Bayesian Assessment of the Distribution of Insurance Claim Counts Using Reversible Jump MCMC

Ioannis Ntzoufras,* Athanassios Katsis,† and Dimitris Karlis‡

Abstract

The aim of this paper is to construct Bayesian model comparison tests between discrete distributions used for claim count modeling in the actuarial field. We use advanced computational techniques to estimate the posterior model odds among different distributions for claim counts. We construct flexible reversible jump Markov Chain Monte Carlo algorithms and implement them in various illustrated examples.

1. Introduction

One of the most important research areas in science and economics has always been the successful modeling of random events such as the level of rainfall, car traffic, or the market demand of a certain commodity. Thus choosing the appropriate statistical distribution is of paramount importance for the accurate representation of the parameter of interest. In actuarial science particularly, the choice of the distribution for the modeling and prediction of the outstanding claims has been extremely important to both practitioners and academics (see, e.g., Makov 2001). Thus the considerable size of the relevant literature comes as no surprise, including the work of Ter Berg (1980) on Poisson and Gamma models, Willmot (1987) on the Poisson inverse Gaussian distribution, Ruohonen (1988) on the Delaporte distribution, Ter Berg (1996) and Scollnik (1998) on the generalized (or Lagrangian) Poisson distribution, Denuit (1997) on the Poisson-Goncharov distribution, and Chaubey et al. (1998) on approximations for aggregate claim distributions. Although the Bayesian approach has been substantially advocated in the outstanding claims problem (see Verrall 1990; de Alba 2002; Ntzoufras and Dellaportas 2002), it has been largely restricted to posterior parameter estimation using mainly the Poisson distribution, rather than testing hypotheses about distributions or estimating model uncertainty. In a general setup, Bayesian inference is based on constructing a model m, its likelihood $f(\boldsymbol{y}|\boldsymbol{\theta}_m, m)$, and the corresponding prior distribution $f(\boldsymbol{\theta}_m|m)$, where $\boldsymbol{\theta}_m$ is a parameter vector under model m and $\boldsymbol{y}$ is the data vector. Inference relies on the posterior distribution $f(\boldsymbol{\theta}_m|\boldsymbol{y}, m)$, whereas quantifying model uncertainty by estimating the posterior model probability $f(m|\boldsymbol{y})$ is an important issue.

For instance, consider two competing models m_0 and m_1. If $f(m)$ is the prior probability of model m, then, using the Bayes theorem, the posterior model odds PO_{01} of model m_0 versus model m_1 is given by

* Ioannis Ntzoufras, PhD, is an Assistant Professor in the Dept. of Statistics, Athens University of Economics and Business, 76 Patision Street, GR-10434, Athens, Greece, ntzoufras@aueb.gr.

† Athanassios Katsis, PhD, is an Assistant Professor in the Dept. of Social and Educational Policy, University of Peloponnese, Damaskinou & Kolokotroni, GR-20100, Korinthos, Greece, katsis@uop.gr.

‡ Dimitris Karlis, PhD, is an Assistant Professor in the Dept. of Statistics, Athens University of Economics and Business, 76 Patision Street, GR-10434, Athens, Greece, karlis@aueb.gr.

$$PO_{01} = \frac{f(m_0|\mathbf{y})}{f(m_1|\mathbf{y})} = \frac{f(\mathbf{y}|m_0)}{f(\mathbf{y}|m_1)} \times \frac{f(m_0)}{f(m_1)} = B_{01} \times \frac{f(m_0)}{f(m_1)},$$

where B_{01} and $f(m_0)/f(m_1)$ are the ''Bayes factor'' and the ''prior model odds'' of model m_0 against model m_1, respectively. Essentially the Bayes factor is the ratio of the posterior model odds over the prior model odds. It is intended to measure the evidence in favor of a certain model with respect to a competing one. Thus, large values of, say, B_{01} (usually greater than 12) indicate strong posterior support of model m_0 against model m_1; for details see Raftery (1996). The quantity $f(\mathbf{y}|m)$ involved in the Bayes factor is called the marginal likelihood of model m and is given by

$$f(\mathbf{y}|m) = \int f(\mathbf{y}|\boldsymbol{\theta}_m, m) f(\boldsymbol{\theta}_m|m) d\boldsymbol{\theta}_m.$$

The Bayes factor B_{10}, of model m_1 against m_0, evaluates the evidence *against* the null hypothesis, which is a familiar concept in classical significance tests; for more details on Bayes factors see Kass and Raftery (1995). Alternatively, when we consider a set of competing models $\mathcal{M} = \{m_1, m_2, \ldots, m_{|\mathcal{M}|}\}$, we focus our interest on the posterior probability of model $m \in \mathcal{M}$, defined as

$$f(m|\mathbf{y}) = \frac{f(\mathbf{y}|m)f(m)}{\sum_{m_l \in \mathcal{M}} f(\mathbf{y}|m_l)f(m_l)} = \left(\sum_{m_l \in \mathcal{M}} PO_{m_l,m} \right)^{-1},$$

where $\mathcal{M}$ and $|\mathcal{M}|$ denote the set and the number of models under consideration, respectively.

Decision theory also can be useful in this case. Under this approach, a utility function for each model is specified, and the model maximizing the expected utility is selected (see for details Bernardo and Smith 1994; Chipman, George, and McCulloch 2001). Strategies based on the maximization of the posterior model probabilities assume zero-one utility functions for choosing the false or the correct model, respectively. This further assumes that the correct model is among the ones under consideration. Within the actuarial context, usually the aim is prediction. In such cases, as one referee pointed out, prediction should be based on the Bayesian model averaging (BMA) technique, which also accounts for model uncertainty (see for details Draper 1995; Chatfield 1995; Clyde 1999; Hoeting et al. 1999; Cairns 2000; Hoeting 2002).

Whatever the final intention is (prediction using BMA or selection of a single model), we need to evaluate posterior model probabilities. The integrals involved in their computations are analytically tractable only in specific examples. Therefore, asymptotic approximations or alternative computational methods must be frequently employed. One of the most popular techniques for calculation of these quantities is the Markov Chain Monte Carlo (MCMC) methodology (see Gilks, Richardson, and Spiegelhalter 1996; Scollnik 2000, 2001) and its recent extensions (reversible jump algorithm or RJMCMC) in varying dimension models (Green 1995); for critical reviews, comparisons, and recent advances see also Carlin and Louis (2000), Chen, Shao, and Ibrahim (2000), Han and Carlin (2001), Chipman, George, and McCulloch (2001), Dellaportas, Forster, and Ntzoufras (2002), and Lopes (2002).

In this paper we focus on the comparison of discrete distributions used for claim counts and the estimation of posterior model probabilities using the Bayesian approach. We do so by employing the RJMCMC algorithm, illustrating its application to the actuarial field using real data sets. The choice of the Bayesian paradigm is beneficial because of the following reasons. First, we can implement model averaging and therefore make more accurate predictions, or alternatively we can use utility functions and base our selection on a fully decision theoretic context. Second, we can compare non-nested models, which is not common practice in the frequentist approach. To improve the exposition, we will focus on three popular discrete distributions; however, the methodology is more general and can be applied to a larger number of candidate distributions. Since this algorithm is computationally advanced, the reader must be familiar with the basic concepts of Bayesian inference and MCMC algorithms.

The paper is organized into five further sections. In Section 2 we present the distributions under consideration. The specification of prior distributions for model comparison, Lindley's paradox, and

their effect on Bayesian model comparisons are discussed in Section 3. In Section 4 we describe in detail the MCMC and reversible jump MCMC algorithms constructed for our model comparisons. Finally, Section 5 includes illustrated examples, while brief closing remarks and conclusions are presented in Section 6.

2. DISTRIBUTIONS FOR CLAIM COUNTS

In this section we describe three widely used distributions for modeling the marginal claim counts, namely, the simple Poisson distribution (Ter Berg 1980), the negative binomial (Verrall 2000), and the generalized Poisson distribution (see, e.g., Consul [1989] for a thorough description of the distribution and Ter Berg [1996] and Scollnik [1998] for its actuarial application). The latter distribution is also known as Lagrangian Poisson. Consequently, the simple Poisson model can be regarded as a special case of either the negative binomial or the generalized Poisson distribution.

Let us assume data y_i, $i = 1, \ldots, n$. Consequently, the simple Poisson model is given by

$$y_i|\lambda \sim Poisson(\lambda)$$

with probability function

$$f(y_i|\lambda) = \frac{\lambda^{y_i} \exp(-\lambda)}{y_i!}.$$

It is well known that, for the Poisson distribution, the mean equals the variance. However, this property is not common in real data sets, where the sample variance usually exceeds the sample mean. For this reason alternative models that allow for overdispersion relative to the simple Poisson model have been considered.

The negative binomial distribution can be constructed by adding a hierarchical step to the simple Poisson model, more specifically

$$y_i|\varepsilon_i, \lambda \sim Poisson(\varepsilon_i\lambda), \quad \varepsilon_i|\vartheta \sim Gamma(\vartheta, \vartheta),$$

where $\vartheta > 0$ and $Gamma(a, b)$ is the gamma distribution with mean a/b and variance a/b^2. The resulting probability function can be written as

$$f(y_i|\lambda, \vartheta) = \frac{\Gamma(y_i + \vartheta)}{\Gamma(y_i + 1)\Gamma(\vartheta)} \left(\frac{\lambda}{\lambda + \vartheta}\right)^{y_i} \left(\frac{\vartheta}{\lambda + \vartheta}\right)^{\vartheta} \tag{2.1}$$

with $E(y_i) = \lambda$ and $V(y_i) = \lambda + \lambda^2/\vartheta$. The Poisson model is a limiting distribution of equation (2.1) for $\vartheta \to \infty$. Here we adopt the parameterization $\phi = \lambda/\vartheta$. In the context of over- or underdispersed distributions the variance-to-mean ratio, called the dispersion index (DI), is usually reported; see, for example, Douglas (1980). For the Poisson random variable the dispersion index is equal to 1. Values far away from 1 indicate that the Poisson assumption is violated. The dispersion index in the negative binomial distribution is equal to

$$DI = \frac{V(y_i)}{E(y_i)} = 1 + \lambda/\vartheta = 1 + \phi.$$

For $\phi \to 0$ the above distribution reduces to the simple Poisson distribution.

The generalized Poisson model with parameters ζ and ω is defined in the following way:

$$f(y_i|\zeta, \omega) = \frac{\zeta(\zeta + \omega y_i)^{y_i - 1}}{y_i!} e^{-(\zeta + \omega y_i)}. \tag{2.2}$$

The mean of this distribution is $E(y_i) = \zeta(1 - \omega)^{-1}$, and the variance is given by $V(y_i) = \zeta(1 - \omega)^{-3}$. According to Ter Berg (1996), valid values for ω are within the interval $[0, 1)$. Typically, the distribution can be also defined for negative values of ω (see for details Scollnik 1998), but such values lead to

underdispersion, which is not a common property of claim count data. For this reason we will not consider this case in the present paper. For $\omega = 0$, the above distribution also reduces to the simple Poisson model with mean ζ.

To enhance comparisons across the three models, we reparametrize the above distribution using $\lambda = \zeta/(1 - \omega)$. Then the reparametrized generalized Poisson distribution has mean $E(y_i) = \lambda$, variance $V(y_i) = \lambda(1 - \omega)^{-2}$, and dispersion index $DI = (1 - \omega)^{-2}$. The probability function is now given by

$$f(y_i|\lambda, \omega) = (1 - \omega)\lambda \frac{\{(1 - \omega)\lambda + \omega y_i\}^{y_i - 1}}{y_i!} e^{-\{(1-\omega)\lambda + \omega y_i\}}. \tag{2.3}$$

This reparametrization is beneficial for the interpretation of the parameters of the three models and will also facilitate the implementation of the MCMC algorithm described below.

3. SPECIFICATION OF THE PRIOR DISTRIBUTIONS

The posterior model odds are very sensitive to the magnitude of their prior variances, tending to increase the support in favor of the simplest model as the prior variances increase (Bartlett 1957; Lindley 1957; see also Sinharay and Stern 2002). Therefore, specifying the prior distribution is pivotal for the a posteriori support of models. In their original publications, Lindley (1957) emphasized the effect of sample size on posterior model odds and the contradicting evidence between Bayesian and classical significance tests, whereas Bartlett (1957) underlined the effect of prior distribution (variance) on the posterior model odds.

Detailed discussion is provided by Shafer (1982), and examples of the sensitivity of posterior model odds to the prior distribution are given by Sinharay and Stern (2002). The problem is even more apparent when comparing two nested models, which is the case here since the Poisson model is nested in both the negative binomial and the generalized Poisson model.

Let us now define our prior distributions, taking this into consideration. We use the following hierarchical structure on model parameters:

$$\begin{aligned} f(\lambda, m_1) &= f(\lambda|m_1)f(m_1), \\ f(\lambda, \vartheta, m_2) &= f(\vartheta|\lambda, m_2)f(\lambda|m_2)f(m_2), \\ f(\lambda, \omega, m_3) &= f(\omega|\lambda, m_3)f(\lambda|m_3)f(m_3). \end{aligned}$$

The usual choice for the prior on model indicator m is the uniform distribution over the parameter space $\mathcal{M}$ resulting in $f(m_i) = \frac{1}{3}$, $i = 1, 2, 3$. This prior can be thought of as noninformative ''in the sense of favoring all models equally'' (Chipman, George, and McCulloch 2001, p. 72). As one referee pointed out, this choice is far from perfect. However, the default prior in the vast majority of Bayesian applications remains the uniform distribution over the model space. Generally, the choice of the prior distribution is a topic of current research investigation extending beyond the scope of introducing the RJMCMC algorithm to an actuarial audience.

In order to be consistent across models with our prior belief, we would like the prior distributions imposed on some common parameters to be the same. For this reason and since the interpretation of λ is the same for all models, we use the same prior distributions $f(\lambda|m)$, which is a $Gamma(a, b)$. Since no prior information is available, the hyperparameters a, b, both can be set equal to a low positive number; in the illustrated examples we use $a = b = 0.0001$ inducing prior mean equal to 1 and variance equal to 10,000. The effect of this choice on the posterior model probabilities will be minimal since parameter λ is present in all models under consideration (see for details Kass and Raftery 1995, sec. 5.3).

Finally, we need to specify the prior distributions $f(\vartheta|\lambda, m_2)$ and $f(\omega|\lambda, m_3)$ such that the prior induced on the dispersion index is the same for both models. For this reason we may specify either $f(\vartheta|\lambda, m_2)$ or $f(\omega|\lambda, m_3)$ and calculate the other distribution deterministically by equating the dispersion indexes

of the distributions under consideration. Using this approach we obtain the same distributions on the dispersion index under both models. By equating the dispersion indexes of the generalized Poisson and negative binomial distributions, we obtain

$$\phi = \lambda/\vartheta = \frac{\omega(2 - \omega)}{(1 - \omega)^2} \quad \text{or} \quad \omega = 1 - \frac{1}{\sqrt{1 + \phi}}. \tag{3.1}$$

Since ω is defined in the interval $[0, 1)$, a usual vague prior for $f(\omega|m_3)$ is the uniform distribution that gives equal probability to any interval of the same range. Therefore, the corresponding prior distribution of ϕ is a Beta type II distribution with parameters 1 and 1/2 and density function

$$f_\phi(\phi|\lambda, m_2) = \frac{1}{2}(1 + \phi)^{-3/2}.$$

Since $\vartheta = \lambda/\phi$, the prior on ϑ is given by

$$f_\vartheta(\vartheta|\lambda, m_2) = f_\phi(\lambda/\vartheta|\lambda, m_2)\lambda\vartheta^{-2} = \frac{1}{2}\lambda\vartheta^{-2}(1 + \lambda/\vartheta)^{-3/2},$$

which is recognized as a scaled Beta type II distribution.

4. Reversible Jump MCMC for Comparison of Claim Count Distributions

In this section we focus on the computation of posterior model odds and probabilities using reversible jump Markov chain Monte Carlo (RJMCMC) methodology introduced by Green (1995). As we have already noted, the integrals involved in the computation of posterior model probabilities have led to the development of various approaches including analytical approximations, Monte Carlo estimates, estimates using the output of MCMC of each model, and recently varying dimensions MCMC (including RJMCMC); for comprehensive reviews and introduction of all these approaches see, for example, Kass and Raftery (1995), Carlin and Louis (2000), Chipman, George, and McCullock (2001), and Lopes (2002). We focus our interest on RJMCMC because of its flexibility and the fact that it can handle a large number of candidate models using a single MCMC run. The latter is evident in variable selection problems (see, e.g., Dellaportas, Forster, and Ntzoufras 2002). Furthermore, since this is an introductory paper on RJMCMC, we illustrate the algorithm using a problem with only three models (Poisson, negative binomial, and generalized Poisson) under consideration. Surely the methodology presented below may be extended to include the evaluation of other competing distributions or models with no additional computational burden.

However, because of the advanced structure of RJMCMC, the researcher needs to be familiar with the fundamental knowledge of Bayesian inference and MCMC algorithms.

4.1 The Reversible Jump MCMC Algorithm

The reversible jump methodology was introduced to the statistical community by Green (1995). It extends the established MCMC techniques since it accounts for comparing models of different dimensions. It is based on creating an irreducible and aperiodic Markov chain that can alternate (*jump*) among various models with parameter spaces of different dimensions, while retaining detailed balance, which ensures the correct limiting distribution. For some critical reviews, comparisons, and recent advances we advise readers to see Carlin and Louis (2000), Chen, Shao, and Ibrahim (2000), Han and Carlin (2001), and Dellaportas, Forster, and Ntzoufras (2002). Further work on the subject can be found in Green and Mira (2001), Rotondi (2002), and Brooks, Guidici, and Roberts (2003).

Let us assume a set of competing models $\mathcal{M}$. A latent variable $m \in \mathcal{M}$ indicates each model, whereas $\boldsymbol{\theta}_m$ denotes the corresponding parameter vector. If the current state of the Markov chain is $(m, \boldsymbol{\theta}_m)$, where $\boldsymbol{\theta}_m$ has dimension d_m, then a general version of the algorithm is the following:

- Propose a new model m' with probability $j(m, m')$.
- Generate $\boldsymbol{u}$ from a specified proposal density $q(\boldsymbol{u}|\boldsymbol{\theta}_m, m, m')$.
- Propose a new vector of parameters $\boldsymbol{\theta}'_{m'}$ by setting $(\boldsymbol{\theta}'_{m'}, \boldsymbol{u}') = g_{m,m'}(\boldsymbol{\theta}_m, \boldsymbol{u})$ where $g_{m,m'}$ is a specified invertible function.
- To achieve the correct limiting distribution, accept the proposed move to model m' with probability

$$\alpha = \min\left(1, \frac{f(\boldsymbol{y}|m', \boldsymbol{\theta}'_{m'})f(\boldsymbol{\theta}'_{m'}|m')f(m')j(m', m), q(\boldsymbol{u}'|\boldsymbol{\theta}_m, m', m)}{f(\boldsymbol{y}|m, \boldsymbol{\theta}_m)f(\boldsymbol{\theta}_m|m)f(m)j(m, m')q(\boldsymbol{u}|\boldsymbol{\theta}_{m'}, m, m')} \left|\frac{\partial g(\boldsymbol{\theta}_m, \boldsymbol{u})}{\partial(\boldsymbol{\theta}_m, \boldsymbol{u})}\right|\right). \tag{4.1}$$

Important features for the efficiency and the practical implementation of the algorithm are the proposal distributions $q(\boldsymbol{u}|\boldsymbol{\theta}_m, m, m')$ and the matching function $g_{m',m}$. The vectors $\boldsymbol{u}$ and $\boldsymbol{u}'$ are used to make the dimensions of the parameter spaces of m and m' equal. The usual practice is to set $d_{\boldsymbol{u}}$ or $d_{\boldsymbol{u}'}$ equal to zero depending on which model has fewer parameters. When $d_m < d_{m'}$, we set $d_{\boldsymbol{u}'} = 0$, generate $\boldsymbol{u}$ as described above, and calculate $\boldsymbol{\theta}'_{m'}$ using the matching function $g_{m,m'}$. Otherwise, when $d_{m'} < d_m$, we set $d_{\boldsymbol{u}'}$ equal to zero and directly calculate $\boldsymbol{\theta}_{m'}$ and $\boldsymbol{u}'$ using the matching function $g_{m,m'}$ since we do not need to generate any additional parameters.

The corresponding proposal distributions are usually constructed by single MCMC runs within each model (see Dellaportas, Forster, and Ntzoufras 2002), while the matching function $g_{m',m}$ is constructed by considering the structural properties of each model and their possible association (for an example see Ntzoufras, Dellaportas, and Forster 2003). Finally, note that because of symmetry, $g_{m',m} = g^{-1}_{m,m'}$.

4.2 Application to the Claim Count Distributions

Let us now construct the RJMCMC algorithm for the claim count data, using the three competing models described in Section 2. The latent model indicator m takes values $m \in \{m_1, m_2, m_3\}$ corresponding to the Poisson, negative binomial, and generalized Poisson distributions, respectively. Moreover, we denote by $\boldsymbol{\theta}_{m_1} = \lambda$, $\boldsymbol{\theta}_{m_2} = (\lambda, \vartheta)^T$, and $\boldsymbol{\theta}_{m_3} = (\lambda, \omega)^T$ the parameters of the Poisson, negative binomial, and generalized Poisson distributions, respectively.

Assuming the current state of the Markov chain, $(m, \boldsymbol{\theta}_m)$, the reversible jump for the comparison of the distributions we are interested in can then be formulated in the following way:

1. Generate model parameters $\boldsymbol{\theta}_m$ from the posterior distribution $f(\boldsymbol{\theta}_m|\boldsymbol{y}, m)$; see Appendix A for the computational details.
2. Propose with probability $j(m, m') = (|\mathcal{M}| - 1)^{-1}$ to jump from m to m', where $m' \neq m$.
3. (a) i. If $m = m_1$ (Poisson) and $m' = m_2$ (negative binomial), then generate a proposed value for ϑ from the proposal distribution $q_\vartheta(\vartheta|m)$.
   ii. If $m = m_1$ (Poisson) and $m' = m_3$ (generalized Poisson), then generate a proposed value for ω from the proposal distribution $q_\omega(\omega|m)$.
   iii. If $m = m_2$ (negative binomial) and $m' = m_3$ (generalized Poisson), then set the proposed value for ω using the matching function

$$\omega = h_{m_2,m_3}(\vartheta) = 1 - (1 + \lambda\vartheta^{-1})^{-1/2},$$

   which is derived by equating the dispersion indexes of the two distributions.
   iv. If $m = m_2$ (negative binomial) or m_3 (generalized Poisson) and $m' = m_1$ (Poisson), then we do not need to generate any additional parameters.
   v. If $m = m_3$ (generalized Poisson) and $m' = m_2$ (negative binomial), then we set

$$\vartheta = h^{-1}_{m_2,m_3}(\omega) = h_{m_3,m_2}(\omega) = \frac{\lambda(1 - \omega)^2}{\omega(2 - \omega)}.$$

   (b) Accept the proposed move with probability $\alpha(m, m') = min\{1, \delta(m, m')\}$, where

$$\delta(m_1, m_2) = \frac{f(\boldsymbol{y}|\lambda, \vartheta, m_2) f(\lambda, \vartheta|m_2) f(m_2)}{f(\boldsymbol{y}|\lambda, m_1) f(\lambda|m_1) f(m_1) q_{\vartheta}(\vartheta|m_1)},$$

$$\delta(m_1, m_3) = \frac{f(\boldsymbol{y}|\lambda, \omega, m_3) f(\lambda, \omega|m_3) f(m_3)}{f(\boldsymbol{y}|\lambda, m_1) f(\lambda|m_1) f(m_1) q_{\omega}(\omega|m_1)},$$

$$\delta(m_2, m_3) = \frac{f(\boldsymbol{y}|\lambda, \omega, m_3) f(\lambda, \omega|m_3) f(m_3)}{f(\boldsymbol{y}|\lambda, \vartheta, m_2) f(\lambda, \vartheta|m_2) f(m_2)} \times \frac{1}{2} (1 + \lambda\vartheta^{-1})^{-3/2} \lambda\vartheta^{-2}.$$

For all other inverse comparisons we have $\delta(m, m') = 1/\delta(m', m)$.

Note that, since a conjugate prior distribution is used for the Poisson model, the posterior of $f(\lambda|\boldsymbol{y}, m_1)$ is a Gamma distribution with parameters $a' = \Sigma_{i=1}^{n} y_i + a$ and $b' = n + b$. Therefore in step 1 above, when $m = 1$, we directly generate values of λ from the posterior distribution.

In the above algorithm, the comparison between models m_2 and m_3 can be also achieved by constructing a Metropolized version of the Carlin and Chib (1995) algorithm (or independence sampler) as described in Dellaportas, Forster, and Ntzoufras (2002). In such a case, we propose additional parameters ϑ and ω by the proposal distributions of steps 3(a)i and 3(a)ii and calculate $\delta(m_2, m_3) = \delta(m_1, m_3)/\delta(m_1, m_2)$. In the algorithm proposed above, we alternatively have constructed a more automatic version of RJMCMC for the comparison of models m_2 and m_3, following the approach of Ntzoufras, Dellaportas, and Forster (2003). Instead of generating the parameters ϑ and ω, we propose the additional parameters needed by matching their dispersion indexes. This results in a one-to-one transformation given by equation (3.1) with the property of retaining the dispersion index constant when a ''jump'' from the negative binomial to the generalized Poisson distribution (or vice versa) is proposed.

Using the proposed prior formulation presented in Section 3 the above ratio degenerates to a simple comparison of the likelihoods (adjusted by the prior model probabilities) in each iteration

$$\delta(m_2, m_3) = \frac{f(\boldsymbol{y}|\lambda, \omega, m_3) f(m_3)}{f(\boldsymbol{y}|\lambda, \vartheta, m_2) f(m_2)}.$$

4.3 Specification of Proposal Distributions

An important aspect regarding the efficiency of the RJMCMC is the careful specification of the proposal distributions $q_{\vartheta}(\vartheta|m)$ and $q_{\omega}(\omega|m)$. The proposal distributions produce values for the additional parameters when a ''jump'' from one model to another is proposed. Hence an efficient RJMCMC algorithm should propose parameter values close to the corresponding posterior distribution of the proposed model m'. If this does not happen, then the proposed values will be frequently rejected, and hence the algorithm will either be concentrated on one model or converge to the correct posterior distribution very slowly.

Here we use pilot runs of MCMC algorithms of length 1,000 iterations within each model. These values serve as rough estimates of the posterior distributions of each model and help us to ensure that the proposed values are close to the target posterior distributions. For the parameter ϑ of the negative binomial model (which takes positive values) a log-normal distribution is employed $q_{\vartheta}(\vartheta|m) = LN(\overline{\log\ \vartheta}, \bar{\sigma}^2_{\log\ \vartheta})$, where $\overline{\log\ \vartheta}$ and $\bar{\sigma}^2_{\log\ \vartheta}$ are the mean and variance of the $\log(\vartheta)$, respectively, estimated using the pilot run; for more details regarding the log-normal density see equation (A.1) given in Appendix A. Alternatively, we may use any other distribution defined in the interval $(0, \infty)$ by matching again their parameters with the posterior mean and variance taken by the pilot runs. The efficiency of each proposal distribution will depend on how close it is to the underlying posterior distribution. We should emphasize that posterior model probabilities should be robust on different choices of proposal distributions provided that good mixing between models is achieved. On the other hand, the efficiency of the algorithm will be sensitive to different choices of proposal distributions, and hence the number of iterations needed to achieve convergence will depend on such a choice; for a detailed discussion and

automatic specification of proposals see Dellaportas, Forster, and Ntzoufras (2002) and Brooks, Giudici, and Roberts (2003).

Similarly, for the parameter ω of the generalized Poisson distribution, which takes values in the interval [0, 1), we utilize $q_\omega(\omega|m) = Beta(\bar{a}, \bar{b})$. The parameters $\bar{a}$ and $\bar{b}$ can be calculated by equating the mean and the variance of the Beta distribution with the estimated mean ($\bar{\omega}$) and variance ($\bar{\sigma}^2_\omega$) of ω from the output of an MCMC pilot run for the GP distribution. Hence

$$\bar{\omega} = \frac{\bar{a}}{\bar{a} + \bar{b}}, \quad \bar{\sigma}^2_\omega = \frac{\bar{a}\bar{b}}{(\bar{a} + \bar{b})^2(\bar{a} + \bar{b} + 1)},$$

which leads to

$$\bar{a} = \bar{\omega}\left(\frac{\bar{\omega}(1 - \bar{\omega})}{\bar{\sigma}^2_\omega} - 1\right), \quad \bar{b} = \bar{a}\,\frac{1 - \bar{\omega}}{\bar{\omega}}.$$

The efficiency of the chain can be improved by suitably increasing or decreasing the variance of the proposal distribution to achieve high acceptance rates.

4.4 Analysis of the RJMCMC Output

After running the MCMC algorithm L iterations we get a sample with values $m^{(k)}$, $\lambda^{(k)}$, $\vartheta^{(k)}$, $\omega^{(k)}$ for each iteration $k = 1, \ldots, L$. The variable m will be an indicator taking values 1, 2, or 3 depending on which model the algorithm visited on the corresponding iteration. When m_1 is visited, then $m = 1$ and $\vartheta = \omega = 0$; when m_2 is visited, then $m = 2$ and $\omega = 0$; when m_3 is visited, then $m = 3$ and $\vartheta = 0$. Note that we discard the first B iterations as a burn-in to eliminate possible effect of initial values.

From the RJMCMC output the posterior model probability $f(m_i|\boldsymbol{y})$ for $i = 1, 2, 3$ is estimated by

$$\hat{f}(m_i|\boldsymbol{y}) = \frac{1}{L - B}\sum_{k=B+1}^{L} I_{m_i}(m^{(k)}),$$

where the indicator function $I_{m_i}(m^{(k)}) = 1$ if $i = m^{(k)}$ and zero otherwise. Then we use the above estimate to calculate the posterior model odds and the Bayes factors given by

$$\widehat{PO}_{ij} = \frac{\hat{f}(m_i|\boldsymbol{y})}{\hat{f}(m_j|\boldsymbol{y})} = \widehat{BF}_{ij}\frac{\hat{f}(m_i)}{\hat{f}(m_j)} \Leftrightarrow,$$

$$\widehat{BF}_{ij} = \widehat{PO}_{ij}\frac{\hat{f}(m_j)}{\hat{f}(m_i)}. \tag{4.2}$$

The posterior distribution of each parameter within each model can be described using simple descriptive measures as the mean derived by

$$\hat{E}(\lambda|\boldsymbol{y}, m_i) = \frac{1}{(L - B)\hat{f}(m_i|\boldsymbol{y})}\sum_{k=B+1}^{L}\{\lambda^{(k)}I_{m_i}(m^{(k)})\}, \quad \text{for} \quad i = 1, 2, 3,$$

$$\hat{E}(\vartheta|\boldsymbol{y}, m_2) = \frac{1}{(L - B)\hat{f}(m_2|\boldsymbol{y})}\sum_{k=B+1}^{L}\{\vartheta^{(k)}I_{m_2}(m^{(k)})\},$$

$$\hat{E}(\omega|\boldsymbol{y}, m_3) = \frac{1}{(L - B)\hat{f}(m_3|\boldsymbol{y})}\sum_{k=B+1}^{L}\{\omega^{(k)}I_{m_3}(m^{(k)})\}.$$

Similarly we can produce Monte Carlo estimates of any descriptive measure or more generally any function $T(\boldsymbol{\theta}_m)$ of the parameters we wish to describe or estimate. A general equation for estimating the posterior expectation of any function $T(\boldsymbol{\theta}_m)$ under model m is given by

$$E\{T(\boldsymbol{\theta}_m)|\boldsymbol{y}, m\} = \frac{1}{(L - B)\hat{f}(m|\boldsymbol{y})} \sum_{k=B+1}^{L} \{T(\boldsymbol{\theta}_m^{(k)}) I_{m_i}(m^{(k)})\}.$$

For example, if we set $T(\boldsymbol{\theta}_m) = \lambda$ or $T(\boldsymbol{\theta}_m) = \{\lambda - \hat{E}(\lambda|\boldsymbol{y}, m)\}^2$, then the above equation produces Monte Carlo estimates of the posterior mean and variance for parameter λ under model m.

The efficiency of such estimates can be quantified by measuring Monte Carlo error (that is, deviance due to the simulation). A popular method is the batch-mean method (see for more details Roberts 1996) in which we split the final sample (output of MCMC) in K subsamples (usually 30–50; in the illustrated examples that follow we have used 50). Each subsample produces an independent estimate of the desired quantity of interest (here λ, ϑ, ω, $f(m_i|\boldsymbol{y})$, or posterior odds). The results of these subsamples can be used to monitor convergence of a quantity (by plotting the evolution of the posterior model probabilities, posterior model odds, or simply the mean of a parameter) or to roughly estimate Monte Carlo error by the standard deviation of the estimates in each subsample.

Bayesian model averaging can be directly imposed by the MCMC output of RJMCMC. For example, we can produce an estimate of the dispersion index using Bayesian model averaging using the following equation:

$$\hat{E}(DI|\boldsymbol{y}) = \frac{1}{L - B} \sum_{i=B+1}^{L} D(\boldsymbol{\theta}_m^{(k)}, m^{(k)}),$$

where $D(\boldsymbol{\theta}_m, m)$ is the function for the dispersion index for each model m as defined in Section 2.

Usually we leave the RJMCMC algorithm to run using the desired prior model probabilities and then directly estimate their posterior values from the output of the chain. In such cases specific models are not visited at all by the RJMCMC algorithm because of their low posterior probability, and hence their relative posterior model odds PO_{ij} cannot be estimated with accuracy. To avoid this we can use the following search algorithm for ''tuning'' the prior model probabilities and estimating posterior odds with precision. The proposed algorithm is suitable for pairwise comparisons, but it can be easily generalized for more than two models. In this algorithm we search for appropriate values of prior model probabilities that result in posterior probabilities for all models that are close to each other. This helps us to estimate posterior model odds (using these ''tuned'' prior probabilities) with precision. Then, using equation (4.2), we can estimate the underlying Bayes factor and, consequently, the corresponding posterior model odds using the prior probabilities we originally accept. The ''search'' algorithm can be summarized using the following steps:

1. Let us denote our desired prior model probabilities by $f(m_1)$ and $f(m_2)$. We start by setting $f^{(1)}(m_1) = f(m_1)$ and $f^{(1)}(m_2) = f(m_2)$, where $f^{(k)}(m)$ are the prior model probabilities used in the k-th iteration of this step.
2. Run the RJMCMC algorithm for a limited number of iterations (say, 1,000) long enough to roughly estimate the magnitude of posterior model probabilities.
3. (a) If the posterior model odds $PO_{12}^{(k)} \notin (0.5 - \xi, 0.5 + \xi)$, then
   i. Set

$$\log \frac{f^{(k+1)}(m_1)}{f^{(k+1)}(m_2)} = \log \frac{f^{(k)}(m_1)}{f^{(k)}(m_2)} - \log \psi, \tag{4.3}$$

   where ψ is usually the posterior model odds $PO_{12}^{(k)}$ unless it takes very small or large values and it is replaced by $\varepsilon/(1 - \varepsilon)$ or $(1 - \varepsilon)/\varepsilon$, respectively. Hence we write that

$$\psi = \min \left\{ \max \left(PO_{12}^{(k)}, \frac{\varepsilon}{1 - \varepsilon} \right), \frac{1 - \varepsilon}{\varepsilon} \right\}, \tag{4.4}$$

   where $PO_{12}^{(k)}$ is the posterior model odds estimated for prior model probabilities $f^{(k)}(m)$ and ε defines an interval of type $(\varepsilon, 1 - \varepsilon)$ outside of which we cannot estimate with precision

the posterior model probabilities. We propose to use $\varepsilon = 0.01$, and hence when $PO_{12}^{(k)} < 0.01$, $\psi = 0.01/0.99 = 0.0101$, while when $PO_{12}^{(k)} > 0.99$, $\psi = 99$.

ii. Return to step 2.

(b) If $PO_{12}^{(k)} \in (0.5 - \xi, 0.5 + \xi)$, then

i. Rerun the RJMCMC with full number of iterations required.

ii. Recalculate $PO_{12}^{(k)}$ from the output of the full RJMCMC.

iii. Calculate the posterior model odds we are interested in using:

$$PO_{12} = PO_{12}^{(k)} \times \frac{f^{(k)}(m_2)}{f^{(k)}(m_1)} \times \frac{f(m_1)}{f(m_2)}. \tag{4.5}$$

We note that the prior model probabilities assume various values not because of any change in the prior beliefs but to ensure that the posterior model odds fall within a specified interval around 0.5. When this happens, we employ equation (4.5) to calculate the posterior model odds for the desired prior model probabilities. In the above algorithm, ξ specifies an interval around 0.5 in which we wish the ''tuned'' posterior model odds to be included. We propose the value $\xi = 0.1$ (hence we use the interval 0.4–0.6) for estimating posterior model odds with increased precision. Alternatively, the value $\xi = 0.2$ (interval 0.3–0.7) also worked well enough in our examples and, on the other hand, considerably reduced the computational effort. Both equations (4.3) and (4.4) can be derived using equation (4.2) assuming that we calculate two posterior model odds using different prior model probabilities and setting their Bayes factors (which do not depend on prior model probabilities) equal. The above process will reduce Monte Carlo error and produce stable estimates for the Bayes factor. When more than two models are under consideration (for example, here we consider three), we may try to tune the posterior model probabilities around $1/|\mathcal{M}|$, and hence they will lie within an interval $[1/|\mathcal{M}| - \xi, 1/|\mathcal{M}| + \xi]$, where $|\mathcal{M}|$ is the number of models under consideration. When the number of models under consideration is large (for example, in the context of variable selection), such action will not be possible. In such cases we propose to search the whole model space $\mathcal{M}$ and then remove all models with low posterior probability and work only on the reduced model space (see for similar techniques Hoeting et al. 1999).

5. ILLUSTRATED EXAMPLES

In this section we employ eight data sets of insurance claims for different countries and years used by Denuit (1997) in order to implement our proposed methodology (see Fig. 1). The first six data sets were also used earlier by Gossiaux and Lemaire (1981), and the first data set (Switzerland, 1961) was used by Ter Berg (1996) as well. All data sets are listed in detail in Appendix B (see Table 7).

Here we present posterior summaries of the parameters of interest for the Poisson (Table 1), negative binomial (Table 2), and generalized Poisson distributions (Table 3). The results concerning the Poisson distribution have been calculated analytically according to Appendix A. We further present boxplots of the posterior distribution of the dispersion index resulting from the negative binomial and the generalized Poisson distributions (see Figure 2). The posterior distribution of the dispersion index is quite close for both distributions.

Regarding the implementation of RJMCMC, we have used pairwise comparisons tuned a priori to achieve posterior model probabilities within the interval 0.40–0.60 (for details see Section 4.4). The emphasis on the estimation of the Bayes factor is important since all the results show that the negative binomial and/or the generalized Poisson distributions are a posteriori much more probable than the simpler Poisson model. In all the chains, we have considered 1,000 iterations as a burn-in and an additional 20,000 iterations for calculation of the posterior distributions. Initial values were calculated using the moment estimates. Parameters of the proposal distributions were based on pilot runs of 2,000 iterations with their variances tuned to achieve high acceptance rates, greater than 80% in all data sets.

Figure 1
Histograms of Denuit (1997) Data

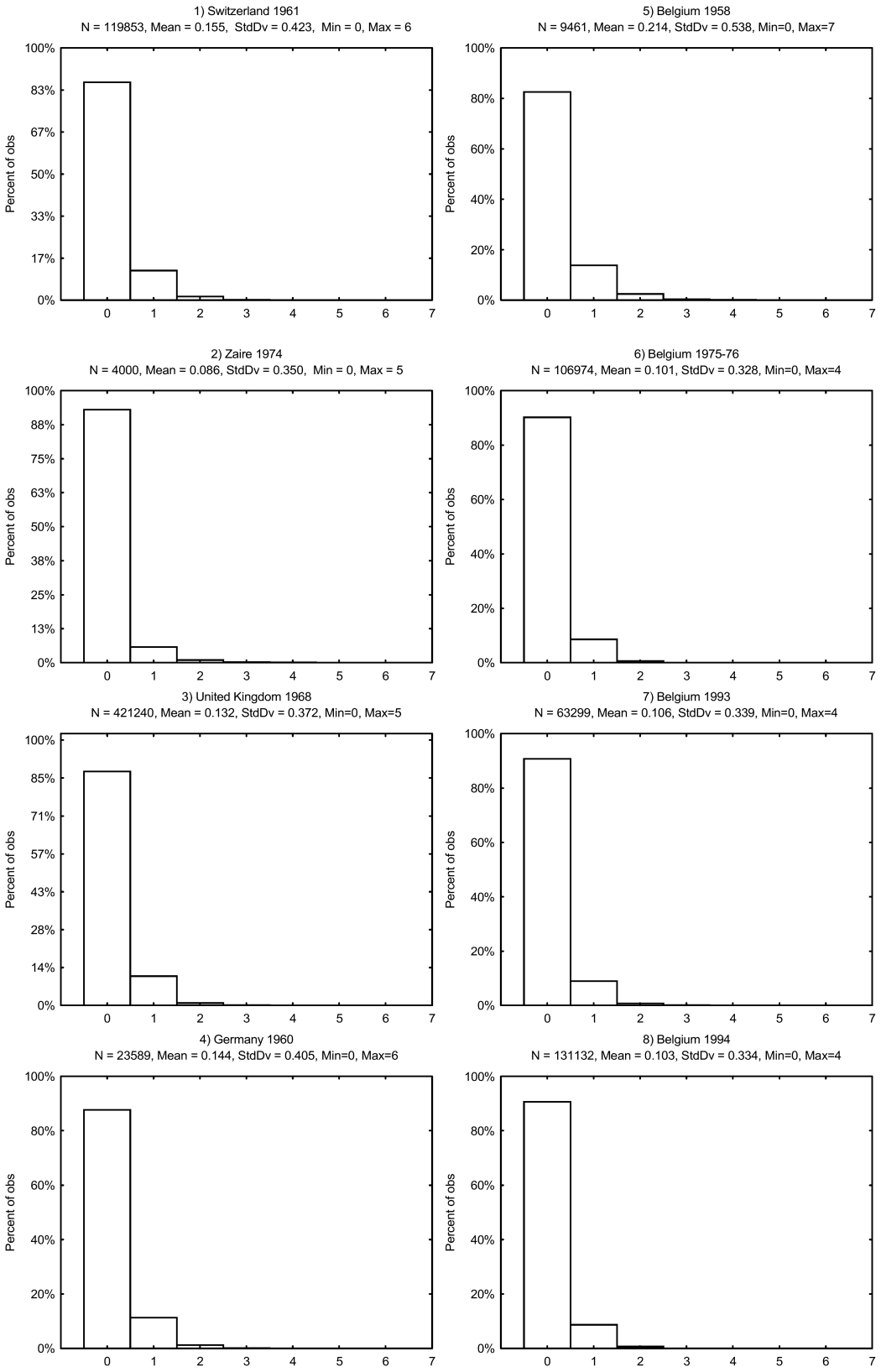


Table 1
Posterior Summaries of λ of Poisson Model for Denuit (1997) Data
$f(\lambda|\boldsymbol{y}) \sim \textit{Gamma}(a', b')$

Data Set	Posterior Values			
	a' $\{\sum_{i=1}^{n} y_i + a\}$	b' $\{n + b\}$	Mean $\{E(\lambda\|\boldsymbol{y}) = a'/b'\}$	Std. Dev. $\{S_{\lambda\|\boldsymbol{y}} = \sqrt{a'}/b'\}$
Switzerland 1961	18,594	119,853	0.155	0.0011
Zaire 1974	346	4,000	0.087	0.0047
United Kingdom 1968	55,493	421,240	0.132	0.0006
Germany 1960	3,402	23,589	0.144	0.0025
Belgium 1958	2,028	9,461	0.214	0.0048
Belgium 1975–76	10,813	106,974	0.101	0.0010
Belgium 1993	6,691	63,299	0.106	0.0013
Belgium 1994	13,594	131,182	0.104	0.0009

Table 2
Posterior Summaries of Parameters of Negative Binomial Distribution for Denuit (1997) Data

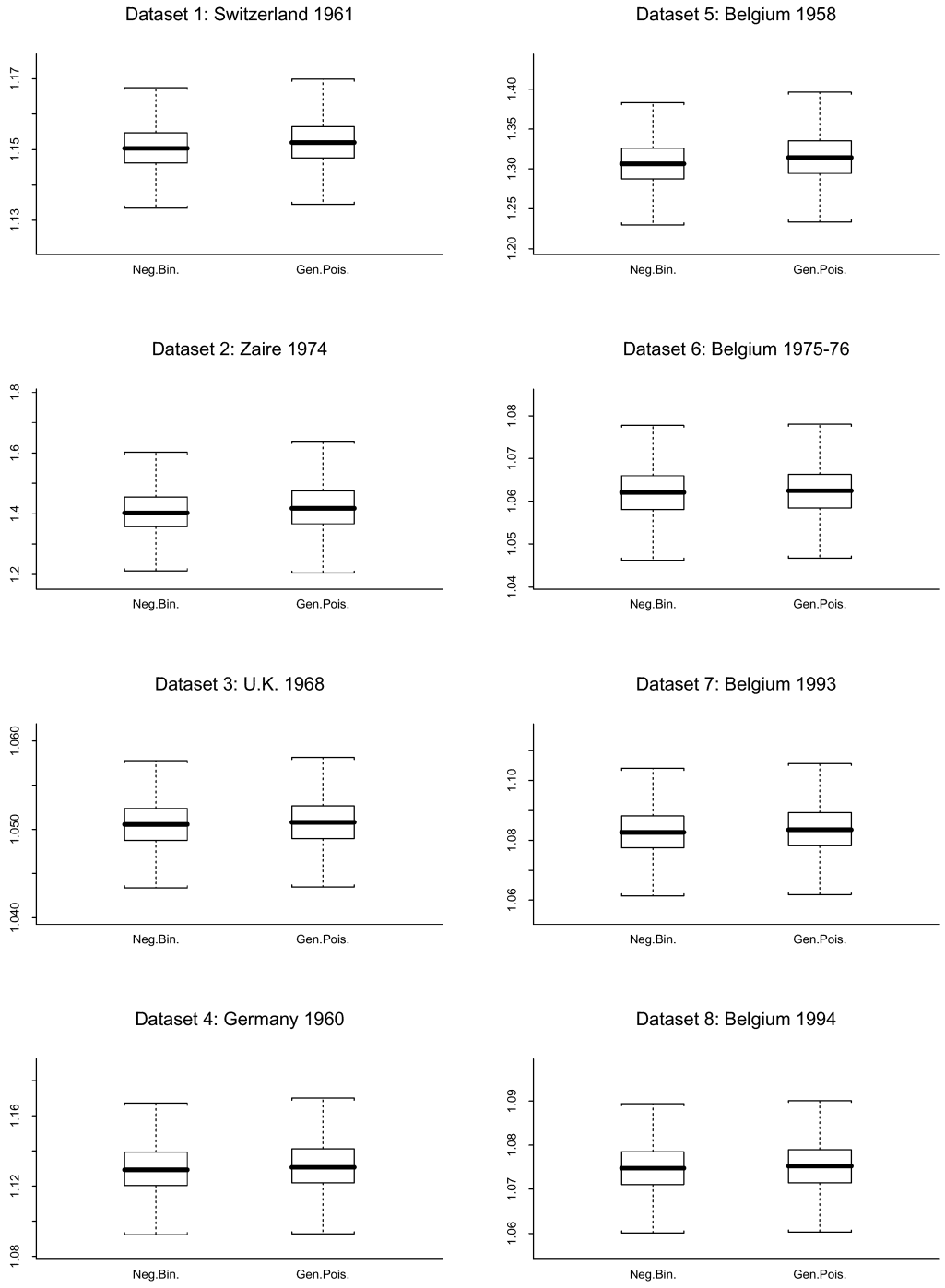
Data Set	Posterior Mean ± Standard Deviation		
	λ	ϑ	Dispersion Index
Switzerland 1961	0.155 ± 0.0012	1.033 ± 0.045	1.151 ± 0.007
Zaire 1974	0.087 ± 0.0056	0.218 ± 0.038	1.410 ± 0.075
United Kingdom 1968	0.132 ± 0.0006	2.607 ± 0.138	1.051 ± 0.003
Germany 1960	0.144 ± 0.0026	1.127 ± 0.126	1.130 ± 0.014
Belgium 1958	0.214 ± 0.0056	0.704 ± 0.062	1.307 ± 0.028
Belgium 1975–76	0.101 ± 0.0010	1.637 ± 0.154	1.062 ± 0.006
Belgium 1993	0.106 ± 0.0013	1.284 ± 0.124	1.083 ± 0.008
Belgium 1994	0.104 ± 0.0009	1.392 ± 0.103	1.076 ± 0.005

The logarithms of the estimated Bayes factors are given in Table 4, and the graphical evolution of the estimated log-Bayes factor is provided in Figure 3 (additional similar figures are available upon request). Estimated Bayes factors for all data sets imply very strong evidence against the simpler Poisson model ($B_{21} > 8.6 \times 10^{25}$ and $B_{31} > 10.97 \times 10^{25}$ for all data sets). The differences between negative binomial and generalized Poisson are small, but, in all data sets, the generalized Poisson model is supported slightly more strongly against the negative binomial model ($1.22 \leq B_{32} \leq 10.80$ and $0.55 \leq f(m_3|\boldsymbol{y}) \leq 0.915$). The support in favor of the generalized Poisson is stronger for data sets 1 and 5 (Switzerland 1961 and Belgium 1958, respectively). Following the interpretation of Kass and Raftery (1995) we have ''positive'' evidence in favor of the generalized Poisson model (against the negative binomial model) only for these two data sets, while for all other comparisons the evidence is low.

Table 3
Posterior Summaries of Parameters of Generalized Poisson Distribution for Denuit (1997) Data

Data Set	Posterior Mean ± Standard Deviation		
	λ	ω	Dispersion Index
Switzerland 1961	0.155 ± 0.0012	0.068 ± 0.0027	1.152 ± 0.007
Zaire 1974	0.087 ± 0.0056	0.161 ± 0.0240	1.425 ± 0.082
United Kingdom 1968	0.132 ± 0.0006	0.025 ± 0.0013	1.051 ± 0.003
Germany 1960	0.144 ± 0.0027	0.060 ± 0.0061	1.131 ± 0.015
Belgium 1958	0.215 ± 0.0056	0.128 ± 0.0099	1.315 ± 0.023
Belgium 1975–76	0.101 ± 0.0010	0.030 ± 0.0027	1.062 ± 0.006
Belgium 1993	0.106 ± 0.0013	0.039 ± 0.0036	1.084 ± 0.008
Belgium 1994	0.104 ± 0.0009	0.036 ± 0.0025	1.077 ± 0.005

Figure 2
Boxplots of Posterior Densities of Dispersion Indexes for Denuit (1997) Data



Predictive distributions are provided for the Switzerland 1961 and Belgium 1958 data in Tables 5 and 6. For the Switzerland 1961 data, we can clearly see some differences between the predictive distribution of the simple Poisson model and the observed data. On the other hand, the differences between the predictive distributions of *NB* and *GP* and the observed data are much smaller. Although differences between *NB* and *GP* are relatively small, the data provide (through the Bayes factor) ''positive'' evidence in favor of the generalized Poisson model when compared to the negative binomial data. The results

Table 4

Estimated Log Bayes Factors and Posterior Model Probabilities for Denuit (1997) Data

Data Set	Log-Bayes Factor			
	m_2 vs. m_1	m_3 vs. m_1	m_3 vs. m_2	$f(m_3\|\boldsymbol{y})$
Switzerland 1961	488.17	490.55	2.38	0.915
Zaire 1974	59.72	59.96	0.23	0.560
United Kingdom 1968	230.45	231.24	0.78	0.688
Germany 1960	70.25	70.82	0.56	0.639
Belgium 1958	139.10	140.97	1.88	0.867
Belgium 1975–76	79.16	79.37	0.20	0.552
Belgium 1993	81.54	82.23	0.69	0.666
Belgium 1994	140.91	141.41	0.49	0.622

Notes: m_1: Poisson distribution; m_2: negative binomial distribution; m_3: generalized Poisson distribution; $f(m_3|\boldsymbol{y})$: posterior probability of model m_3; $f(m_1|\boldsymbol{y}) = 0$ and $f(m_2|\boldsymbol{y}) = 1 - f(m_3|\boldsymbol{y})$ for all data sets.

Figure 3

Ergodic Log Bayes Factors of Generalized Poisson versus Negative Binomial Distribution for Denuit (1997) Data

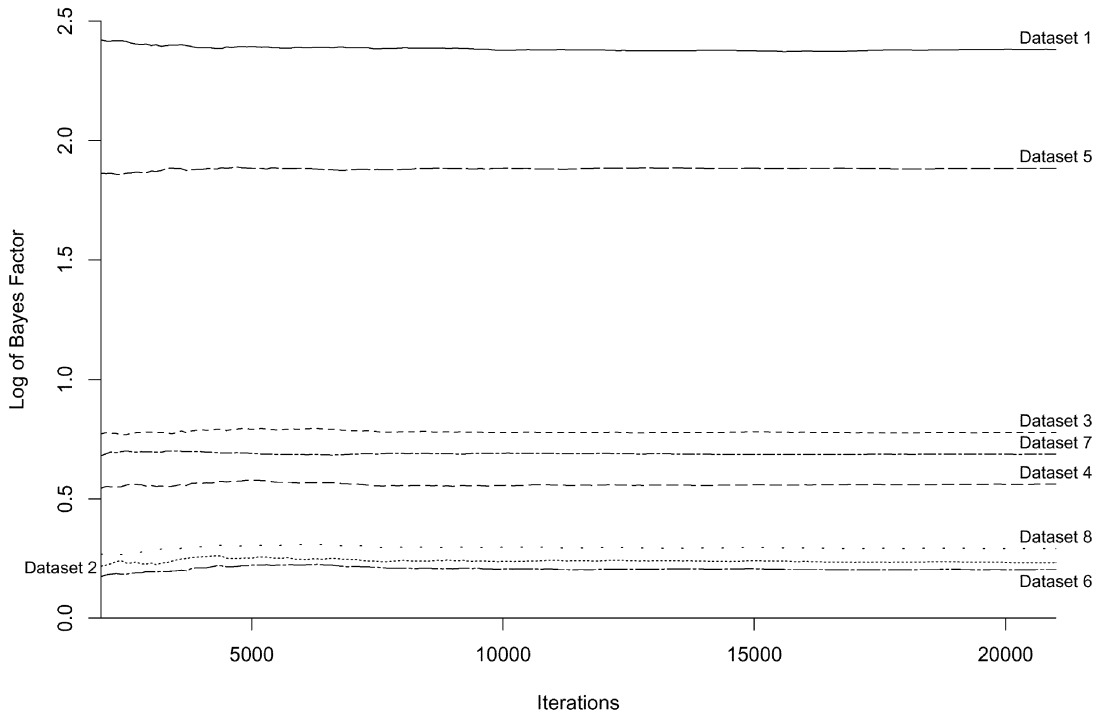


Table 5

Descriptive Summaries of Predictive Distributions for Models 1–3 for Switzerland 1961 Data

Model	Predictive Means of Frequencies (Y)							Mean	DI
	0	1	2	3	4	5	6		
Poisson	102,630.3	15,921.5	1234.9	63.8	2.5	0.1	0.0	0.155	1.000
Negative binomial	103,724.9	13,988.7	1856.5	245.6	32.3	4.3	0.6	0.155	1.150
Generalized Poisson	103,724.1	14,002.1	1837.7	248.7	34.6	4.9	0.7	0.155	1.152
Observed	103,704	14,075	1766	255	45	6	2	0.155	1.156

Table 6
Descriptive Summaries of Predictive Distributions for Models 1–3 for Belgium 1958 Data

Model	Predictive Means of Frequencies (Y)								Mean	DI
	0	1	2	3	4	5	6	7		
Poisson	7636.6	1635.8	175.3	12.5	0.7	0.0	0.0	0.0	0.214	1.000
Negative binomial	7846.4	1288.5	256.6	54.3	11.8	2.6	0.6	0.1	0.214	1.306
Generalized Poisson	7848.4	1290.5	251.4	54.1	12.5	3.1	0.8	0.2	0.214	1.315
Observed	7840	1317	239	42	14	4	4	1	0.214	1.348

for the Belgium 1958 data set are similar. Results of a Bayesian model averaging approach can be easily calculated using weighted means of the quantities reported in Tables 5 and 6.

For the calculations of all the examples presented in this article we have used the Fortran 77 computing language for the generation of the RJMCMC output and then Splus statistical software for its statistical analysis. Any other computing language package can be used without any difficulty. All the computations have been conducted on a Pentium II/600MH personal computer. The code is available on request. We strongly recommend that practitioners develop their own versions of the code to fully grasp the method and be able to adapt it to their own problems. We are now in the process of developing RJMCMC code suited for the R package.

6. DISCUSSION

In this paper we have implemented advanced MCMC techniques in order to assess and compare, using the Bayesian approach, three popular distributions concerning the distribution of claim counts. The use of the Bayesian paradigm is advantageous because it enables model averaging or utility functions as well as comparisons among non-nested models. Results from eight data sets from various countries indicate that the Poisson distribution is not adequate to describe claim count data. The negative binomial and the generalized Poisson distributions are a posteriori much more probable than the Poisson distribution. Between the two, the generalized Poisson is consistently superior a posteriori to the negative binomial distribution, although the strength of this evidence varies from data set to data set.

It must be emphasized that the negative binomial distribution and the generalized Poisson distributions are quite similar for the range of parameters found in the used data sets. In addition, Douglas (1994) pointed out that for count data with small counts, various discrete distributions can fit the data sufficiently well. It is very important that our model choice approach succeeded in choosing between the two models as indicated by the values of the Bayes factors. For larger counts the differences would be more apparent.

Further research may include the implementation of the above methodology in the general claim counts context (see England and Verrall 2002) in order to test if the Poisson distribution is adequate or if more sophisticated models are demanded. Moreover, we may include covariates on λ to treat more sophisticated cases such as the prediction of outstanding claim reserves. Another important issue is the possible extension of RJMCMC methodology in a larger variety of distributions. Using this approach, we may compare different models involved in the claim counts and amounts literature. The extension of the RJMCMC algorithm in such cases is by no means straightforward and needs increased experience and careful construction of each chain. Implementing RJMCMC in a wider variety of actuarial models and problems will enable us to use Bayesian model averaging techniques that increase the predictive ability of any quantity of interest. We hope that a great deal of the relevant research issues will be dealt with in the near future.

APPENDIX A

MCMC WITHIN EACH MODEL

Step 1 of Section 4.2 refers to simple MCMC methods used to generate values of the posterior distribution of a parameter given the model structure. Here we describe how we construct an MCMC algorithm based on the Metropolis-Hastings approach in order to generate parameter values from an a posterior distribution within each model.

Let us assume arbitrary initial values $\boldsymbol{\theta}_m^{(0)}$ and repeat the following steps until convergence is achieved. For our models we have $\boldsymbol{\theta}_{m_1} = \lambda$, $\boldsymbol{\theta}_{m_2} = (\lambda, \vartheta)^T$, and $\boldsymbol{\theta}_{m_3} = (\lambda, \omega)^T$. Initial values are discarded as a burn-in period to eliminate the effect of the initial arbitrary values. For our MCMC algorithm, we propose as initial values $\lambda^{(0)} = \bar{y}$, $\vartheta^{(0)} = \max\{0.01, \bar{y}^2/(s_y^2 - \bar{y})\}$, and $\omega^{(0)} = \max\{0.01, 1 - \sqrt{\bar{y}/s_y^2}\}$, where $\bar{y}$ and s_y^2 are the sample mean and variance of the data.

To implement an MCMC algorithm within a model m we repeat the following steps for $t = 1, \ldots, T$:

Step 1

Set $\boldsymbol{\theta} = \boldsymbol{\theta}_m^{(t)}$.

Step 2

For $j = 1, \ldots, d_m$.

1. Propose θ_j' from $q(\theta_j'|\boldsymbol{\theta}_{\setminus j}, \theta_j)$, where θ_j is the j-th element of vector $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{\setminus j} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_{d_m})^T$.
2. Calculate the acceptance probability

$$\alpha = \min\left\{1, \frac{f(\boldsymbol{y}|\boldsymbol{\theta}', m) f(\boldsymbol{\theta}'|m) q(\theta_j|\boldsymbol{\theta}_{\setminus j}, \theta_j')}{f(\boldsymbol{y}|\boldsymbol{\theta}, m) f(\boldsymbol{\theta}|m) q(\theta_j'|\boldsymbol{\theta}_{\setminus j}, \theta_j)}\right\},$$

   where $\boldsymbol{\theta}' = (\theta_1, \ldots, \theta_{j-1}, \theta_j', \theta_{j+1}, \ldots, \theta_{d_m})^T$.
3. Generate $u \sim U(0, 1)$, where $U(0, 1)$ is the uniform distribution in the interval (0, 1).
4. If $a > u$ then set $\theta_j = \theta_j'$, else leave θ_j unchanged.

Step 3

Set $\boldsymbol{\theta}_m^{(t+1)} = \boldsymbol{\theta}$.

As we have already stated, the posterior distribution of the Poisson model is conjugate, and hence, in RJMCMC, we generate the parameters of λ directly from the posterior distribution.

For the other two models, the proposal distribution concerning the parameter λ was based on the posterior distribution of the parameter when assuming the simpler Poisson model. Hence $q(\lambda'|\lambda)$ was set equal to a Gamma distribution with parameters $a + n\bar{y}$ and $b + n$. This worked sufficiently well for our illustrated data sets.

For the parameters ϑ and ω, we have used variations of the random walk Metropolis algorithms. More specifically, for the parameter ϑ we have used the proposal distribution $q(\vartheta'|\vartheta) = LN(\log \vartheta, C_\vartheta^2)$, where $LN(\mu, s^2)$ is the log-normal distribution with parameters μ and s^2 and density

$$f(x) = \frac{1}{\sqrt{2\pi}sx} \exp\left\{-\frac{1}{2}\left(\frac{\log x - \mu}{s}\right)^2\right\}. \tag{A.1}$$

For ω we have used the proposal distribution $q(\omega'|\omega) = Beta(C_\omega\, \omega/(1 - \omega), C_\omega)$ with mean ω and density given by

$$q(\omega'|\omega) = \frac{\Gamma\left(\frac{C_\omega}{1-\omega}\right)}{\Gamma\left(C_\omega \frac{\omega}{1-\omega}\right)\Gamma(C_\omega)} \omega'^{C_\omega\ \omega/(1-\omega)\ -1}\ (1-\omega')^{C_\omega - 1}.$$

The proposal parameters C_ϑ and C_ω are tuning parameters specified appropriately to achieve an acceptance rate within the range 30–50%.

Following the above comments, the generation of the parameters within model m_2 (negative binomial) has been completed using the following steps (assuming that (λ, ϑ) are the current values of the algorithm):

1. We sample λ from $f(\lambda|\vartheta, \boldsymbol{y}, m_2)$ using the following independent Metropolis scheme:
   (a) Propose new candidate value λ' from $Gamma(n\bar{y} + a, n + b)$.
   (b) Accept the proposed value with probability

$$\alpha = \min\left\{1, \frac{\lambda'}{\lambda}\left(\frac{\lambda+\vartheta}{\lambda'+\vartheta}\right)^{n\bar{y}+n\vartheta+3/2} e^{-n(\lambda-\lambda')}\right\}.$$

2. We sample ϑ from $f(\vartheta|\lambda, \boldsymbol{y}, m_2)$ using the following Metropolis-Hastings scheme:
   (a) Propose new candidate value ϑ' from $LN(\log\vartheta, C_\vartheta^2)$.
   (b) Accept the proposed value with probability $\alpha = \min\{1, A\}$, where A is given by

$$\log A = \sum_{i=1}^{n} \log\frac{\Gamma(y_i+\vartheta')}{\Gamma(y_i+\vartheta)} + n\log\frac{\Gamma(\vartheta)}{\Gamma(\vartheta')} + (n\vartheta'+1/2)\log\vartheta' - (n\vartheta+1/2)\log\vartheta$$
$$+ (n\bar{y}-3/2)\log\frac{\lambda+\vartheta'}{\lambda+\vartheta} + n\vartheta'\log(\lambda+\vartheta') - n\vartheta\log(\lambda+\vartheta).$$

Similarly, the generation of the parameters within model m_3 (generalized Poisson) has been completed using the following steps (assuming that (λ, ω) are the current values of the algorithm):

1. We sample λ from $f(\lambda|\vartheta, \boldsymbol{y}, m_2)$ using the following independent Metropolis scheme:
   (a) Propose new candidate value λ' from $Gamma(n\bar{y} + a, n + b)$.
   (b) Accept the proposed value with probability

$$\alpha = \min\left\{1, \left(\frac{\lambda'}{\lambda}\right)^{n-n\bar{y}} e^{n\omega(\lambda'-\lambda)}\left[\prod_{i=1}^{n}\left(\frac{(1-\omega)\lambda'+\omega y_i}{(1-\omega)\lambda+\omega y_i}\right)^{y_i-1}\right]\right\}.$$

2. We sample ω from $f(\omega|\lambda, \boldsymbol{y}, m_3)$ using the following Metropolis-Hastings scheme:
   (a) Propose new candidate value ω' from $Beta(C_\omega\,\omega/(1-\omega), C_\omega)$.
   (b) Accept the proposed value with probability $\alpha = \min(1, A)$, where A is given by

$$\log A = (n - C_\omega + 1)\log\frac{1-\omega'}{1-\omega} - n(\bar{y}-\lambda)(\omega'-\omega) + \sum_{i=1}^{n}\left((y_i-1)\log\frac{(1-\omega')\lambda+\omega' y_i}{(1-\omega)\lambda+\omega y_i}\right)$$
$$+ \log\frac{\Gamma\left(\frac{C_\omega}{1-\omega'}\right)}{\Gamma\left(\frac{C_\omega}{1-\omega}\right)} + \log\frac{\Gamma\left(C_\omega\frac{\omega}{1-\omega}\right)}{\Gamma\left(C_\omega\frac{\omega'}{1-\omega'}\right)} + \left(C_\omega\frac{\omega'}{1-\omega'}-1\right)\log\omega$$
$$- \left(C_\omega\frac{\omega}{1-\omega}-1\right)\log\omega'.$$

Table 7
Frequency of Automobile Claims for Denuit (1997) Data

Data Set	Number of Accidents							
	0	1	2	3	4	5	6	7
Switzerland 1961	103,704	14,075	1766	255	45	6	2	0
Zaire 1974	3,719	232	38	7	3	1	0	0
Great Britain 1968	370,412	46,545	3935	317	28	3	0	0
Germany 1964	20,592	2,651	297	41	7	0	1	0
Belgium 1978	7,840	1,317	239	42	14	4	4	1
Belgium 1975–76	96,978	9,240	704	43	9	0	0	0
Belgium 1993	57,178	5,618	446	50	8	0	0	0
Belgium 1994	118,700	11,468	930	70	14	0	0	0

APPENDIX B

INSURANCE CLAIMS DATA

The available data are presented in Table 7. The first column corresponds to the number of claims per contract, and the other columns denote the respective frequencies.

REFERENCES

BARTLETT, MAURICE S. 1957. Comment on D. V. Lindley's Statistical Paradox. *Biometrika* 44: 533–34.

BERNARDO, JOSÉ M., AND ADRIAN F. M. SMITH. 1994. *Bayesian Theory*. New York: John Wiley and Sons.

BROOKS, STEPHEN P., PAULO GIUDICI, AND GARETH O. ROBERTS. 2003. Efficient Construction of Reversible Jump Markov Chain Monte Carlo Proposal Distributions (with discussion). *Journal of the Royal Statistical Society, B* 65: 3–56.

CAIRNS, ANDREW J. G. 2000. A Discussion of Parameter and Model Uncertainty in Insurance. *Insurance: Mathematics and Economics* 27: 313–50.

CARLIN, BRADLEY P., AND SIDDHARTHA CHIB. 1995. Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society, B* 57(3): 473–84.

CARLIN, BRADLEY P., AND THOMAS A. LOUIS. 2000. *Bayes and Empirical Bayes Methods for Data Analysis*. 2nd ed. New York: Chapman and Hall.

CHATFIELD, CHRIS. 1995. Model Uncertainty, Data Mining and Statistical Inference (with discussion) *Journal of the Royal Statistical Society, A* 158: 419–66

CHAUBEY, YOGENDA P., JOSÉ GARRIDO, AND SOHIA TRUDEAU. 1998. On the Computation of Aggregate Claims Distributions: Some New Approximations. *Insurance: Mathematics and Economics* 23: 215–30.

CHEN, MING-HUI, QI-MAN SHAO, AND JOSEPH G. IBRAHIM. 2000. *Monte Carlo Methods in Bayesian Computation*. New York: Springer.

CHIPMAN, HUGH, EDWARD I. GEORGE, AND ROBERT E. MCCULLOCH. 2001. The Practical Implementation of Bayesian Model Selection. *IMS Lecture Notes—Monograph Series* 38: 67–120. Discussion: 121–134.

CLYDE, MERLISE A. 1999. Bayesian Model Averaging and Model Search Strategies (with discussion). In *Bayesian Statistics* 6, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, pp. 157–85. Oxford: Oxford University Press.

CONSUL, PREM C. 1989. *Generalized Poisson Distribution: Properties and Applications*. New York: Marcel Decker.

DE ALBA, ENRIQUE. 2002. Bayesian Estimation of Outstanding Claim Reserves. *North American Actuarial Journal* 6(4): 1–20.

DELLAPORTAS, PETROS, JONATHAN J. FORSTER, AND IOANNIS NTZOUFRAS. 2002. On Bayesian Model and Variable Selection Using MCMC. *Statistics and Computing* 12: 27–36.

DENUIT, MICHEL. 1997. A New Distribution of Poisson-Type for the Number of Claims. *ASTIN Bulletin* 27: 229–42.

DOUGLAS, JIM B. 1980. *Analysis of Standard Contagious Distributions*. Statistical Distributions in Science Work Series 4. Fairland, Md.: International Cooperative Publishing House.

———. 1994. Empirical Fitting of Discrete Distributions. *Biometrics* 50: 576–79.

DRAPER, DAVID. 1995. Assessment and Propagation of Model Uncertainty (with discussion). *Journal of the Royal Statistical Society, B* 57: 47–97.

ENGLAND, PETER, AND RICHARD J. VERRALL. 2002. Stochastic Claims Reserving in General Insurance (with discussion). *British Actuarial Journal* 8: 443–518.

GILKS, WALTER R., SYLVIA RICHARDSON, AND DAVID J. SPIEGELHALTER. 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

GOSSIAUX, A. M., AND JEAN LEMAIRE. 1981. Méthodes d'ajustement de distributions de sinistres. *Bulletin de l'Association des Actuaires Suisses* 87–95.

GREEN, PETER J. 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* 82: 711–32.

GREEN, PETER J., AND ANTONIETTA MIRA. 2001. Delayed Rejection in Reversible Jump Metropolis-Hastings. *Biometrika* 88: 1035–53.

HAN, CONG, AND BRADLEY CARLIN. 2001. MCMC Methods for Computing Bayes Factors: A Comparative Review. *Journal of the American Statistical Association* 96: 1122–32.

HOETING, JENNIFER A. 2002. Methodology for Bayesian Model Averaging: An Update. *International Biometrics Conference Proceedings*. Forthcoming.

HOETING, JENNIFER A., DAVID MADIGAN, ADRIAN E. RAFTERY, AND CHRIS T. VOLINSKI. 1999. Bayesian Model Averaging: A Tutorial (with discussion). *Statistical Science* 14: 382–417.

KASS, ROBERT E., AND ADRAIN E. RAFTERY. 1995. Bayes Factors. *Journal of the American Statistical Association* 90: 773–95.

LINDLEY, DENNIS V. 1957. A Statistical Paradox. *Biometrika* 44: 187–92.

LOPES, HEDIBERT FREITAS. 2002. Bayesian Model Selection. Technical Report, Departamento de Métodos Estatísticos Universidade Federal do Rio de Janeiro, Brazil.

MAKOV, UDI E. 2001. Principal Applications of Bayesian Methods in Actuarial Science: A Perspective. *North American Actuarial Journal* 5(4): 53–73.

NTZOUFRAS, IOANNIS, AND PETROS DELLAPORTAS. 2002. Bayesian Modeling of Outstanding Liabilities Incorporating Claim Count Uncertainty (with discussion). *North American Actuarial Journal* 6(1): 113–28.

NTZOUFRAS, IOANNIS, PETROS DELLAPORTAS, AND JONATHAN J. FORSTER. 2003. Bayesian Variable and Link Determination for Generalised Linear Models. *Journal of Statistical Planning and Inference* 111: 165–80.

RAFTERY, ADRIAN E. 1996. Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models. *Biometrika* 83: 251–66.

ROBERTS, GARRETH O. 1996. Markov Chain Concepts Related to Sampling Algorithms. In *Markov Chain Monte Carlo in Practice*, ed. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. London: Chapman and Hall.

ROTONDI, RENATA. 2002. On the Influence of the Proposal Distributions on a Reversible Jump MCMC Algorithm Applied to the Detection of Multiple Change-Points. *Computational Statistics and Data Analysis* 40: 633–53.

RUOHONEN, MATTI. 1988. A Model for the Claim Number Process. *ASTIN Bulletin* 18: 57–68.

SCOLLNIK, DAVID P. M. 1998. On the Analysis of the Truncated Generalized Poisson Distribution Using a Bayesian Method. *ASTIN Bulletin* 28: 135–52.

———. 2000. Actuarial Modeling with MCMC and BUGS: Additional Worked Examples. *Actuarial Research Clearing House* (2): 433–585.

———. 2001. Actuarial Modeling with MCMC and BUGS. *North American Actuarial Journal* 5(2): 96–125.

SHAFER, GLENN. 1982. Lindley's Paradox (with discussion). *Journal of the American Statistical Association* 77: 325–51.

SINHARAY, SANDIP, AND HAL S. STERN. 2002. On the Sensitivity of Bayes Factor to the Prior Distributions. *The American Statistician* 56: 196–201.

TER BERG, PETER. 1980. On the Loglinear Poisson and Gamma Model. *ASTIN Bulletin* 11: 35–40.

———. 1996. A Loglinear Lagrangian Poisson Model. *ASTIN Bulletin* 26: 123–29.

VERRALL, RICHARD J. 1990. Bayes and Empirical Bayes Estimation of the Chain-Ladder Model. *ASTIN Bulletin* 20(2): 217–43.

———. 2000. An Investigation into Stochastic Claims Reserving Models and the Chain-Ladder Technique. *Insurance: Mathematics and Economics* 26: 91–99.

WILLMOT, GORDON E. 1987. The Poisson Inverse Gaussian Distribution as an Alternative to the Negative Binomial. *Scandinavian Actuarial Journal* 12: 113–27.

Discussions on this paper can be submitted until January 1, 2006. The authors reserve the right to reply to any discussion. Please see the Submission Guidelines for Authors on the inside back cover for instructions on the submission of discussions.