Mathematical Statistics
Stockholm University

Modelling Claims Run-off with Reversible Jump Markov Chain Monte Carlo Methods

Richard Verrall, Ola Hössjer, Susanna Björkwall

Research Report 2010:9

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2010:9**,
http://www.math.su.se/matstat

Modelling Claims Run-off with Reversible Jump Markov Chain Monte Carlo Methods

Richard Verrall
Cass Business School, City University, London
r.j.verrall@city.ac.uk

Ola Hössjer
Dept. of Mathematics, Stockholm University

Susanna Björkwall
Dept. of Mathematics, Stockholm University

September 2010

Abstract

In this paper we describe a new approach to modelling the development of claims run-off triangles. This method replaces the usual ad hoc practical process of extrapolating a development pattern to obtain tail factors with an objective procedure. Two possible model classes are suggested, and the results from the example suggest that Model class 2 might be more suitable for practical purposes.

1 Introduction

Stochastic claims reserving methods have received considerable attention in the recent actuarial literature: Wüthrich and Merz (2008) provides a reasonable summary of many of the methods which have been developed. In many cases, the methods discussed have been based on currently used methods which are sometimes described as 'deterministic methods', although they are only deterministic in the sense that a stochastic model has not been specifically written down when they are used. The methods in current use should therefore be regarded as models for the claims data, based on some (possibly implicit) assumptions, for which stochastic models can also be specified. The stochastic models do not in themselves change the reserving results, and there is no reason to argue whether it is right to use a stochastic model as opposed to a deterministic model. A more correct and useful discussion is to decide whether a stochastic model can help in a practical sense, or whether a simple estimate of the outstanding claims by itself is sufficient for all practical purposes. Recent developments in company management and regulatory requirements have increased the need for stochastic methods, and it is often the case that a simple "claims reserve" is no longer sufficient by itself. The first stage in the move towards the widespread application of stochastic reserving methods was to show how the most commonly used practical approaches can be formulated in statistical models. In this context, England and Verrall (2002) has enabled many actuaries to make the first step in applying statistical models. England and Verrall (2002) covered quite a wide range of different approaches, but one of the most commonly used is the chain-ladder technique. The basic premise of the chain-ladder technique is that there is no underlying pattern to the run-off, and that each development year should be allocated a separate parameter. While it can be argued that this means that the chain-ladder technique will be applicable to a wide range of data, it could also be criticised for having too many parameters. It also means that some other assumptions have to be used to model any possible claims development beyond the latest development year already observed. Actuaries often refer to this as modelling the tail, or applying tail factors.

The purpose of this paper is to describe an approach to modelling the tail of the run-off, which uses as its basis some of the methods which are currently applied in practice. England and Verrall (2001) also considered this problem,

although the approach in that paper was quite different. In this paper, we use Bayesian methods, using reversible jump Markov chain Monte Carlo methods in the package winBUGS (Lunn et al, 2000). England and Verrall (2006) provides a useful introduction to the application of Bayesian methods to claims reserving and the use of simulation methods. The application of Bayesian methods has been revolutionised by the use of Markov chain Monte Carlo (MCMC) methods: see, for example Gilks et al (1996). These methods have enabled statisticians to apply complex Bayesian models to a very wide range of applications, and books such as Congdon (2006) contain many such examples. In the actuarial literature, they have been used by Skollnik (2001) and Verrall (2007), for example.

An important extension is the use of reversible jump MCMC (RJMCMC) methods (Green, 1995), which allow the analysis of trans-dimensional models. This means that it is possible to consider models where the number of variables is unknown, or equivalently, a whole class of models, each with a fixed number of variables. This allows us to consider an interesting range of models for claims reserving. The essence of the approach is to start with a run-off pattern which has (like the chain-ladder technique) a separate parameter for each column, and then let the RJMCMC method decide which ones are needed. We also include a simple parametric tail, so that we can also model the tail of the run-off. Thus, our method has a simple parametric model for the run-off, but allows departures from it when the data justify this. At one extreme, it would give a completely smooth run-off, determined by the parametric model, and at the other extreme, a parameter is included for each column and the model reverts to the chain-ladder technique. This is all contained within the model, and it is not necessary to make arbitrary decisions about when to replace the chain-ladder parameters by a parametric tail, as is often done in practice. Our approach is implemented within winBUGS, using the RJMCMC procedures outlined in Lunn et al (2009).

The paper is set out as follows. In section 2, we briefly outline the stochastic model for the claims run-off triangle that will be used in this paper. Section 3 gives an introduction to MCMC methods, and the implementation in winBUGS, as applied to our claims run-off triangle data. Section 4 describes two Bayesian model classes for the which we then illustrate with the examples in section 5. Section 6 contains the conslusions.

2 Stochastic Reserving Models

Without loss of generality, we assume that we have a triangle of data, which is indexed by row (i) and column (j). The row usually refers to the underwriting year or accident year, and the column refers to the delay in receiving the claim. The data consist of aggregated claims, and could be either reported or incurred claims. The cumulative claims are denoted by D_{ij}, and the triangle of data is $\{D_{ij} : 1 \leq i \leq n, 1 \leq j \leq n-i+1\}$. Equivalently, we use the incremental claims, denoted by $\mathcal{I} = \{C_{ij} : 1 \leq i \leq n, 1 \leq j \leq n-i+1\}$, where

$$\begin{aligned} C_{i1} &= D_{i1} \\ C_{ij} &= D_{ij} - D_{i,j-1}, \quad j \geq 2. \end{aligned}$$

The standard chain-ladder development factors, f_j , are usually calculated using the following formula:

$$f_j = \frac{\sum_{i=1}^{n-j+1} D_{ij}}{\sum_{i=1}^{n-j+1} D_{i,j-1}}. \tag{1}$$

Since the development factors are only available up to $j = n$, the chain-ladder technique can only forecast up to delay year n. The usual forecasts of future cumulative claims are $\left\{\widehat{D}_{ij} : i = 2, 3, \ldots, n; j = n-i+2, n-i+3, \ldots, n\right\}$, where

$$\widehat{D}_{ij} = D_{i,n-i+1} \prod_{l=n-i+2}^{j} f_l \tag{2}$$

for $j = n-i+2, n-i+3, \ldots, n$.

In order to include forecasts for outstanding claims beyond delay year n, it is necessary to consider tail factors. This is often done by assuming that the development factors follow some parametric curve beyond a certain delay year, and then extrapolating this forward to later delay years. For example, it is often assumed that the development factors follow an exponential decay pattern: in other words, the log of the development pattern follows a straight line.

There are a number of candidates for the stochastic model that gives the same estimates of outstanding claims as the chain-ladder technique. In this paper, we use the over-dispersed Poisson model suggested by Renshaw and

Verrall (1998); for a more extensive discussion of this model, see England and Verrall (2002). This model is an example of a generalised linear model, which can be expressed in terms of the first two moments only. Thus, for a random variable Y,

$$E[Y] = m \tag{3}$$

and

$$\text{Var}[Y] = \frac{\phi V(m)}{w} \tag{4}$$

where ϕ denotes a scale parameter, $V(m)$ is the so-called variance function (a function of the mean) and w are weights (often set to 1 for all observations). More details of the theory of generalized linear models can be found in McCullagh and Nelder (1989). The over-dispersed Poisson model is similar to a Poisson model, in that the variance function is equal to the mean, but it also includes the dispersion parameter, ϕ. Thus, it is assumed that the incremental claims, C_{ij}, are distributed as independent over-dispersed Poisson random variables, with mean and variance

$$E[C_{ij}] = m_{ij} \tag{5}$$

and

$$\text{Var}[C_{ij}] = \phi m_{ij}. \tag{6}$$

Within this general distributional assumption, many different models can be applied through the choice of the structure for the mean m_{ij}. In order to produce the same forecast values as the chain-ladder technique (under suitable positivity conditions), the mean is chosen so that

$$\log(m_{ij}) = c + \alpha_i + \beta_j \tag{7}$$

Note that constraints have to be applied to the sets of parameters, which could take a number of different forms. In this paper, we use the corner constraints where $\alpha_1 = \beta_1 = 0$.

The scale parameter, ϕ, can be estimated from the data (see England and Verrall, 2002, for details) and it is usually then treated as a "plug-in" estimate

and not counted as a parameter. Allowing for over-dispersion does not affect estimation of the parameters, but has the effect of increasing their standard errors. It is also possible to relax the restriction that the scale parameter is constant for all observations so that it depends on the development period j: see England and Verrall (2006) for more details. However, in the examples in this paper, we use the straightforward over-dispersed Poisson model with a constant scale parameter. It should be noted that the use of this model does not imply that it is only suitable for data consisting exclusively of positive integers. Instead, a "quasi-likelihood" approach is used (see McCullagh and Nelder, 1989), where the likelihood is the same as a Poisson likelihood up to a constant of proportionality. The straightforward application of the over-dispersed Poisson model with the mean structure defined in (7) gives the same estimates of outstanding claims as the chain-ladder technique. As was mentioned in the introduction, the chain-ladder technique has a separate parameter for each delay year, as does the mean structure (7). An alternative to this is to use a parametric curve for the development pattern, such as the so-called Hoerl curve, where the mean structure is:

$$\log(m_{ij}) = c + \alpha_i + \beta \log(j) + \gamma j \tag{8}$$

This curve has a tail which declines exponentially: in other words, $\log(m_{ij})$ follows a straight line for larger values of j. An advantage of using a parametric curve, such as (8), is that it is straightforward to extrapolate forwards and obtain tail factors for the claims run-off. Within this spirit, a hybrid model is sometimes used which follows the chain-ladder model (7) for the early delay years, but then assumes that $\log(m_{ij})$ follows a straight line for the later development years:

$$\log(m_{ij}) = \begin{cases} c + \alpha_i + \beta_j & \text{for } j \leq k \\ c + \alpha_i + \gamma j & \text{for } j > k \end{cases} \tag{9}$$

Clearly the choice of k is important in such as model, and it is usually chosen by ad hoc trial-and-error methods. Björkwall et al. (2010) recently proposed choosing k authomatically in (9) by means of model selection criteria such as AIC, BIC and bootstrap estimates of mean squared error of prediction.

In this paper, we use an approach which is somewhat similar (though not the same as) (9), and employ Bayesian estimation methods. Section 3 contains a brief outline of the Bayesian approach.

3 Bayesian Models, RJMCMC and BUGS

This section contains a very brief overview of the Bayesian techniques which we use to estimate the models which we use for the claims run-off triangles. There is an extensive literature on these modern Bayesian methods, such as the books by Gelman et al (1995) and Congdon (2006), and the web page for the BUGS project contains links to many on-line resources (http://www.mrc-bsu.cam.ac.uk/bugs). Bayesian modeling is based on Bayes theorem, where all parameters are assumed to be unknown random variables. We assume that we have observed data $\mathcal{I}$ whose distribution, $f(\mathcal{I}|\theta, M)$, depends on the model M and a number of parameters θ of that model, with dimensionality depending on M. We assume that M belongs to a class $\mathcal{M}$ of models and that both of M and θ are unknown. Prior distributions, $f(M)$ and $f(\theta|M)$ are assigned to the model and model parameters within a particular model, and the posterior distribution is given by

$$f(M, \theta|\mathcal{I}) \propto f(\mathcal{I}|\theta, M) f(\theta|M) f(M). \tag{10}$$

3.1 Forecasting with trans-dimensional models

In many statistical applications, the main purpose is to identify the best model to use and to use that model to make inferences from the data. In claims reserving, the aim is slightly different in that it is the predictive distribution of the future claims which is of greatest importance (see England and Verrall, 2006, for more details of predictive distributions in the context of Bayesian claims reserving). Conditional on received triangle of claims data $\mathcal{I}$ and model M, the predictive (posterior) distribution for each future incremental claims, C_{il} where $i + l > n - i + 1$, is

$$f(C_{il}|M, \mathcal{I}) = \int f(C_{il}|M, \theta) f(\theta|M, \mathcal{I}) d\theta. \tag{11}$$

To account for model uncertainty, it is possible to take two different approaches within the context of trans-dimensional models. Broadly, these are to choose the most likely model from the Bayesian analysis and use that to produce predictive distributions; or to estimate the predictive distribution by averaging over all models using as weights the posterior probabilities for

the models from the Bayesian analysis. In the first case

$$f\left(C_{il}|\mathcal{I}\right) \approx f\left(C_{il}|M_{\max},\mathcal{I}\right) \tag{12}$$

where $M_{\max} \in \mathcal{M}$ maximises $P\left(M|\mathcal{I}\right)$ among $M \in \mathcal{M}$, and in the second case

$$f\left(C_{il}|\mathcal{I}\right) = \sum_{M\in\mathcal{M}} f\left(C_{il}|M,\mathcal{I}\right) P\left(M|\mathcal{I}\right). \tag{13}$$

In this paper, we use (13) since it is often not clear that one particular model should be preferred outright over all others, and also because we believe that this gives the best reflection of the underlying uncertainty in the predictive distribution. Often the main interest is in predicting cumulative claims, in particular their sum over all accident years i. The procedcure to obtain the predictive distribution is then the same as in (11) and (13), provided we replace C_{il} by the quantity we wish to predict.

3.2 Reversible Jump MCMC

The first step is to combine (11) and (13) and formally rewrite the predictive distribution of an outstanding incremental claim as

$$f\left(C_{il}|\mathcal{I}\right) = \int f(C_{il}|M,\theta) f(\theta,M|\mathcal{I}) d(M,\theta), \tag{14}$$

integrating over the posterior distribution in (10). In some cases, this distribution may be obtained in exact terms, straightforwardly. However, when the model is unknown and the parameter vector is high dimensional, or complex, it is usually not possible to obtain the posterior distribution in closed form. In these cases, simulation methods can prove to be highly effective and the recent advances use simulation based on Markov chains: the so-called Markov chain Monte Carlo methods. The idea is to generate a Markov chain $\{(M^{(b)},\theta^{(b)})\}_{b=1}^{\infty}$ whose equilibrium distribution equals the posterior distribution in (10). One then approximates the predictive distribution (14) by a Monte Carlo average

$$f\left(C_{il}|\mathcal{I}\right) \approx \frac{1}{N}\sum_{a=1}^{N} f(C_{il}|M^{(B+ta)},\theta^{(B+ta)}), \tag{15}$$

where B is the burn-in time (i.e. the time before the Markov chain has converged to its equlibrium distribution) and t a thinning parameter, so that only every tth simulation from the Markov chain is used. Often $t = 1$ is used, but if the serial correlation of the output Markov chain is high, one may reduce it by choosing $t > 1$. The MCMC methodology provides a general framework of generating the Markov chain. Given the current state $(M^{(b)}, \theta^{(b)})$, a subsequent state (M, θ) is drawn from some proposal distribution π and is either accepted or rejected, so that the next state

$$(M^{(b+1)}, \theta^{(b+1)}) = \begin{cases} (M, \theta), & \text{if } (M, \theta) \text{ is accepted,} \\ (M^{(b)}, \theta^{(b)}), & \text{if } (M, \theta) \text{ is rejected.} \end{cases}$$

The ingenious part is to device the acceptance probability of (M, θ) so that a certain reversibility condition is maintained. This virtually means that less likely states should be proposed more seldomly than states with a high probability. If a state is proposed too often, this compensated by a small acceptance probability.

In (blockwise) Gibbs sampling (Geman and Geman, 1984, and Gelfand and Smith, 1990) the model $M = M^{(b)}$ is kept fixed, whereas (a block of) one parameter(s) in θ is updated according to its conditional posterior distbution. In this case all proposals are accepted, since the proposal distribution is the optimal one. When the conditional distribution is difficult to sample from, one employs the more general Metropolis-Hastings (MH) algorithm, which means choosing some other proposal distribution for (a block of) one parameter(s) of θ, still keeping $M = M^{(b)}$ fixed, with

$$P(\text{accept } (M, \theta)) = \min\left(1, \frac{f(\theta|M^{(b)}, \mathcal{I})\pi(\theta^{(b)}|\theta)}{f(\theta^{(b)}|M^{(b)}, \mathcal{I})\pi(\theta|\theta^{(b)})}\right). \tag{16}$$

The reversible jump MCMC is a generalizaton of the MH algortihm which also allows for jumps between different models, i.e. $M \neq M^{(b)}$. Write $\theta = \theta(\theta^{(b)}, u)$ and $\theta^{(b)} = \theta^{(b)}(\theta, v)$ for proposed moves back and forth between $\theta^{(b)}$ and θ. Dimension matching requires that $\dim(M^{(b)}) + \dim(u)$ equals $\dim(M) + \dim(v)$. A proposed move is accepted with probability

$$P(\text{accept } (M, \theta)) = \min\left(1, \frac{f(\theta, M|\mathcal{I})\pi(M^{(b)}, \theta^{(b)}|M, \theta)}{f(M^{(b)}, \theta^{(b)}|\mathcal{I})\pi(M, \theta|M^{(b)}, \theta^{(b)})} \left|\frac{\partial(\theta, v)}{\partial(\theta^{(b)}, u)}\right|\right), \tag{17}$$

where the major difference compared to (16) is the presence of the Jacobian term, which accounts for different parametrizatioins of the two models. In this paper, a proposed jump between models always means adding ($\dim(u) = 1$, $v = 0$) or deleting ($u = 0$, $\dim(v) = 1$) a parameter to the current model.

The better the choice of proposal distribution π, the more candidates are accepted, which implies less serial correlation (smaller burn in time) and faster mixing. In practice, one typically combines Gibbs, MH and RJ moves, where the updates with worst mixing are repeated more frequently.

3.3 Trans-Dimensional Models in BUGS

While it is possible to construct computer programmes separately from first principles for each application, winBUGS (which is freely available) has been designed to be "flexible software for the Bayesian analysis of complex statistical models using Markov chain Monte Carlo (MCMC) methods". We make use of winBUGS, together with the suite of add-ons which allow the application of some reversible jump MCMC methods (these are available from the BUGS project web site, together with "winBUGS Jump Interface: User Manual"). These allow the analysis of trans-dimensional models ($|\mathcal{M}| > 1$), where the structure of the model itself is unknown. In particular, Lunn et al (2009) describe two main classes of models that can be used within winBUGS, one of which is well-suited for the application in this paper. We outline its use for claims reserving in general terms in this section, and specify two possible model classes in section 4.

The trans-dimensional model of Lunn et al (2009) is defined in terms of an unknown number of "entities of interest". In our application, these will be parameters associated with the shape of the run-off. The number of "entities of interest" (which are parameters in our case) is denoted by k, and the prior distribution for k as a binomial distribution with parameters Q and $1/2$. These parameters are denoted by $\eta_1, \eta_2, \ldots, \eta_{k+1}$, and it should be noted that the first one, which is an intercept term, is always present and is not included in the set of up to Q parameters which may or may not be

included in the model. A parameter vector

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_l \end{pmatrix} = \begin{pmatrix} 1 & x_{1v_1} & \cdots & x_{1v_k} \\ 1 & x_{2v_1} & \cdots & x_{2v_k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{lv_1} & \cdots & x_{lv_k} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_{k+1} \end{pmatrix} \tag{18}$$

is then defined. In our case, all the values of x_{ij} will be 1's and 0's (see Section 4). The vector $M = (v_1, \ldots v_k)$ represents the current model, which is a subset of $\{1, \ldots, Q\}$ of size k that determines which 0-1 variables $x_{\cdot j}$ to include, so that there are 2^Q models to choose between in $\mathcal{M}$.

In our case, we have incremental claims data $\mathcal{I}$ assumed to be drawn from an over-dispersed Poisson model with mean $E[C_{ij}] = m_{ij}$ and variance $\mathrm{Var}[C_{ij}] = \phi m_{ij}$. The model for the mean looks similar to (7), with

$$\log(m_{ij}) = c + \alpha_i + \beta_j, \tag{19}$$

where $\alpha_1 = 0$, $\beta_1 = 0$. However, unlike the model for the chain-ladder technique, (7), the parameters $\{\beta_j : j = 2, 3, \ldots, n\}$ will not be estimated as separate, distinct parameters. Instead, we allow for some smoothing of the run-off shape by relating them to the lower-dimensional ψ through

$$\beta_j = \begin{cases} \sum_r Z_{j,r}\psi_r, & \text{for model class 1 and } 2 \le j \le n \\ \beta_2 + \sum_r Z_{j,r}\psi_r, & \text{for model class 2 and } 3 \le j \le n, \end{cases} \tag{20}$$

where $Z = (Z_{j,r})$ is a design matrix, specific to the chosen model class and the model M within that class. Combining (18)-(20), the parameter vector for model M can thus be written

$$\theta = \begin{cases} (c, \alpha_2, \ldots, \alpha_n, \eta_1, \ldots, \eta_{k+1}), & \text{for model class 1,} \\ (c, \alpha_2, \ldots, \alpha_n, \beta_2, \eta_1, \ldots, \eta_{k+1}), & \text{for model class 2,} \end{cases}$$

where the intercept c, the accident year effects α_i, β_2 (for model class 2) and η_1 are included in all models, whereas the k optional parameters $\eta_2, \ldots, \eta_{k+1}$ vary between models. The crucial aspect is that parameters can be included or excluded within the RJ steps (17) of the MCMC sampling procedure by adding or deleting variabels $x_{\cdot j}$ in the design matrix of (18). An important part of the output is the posterior distribution for k, which gives an indication of how many optional parameters should be included, together with the information on which parameters these are.

As mentioned above the prior distribution of M is uniform, $P(M) = 2^{-Q}$ for all members of the model class. Conditional on M, the prior distributions of the optional parameters, $\eta_1, \eta_2, \ldots, \eta_{k+1}$, is set by default such that they are independently normally distributed with

$$E[\eta_1] = m, \mathrm{Var}[\eta_1] = \tau_1 \tag{21}$$

and

$$E[\eta_j] = 0, \mathrm{Var}[\eta_j] = \tau_\eta, j = 2, 3, \ldots, k+1. \tag{22}$$

The values of the hyperparameters m, τ_1 and τ_η are part of the prior specification, and we will draw them from non-informative distributions

$$\begin{aligned} m &= 0, \\ \tau_1 &= 10,000, \\ \tau_\eta^{-1} &\sim \Gamma(0.001, 0.001). \end{aligned}$$

Conditional on M, we use non-informative priors

$$\begin{aligned} c &\sim N(0, 10000), \\ \{\alpha_i : i = 2, 3, \ldots, n\} &\sim \mathrm{IN}(0, 10000), \\ \beta_2 &\sim N(0, 10000), \quad \text{for model class 2.} \end{aligned}$$

for the fixed parameters.

4 Examples of Trans-dimensional Models Classes

In this section, the model classes which are applied to the run-off triangle in section 5 are specified. They are both based on the chain-ladder model and the exponential-tail model briefly described in the introduction.

As outlined in section 2, the underlying assumption that we use in this paper is that the development follows an exponential decay. It will not follow this exactly, because the models will allow departures from this. However, this is the underlying shape of the run-off. This means that the log of the run-off follows a straight line, in which case, the second differences of the parameters $\{\beta_j : j = 2, 3, \ldots, n\}$ will be close to zero. Thus, when applying the RJMCMC model in section 3, we will consider the differences

$$\nabla\beta_j = \beta_j - \beta_{j-1}, \quad j = 2, 3, \ldots, n$$

and second differences

$$\nabla^2\beta_j = \beta_j - 2\beta_{j-1} + \beta_{j-2}, \quad j = 3, 4, \ldots, n$$

of these parameters. Note that $\nabla\beta_j$ measures the gradient of the log of the development pattern, and $\nabla^2\beta_j$ measures the change in the gradient. Thus, if $\nabla^2\beta_j$ is zero, the log development pattern follows a straight line, and non-zero values of $\nabla^2\beta_j$ indicate departures from this. We now define two Bayesian model classes for these parameters.

4.1 Model Class 1

This model is a straightforward model for the second differences of the parameters for the run-off, $\nabla^2\beta_j$. It is also necessary to include the first parameter, β_2, and the initial gradient, $\beta_3 - \beta_2$. Thus, for the full model ($k = Q = n-2$) within this model class, the optional parameters in (18) are defined as

$$\begin{aligned} \eta_1 &= \beta_2, \\ \eta_2 &= \nabla\beta_3 = \beta_3 - \beta_2, \\ \eta_j &= \nabla^2\beta_{j+1} = \beta_{j+1} - 2\beta_j + \beta_{j-1}, \quad j = 3, 4, \ldots, n-1. \end{aligned}$$

Note that η_2 measures the initial gradient of the log of the development pattern, and after that η_j measures the change in the gradient. Thus, if the RJMCMC model indicates that a optional parameter should not be included in the model, it indicates that the log development factors follow a straight line in this region of the run-off.

The first parameter $\eta_1 = \beta_2$ in (4.1) is always included. The other parameters $\beta_3 - \beta_2$ and $\{\nabla^2\beta_{j+1}\}$ are chosen in accordance as specied by the model at hand, where some, or all, of these parameters can be removed from the model in RJMCMC moves (17).

The next step is to specify (18). For the full model ($k = Q = n - 2$) we

do this as

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 1 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & 0 \\ 1 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_2 \\ \beta_3 - \beta_2 \\ \nabla^2 \beta_4 \\ \vdots \\ \nabla^2 \beta_n \end{pmatrix}. \tag{23}$$

with a separate parameter included for the initial gradient and the change in the gradient for each development year from 3 upwards. When $k < Q$, some of these parameters are not included in the model but otherwise the definition of ψ is analogous.

The next step is to express the development parameters β_j in terms of ψ, cf. (20). For the full model, this is achieved by inverting equation (23):

$$\begin{aligned} \beta_2 &= \psi_1 \\ \beta_3 &= \psi_2 \\ \beta_j &= 2\beta_{j-1} - \beta_{j-2} + \psi_{j-1} - \psi_1, \quad j = 4, 5, \ldots, n. \end{aligned}$$

The most obvious way to extrapolate in order to estimate tail factors for the run-off is simply to assume that the second differences of the parameters β_j are 0 beyond the latest development year:

$$\nabla^2 \beta_j = 0 \text{ for } j = n+1, n+2, \ldots$$

With this model for the tail, the parameters can be obtained from

$$\beta_j = 2\beta_{j-1} - \beta_{j-2}, \quad j = n+1, n+2, \ldots.$$

4.2 Model Class 2

This model is again constructed around the premise that the default option should be that the second differences of the parameters $\beta_2, \beta_3, \ldots, \beta_n$ are close to zero. However, in this model, we model the first differences of these parameters directly. The advantage of this is that the model then extropolates using a weighted average of these first differences, perhaps providing

more credible tail factors (as is illustrated in the example in the following section). For ease of exposition, we denote the first differences by

$$b_j = \nabla\beta_j = \beta_j - \beta_{j-1}, \quad j = 3, 4, \ldots, n. \tag{24}$$

The parameters we use in the trans-dimensional model are these first differences, and we model them as follows:

$$\begin{array}{rcl} b_3 & = & \eta_1 \\ b_j & = & b_{j-1} + \delta_j, \quad j = 4, 5, \ldots, n. \end{array} \tag{25}$$

So η_1 is the underlying gradient and δ_j is the change in the gradient for $2 \leq j \leq n$. It is the values of δ_j that may be zero: they are equivalent to the second differences $\nabla^2\beta_j$ in model class 1. From (25), it can be seen that

$$b_j = \eta_1 + \sum_{k=4}^{j} \delta_k, \quad j = 3, 4, \ldots, n. \tag{26}$$

We model β_2 separately as a fixed parameter, it is thus not included among the optional parameters.

For the full model ($k = Q = n - 3$) of this model class, the parameters in (18) are defined in terms of the first differences, and

$$\begin{array}{rcl} \eta_1 & = & b_3 \\ \eta_j & = & \delta_{j+2}, \quad j = 2, 3, \ldots, n-2, \end{array} \tag{27}$$

and in this case (18) becomes

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{n-2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 1 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & 0 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \vdots \\ \eta_{n-2} \end{pmatrix}. \tag{28}$$

It is then straightforward to obtain the parameters $\beta_2, \beta_3, \ldots, \beta_n$ from ψ, since $\beta_j = \beta_{j-1} + b_j = \beta_{j-1} + \psi_{j-2}$, $j = 3, 4, \ldots, n$.

For this model, we extrapolate beyond the latest development year to obtain tail factors by assuming that $\delta_j = 0$ for $j = n+1, n+2, \ldots$, so that $b_j = b_n$ for $j = n+1, n+2, \ldots$ Hence

$$\beta_j = \beta_n + (j - n)\, b_n \text{ for } j = n+1, n+2, \ldots \tag{29}$$

5 Examples

We illustrate this method using the data from Taylor and Ashe (1983), which are shown in Table 1, along with the results from the chain-ladder technique in Tables 2 and 3 for comparison purposes. As was explained in section 3.2, we will use (13) and its approximation (15) to estimate the predictive distributions. Hence, the posterior probablities for each model are of greater interest than which particular model was most likely (a posteriori).

i\j	1	2	3	4	5	6	7	8	9	10
1	357,848	766,940	610,542	482,940	527,326	574,398	146,342	139,950	227,229	67,948
2	352,118	884,021	933,894	1,183,289	445,745	320,996	527,804	266,172	425,046	
3	290,507	1,001,799	926,219	1,016,654	750,816	146,923	495,992	280,405		
4	310,608	1,108,250	776,189	1,562,400	272,482	352,053	206,286			
5	443,160	693,190	991,983	769,488	504,851	470,639				
6	396,132	937,085	847,498	805,037	705,960					
7	440,832	847,631	1,131,398	1,063,269						
8	359,480	1,061,648	1,443,370							
9	376,686	986,608								
10	344,014									

Table 1: Incremental claims data from Taylor and Ashe (1983).

j	f_j
2	3.4906
3	1.7473
4	1.4574
5	1.1739
6	1.1038
7	1.0863
8	1.0539
9	1.0766
10	1.0177

Table 2: Chain-ladder development factors.

j	
2	94,634
3	469,511
4	709,638
5	984,889
6	1,419,459
7	2,177,641
8	3,920,301
9	4,278,972
10	4,625,811
Overall	18,680,856

Table 3: Chain-ladder reserve estimates.

Fitting the chain-ladder over-dispersed Poisson model, (5), (6) and (7), gives the maximum likelihood parameter estimates shown in Table 4.

$\hat{c}$	12.5063		
$\hat{\alpha}_2$	0.3313	$\hat{\beta}_2$	0.9216
$\hat{\alpha}_3$	0.3212	$\hat{\beta}_3$	0.9589
$\hat{\alpha}_4$	0.3060	$\hat{\beta}_4$	1.0261
$\hat{\alpha}_5$	0.2194	$\hat{\beta}_5$	0.4353
$\hat{\alpha}_6$	0.2701	$\hat{\beta}_6$	0.0801
$\hat{\alpha}_7$	0.3723	$\hat{\beta}_7$	-0.0063
$\hat{\alpha}_8$	0.5534	$\hat{\beta}_8$	-0.3944
$\hat{\alpha}_9$	0.3690	$\hat{\beta}_9$	0.0094
$\hat{\alpha}_{10}$	0.2421	$\hat{\beta}_{10}$	-1.3799

Table 4: Maximum likelihood estimates of the parameters of the over-dispersed Poisson model.

It is possible to construct estimates of outstanding liabilities for as many future develop years as are required. In this examples, we will consider 5

more development years beyond the triangle: in other words, we will consider forecasts up to development year 15.

5.1 Model Class 1

We implemented this model using the Reversible Jump add-in of Lunn et al (2000) in winBUGS, and used an initial burn-in of 50,000 updates, followed by a sample of 50,000 updates (with no thinning). The indications are that there are no parameters which should definitely be excluded, showing that the development parameters do not follow an exponential run-off very closely. However, Table 5 lists the optional parameters of the full model and shows that some parameters less likely to be included than others. Parameter number 1 is $\beta_3 - \beta_2$ and parameter number j $(j = 2, 3, \ldots, 8)$ is $\nabla^2 \beta_{j+2}$.

Optional Parameter Number	Marginal Probability
1	0.26022
2	0.30746
3	1
4	0.27546
5	0.73968
6	0.43524
7	0.4741
8	0.57476

Table 5: Posterior marginal probabilities for the optional parameters in Model class 1.

It can be seen from this that the more sensible approach is to use the weighted averages of the models, rather than simply choosing the most likely model. Figure 1 shows the development parameters for the chain-ladder technique, together with the estimates of the means of the posterior distributions of these from Model class 1. As can be seen, these appear to provide a suitable amount of smoothing, together with acceptable extrapolated values.

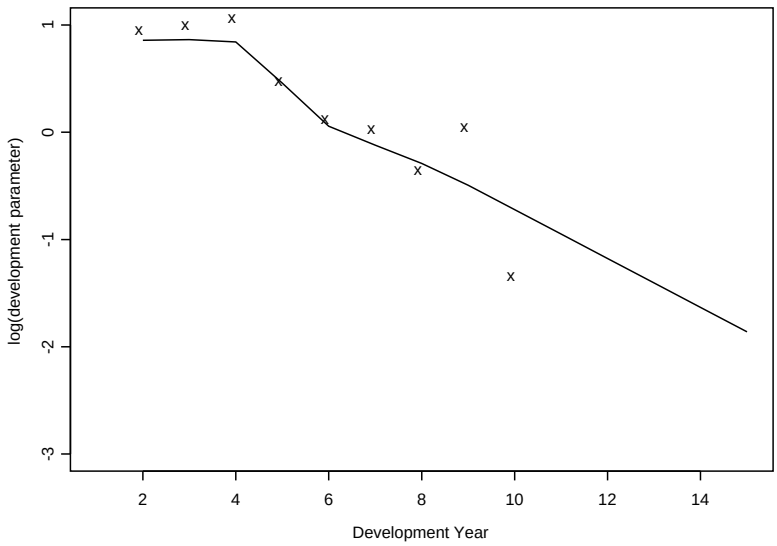

Figure 1: Development parameters for Model class 1, together with chain-ladder estimates, on the log scale

The Bayes estimates of c and α_i, for instance

$$\hat{\alpha}_i = E(\alpha_i(M, \theta)|\mathcal{I}) \approx \frac{1}{N} \sum_{a=1}^{N} \alpha_i(M^{(B+a)}, \theta^{(B+a)}),$$

remain broadly similar to the maximum likelihood estimates for the over-dispersed Poisson model, as shown in Table 6.

	ODP	Model Class 1		ODP	Model class 1
$\hat{c}$	12.5063	12.56771			
$\hat{\alpha}_2$	0.3313	0.3531	$\hat{\beta}_2$	0.9126	0.8584
$\hat{\alpha}_3$	0.3212	0.3184	$\hat{\beta}_3$	0.9589	0.8639
$\hat{\alpha}_4$	0.3060	0.3128	$\hat{\beta}_4$	1.0261	0.8428
$\hat{\alpha}_5$	0.2194	0.2265	$\hat{\beta}_5$	0.4353	0.4579
$\hat{\alpha}_6$	0.2701	0.2818	$\hat{\beta}_6$	0.0801	0.0559
$\hat{\alpha}_7$	0.3723	0.3998	$\hat{\beta}_7$	-0.0063	-0.1197
$\hat{\alpha}_8$	0.5534	0.5449	$\hat{\beta}_8$	-0.3944	-0.2900
$\hat{\alpha}_9$	0.3690	0.3267	$\hat{\beta}_9$	0.0094	-0.4940
$\hat{\alpha}_{10}$	0.2421	0.0978	$\hat{\beta}_{10}$	-1.3799	-0.7217

Table 6: Comparison of estimates of the parameters from Model class 1 and the over-dispersed Poisson model.

Table 7 shows the estimates

$$\hat{D}_{im} - D_{i,n+1-i} = E(D_{im}|\mathcal{I}) - D_{i,n+1-i} \approx \frac{1}{N}\sum_{a=1}^{N} D_{im}^{(B+a)} - D_{i,n+1-i}$$

of outstanding claims together with their prediction errors

$$\text{PE} = \frac{\text{Var}(D_{im}|\mathcal{I})}{\hat{D}_{im} - D_{i,n+1-i}} \approx \frac{\sum_{a=1}^{N}(D_{im}^{(B+a)} - \hat{D}_{im})^2/(N-1)}{\hat{D}_{im} - D_{i,n+1-i}}$$

for the over-dispersed Poisson model and Model class 1 without $(m = n)$ and with $(m = n + 5)$ the tail factors.

	ODP		Model class 1 without tail		Model class 1 with tail	
Row	Estimate	PE	Estimate	PE	Estimate	PE
1					494,057	87%
2	94,634	116%	208,916	33%	917,031	76%
3	469,511	46%	449,830	27%	1,137,396	63%
4	709,638	37%	746,860	23%	1,430,929	52%
5	984,889	31%	1,010,058	20%	1,640,613	44%
6	1,419,459	26%	1,478,397	18%	2,144,523	36%
7	2,177,641	23%	2,349,978	17%	3,099,616	30%
8	3,920,301	20%	3,884,822	16%	4,752,738	25%
9	4,278,972	24%	4,122,598	21%	4,827,815	26%
10	4,625,811	43%	4,295,595	40%	4,894,029	43%
Total	18,680,856	16%	18,547,060	14%	25,338,750	30%

Table 7: Estimates of outstanding claims from Model class 1 and the over-dispersed Poisson model, together with their prediction errors (PE).

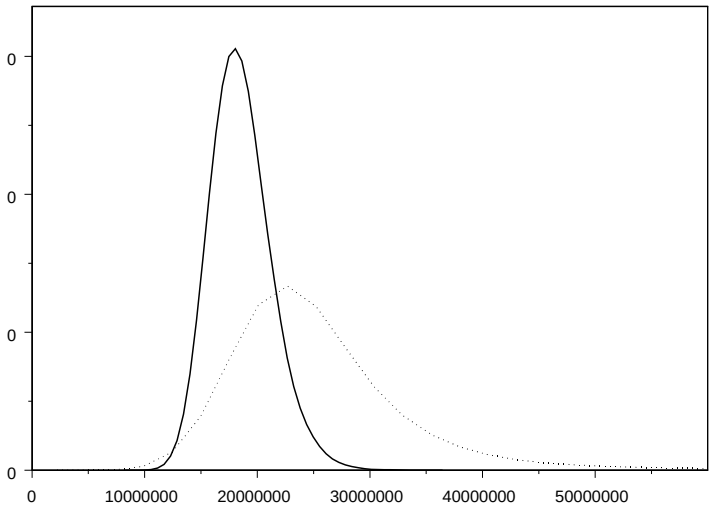

Figure 2: Predictive distribution, with and without the tail, for Model class 1

Figure 2 shows the predictive distribution for total outstanding claims, with and without the tail factors.

5.2 Model class 2

The results from Model class 2 are similar to those from Model class 1, but the parameters may be more straightforward to interpret in a practical context. For example, the estimate of the parameters ψ_j give an immediate idea of the gradient of the log of the development pattern at each point. Again, we implemented this model in winBUGS using an initial burn-in of 50,000 updates, followed by a sample of 50,000 updates without thinning.

As with Model class 1, the results indicate that none of the optional parameters $\delta_4, \ldots, \delta_n$, should definitely be excluded. The posterior marginal probabilities for each parameter, as shown in Table 8 give an indication of where the run-off pattern departs from an exponential decay.

Nr.	Optional parameter	Marginal probability
1	δ_2	0.32748
2	δ_3	0.95402
3	δ_4	0.1842
4	δ_5	0.21566
5	δ_6	0.53034
6	δ_7	0.19792
7	δ_8	0.71406

Table 8: Posterior marginal probabilities for the optional parameters of Model class 2.

Bayes estimates of parameters for Model class 2, together with maximum likelihood estimates of the over-dispersed Poisson chain-ladder model are shown in Table 9.

	ODP	Model class 2		ODP	Model class 2
$\hat{c}$	12.5063	12.4984			
$\hat{\alpha}_2$	0.3313	0.3414	$\hat{\beta}_2$	0.9126	0.9794
$\hat{\alpha}_3$	0.3212	0.3054	$\hat{\beta}_3$	0.9589	0.9248
$\hat{\alpha}_4$	0.3060	0.2982	$\hat{\beta}_4$	1.0261	0.8448
$\hat{\alpha}_5$	0.2194	0.2111	$\hat{\beta}_5$	0.4353	0.5656
$\hat{\alpha}_6$	0.2701	0.2856	$\hat{\beta}_6$	0.0801	0.2861
$\hat{\alpha}_7$	0.3723	0.4147	$\hat{\beta}_7$	-0.0063	0.0066
$\hat{\alpha}_8$	0.5534	0.5376	$\hat{\beta}_8$	-0.3944	-0.2277
$\hat{\alpha}_9$	0.3690	0.3048	$\hat{\beta}_9$	0.0094	-0.4574
$\hat{\alpha}_{10}$	0.2421	0.1677	$\hat{\beta}_{10}$	-1.3799	-0.77912

Table 9: Comparison of estimates of the parameters from Model class 2 and the over-dispersed Poisson model.

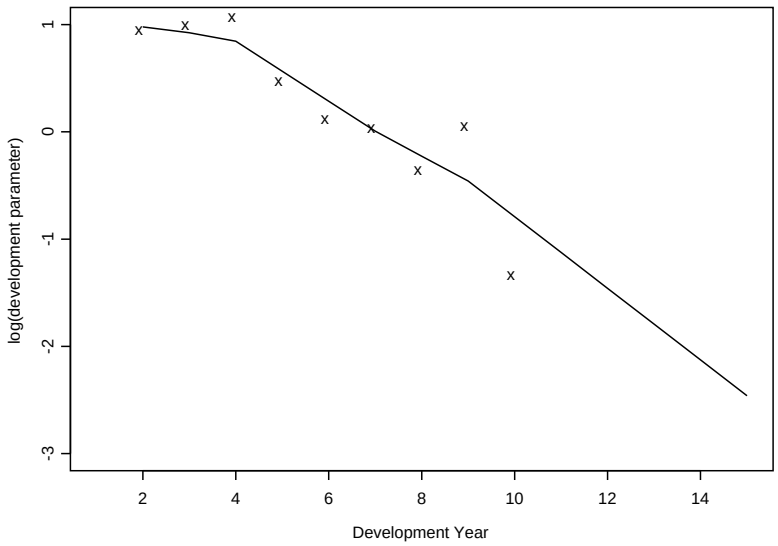

Figure 3: Development Parameters for Model class 1, together with chain-ladder estimates, on the log scale

Figure 3 shows a comparison of the development parameters for Model class 1 and the chain-ladder technique. It is also interesting to compare the

development parameters for models classes 1 and 2, and Figure 4 shows this comparison, together with the estimates from the chain-ladder model, up to development year 10.

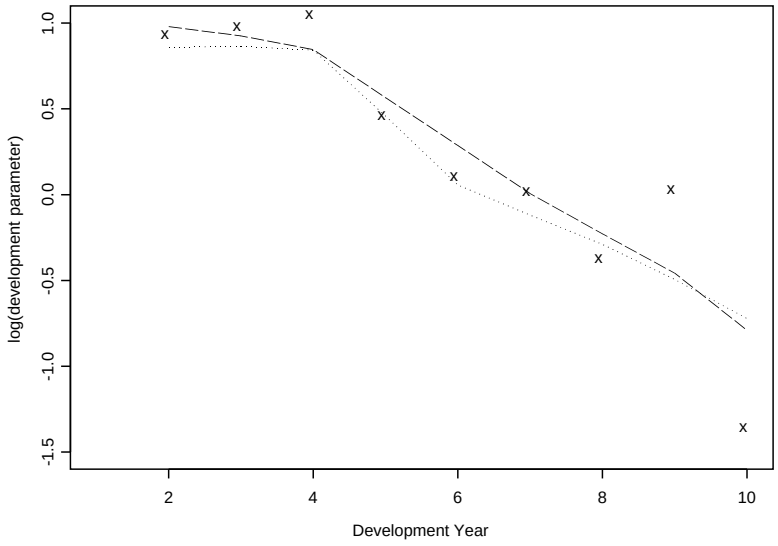

Figure 4: Development Parameters for Model class 1 (dotted line), Model class 2 (dashed line) together with chain-ladder estimates, on the log scale

It is difficult to argue that one model is better than the other. However, the extrapolated values of the parameters can have a significant impact on the tail factors. Since the estimate of β_{10} from the chain-ladder model is quite low, this affects the gradient of the extrapolated values. The effect for Model class 2 is greater than for Model class 1, and it can be seen from Figure 5 that this affects the way the development pattern is extrapolated.

Table 10 shows the estimates of outstanding claims together with their prediction errors, for the over-dispersed Poisson model and Model class 2 (without and with the tail factors).

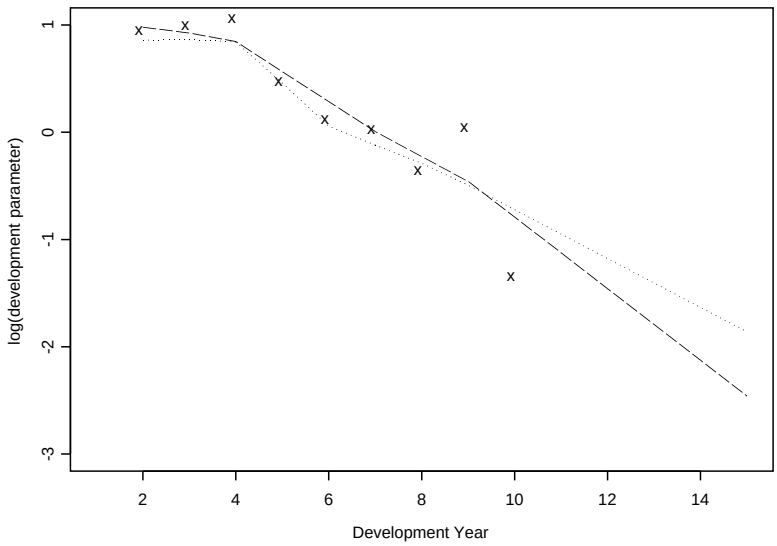

Figure 5: Development Parameters for Model class 1 (dotted line), Model class 2 (dashed line), on the log scale

	ODP		Model class 1 without tail		Model class 1 with tail	
Row	Estimate	PE	Estimate	PE	Estimate	PE
1					268,715	40%
2	94,634	116%	176,280	25%	552,946	34%
3	469,511	46%	406,755	24%	771,530	31%
4	709,638	37%	696,168	21%	1,058,918	26%
5	984,889	31%	976,076	18%	1,308,912	22%
6	1,419,459	26%	1,531,942	17%	1,891,558	20%
7	2,177,641	23%	2,464,137	15%	2,873,533	17%
8	3,920,301	20%	3,870,701	15%	4,333,558	16%
9	4,278,972	24%	4,035,723	21%	4,406,074	21%
10	4,625,811	43%	4,679,039	42%	5,022,804	42%
Total	18,680,856	16%	18,836,820	14%	22,488,550	15%

Table 10: Estimates of outstanding claims from Model class 2 and the over-dispersed Poisson model, together with their prediction errors (PE).

The effect of the different extrapolated values can be seen from the estimates of outstanding claims with the tail factors for Model class 2 in Table 10, which are lower than the corresponding values in Table 7. Figure 6 shows that the estimated predictive distributions of total outstanding claims without tail factors is similar for the two models. However, the difference between the estimates of the predictive distributions of total outstanding claims when tail factors are included are quite different, as shown in Figure 7.

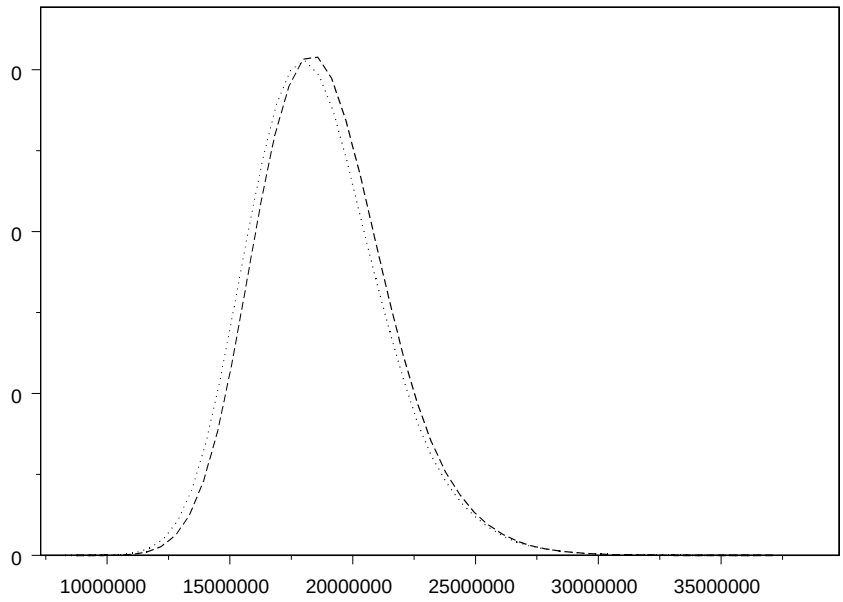

Figure 6: Predictive distribution for total outstanding claims, without tail factors for Model class 1 (dotted line) and Model class 2 (dashed line)

6 Conclusions

This paper has set out a new approach to modelling claims run-off traingles, using reversible jump Markov chain Monte Carlo methods. The advantage of this new method is that it is objective, and can be used to replace the

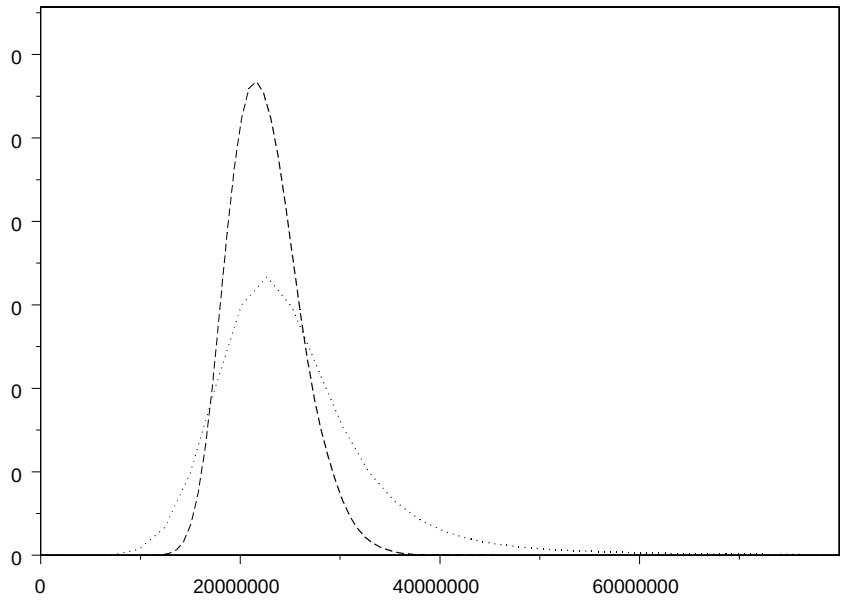

Figure 7: Predictive distribution for total outstanding claims, with tail factors for Model class 1 (dotted line) and Model class 2 (dashed line)

ad hoc procedures used in practice. The indications from the examples were that Model class 2 might be preferable for practical purposes, although it is not clear whether this is simply the case for the particular set of data used in this paper. We believe that this method has great potential for further application in this area.

References

Björkwall, S., Hössjer, O., Ohlson, E. and Verrall, R. (2010). A generalized linear model with smoothing effects for claims reserving. Report 2010:8, Mathematical Statistics, Stockholm University. Submitted.

Congdon, P. (2006). *Bayesian Statistical Modelling*, John Wiley.

England, P.D. and Verrall, R.J. (2002). Stochastic Claims Reserving in General Insurance (with discussion). *British Actuarial Journal*, 8, 443-544

England, P.D. and Verrall, R.J. (2006). Predictive Distributions of Outstanding Liabilities in General Insurance. *Annals of Actuarial Science*, Vol 1, pp 221-270.

Gelfand A.E. and Smith A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85:398-409.

Gelman A., Carlin J.B., Stern H.S., and Rubin D.B. (1995). *Bayesian Data Analysis*. Chapman and Hall, London.

Geman S. and Geman D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6:721-741.

Gilks, W.R., Richardson S., and Spiegelhalter D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* Vol. 82, pp 711-732.

Lunn, D.J., Best, N. and Whittaker, J.C. (2009) Generic reversible jump MCMC using graphical models. *Statistics and Computing*, Vol. 19, pp 395-408.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10, pp325-337.

Renshaw, A.E. and Verrall, R.J. (1998) A stochastic model underlying the chain ladder technique. *British Actuarial Journal*, 4, IV, pp 903-923

Scollnik, D.P.M. (2001). Actuarial modelling with MCMC and BUGS. *North American Actuarial Journal*, 5 (2), 96-125.

Taylor, G.C. and Ashe, F.R. (1983) Second moments of estimates of outstanding claims. *Journal of Econometrics* 23, pp37-61.

Verrall (2007) Obtaining Predictive Distributions for Reserves Which Incorporate Expert Opinion. *Variance* (Casualty Actuarial Society Journal), Vol. 1, pp 53-80.