text{msep}_{R\vert\mathcal{D}_I}(\widehat{R})$ for fitting the first two moments of the Gaussian distribution.

At the first sight, the Gaussian approximations in Figure 6 seem to fit very well. In Figure 7

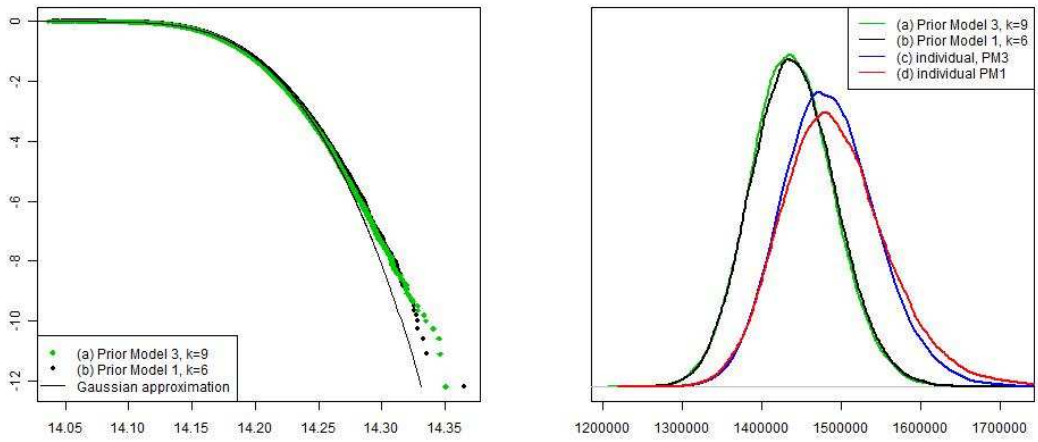

Figure 7: lhs: log-log plot of the empirical survival distributions $x \mapsto \mathbb{P}[R > x|\mathcal{D}_I]$ for (a) Prior Model 3 for truncation index $k = 9$ and (b) Prior Model 1 for truncation index $k = 6$, the line gives the Gaussian approximation; rhs: densities of the 4 models (a)-(d) of Table 1.

(lhs) we present the log-log plot of the empirical survival distributions $x \mapsto \mathbb{P}[R > x|\mathcal{D}_I]$ for models (a) and (b) of Figure 6. We see that in the tails the Gaussian approximation clearly underestimates the potential of large losses because it is less heavy tailed than the distribution of R (which is driven by log-normal distributions).

In Figure 7 (rhs) we plot the resulting empirical densities of models (a)-(d) presented in Table 1. We see that (a) and (b) provide similar results. If we model every development period individually, see models (c) and (d), we obtain the shift seen in Table 1. This shift comes from the fact that the observations for $j = 20, 21$ receive more weight in the latter models, see Figure 5; depending on the data the sign could also go into the other direction. More interestingly, we see that the density is more widely spread the less information we have and the more parameter we have: the least uncertain prediction is obtained in Prior Model 3 with truncation index $k = 9$, the most uncertain in Prior Model 1 with every development period modeled individually. The shift in claims reserves from models (a)-(b) to models (c)-(d) may raise the question whether the tail decay is judged too optimistically under an exponential decay model (since the claims reserves from the individual MLEs modeling are more conservative). In Figure 8 we include in addition to Figure 5 also confidence bounds for the MLEs (symmetric around the posterior estimate $\boldsymbol{\theta}^{\text{post}(k)}$). We observe large volatilities in these MLEs for large development year indexes j and, thus, our model about the exponential decay cannot be rejected in view of Figure 8 because the MLEs are all within the confidence bounds. We also see that the estimates of $\boldsymbol{\sigma}$ could be smoothed to obtain more monotonicity in the confidence bounds. Nevertheless, if the exponential decay is too fast, we could also try to fit a power decay of the form

$$\exp\{\alpha\}\, j^{-\beta} = \exp\{\alpha - \beta \log j\}\,. \tag{6.3}$$

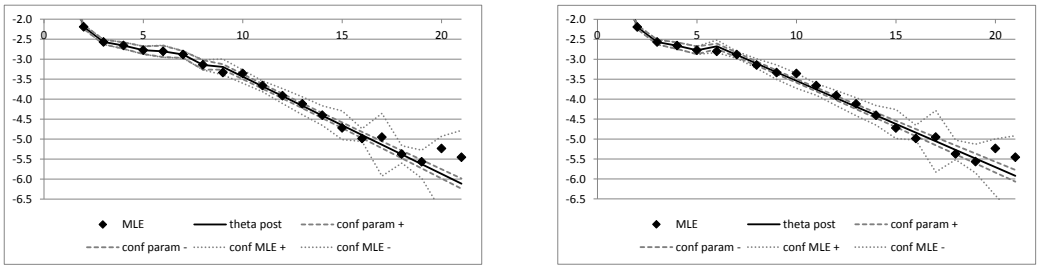

Figure 8: Resulting estimates $\boldsymbol{\theta}^{\text{post}(k)}$ with corresponding intervals of two standard deviations, see (6.1)-(6.2). (a) lhs: Prior Model 3 for truncation index $k = 9$; (b) rhs: Prior Model 1 for truncation index $k = 6$. The weak dotted lines are the confidence bounds for the MLEs $\widehat{\widehat{\boldsymbol{\theta}}}^{(20)}$.

	claims reserves $\widehat{R}$	approximation $\widehat{R}^{\text{approx}}$
(a) Prior Model 3, truncation index $k = 9$	1'438'947	1'438'886
(b) Prior Model 1, truncation index $k = 6$	1'440'738	1'440'545

Table 2: Resulting best-estimate claims reserves $\widehat{R}$ and approximation $\widehat{R}^{\text{approx}}$ for (a) Prior Model 3 with truncation index $k = 9$ and (b) Prior Model 1 with truncation index $k = 6$.

Next we study the first order approximation (and lower bound) presented in Section 5.2, see (5.4). In view of (6.1)-(6.2) we obtain for $j < k$ and $l \geq k$

$$\begin{aligned}
\mathbb{E}\left[e^{\theta_j}\Big|\mathcal{D}_I\right] &= \exp\left\{\boldsymbol{\theta}_j^{\text{post}(k)} + T_{j,j}^{\text{post}(k)}/2\right\}, \\
\mathbb{E}\left[e^{\alpha-l\beta}\Big|\mathcal{D}_I\right] &= \exp\left\{\boldsymbol{\theta}_{k+1}^{\text{post}(k)} - l\boldsymbol{\theta}_{k+2}^{\text{post}(k)} + \frac{T_{k+1,k+1}^{\text{post}(k)} - 2lT_{k+1,k+2}^{\text{post}(k)} + l^2 T_{k+2,k+2}^{\text{post}(k)}}{2}\right\},
\end{aligned}$$

where $\boldsymbol{\theta}_l^{\text{post}(k)}$ is the l-th component of $\boldsymbol{\theta}^{\text{post}(k)}$ and $T_{l,m}^{\text{post}(k)}$ is element (l, m) of covariance matrix $T^{\text{post}(k)}$. This allows approximation $\widehat{R}^{\text{approx}} = \sum_i \widehat{R}_i^{\text{approx}}$ given by (5.4) to be calculated explicitly. In Table 2 we present the corresponding results. We observe that the two values are very close (which also verifies that the simulation algorithm presented in Section 5.1 was implemented correctly).

6.4 Tail factors

Finally, we study the inclusion of tail factors according to Section 5.3. In our example $J_\infty = 50$ is sufficient for capturing the tail, this can be seen by expanding/approximating the tail as in the first order approximation (5.4).

In Figure 9 we plot the resulting reserves $\widehat{R}^{\text{ult}} = \sum_i \widehat{R}_i^{\text{ult}}$ as a function of J_∞, which justifies the choice $J_\infty = 50$. In a similar way to Section 5.1 we simulate payments C_{i,J_∞}, $i \in \{1, \ldots, I\}$, which allow to quantify tail prediction uncertainty within our model $\mathcal{M}^{(k)}$. The results are presented in Table 3. We observe that the predicted claims payments beyond the last observed

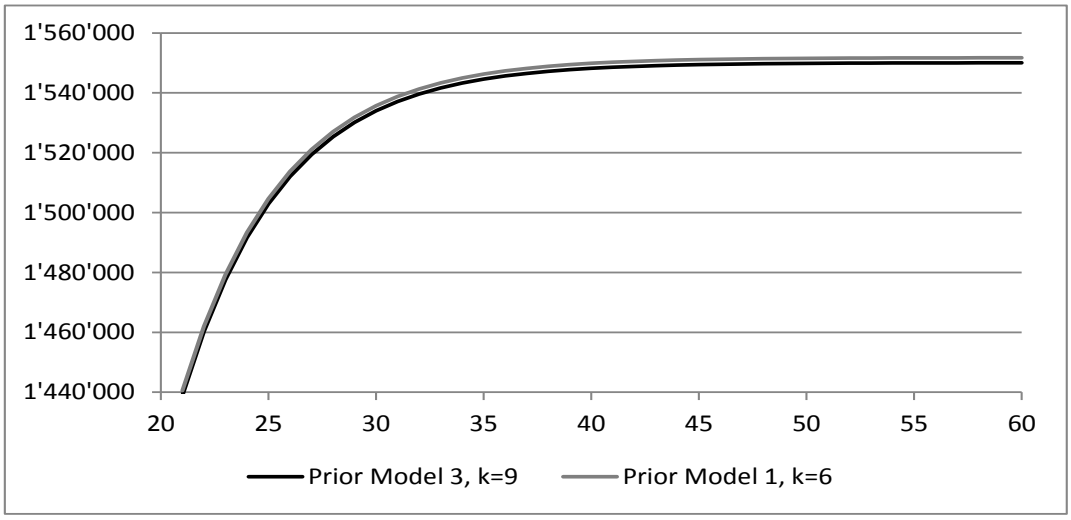

Figure 9: Claims reserves $\widehat{R}^{\text{ult}} = \sum_i \widehat{R}_i^{\text{ult}}$ as a function of $J_\infty = 21, \ldots, 60$ for (a) Prior Model 3 with truncation index $k = 9$ and (b) Prior Model 1 with truncation index $k = 6$.

	reserves $\widehat{R}$	$\text{msep}_{R\vert\mathcal{D}_I}(\widehat{R})^{1/2}$	reserves $\widehat{R}^{\text{ult}}$	cond. MSEP
(a) Prior Model 3, $k = 9$	1'438'947	52'090	1'527'078	55'635
(b) Prior Model 1, $k = 6$	1'440'738	53'321	1'552'331	58'950

Table 3: (a) Prior Model 3 with truncation index $k = 9$ and (b) Prior Model 1 with truncation index $k = 6$: best-estimate claims reserves $\widehat{R}$ and best-estimate claims reserves $\widehat{R}^{\text{ult}}$ including tail factors and corresponding conditional MSEPs for $J_\infty = 50$.

development period $J = 21$ add an additional 7% to the claims reserves $\widehat{R}$. Model (b) is more conservative about this tail development, this comes from the fact that the slope $\mathbb{E}[\beta|\mathcal{D}_I]$ is smaller in Model (b), i.e. 0.22 versus 0.24 in Model (a). The increase in uncertainty (conditional MSEP) is almost 10%. Finally, in Figure 10 we present the corresponding log-log plot and the densities.

7 Conclusion

We consider a Bayesian log-normal model for claims reserving in a chain-ladder framework. We assume that there is a fixed truncation index. Each development period before this truncation index is assumed to have an individual parameter, and development periods after the truncation index are assumed to have a common functional form. We explain how this model can be fit to data and how model selection w.r.t. the truncation index can be done. The advantage of our Bayesian log-normal model is that we do not need involved reversible jump Markov chain Monte Carlo simulation methods as, for instance, used in Verrall-Wüthrich [14]. Once this model is fit to the data, the common functional form above the truncation index gives a natural way to

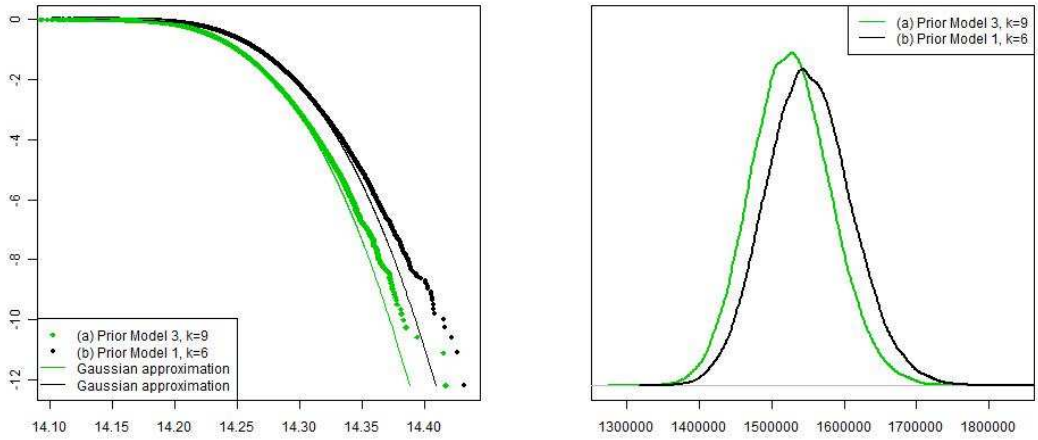

Figure 10: lhs: log-log plot of the survival distributions $x \mapsto \mathbb{P}[R^{\text{ult}} > x|\mathcal{D}_I]$ for (a) Prior Model 3 for truncation index $k = 9$ and (b) Prior Model 1 for truncation index $k = 6$, the line gives the Gaussian approximations; rhs: densities of the 2 models (a) and (b) of Table 3.

estimate tail factors beyond the latest observed development delay.

References

[1] Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control 19/6, 716-723.

[2] Boor, J., (2006). Estimating tail development factors: what to do when the triangle runs out. CAS Forum Winter 2006, 345-390.

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

[4] De Jong, P., Zehnwirth, B. (1983). Claims reserving, state-space models and the Kalman filter. Journal Institute Actuaries 110, 157-182.

[5] England, P.D., Verrall, R.J. (2001). A flexible framework for stochastic claims reserving. Proc. CAS, Vol. LXXXVIII, 1-38.

[6] Gigante, P., Picech, L., Sigalotti, L. (2013). Prediction error for credible claims reserves: an h-likelihood approach. European Actuarial Journal 3/2, 453-470.

[7] Hertig, J. (1985). A statistical approach to the IBNR-reserves in marine insurance. Astin Bulletin 15/2, 171-183.

[8] Johnson, R.A., Wichern, D.W. (1988). Applied Multivariate Statistical Analysis. 2nd edition. Prentice-Hall.

[9] Lee, Y., Nelder, J.A. (1996). Hierarchical generalized linear models. Journal Royal Statistical Society B 58/4, 619-678.

[10] Lee, Y., Nelder, J.A. (2001). Hierarchical generalised linear models: a synthesis of generalised linear models, random-effects models and structured dispersions. Biometrika 88/4, 987-1006.

[11] Merz, M., Wüthrich, M.V., Hashorva, E. (2013). Dependence modelling in multivariate claims run-off triangles. Annals of Actuarial Science 7/1, 3-25.

[12] Shi, P., Basu, S., Meyers, G.G. (2012). A Bayesian log-normal model for multivariate loss reserving. North American Actuarial Journal 16/1, 29-51.

[13] Verrall, R.J. (1996). Claims reserving and generalised additive models. Insurance: Mathematics & Economics 19/1, 31-43.

[14] Verrall, R.J., Wüthrich, M.V. (2012). Reversible jump Markov chain Monte Carlo method for parameter reduction in claims reserving. North American Actuarial Journal 16/2, 240-259.

[15] Wüthrich, M.V. (2012). Discussion of "A Bayesian log-normal model for multivariate loss reserving" by Shi-Basu-Meyers. North American Actuarial Journal 16/3, 398-401.

[16] Wüthrich, M.V. (2013). Non-Life Insurance: Mathematics & Statistics. SSRN Manuscript ID 2319328.

[17] Wüthrich, M.V., Merz, M. (2008). Stochastic Claims Reserving Methods in Insurance. Wiley.

A Data

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	136'367	195'757	213'788	227'545	240'136	249'647	260'260	271'207	285'847	290'814	300'035	305'988	312'930	316'524	319'988	323'268	325'798	326'231	327'722	328'616	330'724	332'137
2	143'135	212'658	233'989	252'155	267'259	280'669	295'863	310'584	322'938	332'021	341'670	349'924	356'908	361'126	367'795	370'548	373'051	376'942	378'314	379'716	381'388	
3	146'469	219'759	241'983	257'063	272'848	285'437	313'398	329'903	340'223	353'075	362'015	371'042	375'094	379'430	382'385	389'057	391'468	395'824	397'531	399'724		
4	158'518	232'128	256'752	276'593	292'807	310'757	322'837	339'751	352'613	366'707	378'735	385'394	394'505	402'618	409'044	412'422	415'624	421'409	424'117			
5	158'633	224'457	249'797	267'676	285'455	303'548	320'282	340'976	352'487	361'300	374'500	388'449	397'848	402'989	408'151	414'016	416'098	419'528				
6	153'215	225'074	249'688	267'753	285'294	307'116	324'791	341'238	353'420	369'549	382'016	390'301	395'206	403'634	406'302	407'819	411'082					
7	153'185	215'699	235'609	255'384	272'749	290'988	304'081	319'717	334'457	352'992	372'879	383'645	394'634	401'194	407'377	410'387						
8	150'974	217'545	242'400	260'473	279'436	299'797	317'991	336'679	352'929	373'339	397'542	407'145	416'136	429'445	435'980							
9	141'432	205'018	225'339	241'315	260'098	277'061	296'286	312'645	330'538	338'629	349'021	357'775	366'468	372'513								
10	141'554	207'510	230'597	250'393	272'538	294'008	321'253	346'836	366'865	381'705	391'678	404'292	411'770									
11	141'899	206'157	229'510	246'710	262'735	280'171	303'956	324'354	343'041	356'874	368'163	380'622										
12	145'037	215'127	240'970	260'457	280'524	304'118	322'331	345'629	357'081	370'673	384'000											
13	135'739	203'999	232'176	250'014	277'500	298'976	323'555	339'853	352'098	364'883												
14	135'350	209'545	236'220	256'710	276'576	293'467	305'436	320'329	336'143													
15	132'847	203'592	227'902	249'914	270'477	286'129	301'347	317'801														
16	135'951	205'450	229'862	250'624	266'371	280'202	300'874															
17	131'151	193'635	215'365	234'202	247'325	262'034																
18	130'188	190'262	213'586	226'115	242'768																	
19	118'505	174'622	192'852	206'808																		
20	118'842	177'671	199'872																			
21	121'011	185'856																				
22	132'116																					

Table 4: Liability insurance run-off, observed cumulative payments $\mathcal{D}_{22} = \{C_{i,j};\ i+j \leq 22,\ 1 \leq i \leq 22,\ 0 \leq j \leq 21\}$.