Annals of Actuarial Science, Vol. 9, part 2, pp. 304–321. © Institute and Faculty of Actuaries 2015
doi:10.1017/S1748499515000019
First published online 16 July 2015

Experience rating with Poisson mixtures

Garfield O. Brown
Statistical Laboratory, Centre for Mathematical Sciences, Cambridge, CB3 0WB, UK
Winston S. Buckley*
Mathematical Sciences, Bentley University, Waltham, MA 02452, USA

Abstract
We propose a Poisson mixture model for count data to determine the number of groups in a Group Life insurance portfolio consisting of claim numbers or deaths. We take a non-parametric Bayesian approach to modelling this mixture distribution using a Dirichlet process prior and use reversible jump Markov chain Monte Carlo to estimate the number of components in the mixture. Unlike Haastrup, we show that the assumption of identical heterogeneity for all groups may not hold as 88% of the posterior probability is assigned to models with two or three components, and 11% to models with four or five components, whereas models with one component are never visited. Our major contribution is showing how to account for both model uncertainty and parameter estimation within a single framework.

Keywords
Experience rating; Poisson mixtures; Reversible jump; MCMC

1. Introduction

Risk is a measure of possible variation of economic outcomes. It is a measure of the variation between the actual outcome and the expected outcome. An individual is exposed to a significant amount of risk associated with perils such as death, disability, fire, and so on. By purchasing an insurance contract, the individual can transfer this risk or variability of possible outcomes to an insurance company in exchange for a set of payments called premiums. This contract gives the individual the right to make a claim on the insurance company to cover losses incurred. In life insurance, the risk is associated with variability in the number of death claims, which is modelled by a probability frequency distribution. In most property/casualty lines of insurance, not only is there frequency distribution of the number of claims, but there is also severity or loss distribution for size of claim.

Claims are triggers that accrue costs against an insurer. The cost of a claim is the magnitude of the effect associated with it. This may also be referred to as loss, size, or amount of damage. Thus, modelling the number of claims is a critical part of loss reserving, pricing, and underwriting in insurance companies. The precision of claims count estimation is therefore crucial to the successful performance of an insurance company. Usually, count regression analysis is used that allows for risk

*Correspondence to: Winston S. Buckley, Mathematical Sciences, Bentley University, Waltham, MA 02452, USA. Tel: 781-891-2000; Fax: 781-891-2457; E-mail: winb365@hotmail.com

factors to be identified, including the expected frequency of claims of policy holders. A common way to calculate the premium is to obtain the conditional expectation of the claims count, given the risk characteristics and combined it with the expected claim size. Jewell (1974) presents the Bayesian credibility model for claim counts with a natural conjugate prior distribution. This model is widely used in non-life insurance companies as part of the process of estimating and predicting the expected claims count in upcoming periods using the past experience of claims of a particular risk class.

In this paper, we consider a mixed Poisson model for claims count data arising in Group Life insurance. We present a Bayesian formulation to determine the number of groups in an insurance portfolio consisting of claim numbers or deaths. We take a non-parametric Bayesian approach to modelling this mixture distribution using a Dirichlet process prior and use reversible jump Markov chain Monte Carlo (RJMCMC) to estimate the number of components in the mixture. The physical interpretation of the model is that heterogeneity is assumed to be drawn from one of a finite number of possible groups, and its proportion will be estimated using Markov chain Monte Carlo (MCMC). Unlike Haastrup (2000), we show that the assumption of identical heterogeneity for all groups may not hold as 88% of the posterior probability is assigned to models with two or three components, and 11% to models with four or five components, whereas models with one component are never visited. We apply different RJMCMC algorithms and show that the birth/death method is preferred. Our major contribution to the actuarial literature is showing how to account for both model uncertainty and parameter estimation within a single framework.

Dellaportas *et al.* (2000, 2003) describe Bayesian model and variable selection using Gibbs sampler and MCMC. Ntzoufras & Dellaportas (2002) model outstanding insurance liabilities incorporating claims count uncertainty using Bayesian analysis. Katsis & Ntzoufras (2005) and Ntzoufras *et al.* (2005) perform Bayesian analysis of insurance claims count distribution using the Gibbs sampler and reversible jump algorithm, respectively. Bermúdez & Karlis (2011) apply MCMC methods to Bayesian multivariate regression Poisson models for the pricing of an insurance contract that contains different types of coverages, which may be dependent, such as automobile insurance. Bermúdez & Karlis (2012) also apply a finite mixture of bivariate Poisson regression models to insurance ratemaking. Verrall & Wüthrich (2012) use the RJMCMC method for parameter reduction in claims reserving. Other Bayesian and MCMC methods with applications to claims reserving are also presented in England *et al.* (2011) and Donnelly & Wüthrich (2012). Streftaris & Worton (2008) apply efficient and approximate Bayesian inference to insurance claims data, whereas Gibbs sampler is used in the Bayesian modelling of financial guarantee insurance in Puustelli *et al.* (2008). The reader is directed to Anastasiadis & Chukova (2012) for an extensive overview of the recent literature on the modelling of insurance claims, and the processes associated with them.

The remainder of this paper is organised as follows: section 2 introduces and extends the Poisson mixture model. Algorithms are presented in section 3. These include the reversible jump model selection and the split and merge method of Dellaportas *et al.* (1997). We apply these algorithms to count data obtained from a major Norwegian insurance company. Numerical results are presented and analysed in section 4. We conclude in section 5.

2. Model and Data

In this section, we present a credibility model for heterogeneity along with data consisting of exposures and deaths/claims count. The data arise from 1,125 groups insured during all, or part of the period 1982–1985, by a major Norwegian insurance company. There are $n = 72$ groups distinguished by occupation category. The heterogeneity model is used to model differences in each

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

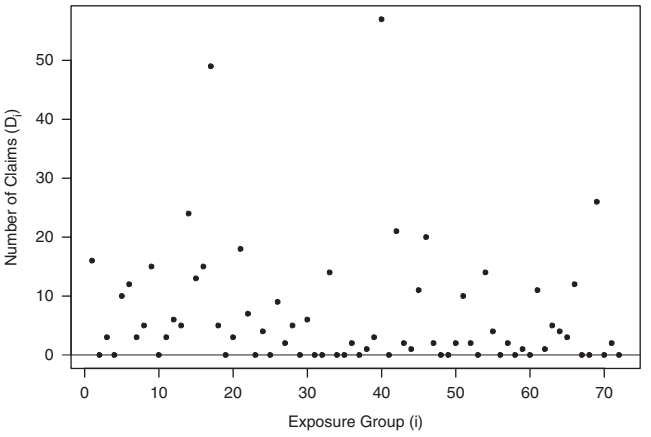

Figure 1. Plot of the number of observed claims for each group.

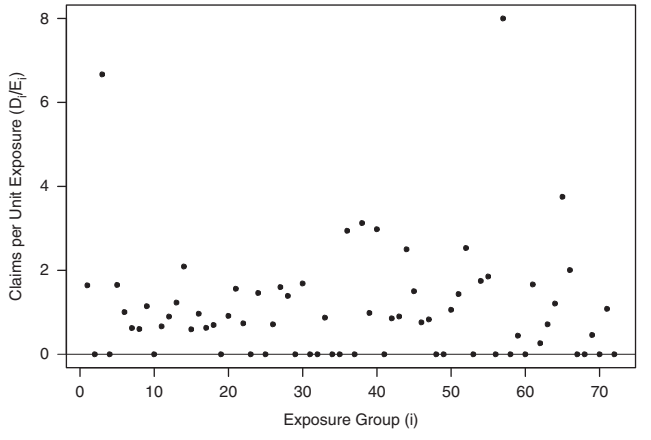

Figure 2. Plot of the number of observed claims per unit exposure.

of the n groups. The ith group has risk exposure E_i, and observed number of deaths D_i. The data are also analysed in Haastrup (2000) and Norberg (1989). Let $D_1, \ldots, D_n$ denote the number of observed deaths in each insured group. Associated with each group is the exposure, denoted as $E_1, \ldots, E_n$, respectively, which is a measure of the propensity of that group to produce claims/deaths. Let D^n denote the collection of all deaths for each group, where

$$D^n = \{D_1, \ldots, D_n\}$$

Similarly, let E^n denote the collection of all exposures for the groups. That is,

$$E^n = \{E_1, \ldots, E_n\}$$

Figure 1 shows a plot of the claim number for each group, whereas Figure 2 shows the claim numbers normalised by their corresponding exposures.

The heterogeneity model is used to model differences in each of the n groups. For each group, the exposures are recorded followed by the resulting number of deaths or claims. Haastrup (2000)

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

assumes that each group i has a unique heterogeneity parameter, denoted λ_i, and that the number of deaths D_i, follows a Poisson distribution with mean $\lambda_i E_i$. The groups are assumed to be mutually independent, given the heterogeneity parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$. Furthermore, Haastrup assumes that this distribution is identical for each group. In practice, large values of E_i will account for large values of D_i, which will lead to similar values of λ_i for each i.

Thus,

$$D_i \sim \text{Poisson}\,(\lambda_i E_i), \qquad i = 1, \ldots, n$$

We take a fully Bayesian approach and assume that λ_i are i.i.d and follow a Gamma distribution with parameters α and β, that is,

$$\lambda_i \sim \text{Gamma}\,(\alpha, \beta)$$

where α and β are also assumed to be unknown.

The advantage of using such mixed distributions is that it allows for overdispersion in the number of occurrences as

$$\mathbb{E}\,(D_i) = \mathbb{E}\,(\mathbb{E}\,(D_i \mid \lambda)) = \mathbb{E}\,(E\lambda) = E\alpha/\beta$$

and

$$\begin{aligned}\mathbb{V}\text{ar}\,(D_i) &= \mathbb{E}\,(\mathbb{V}\text{ar}\,(D_i \mid \lambda)) + \mathbb{V}\text{ar}\,(\mathbb{E}\,(D_i \mid \lambda)) \\ &= E\alpha/\beta + E^2\alpha/\beta^2 > \mathbb{E}\,(D_i)\end{aligned}$$

2.1. Extending the basic model–mixture formulation

We now extend the model by proposing a Poisson mixture model formulation. We assume that D_i, given λ_j, has a Poisson distribution with mean $\lambda_j E_i$. We take a non-parametric Bayesian approach to modelling this mixture distribution using a Dirichlet process prior, and use RJMCMC to estimate the number of components in the mixture. In this case, the physical interpretation of the model is that the heterogeneity is assumed to be drawn from one of k possible components, in proportions $w_1, \ldots, w_k$. The method we describe is essentially a classification problem where we assume that each observed D_i comes from only one of k components, where each component has a Poisson distribution. Thus,

$$D_i \mid \lambda_j \sim \text{Poisson}\,(E_i\lambda_j) \quad j = 1, \ldots, k;\ i = 1, \ldots, n$$

More general forms of the mixture Poisson model with covariates are discussed in Green & Richardson (2002). Mixture models for grouped claim numbers are considered by Tremblay (1992) and Walhin & Paris (1999, 2000). Dellaportas *et al.* (1997) considers count data in finance using split/merge moves, whereas Viallefont *et al.* (2002) provide a more general discussion of mixtures of Poisson distributions using both split/merge moves and birth/death moves. Other methods for determining the number of components in a mixture are discussed by McLachlan & Peel (2000), Phillips & Smith (1996), Carlin & Chib (1995), and Stephens (2000), who use Markov chains to model jointly the number of components and component values. The advantage of the Bayesian formulation is that we can place posterior probabilities on the order of the model.

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

2.2. The likelihood function

Throughout our discussion, n will denote the number of data points and k the number of components in the mixture formulation. For a finite mixture model, the observed likelihood function is

$$L(D^n \mid \lambda, w, E^n) = \prod_{i=1}^{n} \sum_{j=1}^{k} w_j f_j(D_i \mid \lambda_j, E_i) \tag{1}$$

where the weights are non-negative and $\sum_{j=1}^{k} w_j = 1$. Even for moderate values of n and k, this takes a long time to evaluate as there are k^n terms when the inner sums are expanded (Casella *et al.*, 2000). Another form of the likelihood function will be derived shortly. Classical estimation procedures for mixture models are described by Titterington *et al.* (1990) and McLachlan & Peel (2000).

Let z_i be categorical random variables taking values in $\{1, \ldots, k\}$ with probabilities $w_1, \ldots, w_k$, respectively, so that

$$p(z_i = j \mid w) = w_j$$

Let $f_j(\cdot)$ denote a Poisson density with parameter λ_j. Suppose that the conditional distribution of D_i, given $z_i = j$, is Poisson (λ_j), $j = 1, \ldots, k$. Then the unconditional density of D_i is given by

$$\begin{aligned} f(D_i) &= \sum_{j=1}^{k} f_i(D_i \mid z_i = j, \lambda, E^n)\, p(z_i = j) \\ &= \sum_{j=1}^{k} w_j f_i(D_i \mid \lambda_j E_i) \end{aligned} \tag{2}$$

With each pair (D_i, E_i), we associate a latent variable z_i, which is an indicator variable that indicates which component of the mixture is associated with (D_i, E_i). We have $z_i = j$ if the ith data point (D_i, E_i), comes from the jth component of the mixture. Thus, for each i, we have

$$z_i \mid w \sim \mathcal{M}(1; w_1, \ldots, w_k)$$

and

$$D_i \mid z_i \sim \mathcal{P}(\lambda_{zi} E_i)$$

By incorporating the indicator variables z_i, the complete data likelihood is then

$$\begin{aligned} L(D^n \mid z, \lambda, E^n) &= \prod_{i=1}^{n} f(D_i \mid \lambda_{z_i}, E_i) \\ &= \prod_{j=1}^{k} \prod_{\{i:z_i=j\}} f(D_i \mid \lambda_j, E_i) \end{aligned} \tag{3}$$

At times, especially for the fixed k case described below, it is more convenient to work with (3) as it involves multiplications only, rather than additions and multiplications, as in (1). Note that the inclusion of the categorical variables (z_i) does not add to the complexity of the model. Instead, it results in a simplification of the likelihood function from being a product of sums as in (1) to a product only, as in (3).

The convenience of using the missing data formulation is that the posterior conditional distribution of the model parameters would be standard distributions. Moreover, the augmented variables z_i allow us to see the component of the mixture to which the data points are assigned.

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

2.3. Gibbs updates for fixed k

We consider a mixture of Poissons, where conditional on there being k components in the mixture:

$$D_i \sim \sum_{j=1}^{k} w_j f(\cdot \mid \lambda_j, E_i)$$

The weights w_j, sum to 1, and are non-negative. That is,

$$\sum_{j=1}^{k} w_j = 1 \text{ and } w_j \geq 0 \tag{4}$$

$f(D_i \mid \lambda, z_i = j, E_i) \sim$ Poisson $(\lambda_j E_i)$ with allocations $P(z_i = j) = w_j$

and

$$w \sim \mathcal{D}(\delta_1, \ldots, \delta_k)$$

follows a Dirichlet distribution. We also make the additional assumption that the δj's are equal to 1, so that $p(w)$ is a uniform distribution on the space described by (4). For the Poisson parameters λ_j, we take Gamma priors, so that

$$\lambda_j \sim \text{Gamma}(a, b), \qquad j = 1, \ldots, k$$

with the ordering constraint

$$\lambda_1 < \lambda_2 < \cdots < \lambda_k \tag{5}$$

to ensure that the components are identifiable. The ordering constraint is not necessary for the Monte Carlo algorithm to work. However, it does avoid the problem of label switching, as otherwise, any permutation of the indices $\{1, \ldots, k\}$ will result in the same posterior distribution. The choice of objective priors is particularly difficult for finite mixture models, as common improper priors will lead to improper posteriors; see subsection 3.2.2 of Frühwirth-Schnatter (2006) for a discussion. Note that a Dirichlet distribution as the prior on the weights is conjugate so we also have a Dirichlet posterior distribution for the weights but with updated parameters. In addition, we choose identical priors on the hyperparameters λ_j so the model is invariant to ordering.

There are $k!$ ways to order k distinct objects. Because of the ordering constraint in equation (5), and the fact that the λs are i.i.d., the joint density of the collective λ is

$$p(\lambda \mid \alpha, \beta, k) = k!\, p(\lambda_1 \mid \alpha, \beta) \cdots p(\lambda_k \mid \alpha, \beta)\, I_{\lambda_1 < \lambda_2 < \cdots < \lambda_k}(\lambda)$$

When k is fixed and known, the factorial term $k!$ does not affect the MCMC algorithm as it can be absorbed into the normalising constant. However, in the variable k case, it must be noted, as it is a factor in the reversible jump acceptance probability. The joint density of all unknowns is

$$\pi(w, \lambda, z \mid D^n) \propto p(w \mid \delta)\, p(z \mid w)\, p(\lambda \mid \alpha, \beta)\, L(D^n \mid \lambda, z, E^n) \tag{6}$$

With the missing data formulation, the likelihood term $\boldsymbol{L}(D^n \mid \lambda, z, E^n)$ can be written as

$$L(D^n \mid \lambda, z, E) = \prod_{i=1}^{n} \left(\frac{e^{-\lambda_{z_i} E_i} (\lambda_{z_i} E_i)^{D_i}}{D_i!} \right)$$

and

$$p(z \mid w) = \prod_{j=1}^{k} w_j^{n_j}$$

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

where $n_j = \#\{i \mid z_i = j\}$ is the number of observations allocated to component j. The prior distributions are

$$p(w \mid \delta) = \frac{\Gamma\left(\Sigma_{j=1}^{k}\delta_j\right)}{\prod_{j=1}^{k}\Gamma(\delta_j)}\prod_{j=1}^{k} w_j^{\delta_j - 1}$$

$$p(\lambda_i \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\beta\lambda_i}$$

Using Bayes' theorem, we have the following posterior conditional distributions:

$$\pi(\lambda_j) \propto \lambda_j^{\alpha-1}e^{-\beta\lambda_j} \times \lambda_j^{\Sigma_{i\mid z_i=j}D_i}e^{-\lambda_j\Sigma_{i\mid z_i=j}E_i}\, I_{(\lambda_{j-1},\lambda_{j+1})}(\lambda_j)$$

and

$$\pi(w \mid \delta, z) \propto p(w \mid \delta)\, p(z \mid w)$$

so that

$$w \sim \mathcal{D}(\delta_1 + n_1, \ldots, \delta_k + n_k)$$

where $n_j = \#\{i \mid z_i = j\}$. For z, we update the allocations using

$$P(z_i = j) \propto w_j f(D_i \mid \lambda_j, E_i) \quad i = 1, \ldots, n; \quad j = 1, \ldots, k$$

so that

$$p(z_i = j) = \frac{w_j f(D_i \mid \lambda_j, E_i)}{\Sigma_{j=1}^{k} w_j f(D_i \mid \lambda_j, E_i)} \tag{7}$$

This follows from equation (2). The Gibbs algorithm for fixed k is then (Robert & Casella, 1999)

Step 1: Simulate z_i from

$$p(z_i = j) \propto w_j f(D_i \mid \lambda_j, E_i) \text{ for } j = 1, \ldots, k$$

and compute n_j, $n_j\overline{D}_j$, $n_j\overline{E}_j$ from

$$n_j = \sum_{i\mid z_i=j}(1) \quad n_j\overline{D}_j = \sum_{i\mid z_i=j} D_i \quad n_j\overline{E}_j = \sum_{i\mid z_i=j} E_i$$

Step 2: Simulate

$$\lambda_j \sim \text{Gamma}(\alpha + n_j\overline{D}_j, \beta + n_j\overline{E}_j)\, \mathbb{I}_{(\lambda_{j-1},\lambda_{j+1})}(\lambda_j) \text{ for } j = 1, \ldots, k$$

Step 3: Simulate

$$w \sim \mathcal{D}(\delta + n_1, \ldots, \delta + n_k)$$

3. The Algorithms

The analysis in the preceding sections assumed that the number of components in our mixture formulation is fixed and known. We now extend the results to the case where k varies. That is, the number of components of the mixture are not known in advance. This is a model selection problem in which the objective is to select a model with the number of components that best describe our data. The algorithms used in our model selection analysis now follow.

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

3.1. RJMCMC

The reversible jump algorithm is an extension of the Metropolis–Hastings algorithm. We assume there is a countable collection of candidate models, indexed by $M_i \in \mathcal{M} = \{M_1, M_2, \ldots, M_k\}$. We further assume that for each model M_i, there exists an unknown parameter vector $\theta_i \in \mathbb{R}^{n_i}$, where n^i, the dimension of the parameter vector, can vary with i. Typically, we are interested in finding which models have the greatest posterior probabilities, and also to obtain estimates of the parameters. Thus, the unknowns in this modelling scenario will include the model index M_i, as well as the parameter vector θ_i. We assume that the models and corresponding parameter vectors have a joint density $\pi(M_i, \theta_i)$. The reversible jump algorithm constructs a reversible Markov chain on the state space $\mathcal{M} \times \cup_{M_i \in \mathcal{M}} \mathbb{R}^{n_i}$, which has π as its stationary distribution (Green, 1995). In many instances, and in particular for Bayesian problems, this joint distribution is of the form

$$\pi(M_i, \theta_i) = \pi(M_i, \theta_i \mid X) \propto L(X \mid M_i, \theta_i)\, p(M_i, \theta_i)$$

where the prior on (M_i, θ_i) is often of the form

$$p(M_i, \theta_i) = p(\theta_i \mid M_i)\, p(M_i)$$

with $p(M_i)$ being the density of some counting distribution.

Suppose we are at model M_i and a move to model Mj is proposed, with probability r_{ij}. The corresponding move from θ_i to θ_j is achieved by using a deterministic transformation h_{ij}, such that

$$(\theta_j, v) = h_{ij}(\theta_i, u) \tag{8}$$

where u and v are random variables introduced to ensure dimension matching necessary for reversibility. To ensure dimension matching, it is necessary that

$$\dim(\theta_j) + \dim(v) = \dim(\theta_i) + \dim(u)$$

For discussions about possible choices for the function h_{ij}, we refer the reader to Green (1995) and Brooks *et al.* (2003*b*). Let

$$A(\theta_i, \theta_j) = \frac{\pi(M_j, \theta_j)}{\pi(M_i, \theta_i)} \frac{q(v)}{q(u)} \frac{r_{ji}}{r_{ij}} \left| \frac{\partial h_{ij}(\theta_i, u)}{\partial(\theta_i, u)} \right| \tag{9}$$

The acceptance probability for a proposed move from model (M_i, θ_i) to model (M_j, θ_j) is then

$$\min\{1, A(\theta_i, \theta_j)\}$$

where $q(u)$ and $q(v)$ are the respective proposal densities for u and v, and $|\partial h_{ij}(\theta_i, u)/\partial(\theta_i, u)|$ is the Jacobian of the transformation induced by h_{ij}. Green (1995) shows that the algorithm with acceptance probability given above simulates a Markov chain that is reversible and follows from the detailed balance equation

$$\pi(M_i, \theta_i)\, q(u)\, r_{ij} = \pi(M_j, \theta_j)\, q(v)\, r_{ji} \left| \frac{\partial h_{ij}(\theta_i, u)}{\partial(\theta_i, u)} \right|$$

Detailed balance is necessary to ensure reversibility and is a sufficient condition for the existence of a unique stationary distribution. For the reverse move from model M_j to model M_i, it is easy

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

to see that the transformation used is $(\theta_i, u) = h^{-1}(\theta_j, v)$ and the acceptance probability for such a move is

$$\min\left\{1, \frac{\pi(M_i, \theta_i)}{\pi(M_j, \theta_j)} \frac{q(u)}{q(v)} \frac{r_{ij}}{r_{ji}} \left|\frac{\partial h_{ij}(\theta_i, u)}{\partial(\theta_i, u)}\right|^{-1}\right\} = \min\{1, A(\theta_i, \theta_j)^{-1}\}$$

For inference regarding which model has the greater posterior probability, we can base our analysis on a realisation of the Markov chain constructed above. The marginal posterior probability of model M_i is

$$\pi(M_i \mid X) = \frac{p(M_i) f(X \mid M_i)}{\Sigma_{M_j \in \mathcal{M}} p(M_j) f(X \mid M_j)}$$

where

$$f(X \mid M_i) = \int L(X \mid M_i, \theta_i) p(\theta_i \mid M_i) d\theta_i$$

is the marginal density of the data, which is obtained by integrating over the unknown parameters θ. In practice, we estimate $\pi(M_i \mid \boldsymbol{X})$ by counting the number of times the Markov chain visits model M_i in a single long run after reaching stationarity. These between-model moves described in this section are also augmented with within-model Gibbs updates as given in section 2.3 to update model parameters.

To assess convergence of the reversible jump algorithm, we use the method of Brooks & Giudici (1999) in which they propose to run $I \geq 2$ chains in parallel and base their convergence diagnostic on splitting the total variation not just between chains, but also between models. Their method was extended by Brooks *et al.* (2003*a*) to include non-parametric techniques, including χ^2 tests, Kolmogorov–Smirnov tests, and direct convergence rate estimation.

Brooks *et al.* (2003*a*) suggest several methods for assessing convergence within the context of model selection problems. In particular, for reversible jump algorithms, we can have some idea of how fast the simulations approach stationarity by comparing the empirical stationary distribution with the observed model orders. They propose specific test statistics based on the χ^2 distribution and also a Kolmogorov–Smirnov test for goodness of fit. The χ^2 and Kolmogorov–Smirnov compare the stationary distribution of each chain and computes p-values for the computed test statistics. A critical value of 5% is used so that if the χ^2 or Kolmogorov–Smirnov statistic is above this significance level there is no reason to reject the chains as not being from the same stationary distribution. See Brooks *et al.* (2003*a*) for further details.

3.2. Reversible jump model selection

To update the model order, and thereby increase or decrease the number of components in the mixture, we use a combination of birth/death and split/merge moves as described below. We assume a uniform prior on the number of components k, so that

$$k \sim U\{1, \ldots, k_{\max}\}$$

where $k_{\max}$ is chosen to allow the algorithm to explore all feasible models. We set $k_{\max} = 72$, the number of groups, as under our hypothesis, this is the maximum number of components in

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

the mixture. $k = k_{\max}$ only when the groups are all distinct. Setting $k_{\max} = 72$ will allow for direct comparison of the empirical Bayesian and the mixture model approach.

By introducing a prior on the number of components k, we extend the joint density (6) of all parameters. This yields

$$\pi(k, w, z, \lambda \mid D^n) \propto p(k)\, p(w \mid \delta, k)\, p(\lambda \mid \alpha, \beta, k)\, p(z \mid k)\, L(D^n \mid \lambda, z) \tag{10}$$

Note that the densities of the other model parameters now depend on k. In sections 3.3 and 3.5, we describe in detail two algorithms that are used to simulate from this density. These algorithms are then combined with the fixed k updates of section 2.3 to simulate from the density in equation (10). Modelling mixtures with and without the Dirichlet process prior is considered by Green & Richardson (2001), who also consider the case of an unknown number of components. Alternatives to the reversible jump algorithm exist in this context. For example, Dellaportas & Karlis (2001) develop a semi-parametric sample-based method to approximate a mixing density $g(\theta)$, based on the method of moments.

3.3. Split and merge moves

Note that the joint density in equation (10) now depends on k. We use the split/merge method of Dellaportas *et al.* (1997) and Viallefont *et al.* (2002). Suppose we are at a configuration with k components, with

$$\theta_k = \{(\lambda_1, w_1), \ldots, (\lambda_k, w_k)\}$$

and suppose a move to increase the number of components is proposed. We select uniformly one of the current k components to be split. Suppose the jth component (λ_j, w_j) is selected to be split into two components (λ_{j_1}, w_{j_1}) and (λ_{j_2}, w_{j_2}) such that $j_1 = j$ and $j_2 = j+1$, the components originally numbered $j+1, \ldots, k$ are then renumbered $j+2, \ldots, k+1$. The split is also designed so that the first two moments of the split component remain the same as the original component. Thus, we simulate $u1$ and $u2$ from densities defined on the interval [0, 1]. Usually, we use β densities and set

$$\begin{aligned} w_{j_1} &= w_j u_1 \\ w_{j_2} &= w_j (1 - u_1) \\ \lambda_{j_1} &= \lambda_j u_2 \\ \lambda_{j_2} &= \lambda_j (1 - u_1 u_2)/(1 - u_1) \end{aligned}$$

Other choices for splitting and merging components are described in Viallefont *et al.* (2002). The proposed parameter is then

$$\begin{aligned} \theta_{k+1} = \{&(\lambda_1, w_1), \ldots, (\lambda_j - 1, w_{j-1}), (\lambda_{j_1}, w_{j_1}), (\lambda_{j_2}, w_{j_2}) \\ &(\lambda_{j+1}, w_{j+1}), \ldots, (\lambda_k, w_k)\} \end{aligned}$$

If the ordering constraint in equation (5) is not satisfied, then the move is rejected immediately, as the reverse move in which we merge two adjacent components would not be possible. We can compute the Jacobian for this transformation as

$$\left| \frac{\partial \theta_{k+1}}{\partial (\theta_k, u_1, u_2)} \right| = \left| \frac{\partial (w_{j_1}, w_{j_2}, \lambda_{j_1}, \lambda_{j_2})}{\partial (w_j, \lambda_j, u_1, u_2)} \right| = \frac{\lambda_j w_j}{1 - u_1} \tag{11}$$

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

For the reverse move, we select a pair of adjacent components j_1 and j_2. Combining them to get a new component labelled j, we set

$$w_j = w_{j_1} + w_{j_2}, \quad \lambda_j = \frac{w_{j_1}\lambda_{j_1} + w_{j_2}\lambda_{j_2}}{w_{j_1} + w_{j_2}}$$

by keeping the first two moments of the proposed and current configuration constant. We then sample a new set of allocation variables according to equation (7). We also keep track of the probability of each allocation, so that $p_a(z)$ represents the probability of a given allocation. To compute $p_a(z)$, we first simulate z_i using equation (7). For each i, the probability of that allocation is given by

$$p_a(z_i) = \frac{w_{z_i} f(D_i \mid \lambda_{z_i}, E_i)}{\Sigma_{j=1}^{k} w_j f(D_i \mid \lambda_j, E_i)}$$

Finally, we compute the probability of all allocations by

$$p_a(z) = \prod_{i=1}^{n} p_a(z_i)$$

3.4. Acceptance probability

The acceptance probability of a move of type $(k, \theta_k) \Rightarrow (k', \theta_{k'})$ is then $\min\{1, A_{k,\,k'}\}$, where

$$A_{k,k'} = \frac{\pi(k', \theta_{k'})}{\pi(k, \theta_k)} \times \frac{p(k' \Rightarrow k)}{p(k \Rightarrow k')} \times \frac{1}{q(u_1)q(u_2)} \times \left| \frac{\partial \theta_{k'}}{\partial(\theta_k, u_1, u_2)} \right|$$

$$A_{k,k'} = \frac{p(k')p(w' \mid \delta, k')p(\lambda' \mid \alpha, \beta, k')L(D^n \mid \lambda', z')}{p(k)p(w \mid \delta, k)p(\lambda \mid \alpha, \beta, k)L(D^n \mid \lambda, z)} \times \frac{p(z' \mid w', k+1)/p_a(z')}{p(z \mid w, k)/p_a(z)} \times \frac{p(k' \Rightarrow k)}{p(k \Rightarrow k')} \frac{1}{q(u_1)q(u_2)} \left| \frac{\partial \theta_{k'}}{\partial(\theta_k, u_1, u_2)} \right|$$

with $k' = k+1$ this becomes

$$A_{k,k+1} = \frac{p(k+1)}{p(k)} \times \frac{p(w' \mid \delta, k+1)}{p(w \mid \delta, k)} \times \frac{p(\lambda' \mid \alpha, \beta, k+1)}{p(\lambda \mid \alpha, \beta, k)} \times \frac{L(D^n \mid \lambda', w')}{L(D^n \mid \lambda, w)} \times \frac{p(k+1 \Rightarrow k)}{p(k \Rightarrow k+1)} \times \frac{1}{q(u_1)q(u_2)} \left| \frac{\partial \theta_{k+1}}{\partial(\theta_k, u_1, u_2)} \right|$$

Now with a uniform prior on the number of components k and the weights w

$$A_{k,k+1} = \frac{\Gamma(k+1)}{\Gamma(k)} \times \frac{(k+1)p(\lambda_{j_1} \mid \alpha, \beta)p(\lambda_{j_2} \mid \alpha, \beta)}{p(\lambda_j \mid \alpha, \beta)} \times \frac{p(z' \mid w', k+1)/p_a(z')}{p(z \mid w, k)/p_a(z)} \times \frac{L(D^n \mid \lambda', z')}{L(D^n \mid \lambda, z)} \times \frac{m_{k+1}}{s_k} \times \frac{1}{q(u_1)q(u_2)} \times \left| \frac{\partial \theta_{k+1}}{\partial(\theta_k, u_1, u_2)} \right|$$

where the ratio of Gamma terms comes from the ratio of the prior distributions on w' and w and

$$\frac{p(k+1 \Rightarrow k)}{p(k \Rightarrow k+1)} = \frac{m_{k+1}/(k+1-1)}{s_k/k} = \frac{m_{k+1}}{s_k}$$

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

3.5. Birth and death moves

Suppose we are now at model M_k with k components, say

$$\theta_k = \{(\lambda_1, w_1), \ldots, (\lambda_k, w_k)\} \tag{12}$$

If a move is proposed to increase the number of components by one, then we simulate

$$\tilde{w} \sim \text{Beta}(1, k) \text{ and } \tilde{\lambda} \sim \text{Gamma}(a, b)$$

independently. The proposed new component will then have weight $\tilde{w}$ and the other weights are then scaled by a factor of $(1-\tilde{w})$, so that the sum of the weights remain equal to 1. The corresponding Poisson parameter for the proposed component in $\tilde{\lambda}$. Note that $\tilde{\lambda}$ is sampled from its prior distribution. The proposed component is then

$$\theta_{k+1} = (\lambda_1, w_1/(1-\tilde{w})), \ldots, (\lambda_{k,w_k}/(1-\tilde{w})), (\tilde{\lambda}, \tilde{w})\} \tag{13}$$

Using this proposed value of θ_{k+1}, we also simulate proposed values for the allocations z' with model $k+1$. Using the general form of the reversible jump acceptance probability, see, for example, Green (1995), the probability of changing the number of components to $k+1$ is then $\min\{1, A_{k,k+1}\}$, where

$$A_{k,k+1} = \frac{\pi(k+1, \theta_{k+1})}{\pi(k, \theta_k)} \times \frac{p(k+1 \Rightarrow k)}{p(k \Rightarrow k+1)} \times \frac{1}{q(\tilde{w})q(\tilde{\lambda})} \times \left| \frac{\partial \theta_{k+1}}{\partial(\theta_k, \tilde{w}, \tilde{\lambda})} \right|$$

Making the necessary substitutions yield

$$\begin{aligned} A_{k,k+1} = {} & \frac{p(k+1)p(w' \mid \delta, k+1)p(\lambda' \mid \alpha, \beta, k+1)L(D^n \mid \lambda', z')}{p(k)p(w \mid \delta, k)p(\lambda \mid \alpha, \beta, k)L(D^n \mid \lambda, z)} \\ & \times \frac{p(z' \mid w', k+1)/p_a(z')}{p(z \mid w, k+1)/p_a(z)} \times \frac{p(k+1 \Rightarrow k)}{p(k \Rightarrow k+1)} \times \frac{1}{q(\tilde{w})q(\tilde{\lambda})} \\ & \times \left| \frac{\partial \theta_{k+1}}{\partial(\theta_k, \tilde{w}, \tilde{\lambda})} \right| \end{aligned} \tag{14}$$

Using equations (12) and (13) we then have the Jacobian

$$\frac{\partial \theta_{k+1}}{\partial(\theta_k, \tilde{w}, \tilde{\lambda})} = (1-\tilde{w})^{k-1}$$

If we denote the probability of a birth when there are k components by b_k, and the probability of a death by d_k, with $b_k + d_k = 1$, then

$$\frac{p(k+1 \Rightarrow k)}{p(k \Rightarrow k+1)} = \frac{d_{k+1}/(k+1)}{b_k}$$

as for the move to be reversible we would then be able to kill $k+1$ components in the new model, each with equal probability. Substituting these values in equation (14), the ratio $A_{k,k+1}$ reduces to

$$\begin{aligned} A_{k,k+1} = {} & \frac{\Gamma(k+1)}{\Gamma(k)} \times (k+1)p(\tilde{\lambda}) \times \frac{L(D^n \mid \lambda', z')}{L(D^n \mid \lambda, z)} \times \frac{p(z' \mid w', k+1)/p_a(z')}{p(z \mid w, k)/p_a(z)} \\ & \times \frac{d_{k+1}/(k+1)}{b_k} \frac{1}{q(\tilde{w})q(\tilde{\lambda})} \times (1-\tilde{w})^{k-1} \end{aligned}$$

which on substituting $q(\tilde{\lambda}) = p(\tilde{\lambda})$ and $q(\tilde{w}) = k(1-\tilde{w})^{k-1}$ further reduces to

$$A_{k,k+1} = \frac{p(z' \mid w', k+1)/p_a(z')}{p(z \mid w, k)/p_a(z)} \frac{L(D^n \mid \lambda', z')}{L(D^n \mid \lambda, z)} \times \frac{d_{k+1}}{b_k} \tag{15}$$

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

Table 1. Posterior model order.

Model order k	Posterior probability $\pi(k \mid D^n, E^n)$
1	0.00000
2	0.59485
3	0.29058
4	0.08588
5	0.02258
6	0.00448
7	0.00104
8	0.00034
9	0.00026
10	0.00000

For a proposed death move, the acceptance probability is then

$$\min\{1, A_{k,k+1}^{-1}\}$$

Although the algorithm simulates new values of the allocations when proposing to move, it is not necessary to carry the allocations along. For between-model moves, we could replace the missing data formulation by noting that

$$\frac{p(z \mid w, k)}{p_a(z)} L(D^n \mid \lambda, z) = L(D^n \mid \lambda, w)$$

4. Results

We now present some numerical results for this data set based on the model described in section 2.1, and the algorithms described in section 3.2.

Table 1 shows the posterior model probabilities calculated from the reversible jump algorithm by counting the proportion of time that the algorithm visits each model. A plot of the number of components as the chain evolves is shown in Figure 3(a). The results clearly show that the number of components has a posterior mode at $k = 2$. In addition, the model with $k = 1$ component is never visited. Moreover, if the algorithm is started with $k = 1$, then immediately it jumps to $k = 2$, and never returns to $k = 1$. As >88% of the posterior probability mass is placed on the models with two or three components, we discuss those models in detail in section 4.2. The between-model acceptance rates were 7.7% and 5.5% for the birth/death and split/merge moves, respectively. The total acceptance rate when there is equal probability of proposing a birth/death move or a split/merge move, is 6.6%. These results are tabulated in Table 2.

To assess convergence of the algorithm, we simulated four chains using different starting values and different random number seeds for a total of 100,000 iterations. Both the χ^2 and Kolmogorov–Smirnov diagnostics are computed. These diagnostics are plotted in Figure 4.

4.1. Comparing the model move schemes

A comparison of the individual acceptance probabilities shows that the between-model moves are accepted with a larger rate for the birth/death scheme compared with the split/merge scheme.

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

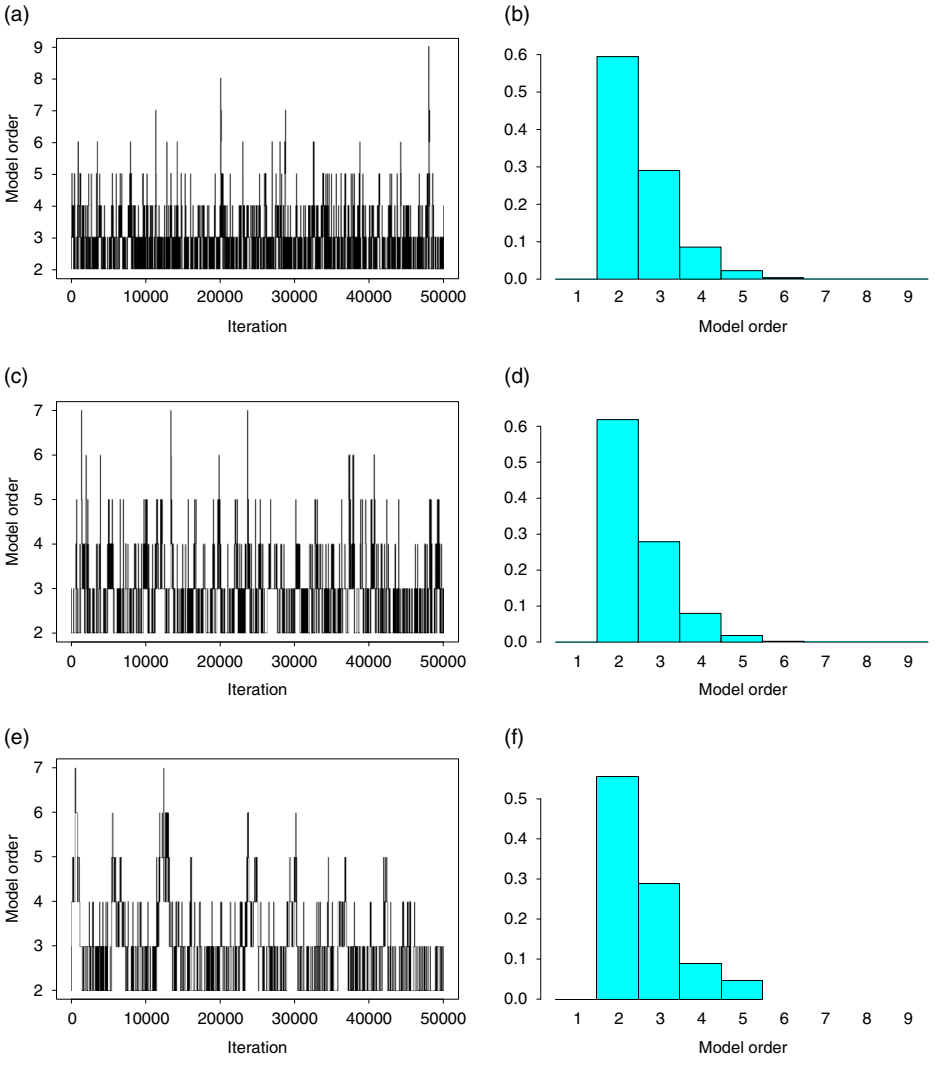

Figure 3. Left: Model trace indicator. Right: Histogram of posterior model order. (a) Birth/death and split/merge model trace. (b) Birth/death and split/merge histogram. (c) Birth/death model trace. (d) Birth/death histogram. (e) Split/merge model trace. (f) Split/merge histogram.

Table 2. Acceptance rates.

Schemes	Acceptance rates
Birth/death	0.077
Split/merge	0.055
Birth/death and split/merge	0.066

This might not always be true, as other split/merge schemes may be proposed (Viallefont *et al.*, 2002). However, it is interesting to note that although the birth and death rates are higher than the split and merge rates, the combined scheme seems to mix better than either scheme implemented alone.

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

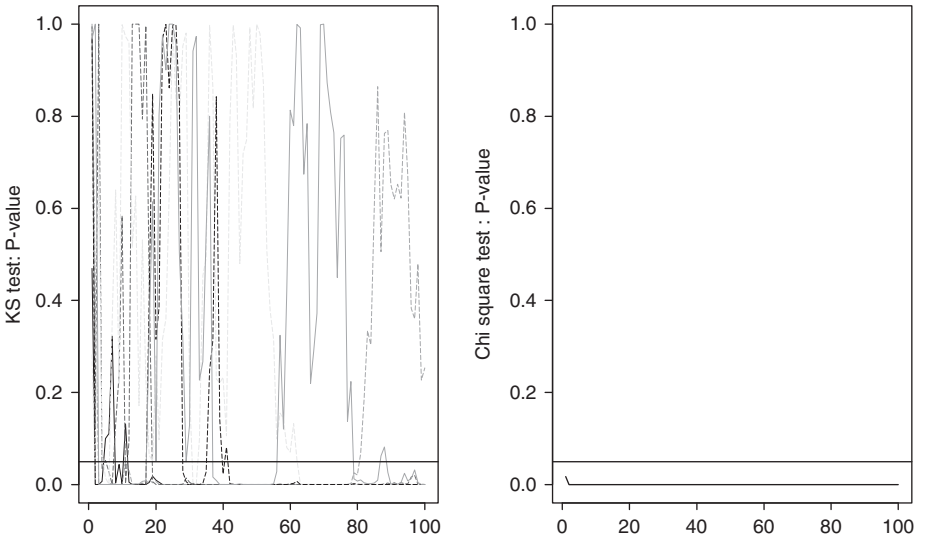

Figure 4. Convergence diagnostics. KS, Kolmogorov–Smirnov.

Table 3. Parameter estimates conditional on $k = 2$.

	Estimates	95% HPD interval
$\lambda 1$	0.731	(0.626, 0.839)
$w1$	0.636	(0.428, 0.821)
$\lambda 2$	1.896	(1.557, 2.249)
$w2$	0.363	(0.178, 0.571)

Note: HPD, highest posterior density.

Although the birth and death scheme has a higher acceptance rate for between-model moves, the excursions away from the values of highest posterior density, $k = 2$ and $k = 3$, are longer than for the combined scheme or the split and merge scheme. This is because when proposing parameters independently from the prior, areas of low probability can be proposed, whereas with the split and merge scheme, areas of low probability mass will generally be rejected. Based on the results presented, the birth/death method would be the preferred algorithm.

4.2. Detailed results for $k = 2$ and $k = 3$

Recall the missing data formulation introduced in section 2.1 for the number of components conditional on $k = 2$. We observed the posterior distribution of z at each iteration when $k = 2$. A study of values of z will tell us how the data have been allocated to the components and therefore, which data points have been generated from either the first Poisson distribution or the second Poisson distribution. This information, along with further information from the portfolio, will help insurance companies classify groups of life insurance portfolios. The parameter estimates are shown in Table 3.

Similar results for the posterior parameter estimates, conditional on there being three components in the mixture, are given in Table 4.

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

Table 4. Parameter estimates conditional on $k = 3$.

	Estimates	95% HPD interval
$\lambda 1$	0.462	(0.000, 0.770)
$w1$	0.299	(0.000, 0.656)
$\lambda 2$	1.115	(0.625, 1.692)
$w2$	0.495	(0.157, 0.797)
$\lambda 3$	2.481	(1.570, 3.366)
$w3$	0.204	(0.002, 0.464)

Note: HPD, highest posterior density.

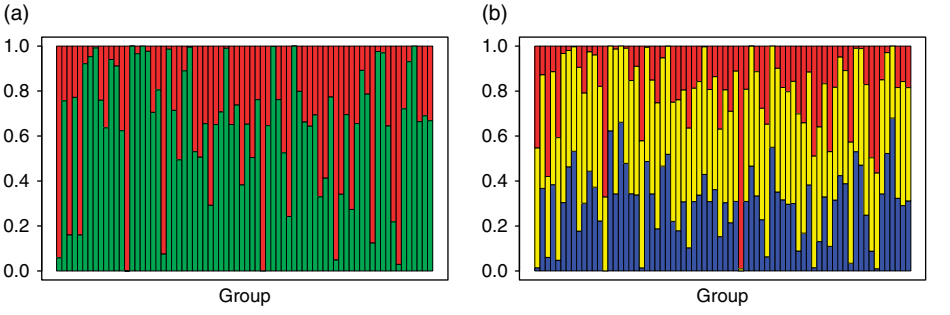

Figure 5. Probability (vertical axis) of data from group i (horizontal axis) being assigned to individual components conditional on $k = 2$ (a) and $k = 3$ (b).

Figure 5 shows the posterior probability of each data point being allocated to a particular component of the mixture, conditional on $k = 2$ and $k = 3$, respectively.

5. Conclusion

We present a model for heterogeneity in Group Life insurance by proposing a Poisson mixture model for count data to determine the number of groups in a portfolio consisting of claim numbers or deaths. We take a non-parametric Bayesian approach to modelling this Poisson mixture distribution using a Dirichlet process prior and use RJMCMC to estimate the number of components in the mixture. We show that a mixture with two or three components works best, in that they result in the highest posterior probabilities for parameters. In particular, 88% of the posterior probability is assigned to models with two or three components, and 11% to models with four or five components, whereas models with one component are never visited. In contrast to Haastrup (2000), we show that the assumption of identical heterogeneity for all groups may not be valid. In this case, it is necessary to put similar groups together for further analysis.

We contribute to the actuarial literature by showing how to account for both model uncertainty and parameter estimation within a single framework. This research can be extended to the case where claims are grouped, such as in Walhin & Paris (1999, 2000).

There are two main advantages of our model. First, it allows us to use the data to estimate the number of groups within observed count data. This is much better than a priori assuming a fixed number of groups. Our example confirms the absence of heterogeneity by proving that two or three components are much

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press

more likely than assuming a single heterogeneous group. The second advantage of our model is that it allows ratemaking and further risk calculations to be done for each group.

Acknowledgement

The authors sincerely thank the two anonymous referees for their comments and suggestions which greatly helped to improve the manuscript.

References

Anastasiadis, S. & Chukova, S. (2012). Multivariate insurance models: an overview. *Insurance: Mathematics and Economics*, **51**(1), 222–227.

Bermúdez, L. & Karlis, D. (2011). Bayesian multivariate Poisson models for insurance ratemaking. *Insurance: Mathematics and Economics*, **48**(2), 226–236.

Bermúdez, L. & Karlis, D. (2012). A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking. *Computational Statistics and Data Analysis*, **56**(12), 3988–3999.

Brooks, S.P. & Giudici, P. (1999). Diagnosing convergence of reversible jump MCMC algorithms. In J.M. Bernardo, J.O. Berger, A.P. Dawid & A.F.M. Smith, (Eds.), *Bayesian Statistics 6* (pp. 733–742). Oxford University Press, Oxford.

Brooks, S.P., Giudici, P. & Philippe, A. (2003*a*). Nonparametric convergence assessment for MCMC model selection. *Journal of Computational and Graphical Statistics*, **12**, 1–22.

Brooks, S.P., Giudici, P. & Roberts, G.O. (2003*b*). Efficient construction of reversible jump MCMC proposal distributions (with discussion). *Journal of the Royal Statistical Society, Series B*, **65**(1), 3–55.

Carlin, B.P. & Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, **57**, 473–484.

Casella, G., Robert, C.P. & Wells, M.T. (2000). Mixture models, latent variables and partitioned importance sampling, technical report no. 2000–03, CREST, INSEE, Paris.

Dellaportas, P., Forster, J.J. & Ntzoufras, I. (2000). Bayesian variable selection using the Gibbs sampler. In D.K. Dey, S. Ghosh & B. Mallick (Eds.), *Generalized Linear Models: A Bayesian Perspective* (pp. 271–286). New York: Marcel Dekker.

Dellaportas, P., Forster, J.J. & Ntzoufras, I. (2003). On Bayesian model and variable selection using MCMC. *Statistics and Computing*, **12**, 27–36.

Dellaportas, P. & Karlis, D. (2001). A simulation approach to nonparametric empirical Bayes analysis. *International Statistical Review*, **69**(1), 63–79.

Dellaportas, P., Karlis, D. & Xekalaki, E. (2011). Bayesian analysis of finite Poisson mixtures. *British Journal of Science*, **1**(1), 96–110.

Donnelly, C. & Wüthrich, M. (2012). Bayesian prediction of disability insurance frequencies using economic indicators. *Annals of Actuarial Science*, **6**(2), 381–400.

England, P.D., Verrall, R.J., & Wüthrich, M.V. (2012). Bayesian Overdispersed Poisson Model and the Bornhuetter-Ferguson Claim Reserving Method. *Annals of Actuarial Science*, **6**(2), 258–283.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. New York: Springer, ISBN-10: 0-387-32909-9.

Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**(4), 711–732.

Green, P.J. & Richardson, S. (2001). Modelling heterogeneity with and without the Dirichlet process. *Scandinavian Journal of Statistics*, **28**(2), 355–375.

Green, P.J. & Richardson, S. (2002). Hidden Markov models and disease mapping. *Journal of the American Statistical Society*, **97**(460), 1055–1070.

Haastrup, S. (2000). Comparison of some Bayesian analyses of heterogeneity in group life insurance. *Scandinavian Actuarial Journal*, **2000**, 2–16.

Jewell, W.S. (1974). Credible means are exact Bayesian for exponential families. *ASTIN Bulletin*, **8**(1), 77–90.

Katsis, A. & Ntzoufras, I. (2005). Bayesian hypothesis testing for the distribution of insurance claim counts using the Gibbs sampler. *Journal of Computational Methods in Science and Engineering*, **5**, 201–214.

McLachlan, G. & Peel, D. (2000). *Finite Mixture Models*. Wiley-Interscience.

Norberg, R. (1989). Experience rating in group life insurance. *Scandinavian Actuarial Journal*, **1989**, 194–224.

Ntzoufras, I. & Dellaportas, P. (2002). Bayesian modelling of outstanding liabilities incorporating claim count uncertainty. *North American Actuarial Journal*, **6**(1), 113–128.

Ntzoufras, I., Katsis, A. & Karlis, A. (2005). Bayesian assessment of the distribution of insurance claim counts using the reversible jump algorithm. *North American Actuarial Journal*, **9**, 90–105.

Phillips, D.B. & Smith, A.F.M. (1996). *Bayesian model comparison via jump diffusions*. In. W.R. Gilks, S. Richardson & D.J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice* (pp. 215–239). Chapman and Hall.

Puustelli, A., Koskinen, L. & Luoma, A. (2008). Bayesian modelling of financial guarantee insurance. *Insurance: Mathematics and Economics*, **43**(2), 245–254.

Robert, C.P. & Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer.

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components – an alternative to reversible jump methods. *Annals of Statistics*, **28**(1), 40–74.

Streftaris, G. & Worton, B. (2008). Efficient and accurate approximate Bayesian inference with an application to insurance data. *Computational Statistics and Data Analysis*, **52**(5), 2604–2622.

Titterington, D.M., Smith, A.F.M. & Makov, U.E. (1990). *Statistical Analysis of Finite Mixture Distributions*. Wiley, New York, NY.

Tremblay, L. (1992). Using the Poisson inverse Gaussian in bonus-malus systems. *ASTIN Bulletin*, **22**(1), 97–106.

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

Viallefont, V., Richardson, S. & Green, P.J. (2002). Bayesian analysis of Poisson mixtures. *Journal of Nonparametric Statistics*, **14**(1–2), 181–202.

Walhin, J.F. & Paris, J. (1999). Using mixed Poisson processes in connection with bonus-malus systems. *ASTIN Bulletin*, **29**(1), 81–99.

Walhin, J.F. & Paris, J. (2000). The true claim amount and frequency distributions within a bonus-malus system. *ASTIN Bulletin*, **30**(2), 391–403.

https://doi.org/10.1017/S1748499515000019 Published online by Cambridge University Press