Reversible Jump Markov Chain Monte Carlo Method for Parameter Reduction in Claims Reserving

Richard J. Verrall* Mario V. Wüthrich†

Abstract

This paper presents an application of reversible jump Markov chain Monte Carlo (RJMCMC) methods to the important problem of setting claims reserves in general insurance business. These reserves are necessary because the premium is received early, but claims may take years to be reported and settled. A measure of the uncertainty in these reserves estimates is also needed for solvency purposes. The RJMCMC methods described in this paper represent an improvement over the manual processes often employed in practice.

1 Introduction

The process of setting provisions for outstanding claims is an important part of the management of general insurance companies. It is important that sufficient capital is held in order to pay any emerging liabilities, but it is also important for the company not to over-reserve, in order that the shareholders can benefit appropriately from the profits of the company. This process requires skillful technical analysis, but it is also necessary to use sensible management judgment. This analysis and judgment appropriately combined should result in claims reserves which are sufficient for paying subsequent claims developments and which are relatively stable over time. The setting of claims reserves usually begins with a technical method applied to data in a straightforward way, followed by adjustments and fine-tuning in order to generate claims reserves that are satisfactory in all senses. Often, the starting point is a simple method such as the chain-ladder method, and one of the important aspects that then has to be further considered are tail factors to estimate the future run-off of claims beyond the latest development year so far observed. To be more specific, claims data is often studied in the form of a so-called claims

*Cass Business School, City University, 106 Bunhill Row, London EC1Y 8T2, UK

†ETH Zurich, RiskLab Switzerland, Department of Mathematics, 8092 Zurich, Switzerland

development triangle, where the rows of this triangle correspond to accident years $i \in \{0, \ldots, I\}$ and the columns correspond to development years $j \in \{0, \ldots, I\}$. Claims observations at time I are then collected in the upper claims development triangle

$$\mathcal{D}_I = \{X_{i,j};\ i + j \leq I,\ 1 \leq i \leq I,\ 1 \leq j \leq I\}, \tag{1.1}$$

and one aims to predict the lower triangle $\mathcal{D}_I^c = \{X_{i,j};\ i + j > I,\ 0 \leq i \leq I,\ 0 \leq j \leq I\}$. As mentioned above, a common starting point is a model which specifies row parameters μ_i and column parameters γ_j and then assumes a cross-classified model of the type

$$\mathbb{E}_{\mu_i,\gamma_j}[X_{i,j}] = \mu_i\ \gamma_j, \qquad \text{for all } i, j \in \{0, \ldots, I\}. \tag{1.2}$$

In practice this means that we need to estimate $2(I + 1)$ parameters $\mu_0, \ldots, \mu_I, \gamma_0, \ldots, \gamma_I$ from $(I + 1)(I + 2)/2$ observations $X_{i,j} \in \mathcal{D}_I$. For example, if we have $I = 9$ we need to estimate 20 parameters from 55 observations. In classical statistical theory, one would say that this model is over-parametrized. The over-parametrization may lead to a good alignment of the parameters to past observations $\mathcal{D}_I$ but therefore the predictive power for the lower triangle $\mathcal{D}_I^c$ is quite low because we have a large parameter uncertainty. The natural thing to do is to reduce the number of parameters by fitting a parametric curve to (some of) the parameters: for example De Jong-Zehnwirth [3] used a curve which is equivalent to assuming that the shape of the development pattern γ_j is a gamma curve. Other papers have also investigated parametric curves to model the development pattern γ_j, but it is often found that in practice these curves are not sufficiently flexible for most triangles of data. England-Verrall [4] suggested using generalized additive models, which apply local smoothing methods within the framework of the over-dispersed Poisson model discussed below. However, the usual (manual) process used in practice is to try to find a very simple model for the tail of the run-off, and use separate parameters for the early development years. In this paper we suggest a method to replace this judgmental process with a more automatic mathematical procedure that should produce results which are at least as good as those from the manual process. Björkwall et al. [1] have a very similar aim, but that paper uses bootstrapping methods and we believe that the method presented in this paper has a number of advantages. Thus, the aim is to model $\gamma_0, \ldots, \gamma_{k-1}$ individually up to a truncation index $k \in \{1, \ldots, I\}$ and then fit an exponential decay to the remaining development parameters, i.e. for $j \in \{k, \ldots, I\}$ we set

$$\gamma_j = \exp\{\alpha - j\beta\},$$

for $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}_+$. It is this choice of the truncation index k which is often done rather subjectively by the modeler. Our aim in this paper is to replace this using a Bayesian claims

reserving model. This model will decide itself which the truncation index k is, only by doing Bayesian inference from the data $\mathcal{D}_I$. That is, our Bayesian modeling framework will avoid any subjective choice of k but will only let the data speak. While we believe that there are some parts of the process that require expert judgment, it is also our opinion that there are times when the use of a manual procedure is inefficient and can, perhaps, give less reliable results. Thus, we believe that it is helpful to consider replacing this part of the process with our more automated method.

We study a whole family of models $\mathcal{M}_k$, $k \in \{1, \ldots, I\}$, all of which have a different number of parameters. We will let the data decide which model, $\mathcal{M}_k$, it prefers (is the most likely for the particular data set). This problem can be solved numerically with the reversible jump Markov chain Monte Carlo (RJMCMC) method which is a particular Markov chain Monte Carlo (MCMC) method that can jump between different models $\mathcal{M}_k$ that have different parameter sets. RJMCMC methods were developed by Green [9, 10], and these days we realize how powerful they are for model selection purposes. Verrall et al. [15] also considered a Bayesian model for the development pattern, using a different application of RJMCMC methods. Although, Verrall et al. [15] presents an interesting model, which may well be of use in practice, the model presented in this paper is much closer to the current actuarial practice. Having both of these methods available should be of benefit to actuaries, and it is possible that either one of them will become more widely used.

Organization of this paper. In Section 2 we introduce the Bayesian over-dispersed Poisson (ODP) model. The crucial definition will be that we define a whole family of prior distributions for the parameters. In Section 3 we describe in careful detail the RJMCMC method and its application to our Bayesian ODP model. Finally, in Section 4 we provide examples and several remarks. Moreover, we explain how the model is extended for tail factor estimation.

2 Bayesian over-dispersed Poisson model

In this section, we define the Bayesian ODP model which is used in this study. The classical ODP model introduced by Renshaw-Verrall [14] and England-Verrall [5] is one of the most popular stochastic claims reserving models. On the one hand it provides the chain-ladder reserves and on the other hand it is very easy to generate bootstrap samples from. For the calculation of predictive distributions of the lower triangle and for the study of prediction and model uncertainty it is

often advantageous to introduce prior distributions for the parameters (including any available expert knowledge on them), see Gisler-Wüthrich [8] and Bühlmann et al. [2]. A Bayesian ODP model was briefly covered in England-Verrall [5] and discussed in detail in England et al. [6]. The latter paper only considers cross-classified models with separate parameters for each development year j, whereas we assume an exponential decay for the column parameters γ_j beyond the truncation index, $j \in \{k, \ldots, I\}$. This leads to a whole family of models $\mathcal{M}_k$, $k \in \{1, \ldots, I\}$, and Bayesian inference method will provide posterior probabilities for these models (from which model selection can be done). This is set out formally in Model Assumptions 2.1.

Model Assumption 2.1 (Bayesian ODP model) *Choose a fixed $k \in \{1, \ldots, I\}$. Then the Bayesian ODP model $\mathcal{M}_k$ is given by the following model assumptions.*

- *Conditionally, given parameters $\boldsymbol{\vartheta} = (\mu_0, \ldots, \mu_I, \gamma_0, \ldots, \gamma_I, \varphi)$, $X_{i,j}$ are independent random variables with*
$$\left.\frac{X_{i,j}}{\varphi}\right|_{\boldsymbol{\vartheta}} \sim \text{Poi}\,(\mu_i \gamma_j / \varphi).$$
- *Assume that $\varphi > 0$ is a given constant and that the parameter vector*
$$\theta_k = (\alpha, \beta, \mu_0, \ldots, \mu_I, \gamma_0, \ldots, \gamma_{k-1})$$
has prior distribution $p_k(\theta_k)$ with independent components satisfying
$$\begin{aligned} \mu_i &\sim \Gamma(s, s/m_i) && \text{for } i = 0, \ldots, I, \\ \gamma_j &\sim \Gamma(v, v/c_j) && \text{for } j = 0, \ldots, k-1, \\ \alpha &\sim \mathcal{N}(a, \sigma^2), && \beta \sim \mathcal{N}(b, \tau^2), \end{aligned}$$
for given prior parameters $m_i, s, c_j, v, \sigma, \tau > 0$ and $a, b \in \mathbb{R}$. Moreover, for $j \in \{k, \ldots, I\}$
$$\gamma_j = \exp\{\alpha - j\beta\}. \tag{2.1}$$

Note that we model the incremental development pattern γ_j individually for $j \in \{0, \ldots, k-1\}$ by gamma prior distributions, whereas for latter development periods $j \in \{k, \ldots, I\}$ we impose an exponential decay for $\beta > 0$. In this way we obtain a whole family of models $\mathcal{M}_k$ with $k \in \{1, \ldots, I\}$ and we seek for the optimal truncation index k. Note that the choice of an exponential decay (2.1) is not really crucial for applying the RJMCMC method and we could replace this with other choices.

Model Assumptions 2.1 give a cross-classified over-dispersed model with the first two conditional moments given by, see also (1.2),

$$\mathbb{E}\left[X_{i,j}|\,\theta_k\right] = \mu_i\ \gamma_j \qquad \text{and} \qquad \mathrm{Var}\left(X_{i,j}|\,\theta_k\right) = \varphi\ \mu_i\ \gamma_j.$$

That is, conditional on the parameters θ_k we obtain the classical multiplicative structure, see for example Section 2.3 in England-Verrall [5].

Assume $\mathcal{A} \subset \{0,\ldots,I\} \times \{0,\ldots,I\}$ is a non-empty set of indexes (i,j). The joint density in model $\mathcal{M}_k$ of the data $(X_{i,j})_{(i,j)\in\mathcal{A}}$ and the parameter vector θ_k is given by

$$\begin{aligned} f_k\left((X_{i,j})_{(i,j)\in\mathcal{A}},\theta_k\right) &= f_k\left((X_{i,j})_{(i,j)\in\mathcal{A}}\big|\,\theta_k\right)\ p_k(\theta_k) \qquad (2.2)\\ &\propto \prod_{(i,j)\in\mathcal{A}} e^{-\frac{\mu_i\gamma_j}{\varphi}}\ \frac{\left(\frac{\mu_i\gamma_j}{\varphi}\right)^{X_{i,j}/\varphi}}{(X_{i,j}/\varphi)!}\ \prod_{i=0}^{I}\mu_i^{s-1}e^{-\frac{s}{m_i}\ \mu_i}\ \prod_{j=0}^{k-1}\gamma_j^{v-1}e^{-\frac{v}{c_j}\ \gamma_j}\\ &\qquad \times\ \exp\left\{-\frac{1}{2\sigma^2}\left(\alpha-a\right)^2\right\}\ \ \exp\left\{-\frac{1}{2\tau^2}\left(\beta-b\right)^2\right\}. \end{aligned}$$

The sign $\propto$ means up to normalizing constants. The first product on the right-hand side is the likelihood of the data $\{X_{i,j};\ (i,j)\in\mathcal{A}\}$, given the parameter θ_k, the second and third products are the prior densities of $\mu_0,\ldots,\mu_I$ and $\gamma_0,\ldots,\gamma_{k-1}$, respectively. The last line in (2.2) describes the prior densities of the regression parameters α and β.

We now analyze the posterior densities of θ_k, see also (3.1) below. Bayes' theorem gives within model $\mathcal{M}_k$

$$f_k\left(\theta_k\,\big|(X_{i,j})_{(i,j)\in\mathcal{A}}\right)\ \ \propto\ \ f_k\left((X_{i,j})_{(i,j)\in\mathcal{A}},\theta_k\right).$$

Moreover, we compare the different models $\mathcal{M}_k$ using the RJMCMC method which attaches posterior probabilities also to the different models $\mathcal{M}_k$.

3 Bayesian model selection

3.1 Reversible jump Markov chain Monte Carlo method

In this outline we closely follow Johansen et al. [12]. Assume we have I different models $\mathcal{M}_k = \{f_k\left(\cdot|\,\theta_k\right);\ \theta_k\in\mathbf{\Theta}_k\}$, where $\mathbf{\Theta}_k$ denotes the parameter space of model $k\in\{1,\ldots,I\}$ and $f_k\left(\cdot|\,\theta_k\right)$ is the probability density in model k for given parameter $\theta_k\in\mathbf{\Theta}_k$.

Assume we have observation y from model $\mathcal{M}_k$ but we do not know from which underlying parameter $\theta_k\in\mathbf{\Theta}_k$ this observation was generated. Using Bayesian inference methods we choose

a prior density $p_k(\theta_k)$ on the parameter space $\mathbf{\Theta}_k$ and then calculate the posterior distribution of θ_k, given the observation y. This posterior distribution is given by the density

$$p_k\left(\theta_k|\, y\right) \;\propto\; f_k\left(y|\, \theta_k\right)\; p_k(\theta_k), \tag{3.1}$$

where the proportionality sign $\propto$ means up to the normalizing constant. The posterior density (3.1) describes how we learn from an observation y for the true underlying parameter $\theta_k \in \mathbf{\Theta}_k$. The situation now becomes more involved if we do not know from which model $\mathcal{M}_k$, $k \in \{1, \ldots, I\}$, this data y was generated. Therefore, we do not only make Bayesian inference on the model parameter $\theta_k \in \mathbf{\Theta}_k$ but also on the model $\mathcal{M}_k$, $k \in \{1, \ldots, I\}$, itself. We choose a prior distribution on the model space and we denote this prior distribution by

$$p(\mathcal{M}_k) > 0 \qquad \text{with} \qquad \sum_{k=1}^{I} p(\mathcal{M}_k) = 1.$$

The posterior distribution on the model and parameter space, given observation y, is given by

$$p\left(\mathcal{M}_k, \theta_k|\, y\right) \;\propto\; f_k\left(y|\, \theta_k\right)\; p_k(\theta_k)\; p(\mathcal{M}_k). \tag{3.2}$$

If we are only interested in the model selection problem, Bayesian inference provides the posterior density

$$p\left(\mathcal{M}_k|\, y\right) = \int_{\mathbf{\Theta}_k} p\left(\mathcal{M}_k, \theta_k|\, y\right)\; d\theta_k \;\;\propto\;\; p(\mathcal{M}_k) \int_{\mathbf{\Theta}_k} f_k\left(y|\, \theta_k\right)\; p_k(\theta_k)\; d\theta_k.$$

We can now either choose the model $\mathcal{M}_k$ with the highest posterior probability $p\left(\mathcal{M}_k|\, y\right)$ for further modeling or we can perform model averaging using these posterior probabilities as weights. MCMC methods are powerful simulation techniques which help to find these posterior distributions $p\left(\mathcal{M}_k|\, y\right)$ and $p\left(\mathcal{M}_k, \theta_k|\, y\right)$ numerically, see Gilks et al. [7]. Mainly, the Metropolis-Hastings (MH) [13, 11] algorithm and the Gibbs sampler (which is a special case of the MH block sampler) are applied. However, the MCMC methods are more involved if we allow the algorithm to jump between different models (trans-dimensional simulations). One reason therefore is that the parameter spaces $\mathbf{\Theta}_k \subset \mathbb{R}^{d_k}$ may have different dimensions d_k, and that the space

$$\mathbf{\Theta} = \bigcup_{k \in \{1,\ldots,I\}} \{k\} \times \mathbf{\Theta}_k \tag{3.3}$$

may not be well-behaved. Green [9, 10] has developed the RJMCMC method which can deal with such situations (see also Johansen et al. [12], Chapter 6). The idea is to generate a Markov chain $(\Theta^{(t)})_{t\geq 0} = (k^{(t)}, \theta^{(t)}_{k^{(t)}})_{t\geq 0}$ whose stationary limit distribution is given by (3.2). In order to make this work we need to carefully model such a Markov chain on an extension of $\mathbf{\Theta}$ that

takes care of the different dimensions d_k of the models $\mathcal{M}_k$. The RJMCMC algorithm goes as follows:

Algorithm 3.1 (RJMCMC algorithm)

(1) Initialize $\Theta^{(0)}$ (fix one $k^{(0)}$ and then e.g. choose the MLE for $\theta^{(0)}_{k^{(0)}}$ in this particular model).

(2) For $t \geq 0$ do

(a) select a model $\mathcal{M}_{k^*}$ with proposal probability $k^* \sim q\left(\cdot \left| k^{(t)}\right.\right)$;

(b) if $k^* = k^{(t)}$ then update the parameters using the classical MH algorithm which provides $\theta^{(t+1)}_{k^*}$ (from $\theta^{(t)}_{k^*}$) and set $\Theta^{(t+1)} = (k^*, \theta^{(t+1)}_{k^*})$; go to item (d);

(c) otherwise (if $k^* \neq k^{(t)}$) then

- generate $u^{(t)}$ from distribution $g_{k^{(t)}\to k^*}\left(\cdot \left| \theta^{(t)}_{k^{(t)}}\right.\right)$,
- set $(\theta^*_{k^*}, u^*) = T_{k^{(t)}\to k^*}\left(\theta^{(t)}_{k^{(t)}}, u^{(t)}\right)$,
- calculate the acceptance probability $\alpha(t \to *)$ which is the minimum of 1 and

$$\frac{p\left(\mathcal{M}_{k^*}, \theta^*_{k^*} \middle| y\right)}{p\left(\mathcal{M}_{k^{(t)}}, \theta^{(t)}_{k^{(t)}} \middle| y\right)} \frac{q\left(k^{(t)} \middle| k^*\right)}{q\left(k^* \middle| k^{(t)}\right)} \frac{g_{k^*\to k^{(t)}}\left(u^* \middle| \theta^*_{k^*}\right)}{g_{k^{(t)}\to k^*}\left(u^{(t)} \middle| \theta^{(t)}_{k^{(t)}}\right)} \left| \frac{\partial T_{k^{(t)}\to k^*}\left(\theta^{(t)}_{k^{(t)}}, u^{(t)}\right)}{\partial\left(\theta^{(t)}_{k^{(t)}}, u^{(t)}\right)} \right|, \tag{3.4}$$

- with acceptance probability $\alpha(t \to *)$ set $\Theta^{(t+1)} = \Theta^{(t)} = (k^*, \theta^*_{k^*})$ and otherwise keep $\Theta^{(t+1)} = (k^{(t)}, \theta^{(t)}_{k^{(t)}})$;

(d) iterate this procedure under item (2).

Under appropriate assumptions this algorithm provides a Markov chain $(\Theta^{(t)})_{t\geq 0} = (k^{(t)}, \theta^{(t)}_{k^{(t)}})_{t\geq 0}$ whose stationary limit distribution is the target posterior distribution $p\left(\mathcal{M}_k, \theta_k \middle| y\right)$. These assumptions require a careful choice of the proposal distributions, essentially they should be such that we obtain an irreducible and aperiodic Markov chain. Moreover, the choices of the distributions $g_{k\to k^*}(\cdot|\cdot)$ and the functions $T_{k\to k^*}(\cdot,\cdot)$ need to fulfill the following requirement:
The functions $T_{k\to k^*}$ should be a diffeomorphism with $T_{k^*\to k} = T^{-1}_{k\to k^*}$. This implies that $(\theta^*_{k^*}, u^*)$ and $(\theta^{(t)}_{k^{(t)}}, u^{(t)})$ need to have the same dimension. We describe our choices in detail in the next subsection.

3.2 Application of the reversible jump algorithm

Our aim now is to apply the RJMCMC algorithm to the Bayesian ODP Model 2.1 and optimize over the different truncation indexes $k \in \{1, \ldots, I\}$. Note that model $\mathcal{M}_k$ has a $(2 + I + 1 + k)$-

dimensional parameter space $\mathbf{\Theta}_k$. Having no prior preference for one of the models we set

$$p(\mathcal{M}_k) = 1/I \qquad \text{for all } k \in \{1, \ldots, I\}. \tag{3.5}$$

We insert this choice (3.5) together with data $\mathcal{D}_I$, see (1.1), and we see that the last term in (3.2) disappears in the proportionality sign and we are left with (2.2) setting $\mathcal{A} = \{(i,j);\ i+j \le I\}$ for the posterior distribution, i.e. the posterior density is given by

$$p\left(\mathcal{M}_k, \theta_k | \mathcal{D}_I\right) \ \propto \ f_k\left(\mathcal{D}_I | \theta_k\right) \ p_k(\theta_k).$$

We now describe all steps in the RJMCMC Algorithm 3.1 for our Bayesian ODP Model 2.1.

Step (2a) of RJMCMC. We need to choose the proposal distributions $q(\cdot|\cdot)$. For $k \in \{2, \ldots, I-1\}$ we choose $q(k-1|k) = q(k|k) = q(k+1|k) = 1/3$. Moreover, $q(1|1) = 2/3$ and $q(2|1) = 1/3$ as well as $q(I|I) = 2/3$ and $q(I-1|I) = 1/3$. These choices imply that we only do model jumps to next neighbor models $k \to k+1$ and $k \to k-1$, and moreover these jump probabilities are chosen such that the second term in the acceptance probability $\alpha(t \to *)$ cancels, see (3.4).

Step (2b) of RJMCMC. If $k^* = k^{(t)}$ we set $k^{(t+1)} = k^{(t)}$ and we apply the MH block sampler which consists of three steps:
Step 1: We update $(\mu_0^{(t)}, \ldots, \mu_I^{(t)})$ using the Gibbs sampler. Note that within model $\mathcal{M}_{k^{(t)}}$

$$p\left(\mu_0, \ldots, \mu_I \,\middle|\, \alpha^{(t)}, \beta^{(t)}, \gamma_0^{(t)}, \cdots, \gamma_{k^{(t)}-1}^{(t)}, \mathcal{D}_I\right)$$

are independent (in μ_i) gamma densities with parameters

$$s \ \mapsto \ s_i^{post} = s + \frac{1}{\varphi} \sum_{j=0}^{I-i} X_{i,j} \qquad \text{and} \qquad \frac{s}{m_i} \ \mapsto \ \left(\frac{s}{m_i}\right)_i^{post} = \frac{s}{m_i} + \frac{1}{\varphi} \sum_{j=0}^{I-i} \gamma_j^{(t)},$$

where we set $\gamma_j^{(t)} = \exp\left\{\alpha^{(t)} - j\beta^{(t)}\right\}$ for $j \in \{k^{(t)}, \ldots, I\}$. Therefore, we can directly generate

$$\mu_0^{(t+1)}, \ldots, \mu_I^{(t+1)} \ \sim \ p\left(\cdot \,\middle|\, \alpha^{(t)}, \beta^{(t)}, \gamma_0^{(t)}, \cdots, \gamma_{k^{(t)}-1}^{(t)}, \mathcal{D}_I\right).$$

This updates the parameters $\mu_0, \ldots, \mu_I$.
Step 2: We update $(\gamma_0^{(t)}, \ldots, \gamma_{k^{(t)}-1}^{(t)})$ also using the Gibbs sampler. Note that within model $\mathcal{M}_{k^{(t)}}$

$$\gamma_0^{(t+1)}, \ldots, \gamma_{k^{(t)}-1}^{(t+1)} \ \sim \ p\left(\cdot \,\middle|\, \alpha^{(t)}, \beta^{(t)}, \mu_0^{(t+1)}, \ldots, \mu_I^{(t+1)}, \mathcal{D}_I\right)$$

are independent (in $\gamma_j^{(t+1)}$) gamma densities with parameters

$$v \;\mapsto\; v_j^{post} = v + \frac{1}{\varphi}\sum_{i=0}^{I-j} X_{i,j} \qquad \text{and} \qquad \frac{v}{c_j} \;\mapsto\; \left(\frac{v}{c_j}\right)_j^{post} = \frac{v}{c_j} + \frac{1}{\varphi}\sum_{i=0}^{I-j}\mu_i^{(t+1)}.$$

Therefore, we can directly generate the updated parameters $\gamma_0^{(t+1)},\ldots,\gamma_{k^{(t)}-1}^{(t+1)}$ from these gamma densities.

Step 3. We update $(\alpha^{(t)},\beta^{(t)})$. This update is done using the classical MH algorithm, see for example Section 4.4 in Wüthrich-Merz [16]. The relevant density in (α,β) for this update is proportional to (as a function of α and β)

$$\prod_{j=k^{(t)}}^{I}\left[e^{-\exp\{\alpha-j\beta\}\sum_{i=0}^{I-j}\frac{\mu_i^{(t+1)}}{\varphi}}\left(\exp\{\alpha-j\beta\}\right)^{\sum_{i=0}^{I-j}\frac{X_{i,j}}{\varphi}}\right]\exp\left\{-\frac{1}{2\sigma^2}\left(\alpha-a\right)^2\right\}\exp\left\{-\frac{1}{2\tau^2}\left(\beta-b\right)^2\right\}.$$

Thus, we have an explicit form for which the MH algorithm provides the update $(\alpha^{(t)},\beta^{(t)})\to(\alpha^{(t+1)},\beta^{(t+1)})$.

Combining Steps 1-3 provides the updated parameter

$$\Theta^{(t+1)} = (k^{(t+1)},\theta_{k^{(t+1)}}^{(t+1)}) = \left(k^{(t+1)},(\alpha^{(t+1)},\beta^{(t+1)},\mu_0^{(t+1)},\ldots,\mu_I^{(t+1)},\gamma_0^{(t+1)},\ldots,\gamma_{k^{(t)}-1}^{(t+1)})\right).$$

Step (2c) of RJMCMC. We need to choose the proposal distributions $g_{k\to k^*}(\cdot|\theta_k)$ and the functions $T_{k\to k^*}(\cdot,\cdot)$. Since we only allow for next neighbor jumps we have that $k^*=k\pm1$.

Case 1: Assume $k^{(t)}<I$ and $k^*=k^{(t)}+1$, i.e. we add a parameter $\gamma_{k^{(t)}}$ to the model. In this case we choose

$$u^{(t)} \;\sim\; g_{k^{(t)}\to k^{(t)}+1}(\cdot|\theta_{k^{(t)}}^{(t)}) \;\overset{(d)}{=}\; \Gamma\left(v^*, v^*/\exp\left\{\alpha^{(t)}-k^{(t)}\beta^{(t)}\right\}\right),$$

for $\theta_{k^{(t)}}^{(t)}=(\alpha^{(t)},\beta^{(t)},\mu_0^{(t)},\ldots,\mu_I^{(t)},\gamma_0^{(t)},\ldots,\gamma_{k^{(t)}-1}^{(t)})$ and given $v^*>0$. We set

$$\theta_{k^*}^* = T_{k^{(t)}\to k^{(t)}+1}\left(\theta_{k^{(t)}}^{(t)},u^{(t)}\right) = \left(\theta_{k^{(t)}}^{(t)},u^{(t)}\right) = \left(\alpha^{(t)},\beta^{(t)},\mu_0^{(t)},\ldots,\mu_I^{(t)},\gamma_0^{(t)},\ldots,\gamma_{k^{(t)}-1}^{(t)},u^{(t)}\right),$$

i.e. $T_{k^{(t)}\to k^{(t)}+1}$ is simply the identity matrix from $\mathbb{R}^{2+(I+1)+k^{(t)}+1}\to\mathbb{R}^{2+(I+1)+k^{(t)}+1}$. This implies that the matrix $T_{k^{(t)}+1\to k^{(t)}}=T_{k^{(t)}\to k^{(t)}+1}^{-1}$ is also the identity matrix and the determinant of the Jacobian (last term in the acceptance probability $\alpha(t\to*)$, see (3.4)) disappears. Note that the only parameter that differs is the one corresponding to development period $k^{(t)}$. Therefore, the acceptance probability has quite a simple form (see also (2.2), (3.2) and step (2a)): set $k=k^{(t)}$, $\gamma_k^{(t)}=\exp\left\{\alpha^{(t)}-k^{(t)}\beta^{(t)}\right\}$ and $\gamma_k^*=u^{(t)}$ then we have acceptance probability

$$\alpha(t\to*)=\min\left\{1\;,\;\prod_{i=0}^{I-k}\left[\frac{e^{-\frac{\mu_i^{(t)}\gamma_k^*}{\varphi}}\left(\gamma_k^*\right)^{\frac{X_{i,k}}{\varphi}}}{e^{-\frac{\mu_i^{(t)}\gamma_k^{(t)}}{\varphi}}\left(\gamma_k^{(t)}\right)^{\frac{X_{i,k}}{\varphi}}}\right]\;\frac{\frac{(v/c_k)^v}{\Gamma(v)}(\gamma_k^*)^{v-1}e^{-\frac{v}{c_k}\gamma_k^*}}{\frac{(v^*/\gamma_k^{(t)})^{v^*}}{\Gamma(v^*)}(\gamma_k^*)^{v^*-1}e^{-\frac{v^*}{\gamma_k^{(t)}}\gamma_k^*}}\right\}.$$

Case 2: Assume $k^{(t)} > 2$ and $k^* = k^{(t)} - 1$, i.e. we subtract one parameter $\gamma_{k^{(t)}-1}$ from the model. In this case we set

$$(\theta^*_{k^*}, u^*) = T_{k^{(t)} \to k^{(t)}-1}\left(\theta^{(t)}_{k^{(t)}}\right) = \theta^{(t)}_{k^{(t)}} = \left(\alpha^{(t)}, \beta^{(t)}, \mu^{(t)}_0, \ldots, \mu^{(t)}_I, \gamma^{(t)}_0, \ldots, , \gamma^{(t)}_{k^{(t)}-2}, \gamma^{(t)}_{k^{(t)}-1}\right),$$

i.e. $u^* = \gamma^{(t)}_{k^{(t)}-1}$. Note that the only parameter that is relevant is the one corresponding to development period $k^{(t)} - 1$. Therefore, the acceptance probability has again a simple form: $k^* = k^{(t)} - 1$ set $\gamma^*_{k^*} = \exp\left\{\alpha^{(t)} - k^*\beta^{(t)}\right\}$ then we have acceptance probability

$$\alpha(t \to *) = \min\left\{1\ ,\ \prod_{i=0}^{I-k^*}\left[\frac{e^{-\frac{\mu^{(t)}_i \gamma^*_{k^*}}{\varphi}}\ (\gamma^*_{k^*})^{\frac{X_{i,k^*}}{\varphi}}}{e^{-\frac{\mu^{(t)}_i \gamma^{(t)}_{k^*}}{\varphi}}\left(\gamma^{(t)}_{k^*}\right)^{\frac{X_{i,k^*}}{\varphi}}}\right]\ \frac{\frac{(v^*/\gamma^*_{k^*})^{v^*}}{\Gamma(v^*)}(\gamma^{(t)}_{k^*})^{v^*-1}e^{-\frac{v^*}{\gamma^*_{k^*}}\gamma^{(t)}_{k^*}}}{\frac{(v/c_{k^*})^{v}}{\Gamma(v)}(\gamma^{(t)}_{k^*})^{v-1}e^{-\frac{v}{c_{k^*}}\gamma^{(t)}_{k^*}}}\right\}.$$

This details all the steps that are necessary for the implementation of the RJMCMC Algorithm 3.1 in the Bayesian ODP Model 2.1 for the prior choices (3.5).

4 Examples

4.1 Predictive modeling

The ultimate goal is to predict the lower triangle $\mathcal{D}^c_I$ based on the information collected in the upper triangle $\mathcal{D}_I$. As described in Verrall et al. [15], formulas (12)-(13), there are two different ways to consider predictive distributions of $\mathcal{D}^c_I$, based on $\mathcal{D}_I$: (i) either we choose the model $\mathcal{M}_k$ that has the largest posterior probability $p(\mathcal{M}_k|\mathcal{D}_I)$ and then make predictions within this model; (ii) or we make model averaging over all models $\mathcal{M}_k$, $k \in \{1, \ldots, I\}$, according to their posterior distributions. We can either study $f\left(\mathcal{D}^c_I|\,\mathcal{D}_I, \mathcal{M}_k\right)$ for the model $\mathcal{M}_k$ with maximal posterior probability $p(\mathcal{M}_k|\mathcal{D}_I)$ or we can study the predictive density (for model averaging) given by

$$f\left(\mathcal{D}^c_I|\,\mathcal{D}_I\right) = \sum_{k\in\{1,\ldots,I\}} f\left(\mathcal{D}^c_I|\,\mathcal{D}_I, \mathcal{M}_k\right)\ p\left(\mathcal{M}_k|\,\mathcal{D}_I\right).$$

If we use our Bayesian ODP Model Assumptions 2.1 this can be rewritten as

$$f\left(\mathcal{D}^c_I|\,\mathcal{D}_I\right) = \sum_{k\in\{1,\ldots,I\}} \left(\int_{\mathbf{\Theta}_k} f_k\left(\mathcal{D}^c_I|\,\theta_k\right)\ p_k\left(\theta_k|\,\mathcal{D}_I\right)\ d\theta_k\right)\ p\left(\mathcal{M}_k|\,\mathcal{D}_I\right).$$

This last equality exactly gives the connection between model averaging and predictive modeling in fixed models $\mathcal{M}_k$.

This density is now approximated with the help of the empirical distribution of $(\Theta^{(t)})_{t\geq T+1}$ from the RJMCMC Algorithm 3.1, where T denotes the burn-in cost until the Markov chain $(\Theta^{(t)})_{t\geq 0}$ has sufficiently converged to the stationary limit distribution $p(\mathcal{M}_k, \theta_k|\mathcal{D}_I)$. Then approximate

$$f\left(\mathcal{D}_I^c|\,\mathcal{D}_I\right) \;\approx\; \frac{1}{N}\sum_{t=T+1}^{T+N} f_k\left(\mathcal{D}_I^c\,\middle|\theta_{k^{(t)}}^{(t)}\right), \tag{4.1}$$

and similarly we choose the marginal distribution for fixed k for $f\left(\mathcal{D}_I^c|\,\mathcal{D}_I, \mathcal{M}_k\right)$.
This means that we have approximated the full predictive distribution from which we can calculate any key figure and risk measure. For simplicity we concentrate in this article on the claims reserves and the prediction variance. The claims reserves are given by

$$R \;=\; \sum_{i=1}^{I}\sum_{j=I-i+1}^{I}\mathbb{E}\left[X_{i,j}|\,\mathcal{D}_I\right] \;=\; \sum_{i+j>I}\mathbb{E}\left[\mu_i\gamma_j|\,\mathcal{D}_I\right] \;\approx\; \frac{1}{N}\sum_{t=T+1}^{T+N}\left[\sum_{i+j>I}\mu_i^{(t)}\gamma_j^{(t)}\right] \overset{def.}{=} \widehat{R}, \tag{4.2}$$

where $\mu_i^{(t)}$ and $\gamma_j^{(t)}$ is extracted from $\Theta^{(t)}$. Note that we write $\{i+j>I\}$ as abbreviation for $\{i+j>I, 1\leq i\leq I,\ 1\leq j\leq I\}$. Similar to (4.1) in England et al. [6] we obtain for the prediction variance (conditional mean square error of prediction)

$$\operatorname{msep}_{\sum_{i+j>I} X_{i,j}|\mathcal{D}_I}(R) = \operatorname{Var}\left(\sum_{i+j>I} X_{i,j}\,\middle|\,\mathcal{D}_I\right) = \varphi\ R + \operatorname{Var}\left(\sum_{i+j>I}\mu_i\gamma_j\,\middle|\,\mathcal{D}_I\right).$$

This is approximated by

$$\widehat{\operatorname{msep}}_{\sum_{i+j>I} X_{i,j}|\mathcal{D}_I}\left(\widehat{R}\right) = \varphi\ \widehat{R} + \frac{1}{N}\sum_{t=T+1}^{T+N}\left[\sum_{i+j>I}\left(\mu_i^{(t)}\gamma_j^{(t)}\right)^2\right] - (\widehat{R})^2. \tag{4.3}$$

Completely analogously we obtain the predictor $\widehat{R}^k$ and the corresponding conditional MSEP estimator $\widehat{\operatorname{msep}}^k_{\sum_{i+j>I} X_{i,j}|\mathcal{D}_I}\left(\widehat{R}^k\right)$ within model $\mathcal{M}_k$, $k\in\{1,\ldots,I\}$.

4.2 Synthetic example 1

The aim in our first synthetic example is to generate simulated data from given parameters $\boldsymbol{\vartheta}$ according to the Bayesian ODP Model 2.1 for a fixed truncation index k and to see whether the RJMCMC algorithm can detect the true truncation index.
We choose $I=9$, $k=4$, $\varphi=25,000$ and

$$\begin{aligned}\theta_k \;=\; \Big(&\alpha=-1.6159,\ \beta=0.2,\ \mu_0=(1.02)^0\cdot 10^7,\ \ldots,\mu_I=(1.02)^I\cdot 10^7,\\ &\gamma_0=15.9\%,\ \gamma_1=17.9\%,\ \gamma_2=17.9\%,\ \gamma_3=13.9\%\Big).\end{aligned}$$

Note that we choose a fixed row exposure of $\mu_0 = 10^7$ and we expand this with a constant inflation rate of 2%. The column parameters (especially α) are chosen such that we obtain $\sum_{j=0}^{9} \gamma_j = 100\%$, i.e. we have a normalized claims development pattern. From these parameters we generate a set of observations $\mathcal{D}_I$. The data is provided in Table 1.

i/j	0	1	2	3	4	5	6	7	8	9
0	1,619,686	1,605,365	1,789,134	1,204,245	701,947	644,759	741,318	606,242	419,492	433,979
1	2,096,654	1,828,792	1,912,791	1,283,858	1,172,218	714,845	707,710	525,486	397,655	
2	2,013,759	1,736,024	1,684,836	1,885,917	852,724	612,746	734,329	566,127		
3	1,508,176	1,748,213	1,714,909	1,453,170	890,544	669,430	480,932			
4	1,565,750	1,640,238	2,160,561	1,985,603	1,127,682	871,973				
5	2,199,588	2,158,210	1,854,062	1,792,833	852,008					
6	2,065,669	2,217,071	1,818,990	1,270,410						
7	1,858,507	2,625,637	1,863,646							
8	1,783,441	1,900,148								
9	1,848,961									

Table 1: Synthetic example 1: data $\mathcal{D}_I$.

In a first step we apply the classical ODP model (see England-Verrall [5] and Wüthrich-Merz [16], Sections 2.3 and 6.4) and derive the MLEs for μ_i and γ_j irrespective of the truncation index. This provides the results in Table 2. From this we can calculate the "true" claims

$j, i =$	0	1	2	3	4	5	6	7	8	9
$\widehat{\gamma}_j^{MLE}$	16.7%	17.4%	16.6%	14.2%	8.5%	6.6%	6.4%	5.3%	3.9%	4.4%
$\widehat{\mu}_i^{MLE}$	$9.8{\cdot}10^6$	$11.1{\cdot}10^6$	$11.0{\cdot}10^6$	$9.8{\cdot}10^6$	$11.7{\cdot}10^6$	$12.1{\cdot}10^6$	$11.4{\cdot}10^6$	$12.5{\cdot}10^6$	$10.8{\cdot}10^6$	$11.8{\cdot}10^6$

Table 2: Synthetic example 1: MLE for data $\mathcal{D}_I$ given in Table 1.

reserves (because we know the true parameters θ_k) and the ones provided by the MLE method. They are given by

$$\begin{aligned} \widehat{R}^{\theta_k} &= \sum_{i+j>I} \mathbb{E}\left[X_{i,j} \middle| \theta_k\right] = \sum_{i+j>I} \mu_i \gamma_j = 34,409,274, \\ \widehat{R}^{MLE} &= \sum_{i+j>I} \widehat{\mu}_i^{MLE} \widehat{\gamma}_j^{MLE} = 34,855,692. \end{aligned}$$

We see that in this particular example the MLE method over-estimates the true claims reserves by 446,418 (approximately 1.3%).

In order to apply the Bayesian ODP Model 2.1 we need to specify the parameters and the prior distributions. We set $\varphi = 25,000$, i.e. to the true value. The prior means of μ_i and γ_j are specified as follows

$$\mathbb{E}\left[\mu_i\right] = m_i \stackrel{!}{=} \widehat{\mu}_i^{MLE} \qquad \text{and} \qquad \mathbb{E}\left[\gamma_j\right] = c_j \stackrel{!}{=} \widehat{\gamma}_j^{MLE}. \tag{4.4}$$

Note that we have for the coefficient of variations

$$\mathrm{Vco}\,(\mu_i) = \frac{\mathrm{Var}\,(\mu_i)^{1/2}}{\mathbb{E}\,[\mu_i]} = s^{-1/2} \qquad \text{and} \qquad \mathrm{Vco}\,(\gamma_j) = \frac{\mathrm{Var}\,(\gamma_j)^{1/2}}{\mathbb{E}\,[\gamma_j]} = v^{-1/2}. \tag{4.5}$$

We assume that the expert can specify the exposure μ_i with a coefficient of variation of 10% which corresponds to the choice $s = 100$. The priors for γ_j are chosen rather non-informatively with $v = 1$, which corresponds to a coefficient of variation of 100%. Finally, we make the following expert choices for the priors of α and β (these are based on the MLE observations) $a = -1$, $\sigma = 10$, $b = 0.5$, $\tau = 10$. Model 2.1 is then completely specified and posterior distributions can be calculated for fixed truncation index k.

In order to apply the RJMCMC Algorithm 3.1 there remains the choice of the proposal distribution for the update of $(\alpha^{(t)}, \beta^{(t)})$ in Step (2b) of the RJMCMC algorithm, and the choice of the parameter v^* in Step (2c). We choose $v^* = 100$ which provides a jump rate between the models of 10.8% (within RJMCMC Step (2c)). As proposal distributions in Step (2b) we choose Gaussian distributions with

$$\alpha^*|_{\alpha^{(t)}} \sim \mathcal{N}(\alpha^{(t)}, 0.05) \qquad \text{and} \qquad \beta^*|_{\beta^{(t)}} \sim \mathcal{N}(\beta^{(t)}, 0.05).$$

This choice provides an average acceptance rate of 18.4%.

With these specifications we can run the RJMCMC algorithm. As initial state we choose $k^{(0)} = 2$ and then we have run $N = 1,000,000$ simulations after subtraction of burn-in costs $T = 10,000$. Figure 1 shows how we jump between the models $\mathcal{M}_k$ (during the burn-in). The 1,000,000

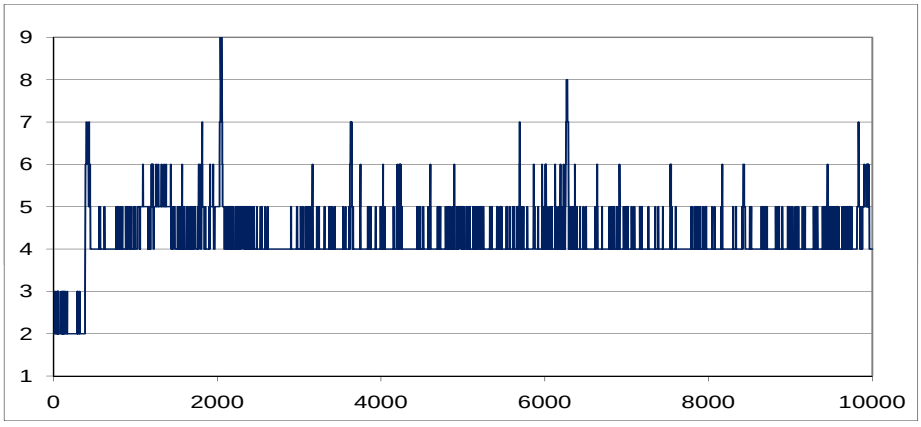

Figure 1: Model selection $k^{(t)}$ for the first 10,000 RJMCMC simulations $t = 0, \ldots, 10,000$.

RJMCMC samples provide the posterior model probabilities $p(\mathcal{M}_k|\mathcal{D}_I)$ given in Figure 2. We

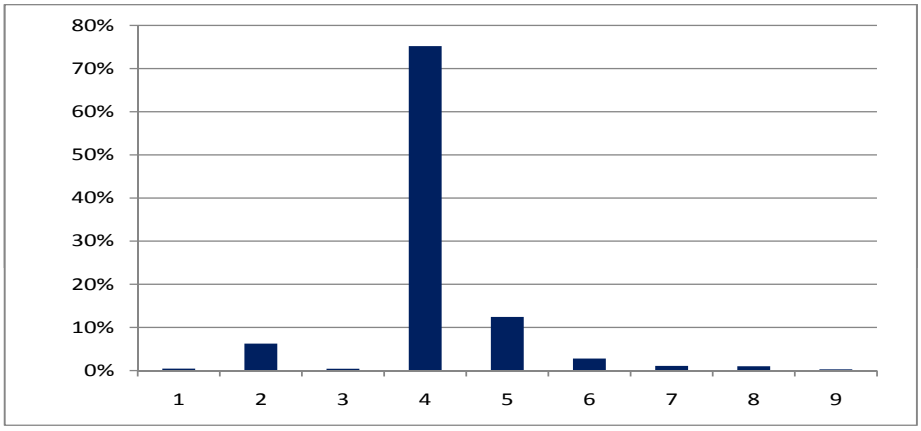

Figure 2: Synthetic example 1: posterior model probabilities $p(\mathcal{M}_k|\mathcal{D}_I)$ from 1,000,000 simulations.

observe that the RJMCMC algorithm strongly favors model $\mathcal{M}_4$ (with posterior probability 75%) which corresponds to the correct truncation index $k = 4$. In Figure 3 we provide the estimates of the development pattern $\log(\gamma_j)$. We see that the RJMCMC pattern $\log(\mathbb{E}[\gamma_j|\mathcal{D}_I])$

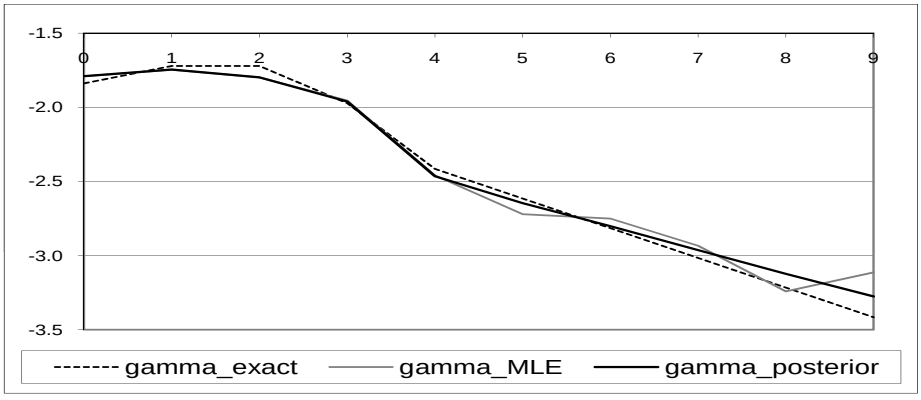

Figure 3: Synthetic example 1: true development pattern $\log(\gamma_j)$, MLE development pattern $\log(\widehat{\gamma}_j^{MLE})$ and posterior means $\log(\mathbb{E}[\gamma_j|\mathcal{D}_I])$ for $j = 0, \ldots, 9$.

smooths the MLE pattern $\log(\widehat{\gamma}_j^{MLE})$. It is basically a straight line after the truncation index $k = 4$. It has expected posterior slope $\mathbb{E}[\beta|\mathcal{D}_I] = 0.16$ which is slightly less than the true value of 0.2.

Finally, we calculate the claims reserves and the prediction variance, the results are provided in Table 3. The resulting claims reserves and prediction variance are compared for three different

	claims reserves	$\text{msep}^{1/2}$
full model $f(\mathcal{D}_I^c\|\mathcal{D}_I)$	34,507,772	2,038,016
maximal posterior probability model $f_4(\mathcal{D}_I^c\|\mathcal{D}_I, \mathcal{M}_4)$	34,515,029	1,991,573
Bayesian ODP model of England et al. [6]	34,887,898	2,083,059

Table 3: Synthetic example 1: resulting claims reserves and prediction variance for the three models $f(\mathcal{D}_I^c|\mathcal{D}_I)$ (full model), $f_4(\mathcal{D}_I^c|\mathcal{D}_I, \mathcal{M}_4)$ (model with the highest posterior probability $p(\mathcal{M}_k|\mathcal{D}_I)$) and $f_9(\mathcal{D}_I^c|\mathcal{D}_I, \mathcal{M}_9)$ (model with no smoothing of column parameters γ_j according to England et al. [6]).

models: (1) full model $f(\mathcal{D}_I^c|\mathcal{D}_I)$; (2) model $f_4(\mathcal{D}_I^c|\mathcal{D}_I, \mathcal{M}_4)$ with the highest posterior probability $p(\mathcal{M}_k|\mathcal{D}_I)$; (3) the individual column parameters γ_j model $f_9(\mathcal{D}_I^c|\mathcal{D}_I, \mathcal{M}_9)$, according to England et al. [6].

Observations.

We observe that the full model and the maximal posterior probability model $\mathcal{M}_4$ give very similar claims reserves. This is clear because $\mathcal{M}_4$ has a posterior probability of 75% and hence is the dominant sub model in the full model. The Bayesian ODP model clearly deviates from the other two models, it gives different reserves and has a higher conditional MSEP (which partly comes from the fluctuation of the posterior development pattern).

If we consider the conditional MSEPs we see that $\mathcal{M}_4$ gives a lower value. This low value is based on the fact that $k = 4$ is an optimal truncation level for the parameter reduction. Whereas the full model also averages over less optimal truncation levels which have a higher conditional MSEP. Henceforth, we can either go for $\mathcal{M}_4$ or we can go for the full model (and then we also account for some model uncertainty in the prediction uncertainty analysis).

Finally, we also remark that the RJMCMC algorithm gives the full posterior distribution (4.1). Therefore we could calculate any other risk measure besides the conditional MSEP.

4.3 Synthetic example 2

For the second synthetic example we basically choose the same set up as in the previous example, i.e. $I = 9$, $\varphi = 25,000$ and $\mu_i = (1.02)^i \cdot 10^7$ for $i \in \{0, \ldots, I\}$. For the column development

factors γ_j we choose an exponential decay

$$\gamma_j = \exp\{\alpha - j\beta\}, \qquad \text{for } j \in \{0, \ldots, 3, 5, \ldots, 9\},$$
$$\text{with } \alpha = -1.5935 \text{ and } \beta = 0.2,$$

and $\gamma_4 = 12.2\%$. That is, we choose a straight line for $\log \gamma_j$ but one column parameter deviates from this straight line. We then generate data $\mathcal{D}_I$ which provides Table 4. For this data we

i/j	0	1	2	3	4	5	6	7	8	9
0	1,794,068	1,428,926	1,278,271	1,285,400	1,509,556	693,145	596,088	420,823	496,766	257,272
1	1,712,508	2,038,374	1,493,193	982,989	1,289,018	722,341	745,301	497,190	539,322	
2	2,127,632	1,728,955	1,136,952	1,207,146	1,463,763	761,753	483,590	546,506		
3	2,412,364	1,964,107	1,477,259	1,276,313	1,281,865	592,705	722,250			
4	2,287,037	1,792,895	1,737,740	1,265,971	1,260,319	876,562				
5	2,265,176	2,005,222	1,635,869	1,137,083	1,314,653					
6	2,290,865	1,612,424	1,674,967	1,368,915						
7	2,268,155	2,527,768	1,605,586							
8	2,132,379	1,727,481								
9	2,163,595									

Table 4: Synthetic example 2: data $\mathcal{D}_I$.

can calculate the maximal posterior probability which provides Figure 4. We observe that

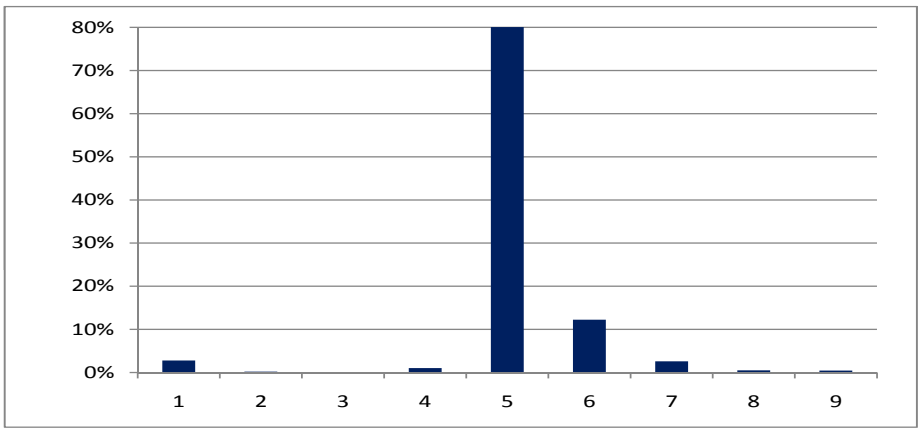

Figure 4: Synthetic example 2: posterior model probabilities $p(\mathcal{M}_k|\mathcal{D}_I)$ from 1,000,000 simulations.

the RJMCMC algorithm clearly detects that there is something wrong in γ_4 and it attaches with posterior probability of 80% a straight line after that development period, i.e. it strongly favors model $\mathcal{M}_5$. In Figure 5 we provide the estimates of the development pattern $\log(\gamma_j)$. In this example the posterior means match the true values very well and we clearly see that

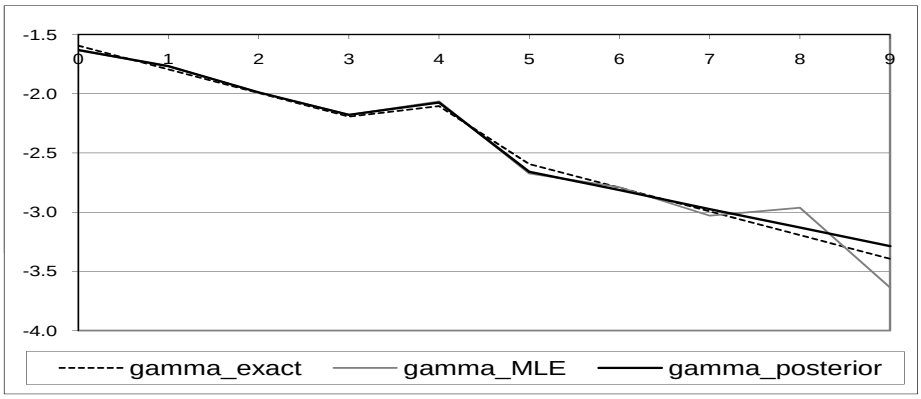

Figure 5: Synthetic example 2: true development pattern $\log(\gamma_j)$, MLE development pattern $\log(\widehat{\gamma}_j^{MLE})$ and posterior means $\log(\mathbb{E}[\gamma_j|\mathcal{D}_I])$ for $j = 0, \ldots, 9$.

$\log \gamma_4$ deviates from the straight line. The MLE has, not surprisingly, more deviations from the straight line, especially for high development periods where we have only few observations.

4.4 Real data example

We choose a liability insurance run-off portfolio where we have data for 22 development years j and accident years i. The data is provided in Table 7. Note that we have a very slow run-off of the liabilities and we would like to fit an exponential decay to late development year parameters γ_j for large j.

We choose similar prior distributions of μ_i and γ_j to the previous examples, i.e. as prior mean we choose the MLEs $\widehat{\mu}_i^{MLE}$ and $\widehat{\gamma}_j^{MLE}$ with prior uncertainties $s = 100$ and $v = 1$, see (4.4)-(4.5). The distributions of α and β are as in synthetic example 1, i.e. $a = -1$, $\sigma = 10$, $b = 0.5$ and $\tau = 10$. The jump rate between the models is driven by $v^* = 100$ which provides an average model jump rate of 14.5% and the proposal parameters in Step (2b) are chosen as

$$\alpha^*|_{\alpha^{(t)}} \sim \mathcal{N}(\alpha^{(t)}, 0.02) \qquad \text{and} \qquad \beta^*|_{\beta^{(t)}} \sim \mathcal{N}(\beta^{(t)}, 0.02).$$

This choice provides an average acceptance rate of 14.6%. Finally, we need to choose the dispersion parameter φ. For this parameter we do an ad-hoc choice using Pearson's residuals, see (6.58) in Wüthrich-Merz [16]. This provides $\widehat{\varphi} = 631.8$. Now we have specified all parameters so that we can apply the RJMCMC Algorithm 3.1 for the Bayesian ODP Model 2.1. We again run $N = 1,000,000$ simulations after the subtraction of the burn-in costs of $T = 10,000$.

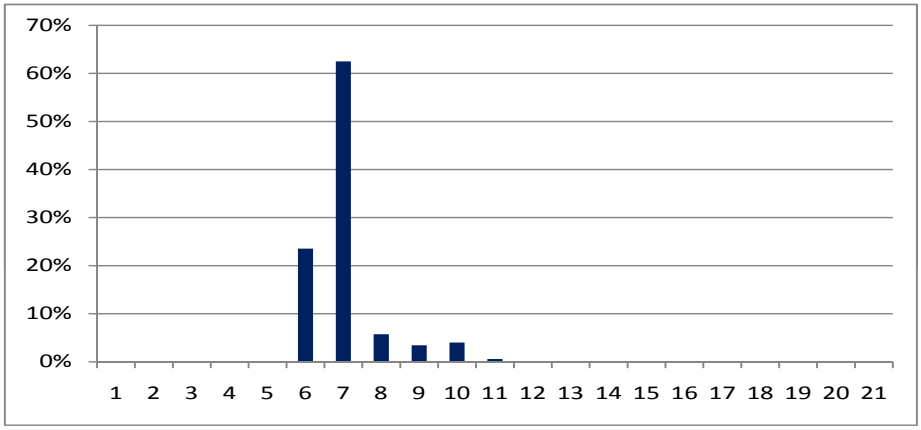

Figure 6: Real data example: posterior model probabilities $p(\mathcal{M}_k|\mathcal{D}_I)$ from 1,000,000 simulations.

This RJMCMC simulation provides the posterior model probabilities in Figure 6. Figure 6 gives a rather clear picture that we have an exponential decay in γ_j after development period $k = 7$, i.e. the RJMCMC algorithm favors model $\mathcal{M}_7$ with posterior probability of 63%. The reason for this strong preference can also be seen in Figure 7 where we plot $\log \gamma_j$. In the real data

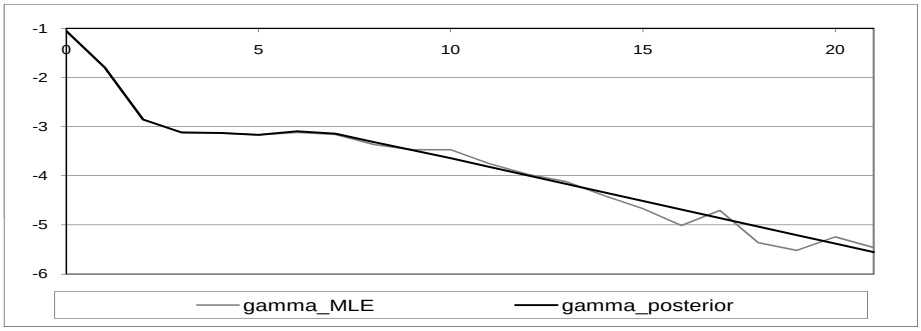

Figure 7: Real data example: MLE development pattern $\log(\widehat{\gamma}_j^{MLE})$ and posterior means $\log(\mathbb{E}[\gamma_j|\mathcal{D}_I])$ for $j = 0, \ldots, 21$.

example we then obtain the claims reserves provided in Table 5.

Observations.

We see that the claims reserves from the maximal posterior probability model $\mathcal{M}_7$ are approxi-

	claims reserves	$\text{msep}^{1/2}$
full model $f(\mathcal{D}_I^c\|\mathcal{D}_I)$	1,475,438	55,810
maximal posterior probability model $f_7(\mathcal{D}_I^c\|\mathcal{D}_I, \mathcal{M}_7)$	1,476,301	54,073
Bayesian ODP model of England et al. [6]	1,463,376	56,247

Table 5: Real data example: resulting claims reserves and prediction variance for the three models $f(\mathcal{D}_I^c|\mathcal{D}_I)$ (full model), $f_7(\mathcal{D}_I^c|\mathcal{D}_I, \mathcal{M}_7)$ (model with the highest posterior probability $p(\mathcal{M}_k|\mathcal{D}_I)$) and $f_{21}(\mathcal{D}_I^c|\mathcal{D}_I, \mathcal{M}_{21})$ (model with no smoothing of column parameters γ_j according to England et al. [6]).

mately the same as the one from the full model. This comes from the fact that in model $\mathcal{M}_7$ we obtain $\mathbb{E}[\beta|\mathcal{D}_I, \mathcal{M}_7] = 0.1746$ which is very similar compared to the full model $\mathbb{E}[\beta|\mathcal{D}_I] = 0.1765$. That is, we obtain about the same decay of the outstanding loss liabilities in model $\mathcal{M}_7$ as in the full model. Moreover, the maximal posterior probability model $\mathcal{M}_7$ is rather dominant. The resulting conditional MSEP figures are also very similar between the maximal posterior model and the full model. The slight difference should not be over-stated.

The Bayesian ODP model of England et al. [6] gives a lower result for the claims reserves which is mainly caused by the low development parameters γ_j in periods $j = 16, 18, 19$ (see Figure 7). Because the full model gives such a clear picture about the truncation index $k = 7$ we believe that the Bayesian ODP model under-estimates the outstanding loss liabilities.

Tail factors.

The data in Table 7 also suggests that there is a claims development beyond development period $j = 21$, i.e. we need to estimate a tail factor that accounts for the payments after $j = 21$. Using the exponential decay, we can now easily fit a tail factor to the observed triangle. We denote the tail factor by

$$\gamma_+ \;=\; \sum_{j>21} \gamma_j \;=\; \sum_{j>21} \exp\{\alpha - j\beta\}\,.$$

If we assume that, conditionally given all parameters, the payments beyond development year $j = 21$ have independent over-dispersed Poisson distributions, i.e.

$$\left.\frac{\sum_{j>21} X_{i,j}}{\varphi}\right|_{\boldsymbol{\vartheta}} \sim \text{Poi}\,(\mu_i\,\gamma_+/\varphi),$$

then we get the predictive distribution including tail factors. The RJMCMC provides then an empirical approximation to the conditional posterior distribution

$$f\,(X_{i,j};\; i+j > I,\; 0 \le i \le I,\; j \ge 1|\,\mathcal{D}_I)\,.$$

In view of (4.2) we obtain the reserves including the tail factor

$$R^+ \;=\; \sum_{i=0}^{I}\sum_{j=1}^{\infty}\mathbb{E}\left[X_{i,j}\middle|\,\mathcal{D}_I\right] \;=\; \sum_{i=0}^{I}\sum_{j=1}^{\infty}\mathbb{E}\left[\mu_i\gamma_j\middle|\,\mathcal{D}_I\right] \;=\; R+\sum_{i=0}^{I}\mathbb{E}\left[\mu_i\gamma_+\middle|\,\mathcal{D}_I\right], \tag{4.6}$$

which is obtained from the RJMCMC sample as in (4.2). In a similar fashion we also obtain the corresponding conditional MSEP. If we restrict Model Assumptions 2.1 for β to distributions that are supported on the positive real line, we obtain posterior tail factor

$$\sum_{i=0}^{I}\mathbb{E}\left[\mu_i\gamma_+\middle|\,\mathcal{D}_I\right]=\sum_{i=0}^{I}\mathbb{E}\left[\mu_i\sum_{j>21}\exp\left\{\alpha-j\beta\right\}\middle|\,\mathcal{D}_I\right]=\sum_{i=0}^{I}\mathbb{E}\left[\mu_i\;\frac{\exp\left\{\alpha-22\cdot\beta\right\}}{1-\exp\left\{-\beta\right\}}\middle|\,\mathcal{D}_I\right],$$

which is obtained numerically. The results are presented in Table 6.

	no tail factor		with tail factor	
	reserves R	msep$^{1/2}$	reserves R^+	msep$^{1/2}$
full model	1,475,438	55,810	1,652,599	84,556
maximal posterior probability model	1,476,301	54,073	1,654,926	72,932
model of England et al. [6]	1,463,376	56,247	not available	

Table 6: Real data example: resulting claims reserves and prediction variance for the three models: full model, model $\mathcal{M}_7$) with the maximal posterior probability, and model of England et al. [6]).

Observations.

We see that the tail factor substantially increases both the claims reserves and the prediction uncertainty. The claims reserves are increased by $R^+ - R \approx 180,000$ (or 12% in terms of R). This gives roughly 8,000 per accident year $i \in \{0,\ldots,21\}$. The increase in the predictive confidence interval is even 50% or 35%, respectively! We conclude that the tail factor is an important quantity for this data set, both for the claims reserves but also for the uncertainty in these claims reserves.

5 Conclusions

This paper has presented an application of RJMCMC methods to an important problem in the management of general insurance business. The method works very well and allows an automatic choice to be made, replacing a manual procedure. We believe that this is advantageous and elegant, since it reduces the need for time consuming analysis which may even introduce a bias.

References

[1] Björkwall, S., Hössjer, O.,Ohlsson, E., Verrall, R.J. (2010). A generalized linear model with smoothing effects for claims reserving. Preprint.

[2] Bühlmann, H., De Felice, M., Gisler, A., Moriconi, F., Wüthrich, M.V. (2009). Recursive credibility formula for chain ladder factors and the claims development result. Astin Bulletin 39/1, 275-306.

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

[4] England, P.D., Verrall, R.J. (2001). A flexible framework for stochastic claims reserving. Proc. CAS 88, 1-38.

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

[6] England, P.D., Verrall, R.J., Wüthrich, M.V. (2010). Bayesian overdispersed Poisson model and the Bornhuetter-Ferguson claims reserving method. Preprint.

[7] Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (1996). Markov Chain Monte Carlo in Practice. Chapman & Hall.

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

[9] Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82/4, 711-732.

[10] Green, P.J. (2003). Trans-dimensional Markov chain Monte Carlo. In: Highly Structured Stochastic Systems, P.J. Green, N.L. Hjort, S. Richardson (eds.), Oxford Statistical Science Series, 179-206. Oxford University Press.

[11] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97-109.

[12] Johansen, A.M., Evers, L., Whiteley, N. (2010). Monte Carlo Methods. Lecture notes, Department of Mathematics, University of Bristol.

[13] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. (1953). Equation of state calculations by fast computing machines. J. Chem. Phys. 21/6, 1087-1092.

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

[15] Verrall, R.J., Hössjer, O., Björkwall, S. (2010). Modelling claims run-off with reversible jump Markov chain Monte Carlo Methods. Preprint.

[16] 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
0	136,367	59,390	18,031	13,757	12,591	9,511	10,613	10,947	14,640	4,967	9,221	5,953	6,942	3,594	3,464	3,280	2,530	433	1,491	894	2,108	1,413
1	143,135	69,523	21,331	18,166	15,104	13,410	15,194	14,721	12,354	9,083	9,649	8,254	6,984	4,218	6,669	2,753	2,503	3,891	1,372	1,402	1,672	
2	146,469	73,290	22,224	15,080	15,785	12,589	27,961	16,505	10,320	12,852	8,940	9,027	4,052	4,336	2,955	6,672	2,411	4,356	1,707	2,193		
3	158,518	73,610	24,624	19,841	16,214	17,950	12,080	16,914	12,862	14,094	12,028	6,659	9,111	8,113	6,426	3,378	3,202	5,785	2,708			
4	158,633	65,824	25,340	17,879	17,779	18,093	16,734	20,694	11,511	8,813	13,200	13,949	9,399	5,141	5,162	5,865	2,082	3,430				
5	153,215	71,859	24,614	18,065	17,541	21,822	17,675	16,447	12,182	16,129	12,467	8,285	4,905	8,428	2,668	1,517	3,263					
6	153,185	62,514	19,910	19,775	17,365	18,239	13,093	15,636	14,740	18,535	19,887	10,766	10,989	6,560	6,183	3,010						
7	150,974	66,571	24,855	18,073	18,963	20,361	18,194	18,688	16,250	20,410	24,203	9,603	8,991	13,309	6,535							
8	141,432	63,586	20,321	15,976	18,783	16,963	19,225	16,359	17,893	8,091	10,392	8,754	8,693	6,045								
9	141,554	65,956	23,087	19,796	22,145	21,470	27,245	25,583	20,029	14,840	9,973	12,614	7,478									
10	141,899	64,258	23,353	17,200	16,025	17,436	23,785	20,398	18,687	13,833	11,289	12,459										
11	145,037	70,090	25,843	19,487	20,067	23,594	18,213	23,298	11,452	13,592	13,327											
12	135,739	68,260	28,177	17,838	27,486	21,476	24,579	16,298	12,245	12,785												
13	135,350	74,195	26,675	20,490	19,866	16,891	11,969	14,893	15,814													
14	132,847	70,745	24,310	22,012	20,563	15,652	15,218	16,454														
15	135,951	69,499	24,412	20,762	15,747	13,831	20,672															
16	131,151	62,484	21,730	18,837	13,123	14,709																
17	130,188	60,074	23,324	12,529	16,653																	
18	118,505	56,117	18,230	13,956																		
19	118,842	58,829	22,201																			
20	121,011	64,845																				
21	132,116																					

Table 7: Liability insurance real data example.