City Research Online

City, University of London Institutional Repository

Citation: Verrall, R. J., Hossjer, O. & Bjorkwall, S. (2012). Modelling Claims Run-off with Reversible Jump Markov Chain Monte Carlo Methods. ASTIN Bulletin, 42(1), pp. 35-58. doi: 10.2143/AST.42.1.2160711

This is the unspecified version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/3800/

Link to published version: https://doi.org/10.2143/AST.42.1.2160711

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

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
and
Susanna Björkwall
Dept. of Mathematics, Stockholm University

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. An example is given, illustrating the results in a practical context, and the WinBUGS code is supplied.

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). Makov (2001) provides a review of Bayesian methods in actuarial science and England and Verrall (2006) is 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) and Ntzoufras (2009) contain many such examples. In the actuarial literature, they have been used by Scollnik (2001), Ntzoufras and Dellaportas (2002) (see also the discussion by Scollnik, 2002, and de Alba, 2002), Verrall (2007) and Wüthrich (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).

Other papers to have considered RJMCMC methods and related issues include Ntzoufras et al. (2004), Katsis and Ntzoufras (2005), Peters et al (2009) and Verrall and Wüthrich (2010). This latter paper also considers modelling the run-off shape in claims reserving but uses a different approach from the current paper. There are two main differences between the approach adopted in this paper and that of Verrall and Wüthrich (2010). The first is that this paper uses winBUGS rather than a specially written programme for this application. The advantage of using winBUGS is that it makes it easier for users to consider how to adapt the approach to their own needs; the disadvantage is that it is more of a "black box" and does not facilitate changing things such as the sampler used. Thus, while we acknowledge that it is important to be aware of the potential difficulties associated with "mixing" between model subspaces in a RJMCMC chain, we simply have to accept the sampling approach which winBUGS employs. The other difference is in the actual model of the run-off shape. In this paper the underlying model is piece-wise linear (on the log scale), but can be composed of a number of "pieces", whereas in Verrall and Wüthrich (2010), it is assumed that the run-off is simply linear beyond a certain development period.

The paper is set out as follows. In Section 2, we 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 contains an example to illustrate the method, and Section 5 contains the conclusions.

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.

2.1 Stochastic models for claims data in a run-off triangle

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, in general terms, 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). Of course, it would also be possible to use other distributions for the data, but the over-dispersed Poisson distribution is sufficient for the purposes of this paper and has the advantage of giving the same estimates of outstanding claims if the appropriate model is used for the mean. 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

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

For the Bayesian approach used in this paper, it is necessary to specify the full distributional assumptions. For the over-dispersed Poisson model, this can be expressed as

$$\frac{C_{ij}}{\phi}|m_{ij} \sim \text{Poisson}\left(\frac{m_{ij}}{\phi}\right) \tag{7}$$

where $\phi > 0$ is assumed to be a known constant.

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

$$\log\left(m_{ij}\right) = c + \alpha_i + \beta_j \tag{8}$$

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 $\alpha_1 = \beta_1 = 0$. Throughout this paper, we are concerned with the modelling of the "development pattern". The development pattern can be considered through a number of different parameterisations, but in this paper we are considering specifically the log-development parameters, β_j.

The scale parameter, ϕ is usually then treated as a "plug-in" estimate and not counted as a parameter. In a fully Bayesian model, it might be considered appropriate to specify distributional assumptions for ϕ, but this is not something addressed in this paper. For this reason, we follow what is usually done with classical estimation methods and simply replace the dispersion parameter by an estimate. 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 (8) 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 (8). 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:

$$\log(m_{ij}) = c + \alpha_i + \beta \log(j) + \gamma j \tag{9}$$

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 (9), 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 (8) for the early delay years, but then assumes that $\log(m_{ij})$ follows a straight line for the later development years:

$$\log(m_{ij}) = \begin{cases} c + \alpha_i + \beta_j & \text{for } j \leq k \\ c + \alpha_i + \gamma j & \text{for } j > k \end{cases} \tag{10}$$

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. (2011) recently proposed choosing k automatically in (10) 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 to (though not the same as) (10), and employ Bayesian estimation methods. The models used in this paper for the run-off pattern are defined in Section 3.3. However, before we can define the specific models used, we need to consider the parameters for the run-off shape, $\{\beta_j : j = 2, 3, \ldots, n\}$, and we do this in the Section 2.2.

2.2 A re-parameterisation of the run-off shape

In this section, we re-parameterise the run-off shape so that the trans-dimensional approach in Section 3.3 can be applied. This model is based on the chain-ladder model with an underlying exponentially-decaying tail. Although this is the basis for the model, the tail will not follow this exactly, because the Bayesian model 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. Note that the requirement for the application of this approach is that the parameters should be "tested" to see whether they could assumed to be zero. Thus, any parameterisation which had parameters that could be compared with zero would be appropriate. In this section we use what we believe is the most obvious choice for this parameterisation, but it is possible that there are other such parameterisations which could also be investigated. Thus, we 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. A matrix representation of the relationship between the first and second differences of the development parameters is given below.

$$\begin{pmatrix} \nabla\beta_3 \\ \nabla\beta_4 \\ \vdots \\ \nabla\beta_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 1 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & 1 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & 0 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix} \begin{pmatrix} \nabla\beta_3 \\ \nabla^2\beta_4 \\ \nabla^2\beta_5 \\ \vdots \\ \nabla^2\beta_n \end{pmatrix}. \tag{11}$$

This representation will be used in the trans-dimensional Bayesian model and coincides with Lunn et al. (2009). Note also that it is straightforward to obtain the parameters $\beta_3, \beta_4, \ldots, \beta_n$, since $\beta_j = \beta_{j-1} + \nabla\beta_j$, $j = 3, 4, \ldots, n$.

For forecasting purposes, we will extrapolate beyond the latest development year to obtain tail factors by assuming that $\nabla^2\beta_j = 0$ for $j = n+1, n+2, \ldots$, so that

$$\beta_j = 2\beta_{j-1} - \beta_{j-2} \text{ for } j = n+1, n+2, \ldots \tag{12}$$

For estimation purposes, it is sufficient to estimate either $\{\beta_j\}_{j=2}^n$ or β_2, $\nabla\beta_3$ and $\{\nabla^2\beta_j\}_{j=3}^n$: if maximum likelihood estimation were used, we would get exactly the same results for outstanding claims etc. The reason for considering the latter set of parameters is that it means that it is possible to apply the trans-dimensional Bayesian models to $\{\nabla^2\beta_j\}_{j=3}^n$. This is considered in Section 3.3, but first an introduction to Bayesian modelling using MCMC methods is given in the following Section.

3 Bayesian Models, RJMCMC and WinBUGS

This section contains a very brief overview of the Bayesian techniques used to give priority between models, which we then use to estimate the models which we apply to 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

$$f(M, \theta|\mathcal{I}) \propto f(\mathcal{I}|\theta, M)\, f(\theta|M)\, f(M). \tag{13}$$

In the context of Section 2.2, model M could be indentified with a subset of the integers $\{4, 5, \ldots, n\}$ and the parameter vector for model M is

$$\theta = (c, \alpha_2, \ldots, \alpha_n, \beta_2, \nabla\beta_3, \nabla^2\beta_j; j \in M),$$

of which c, $\{\alpha_i; i = 2, 3, \ldots, n\}$, β_2 and $\nabla\beta_3$ are always included, whereas each one of the second difference parameters could either be included or excluded. The model class $\mathcal{M}$ thus consists of 2^{n-3} models, depending on which second difference parameters we choose to include or exclude.

3.1 Forecasting with trans-dimensional models

In many statistical applications, the main purpose is to identify the best model 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

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

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. The first case is known as the "maximum a posteriori" (MAP) estimator, and the forecast distribution is given by

$$f\left(C_{il}|\mathcal{I}\right) \approx f\left(C_{il}|M_{\max}, \mathcal{I}\right) \tag{15}$$

where $M_{\max} \in \mathcal{M}$ maximises $P\left(M|\mathcal{I}\right)$ among $M \in \mathcal{M}$. The second case, which can be referred to as "Bayesian model averaging" (BMA), gives the following forecast distribution

$$f\left(C_{il}|\mathcal{I}\right) = \sum_{M \in \mathcal{M}} f\left(C_{il}|M, \mathcal{I}\right) P\left(M|\mathcal{I}\right). \tag{16}$$

In this paper, we use (16) 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. For a more extensive discussion of the relative advantages of BMA, see Hoeting et al. (1999). Often the main interest is in predicting cumulative claims, in particular their sum over all accident years i. The procedure to obtain the predictive distribution is then the same as in (14) and (16), provided we replace C_{il} by the quantity we wish to predict.

3.2 Reversible Jump MCMC

The first step is to combine (14) and (16) and formally rewrite the predictive distribution of an outstanding incremental claim as

$$f\left(C_{il}|\mathcal{I}\right) = \int f(C_{il}|M,\theta) f(M,\theta|\mathcal{I}) d(M,\theta), \tag{17}$$

integrating over the posterior distribution in (13). 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 (13). One then approximates the predictive distribution (17) by a Monte Carlo average

$$f\left(C_{il}|\mathcal{I}\right) \approx \frac{1}{N}\sum_{a=1}^{N} f(C_{il}|M^{(B+ta)},\theta^{(B+ta)}), \tag{18}$$

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 is 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

$$P(\text{accept } (M, \theta)) = \min\left(1, \frac{f(\theta|M^{(b)}, \mathcal{I})\pi(\theta^{(b)}|\theta)}{f(\theta^{(b)}|M^{(b)}, \mathcal{I})\pi(\theta|\theta^{(b)})}\right). \tag{19}$$

The reversible jump(RJ) 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

$$P(\text{accept } (M, \theta)) = \min\left(1, \frac{f(M, \theta|\mathcal{I})\pi(M^{(b)}, \theta^{(b)}|M, \theta)}{f(M^{(b)}, \theta^{(b)}|\mathcal{I})\pi(M, \theta|M^{(b)}, \theta^{(b)})} \left|\frac{\partial(\theta, v)}{\partial(\theta^{(b)}, u)}\right|\right), \tag{20}$$

where the major difference compared to (19) is the presence of the Jacobian term, which accounts for different parameterisations 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. For instance, Roberts and Rosenthal, (2007) have shown that any such adaptive MCMC algorithms that satisfy certain bounded convergence and diminishing adaptation conditions yield ergodic Markov chains and convergent estimators. Roberts and Rosenthal (2009) investigate a number of adaptive MCMC algorithms by simulation and notice that they generally have better performance than the nonadaptive ones. The trandimensional adaptive learning algorithm of Nevat et al. (2009) successfully employs stochastic approximation in order to improve mixing of the RJMCMC chain. Fan et al. (2009) propose the use of marginal density estimators to construct between-model proposal distributions for moves that alter M.

3.3 Trans-Dimensional Models in WinBUGS

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 and which is described in this section.

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. Thus, we use the re-parametrisation of the run-off shape specified in Section 2.2, and use $\{\nabla^2\beta_j : j = 4, 5, \ldots, n\}$ as the "entities of interest". The idea of trans-dimensional models is that the number of these parameters which should be included in the model is not known a priori. Thus, the number of parameters to be included is denoted by k, and is regarded as another parameter whose distribution must be assigned as part of the prior specification. A priori,

each of the second differences is equally likely to be included or excluded, and hence the prior distribution of k is specified as a binomial distribution with parameters $n-3$ and $1/2$. 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.

With the parameterisation of the model given in Section 2.2 we define ψ as follows:

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{n-2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 1 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & 1 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & 0 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix} \begin{pmatrix} \nabla\beta_3 \\ \nabla^2\beta_4 \\ \nabla^2\beta_5 \\ \vdots \\ \nabla^2\beta_n \end{pmatrix}. \tag{21}$$

Note that this notation, in particular ψ, has been chosen to coincide with Lunn et al. (2009) enabling a direct comparison to be made with that paper. Note also that $\psi_j = \nabla\beta_{j+2}$ for $j = 1, 2, \ldots, n-2$ and the parameters $\beta_3, \beta_4, \ldots, \beta_n$ can be obtained from ψ as $\beta_j = \beta_{j-1} + \psi_{j-2}$, $j = 3, 4, \ldots, n$.

With this parameterisation, it is possible to construct a Bayesian model using trans-dimensional models in WinBUGS where each second difference of the run-off parameters, $\{\nabla^2\beta_j : j = 4, 5, \ldots, n\}$, is treated as an "optional parameter" which can either be included or excluded. In more detail, the algorithm alternates between the following three types of moves:

Dimension moves:	Propose new $\lvert M\rvert$, M and θ in this order.
Configuration moves:	Propose new M and θ in this order, with $\lvert M\rvert$ fixed.
Coefficients moves:	Propose a new θ, with M fixed.

The model is completed by the specification of the prior distributions, which are given in the following section.

3.4 Specification of the Prior Distributions

The model consists of the distribution of the data, given the parameters,(7), together with prior distributions for all the parameters. For a trans-dimensional model, it is necessary first to condition on the current model, M. As mentioned above, the prior distribution of M is uniform, $P(M) = 2^{-(n-3)}$ for all members of the model class. Since we are using the trans-dimensional modelling procedure specified by Lunn et al. (2009) and implemented in WinBUGS, some of the prior distributions are specified by default. In particular, conditional on M, the prior distributions of the optional parameters, $\{\nabla^2\beta_j : j = 4, 5, \ldots, n\}$, is set by default such that they are independently normally distributed with

$$E\left[\nabla^2\beta_j\right] = 0, \mathrm{Var}\left[\nabla^2\beta_j\right] = \tau. \tag{22}$$

We specify a prior distribution for the hyperparameter τ which is an inverse-gamma distribution whose variance, 1000, is large:

$$\tau^{-1} \sim \Gamma\left(0.001, 0.001\right). \tag{23}$$

Conditional on M, we use non-informative priors distributions for the remaining parameters which are all normal with mean zero and variance 10,000:

$$\begin{array}{rcl} c & \sim & N(0, 10,000), \\ \{\alpha_i : i = 2, 3, \ldots, n\} & \sim & \text{independent N}(0, 10,000), \\ \beta_2 & \sim & N(0, 10,000), \\ \nabla\beta_3 & \sim & N(0, 10,000). \end{array} \tag{24}$$

Whenever a Bayesian approach is used, it is necessary to ensure that the prior distributions are chosen appropriately. In many cases, as here, non-informative prior distributions are used and it is expected that these will therefore not influence the results, which should just depend on the data and the structure of the model. However, it is well-known that even a seemingly non-informative prior distribution can have some effect on the data. This is known as the Jeffreys-Bartlett-Lindley paradox and it means that, even though a non-informative prior distribution is used, the form of the prior distribution chosen can result in slightly different results being obtained. For

a more detailed discussion in the context of claims reserving and the chain-ladder model, we refer to England et al. (2010), which investigated various forms of the prior distributions and showed that these could result in small differences in the estimates of outstanding claims. In the context of this paper, we believe that the choices of prior distributions are appropriate to illustrate the method. Finally, it should be noted that the variance of the prior normal distributions is chosen to be relatively large. Clearly, the results would be altered if this value was to be significantly reduced since the prior distributions would then be expected to influence the results. It would be possible to make the prior variance even larger (100,000, say), but we have found that this does not have any significant effect on the results. We believe that 10,000 is a sufficiently large value for all practical purposes, but the reader can experiment with the WinBUGS code supplied in the Appendix.

3.5 Convergence Diagnostics

In order for the MCMC procedure to work, it is crucial that the Markov chain $\{(M^{(b)}, \theta^{(b)})\}$ reaches the equlibrium posterior distribution (13), and, after that, is run for a sufficiently long time. This is achieved by appropriate choice of burn-in time B, thinning parameter t and chain length N in (18). In most applications, it is very difficult to estimate the theoretical rate of convergence of the Markov chain towards equlibrium. For this reason, a number of *convergence diagnostics* tools have been developed. For instance, Geweke (1991) considers Gibbs sampling and data augmentation, and suggests testing differences between the posterior means of the early and late parts of a single Markov chain by means of spectral methods from time series analysis. Since then, many different methods of convergence diagnostics have been proposed, as reviewed by Cowles and Carlin (1996) and Brooks and Roberts (1998).

There is different software available for monitoring MCMC oonvergence, and we will use the R package boa (Smith, 2007). Although several convergence criteria are included in boa, we will focus on a method originally proposed by Gelman and Rubin (1992). First, a certain *variable* is selected, which is some scalar function of the parameter vector θ that can be computed regardless of model M. For instance, this could be α_i, β_j or the outstanding

reserve for accident year i. Here we will mainly focus on the total outstanding reserve though, with or without a tail assumption. Secondly, $m \geq 2$ parallell chains are run and the variable of interest is monitored. After burn in and possible thinning, the length of all chains is N. Thirdly, the within sample variance W of the m chains is compared with the between sample variance B/N for the chosen variable, by means of the *potential scale reduction factor*

$$\text{PSRF} = \sqrt{\frac{N-1}{N} + \frac{m+1}{mN}\frac{B}{W}}, \tag{25}$$

which should be close to 1 in order to indicate convergence. Thirdly, a corrected scale reduction factor

$$\text{CSRF} = \sqrt{\frac{\text{df}+3}{\text{df}+1}}$$

is computed, to account for sampling variability in the estimate of the true variance of the variable of interest, where df is a method of moments estimate of the degrees of freedom, based on a t approximation in the posterior inference. Quantiles can be computed for CSRF, assuming a normal distribution. The boa software outputs the 0.5 and 0.975 quantiles of CSRF for the chosen variable(s). As a rule of thumb, a 0.975 quantile greater than 1.20 is interpreted as evidence of non-convergence for a variable (Smith, 2007).

However, it is generally a good idea to provide additional diagnostics for transdimensional models, due to the well known difficulties of attaining convergence, see for instance the discussion in subsection 3.2 of this paper or subsection 2.5 of the RJMCMC review paper by Sisson (2005). For this reason, we also consider a more robust convergence diagnostic, the multivariate potential scale reduction factor (MPSRF) of Brooks and Gelman (1998). It can be used to monitor *simultaneous* convergence of a prechosen set of $p \geq 1$ variables. Essentially, B and W become $p \times p$ matrices, and B/W is replaced by the largest eigenvalue of $W^{-1}B$ in (25). The MPSRF has the property of being at least as large as the maximum PSRF of the variables to be examined, and it should also be close to 1.

4 Example

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 (16) and its approximation (18) to estimate the predictive distribution. Hence, the posterior probability is of greater interest than which particular model was most likely (a posteriori). The WinBUGS code for this example is supplied in the Appendix.

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 (8), gives the maximum likelihood parameter estimates shown in Table 4.

$\hat{c}$	12.5063		
$\hat{\alpha}_2$	0.3313	$\hat{\beta}_2$	0.9126
$\hat{\alpha}_3$	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 development years as are required. In this example, we will consider 5

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

We implemented the Bayesian model using the Reversible Jump add-in of Lunn et al (2009) in WinBUGS. In order to determine the appropriate burn-in time B, thinning parameter t and chain length N, we first performed convergence diagnostic tests, as described in section 3.5. The results are summarized in Table 5 for different runs of m parallell Markov chains of various lengths:

Chain parameters				Reserve without tail factors			Reserve with tail factors			Multivar.
m	$B+tN$	B	t	PSRF	CSRF50	CSRF975	PSRF	CSRF50	CSRF975	MPSRF
4	61 699	32 349	1	1.049	1.052	1.151	1.212	1.261	1.674	1.643
4	61 699	32 349	5	1.049	1.052	1.151	1.208	1.256	1.661	1.644
5	100 000	50 000	5	1.124	1.139	1.339	2.074	2.348	3.955	2.927
5	200 000	100 000	1	1.010	1.010	1.028	1.267	1.325	1.776	1.528
5	804 996	405 000	1	1.048	1.052	1.133	1.328	1.585	4.261	1.399
4	804 996	405 000	1	1.004	1.005	1.015	1.089	1.157	1.445	1.182
5	1 000 000	502 500	1	1.029	1.030	1.080	1.213	1.352	2.126	1.293
4	1 000 000	502 500	1	1.010	1.011	1.033	1.138	1.210	1.618	1.284

Table 5: Convergence diagnostics for the total outstanding reserve with and without tail factors, using the potential scale reduction factor (PSRF), the median (CSRF50) and 97.5% quantile (CSRF975) of the corrected scale reduction factor distribution, and the multivariate potential scale reduction factor (MPSRF). Run 1 and 2 are based on the same RJMCMC output, with different thinning. Run 6 (8) is obtained from Run 5 (7) by discarding one of the 5 parallell chains.

We see from Table 5 that the outstanding reserve values converge well when tail factors are excluded, but less well when the tail factors are included. For this reason, results for the outstanding reserve with tail factors included should be interpreted with some caution, even if one believes the model for the tail factors to be correct. We also included MPSRF based on $p = 2$ variables, the outstanding reserve with or without a tail assumption. The fact that MPSRF is rather high indicates the convergence problems for the outstanding reserve with a tail. An even more robust version of MPSRF could be defined by adding, for instance, several runoff parameters β_j.

Based on the results in Table 5, we decided to proceed with an initial burn-in of 500,000 updates, followed by a sample of 500,000 updates and no thinning, i.e. $t = 1$. The results indicate that none of the optional parameters $\nabla^2\beta_4, \ldots, \nabla^2\beta_n$, should definitely be excluded. The posterior marginal probabilities for each parameter, as shown in Table 6 give an indication of where the run-off pattern departs from an exponential decay.

Nr.	Optional parameter	Marginal probability
1	$\nabla^2\beta_4$	0.2610
2	$\nabla^2\beta_5$	1.000
3	$\nabla^2\beta_6$	0.1775
4	$\nabla^2\beta_7$	0.7597
5	$\nabla^2\beta_8$	0.2782
6	$\nabla^2\beta_9$	0.3202
7	$\nabla^2\beta_{10}$	0.3499

Table 6: Posterior marginal probabilities for the optional parameters.

Estimates of parameters for the Bayesian model, together with maximum likelihood estimates of the over-dispersed Poisson chain-ladder model are shown in Table 7. For the Bayesian approach, we approximate the parameter estimates by $\hat{c} = E(c|\mathcal{I})$, $\hat{\alpha}_i = E(\alpha_i|\mathcal{I})$ and $\hat{\beta}_j = E(\beta_j|\mathcal{I})$ from the MCMC output, similarly as in (18).

	ODP	Posterior mean	Posterior PI		ODP	Posterior mean	Posterior PI
$\hat{c}$	12.5063	12.4807	(12.1316,12.4834)				
$\hat{\alpha}_2$	0.3313	0.3539	(0.0558,0.6600)	$\hat{\beta}_2$	0.9126	0.9322	(0.6531,1.2250)
$\hat{\alpha}_3$	0.3212	0.3200	(0.0119,0.6348)	$\hat{\beta}_3$	0.9589	0.9248	(0.6897,1.2347)
$\hat{\alpha}_4$	0.3060	0.3165	(0.0006,0.6386)	$\hat{\beta}_4$	1.0261	0.9678	(0.6603,1.2746)
$\hat{\alpha}_5$	0.2194	0.2256	(-0.1091,0.5611)	$\hat{\beta}_5$	0.4353	0.5448	(0.2328,0.8545)
$\hat{\alpha}_6$	0.2701	0.2797	(-0.0546,0.6179)	$\hat{\beta}_6$	0.0801	0.1106	(-0.3182,0.5261)
$\hat{\alpha}_7$	0.3723	0.4008	(0.0605,0.7440)	$\hat{\beta}_7$	-0.0063	-0.0339	(-0.3640,0.3101)
$\hat{\alpha}_8$	0.5534	0.5600	(0.1959,0.9225)	$\hat{\beta}_8$	-0.3944	-0.1904	(-0.6273,0.2453)
$\hat{\alpha}_9$	0.3690	0.3588	(-0.1183,0.8147)	$\hat{\beta}_9$	0.0094	-0.4025	(-1.0180,0.1602)
$\hat{\alpha}_{10}$	0.2421	0.1922	(-0.7360,0.9943)	$\hat{\beta}_{10}$	-1.3799	-0.6375	(-1.5143,0.1285)

Table 7: Comparison of estimates of the parameters from the over-dispersed Poisson model and the Bayesian model, with 95% prediction intervals for the latter.

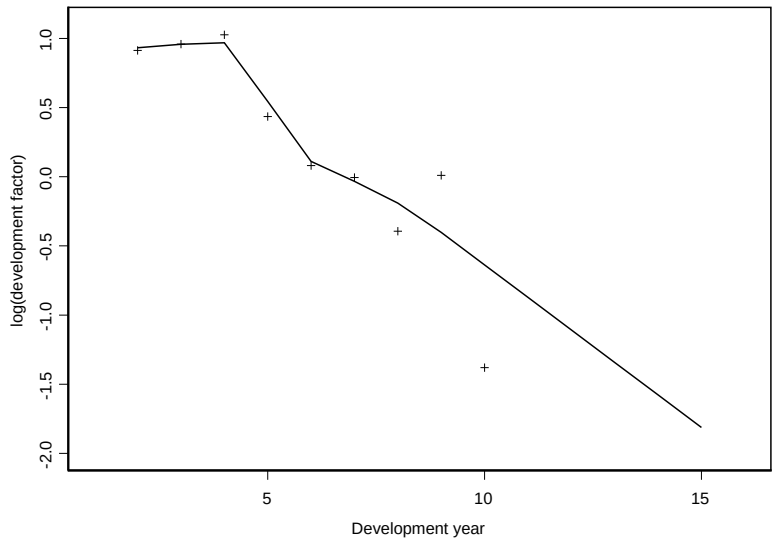

Figure 1: Development parameters for the Bayesian model, together with chain-ladder estimates, on the log scale

Figure 1 shows a comparison of the development parameters for the Bayesian model and the chain-ladder technique.

Table 8 shows the estimates of outstanding claims together with their prediction errors (PE), for the over-dispersed Poisson model and the Bayesian model (without and with the tail factors). The prediction error is the standard deviation of the predictive distribution, expressed as a percentage of the estimate of outstanding claims. Note that the Bayesian prediction errors are, in general, lower than those from the standard ODP model (particularly for the earlier accident years). This is a reflection of the smoothing of the run-off shape inherent in the Bayesian method.

	ODP		Bayesian Model without tail		Bayesian Model with tail	
Row	Estimate	PE	Estimate	PE	Estimate	PE
1					516,379	79%
2	94,634	116%	217899	42%	960,835	79%
3	469,511	46%	466,227	36%	1,191,371	63%
4	709,638	37%	772,393	30%	1,499,630	54%
5	984,889	31%	1,029,486	24%	1,694,783	45%
6	1,419,459	26%	1,483,139	18%	2,180,434	35%
7	2,177,641	23%	2,360,039	16%	3,141700	27%
8	3,920,301	20%	4,002,315	16%	4,919,216	23%
9	4,278,972	24%	4,313,685	21%	5,070,645	25%
10	4,625,811	43%	4,781,149	42%	5,468,297	44%
Total	18,680,856	16%	19,426,333	15%	26,643,290	28%

Table 8: Estimates of outstanding claims from the Bayesian model and the over-dispersed Poisson model, together with their prediction errors (PE).

One of the advantages of stochastic models is that it is possible in many cases to estimate the full predictive distribution of outstanding claims. This is often done using simulation methods, such as bootstrapping. When MCMC methods are used, which are inherently based on simulation, it is completely straightforward to estimate the predictive distribution of any desired quantity. To illustrate this, Figure 2 shows the estimated predictive distributions of total outstanding claims with and without tail factors for the Bayesian model.

5 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 ad hoc procedures used in practice. We believe that this method has great potential for further application in this area.

One of the aspects of winBUGS which can be seen as either a positive

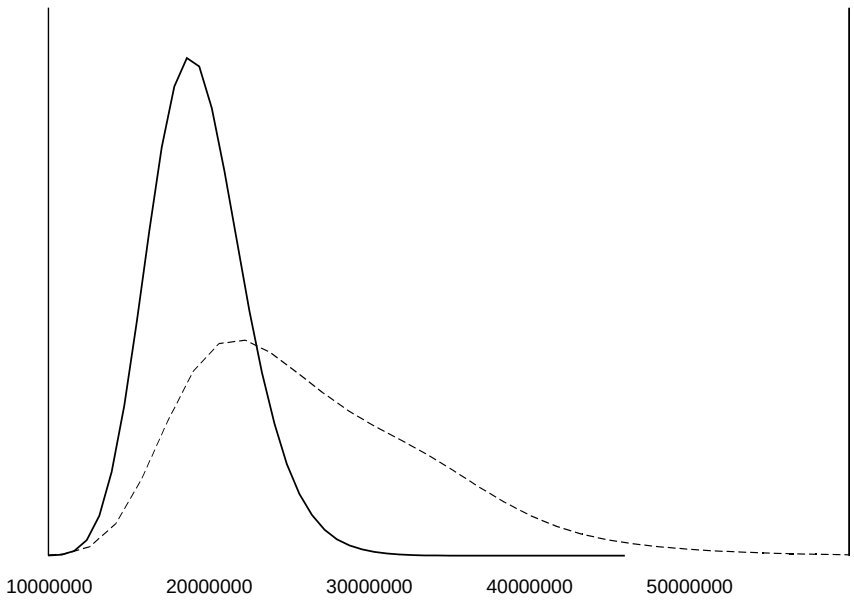

Figure 2: Predictive distribution for total outstanding claims for the Bayesian model, without tail factors (solid line) and with tail factors (dashed line)

feature or a negative feature is that it is not necessary to specify the sampling procedures in detail. This makes it easier to reach a point where a model of interest can be implemented, but it also means that it is not possible to intervene directly in the proposal method. The approach taken in this paper is to use winBUGS because we believe that its accessibility is important for non-expert users who are interested in exploring the feasibility of the models which might be applied. However, we also acknowledge the potential difficulties this necessarily brings with it, such as issues around mixing and between-model moves. Indeed, in section 4 we found that the predictive distribution of the total reserve with tail factors to be a bit unreliable due to non-convergence of the Markov chains. We would anticipate that a specially written programme, based on adaptive RJMCMC (see subsection 3.2) might be deemed more appropriate than winBUGS if such a method, or a closely related method were to be adopted for a practical application.

References

Björkwall, S., Hössjer, O., Ohlsson, E. and Verrall, R. (2011). A generalized linear model with smoothing effects for claims reserving. *Insurance, Mathematics and Economics* **49**, 27-37.

Brooks, S. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434-455.

Brooks S.P. and Roberts, G.O. (1998). Convergence Assessment Techniques for Markov Chain Monte Carlo. *Statistics and Computing*, 8(4), 319-335.

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

Cowles M.K. and Carlin, B.P. (1996). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association*, 91, 883-904.

de Alba, E. (2002). Bayesian Estimation of Outstanding Claim Reserves. *North American Actuarial Journal*, 6, 1-20.

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*, 1, 221-270.

England, P.D. Wüthrich, M.V., and Verrall, R.J. (2010). Bayesian Overdispersed Poisson Model and the Bornhuetter-Ferguson Claims Reserving Method. Pre-print.

Fan, Y., Peters, G.W. and Sisson, S.A. (2009). Automating and evaluating reversible jump MCMC proposal. *Statistics and Computing* 19, 409-421. distributions

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.

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.

Gelman, A. and Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7, 457-511.

Geweke, J. (1991). Evaluating the Accuracy of Sampling Based Approaches to the Calculation of Posterior Moments. Federal Reserve Bank of Minneapolis. Research Department Staff Report 148.

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* 82, 711-732.

Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science* 14(4), 382âĂŞ417.

Katsis, A. and Ntzoufras, I. (2005). Testing Hypotheses for the Distribution of Insurance Claim Counts Using the Gibbs Sampler. *Journal of Computational Methods in Sciences and Engineering*, 5, 201-214.

Lunn, D.J., Best, N. and Whittaker, J.C. (2009). Generic Reversible Jump MCMC Using Graphical Models. *Statistics and Computing* 19, 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, 325-337.

Makov, U. (2001). Principal Applications of Bayesian Methods in Actuarial Science: A Perspective. *North American Actuarial Journal*, 5, 53-60.

Nevat, I., Peters, G.W. and Yuan, J. (2009). Channel Estimation in OFDM Systems with Unknown Power Delay Profile Using Transdimensional MCMC via Stochastic Approximation. *69th Vehicular Technology Conference, VTC, Spring 2009*, 1-6.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*, John Wiley.

Ntzoufras, I. and Dellaportas, P. (2002). Bayesian Modelling of Outstanding Liabilities Incorporating Claim Count Uncertainty (with discussion). *North American Actuarial Journal*, 6, 113-128.

Ntzoufras, I. Katsis, A. and Karlis, D. (2005). Bayesian Assessment of the Distribution of Insurance Claim Counts Using Reversible Jump MCMC. *North American Actuarial Journal*, 9, 90-108.

Peters, G.W., Shevchenko, M.V. and Wüthrich, P.V. (2009). Model Uncertainty in Claims Reserving within Tweedie's Compound Poisson Models. *Astin Bulletin*, 39(1), 1-33.

Renshaw, A.E. and Verrall, R.J. (1998). A Stochastic Model Underlying the Chain Ladder Technique. *British Actuarial Journal*, 4, IV, 903-923.

Roberts, G.O. and Rosenthal, J.S. (2007). Coupling and Ergodicity of Adaptive Markov Chain Monte Carlo Algorithms. *Journal of Applied Probability* 44(2), 458-475.

Roberts, G.O. and Rosenthal, J.S. (2009). Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics* 18(2), 349-367.

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

Scollnik, D.P.M. (2002). Implementation of Four Models for Outstanding Liabilities in WinBUGS: A discussion of a paper by Ntzoufras and Dellaportas, *North American Actuarial Journal*, 6, 128-136.

Sisson, S.A. (2005). Transdimensional Markov Chains. *Journal of the American Statistical Association*, 100(471), 1077-1089.

Smith, B.J. (2007). boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. *Journal of Statistical Software*, 21(11), 1-37.

Taylor, G.C. and Ashe, F.R. (1983). Second Moments of Estimates of Outstanding Claims. *Journal of Econometrics*, 23, 37-61.

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

Verrall, R.J. and Wüthrich, M.V. (2010). Reversible Jump Markov Chain Monte Carlo Method for Parameter Reduction in Claims Reserving

Wüthrich, M.V. (2007). Using a Bayesian Approach for Claims Reserving. *Variance* 1(2), 292-301.

Wüthrich, M.V. and Merz, M. (2008). *Stochastic Claims Reserving in Insurance*. John Wiley.

Acknowledgement
The authors wish to thank two anonymous referee for helpful comments that considerably improved the manuscript. Ola Hössjer's work was supported by the Swedish Research Council, grant 621-2008-4946, and the Gustafsson Foundation for Research in Natural Sciences and Medicine.

Appendix

This Appendix contains the WinBUGS code used for the the example. Note that it is first necessary to download and install the Jump Interface.

Note that "scale" is the plugged in value of the scale parameter, ϕ; "tau" is the reciprocal of the hyperparameter τ; and "k1" is the parameter k in the text of the paper.

```
model
{
# The likelihood is constructed using the zeros trick. The data are first divided
# by 1000 for computational efficiency.
for( i in 1 : 55 ) {
Z[i] <- Y[i]/1000
log(mu[i]) <- cons+alpha[row[i]] + beta[col[i]];
zeros[i]<- 0
zeros[i] ~ dpois(phi[i])
phi[i] <- ((mu[i]-Z[i])-Z[i]*log(mu[i]/Z[i]))/scale   # MINUS log likelihood
}

# psi is not available directly, and so we create an artificial variable, b1,
#which is essentially equal to psi.
for (i in 1:8) {
b1[i]~dnorm(psi[i],100000)
 }

# This section sets up the trans-dimensional model for the run-off parameters
beta[1]<-0
beta[2]<-beta2
beta2~dnorm(0,0.0001)
for (i in 1:8) {beta[i+2]<-beta[i+1]+b1[i]}
for (i in 1:5) {beta[10+i]<-beta[9+i]+b1[8]}
psi[1:8]<-jump.lin.pred.int(X[1:8,1:7],k1,tau,0,0.0001)
tau~dgamma(0.001,0.001)

id<-jump.model.id(psi[1:8])
k1~dbin(0.5,7)
```

```
# As suggested by England and Verrall (2006), we use a gamma
#distribution with the same mean and variance as the ODP for forecasting.
for( i in 56 : 100 ) {
log(mu[i]) <- cons+alpha[row[i]] + beta[col[i]];
fa[i] <-max(0.01,1000*mu[i]/scale)
fb[i] <- 1/scale
Z[i] ~ dgamma(fa[i], fb[i])
}

for( i in 1 : 100 ) {
fit[i] <- Z[i]
}

for(i in 1:50) {
log(muT[i])<- cons+alpha[rowT[i]]+beta[colT[i]]
faT[i]<-max(0.01,1000*muT[i]/scale)
fbT[i]<-1/scale
ZT[i]~dgamma(faT[i],fbT[i])
}

for (i in 1:10) {
Tail[i]<-sum(ZT[5*(i-1)+1:5*i])
}

scale <- 52.601
cons~dnorm(0.0,0.0001)
alpha[1]<-0
for (k in 2:10) {alpha[k]~ dnorm(0.0,0.0001)}

R[1] <- 0
R[2] <- fit[56]
R[3] <- sum(fit[57:58])
R[4] <- sum(fit[59:61])
R[5] <- sum(fit[62:65])
R[6] <- sum(fit[66:70])
R[7] <- sum(fit[71:76])
```

```
R[8] <- sum(fit[77:83])
R[9] <- sum(fit[84:91])
R[10] <- sum(fit[92:100])
Total <- sum(R[2:10])

for (i in 1:10) {
RT[i]<-R[i]+Tail[i]
}
TotalT<-sum(RT[1:10])
}


#INITIAL VALUES
list(alpha = c(NA,0,0,0,0,0,0,0,0,0), b1 = c(0,0,0,0,0,0,0,0,0),
cons=0, tau=10, beta2=1)

# DATA
list(row=c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,
4,4,4,4,4,4,4,5,5,5,5,5,5,6,6,6,6,6,7,7,7,7,8,8,8,9,9,10,
2,3,3,4,4,4,5,5,5,5,6,6,6,6,6,7,7,7,7,7,7,8,8,8,8,8,8,8,
9,9,9,9,9,9,9,9,10,10,10,10,10,10,10,10,10),
col=c(1,2,3,4,5,6,7,8,9,10,1,2,3,4,5,6,7,8,9,1,2,3,4,5,6,7,8,
1,2,3,4,5,6,7,1,2,3,4,5,6,1,2,3,4,5,1,2,3,4,1,2,3,1,2,1,
10,9,10,8,9,10,7,8,9,10,6,7,8,9,10,5,6,7,8,9,10,
4,5,6,7,8,9,10,3,4,5,6,7,8,9,10,2,3,4,5,6,7,8,9,10),
Y=c(357848,766940,610542,482940,527326,574398,
146342,139950,227229,67948,
352118,884021,933894,1183289,445745,320996,
527804,266172,425046,
290507,1001799,926219,1016654,750816,146923,
495992,280405,
310608,1108250,776189,1562400,272482,352053,
206286,
443160,693190,991983,769488,504851,470639,
396132,937085,847498,805037,705960,
440832,847631,1131398,1063269,
359480,1061648,1443370,
376686,986608,
```

```
344014,
NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA),
ZT=c(NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA),
rowT=c(
1,1,1,1,1,
2,2,2,2,2,
3,3,3,3,3,
4,4,4,4,4,
5,5,5,5,5,
6,6,6,6,6,
7,7,7,7,7,
8,8,8,8,8,
9,9,9,9,9,
10,10,10,10,10),
colT=c(
11,12,13,14,15,
11,12,13,14,15,
11,12,13,14,15,
11,12,13,14,15,
11,12,13,14,15,
11,12,13,14,15,
11,12,13,14,15,
11,12,13,14,15,
11,12,13,14,15,
11,12,13,14,15),
X=structure(
.Data = c(
0,0,0,0,0,0,0,
1,0,0,0,0,0,0,
1,1,0,0,0,0,0,
1,1,1,0,0,0,0,
1,1,1,1,0,0,0,
1,1,1,1,1,0,0,
1,1,1,1,1,1,0,
```

```
1,1,1,1,1,1,1),
.Dim = c(8,7)))
```