MODEL UNCERTAINTY IN CLAIMS RESERVING WITHIN TWEEDIE'S COMPOUND POISSON MODELS

BY

GARETH W. PETERS, PAVEL V. SHEVCHENKO AND MARIO V. WÜTHRICH

ABSTRACT

In this paper we examine the claims reserving problem using Tweedie's compound Poisson model. We develop the maximum likelihood and Bayesian Markov chain Monte Carlo simulation approaches to fit the model and then compare the estimated models under different scenarios. The key point we demonstrate relates to the comparison of reserving quantities with and without model uncertainty incorporated into the prediction. We consider both the model selection problem and the model averaging solutions for the predicted reserves. As a part of this process we also consider the sub problem of variable selection to obtain a parsimonious representation of the model being fitted.

KEYWORDS

Claims reserving, model uncertainty, Tweedie's compound Poisson model, Bayesian analysis, model selection, model averaging, Markov chain Monte Carlo.

1. CLAIMS RESERVING

Setting appropriate claims reserves to meet future claims payment cash flows is one of the main tasks of non-life insurance actuaries. There is a wide range of models, methods and algorithms used to set appropriate claims reserves. Among the most popular methods there is the chain-ladder method, the Bornhuetter-Ferguson method and the generalized linear model methods. For an overview, see Wüthrich and Merz (2008) and England and Verrall (2002).

Setting claims reserves includes two tasks: estimate the mean of future payments and quantify the uncertainty in this prediction for future payments. Typically, quantifying the uncertainty includes two terms, namely the so-called process variance and the (parameter) estimation error. The process variance reflects that we predict random variables, i.e. it describes the pure process uncertainty. The estimation error reflects that the true model parameters need to be estimated and hence there is an uncertainty in the reliability of these estimates. In this paper, in addition to these two terms, we consider a third source of

Astin Bulletin 39(1), 1-33. doi: 10.2143/AST.39.1.2038054

error/uncertainty, namely, we analyze the fact that we could have chosen the wrong model. That is, we select a family of claims reserving models and quantify the uncertainty coming from a possibly wrong model choice within this family of models.

Such an analysis is especially important when answering solvency questions. A poor model choice may result in a severe shortfall in the balance sheet of an insurance company, which requires under a risk-adjusted solvency regime an adequate risk capital charge. We analyze typical sizes of such risk capital charges within the family of Tweedie's compound Poisson models, see Tweedie (1984), Smyth and Jørgensen (2002) and Wüthrich (2003).

Assume that $Y_{i,j}$ are incremental claims payments with indices $i, j \in \{0, \ldots, I\}$, where i denotes the accident year and j denotes the development year. At time I, we have observations

$$\mathcal{D}_I = \{Y_{i,j};\ i + j \le I\} \tag{1.1}$$

and for claims reserving at time I we need to predict the future payments

$$\mathcal{D}_I^c = \{Y_{i,j};\ i + j > I,\ i \le I\}, \tag{1.2}$$

see Table 1. Hence, the outstanding claims payment at time I is given by

$$R = \sum_{i=1}^{I} R_i = \sum_{i+j>I} Y_{i,j}. \tag{1.3}$$

Its conditional expectation at time I is given by

$$E\left[R \,\middle|\, \mathcal{D}_I\right] = \sum_{i=1}^{I} E\left[R_i \,\middle|\, \mathcal{D}_I\right] = \sum_{i+j>I} E\left[Y_{i,j} \,\middle|\, \mathcal{D}_I\right]. \tag{1.4}$$

Hereafter, the summation $i + j > I$ is for $i \le I$. Therefore, we need to predict R and to estimate $E[R \,|\, \mathcal{D}_I]$. Assume that $\widehat{R}$ is an appropriate $\mathcal{D}_I$-measurable predictor for R and $\mathcal{D}_I$-measurable estimator for $E[R \,|\, \mathcal{D}_I]$. Then, $\widehat{R}$ is used to predict the future payments and is the amount that is put aside in the balance sheet of the insurance company for these payments.

Prediction uncertainty is then often studied with the help of the (conditional) mean square error of prediction (MSEP) which is defined by

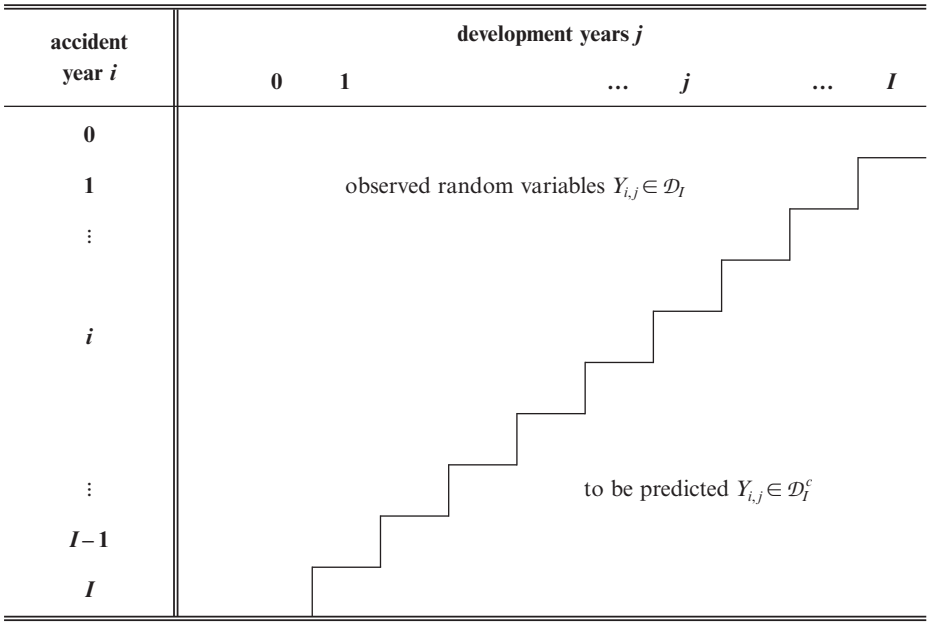
$$\operatorname{msep}_{R|\mathcal{D}_I}(\widehat{R}) = E\left[(R - \widehat{R})^2 \,\middle|\, \mathcal{D}_I\right]. \tag{1.5}$$

If $\widehat{R}$ is $\mathcal{D}_I$-measurable, the conditional MSEP can easily be decoupled as follows, see Wüthrich and Merz (2008), section 3.1:

$$\begin{aligned} \operatorname{msep}_{R|\mathcal{D}_I}(\widehat{R}) &= \operatorname{Var}(R \,|\, \mathcal{D}_I) + (E[R \,|\, \mathcal{D}_I] - \widehat{R})^2 \\ &= \text{process variance} + \text{estimation error}. \end{aligned} \tag{1.6}$$

TABLE 1

CLAIMS DEVELOPMENT TRIANGLE.



It is clear that the consistent estimator $\widehat{R}$ which minimizes the conditional MSEP is given by $E[R|\mathcal{D}_I]$ and is used, hereafter, as the "best estimate" for reserves. Assuming the model is parameterized by the parameter vector $\boldsymbol{\theta}$, $\mathrm{Var}(R|\mathcal{D}_I)$ can be decomposed as

$$\begin{aligned}\mathrm{Var}(R|\mathcal{D}_I) &= E[\mathrm{Var}(R|\boldsymbol{\theta}, \mathcal{D}_I)|\mathcal{D}_I] + \mathrm{Var}(E[R|\boldsymbol{\theta}, \mathcal{D}_I]|\mathcal{D}_I) \\ &= \text{average process variance} + \text{parameter estimation error.}\end{aligned} \tag{1.7}$$

These are the two terms that are usually studied when quantifying prediction uncertainties in a Bayesian context, where the unknown parameters $\boldsymbol{\theta}$ are modelled stochastically. That is, we obtain in the Bayesian context a similar decomposition as in the frequentist estimation (1.6). In the frequentist approach, the second term in (1.6) is often estimated by $\mathrm{Var}(\widehat{R})$, see for example section 6.4.3 in Wüthrich and Merz (2008).

As discussed in Cairns (2000), in full generality one could consider several sources of model uncertainty, however unlike Cairns (2000) we focus on a specific class of models. We consider the setting discussed in Bernardo and Smith (1994) termed M Complete modelling. In such a setting the premise is that one considers a set of models in which the "truth" exists but is unknown *a priori*. In this setting we demonstrate the risk associated with the model

uncertainty which we analyze jointly as a decomposition into two main parts. The first involves the uncertainty in the parameterization of the model, this is a variable selection problem within a nested model structure in the same vein as discussed in Cairns (2000). It relates to finding a trade-off between parsimony and accuracy in the estimation. The second source of model uncertainty that we study involves the choice of a parameter which determines membership from a spectrum of possible models within the Tweedie's compound Poisson family of models. We restrict the analysis to Tweedie's compound Poisson models and justify this by assuming we are working in the M Complete setting. If we relaxed this assumption and therefore consider competing models not in this family, then the analysis would be difficult to interpret and analyze in the manner we develop in this paper. The second source of model uncertainty will be considered under both a model selection and a model averaging setting, given the first "variable selection" uncertainty is resolved. As mentioned in Cairns (2000) achieving such an analysis requires advanced simulation methodology. Note, in future work we would also consider the M Open modeling framework of Bernardo and Smith (1994) which relaxes the belief that the truth lies in the set of models considered and hence introduces additional uncertainty associated with the family of models considered. The advanced sampling methodology required to study the M Open model setting will be briefly discussed.

The paper is organised as follows. In section 2, we present Tweedie's compound Poisson model and section 3 considers parameter estimation in the model, using the maximum Likelihood and Bayesian Markov chain Monte Carlo approaches for a real data set. Having addressed the variable selection question in section 4, we then analyze claims reserve estimation and model uncertainty in both a frequentist and Bayesian setting in section 5. We finish with conclusions from our findings.

2. Tweedie's compound Poisson model

We assume that $Y_{i,j}$ belongs to the family of Tweedie's compound Poisson models. Below we provide three different parameterizations for Tweedie's compound Poisson models, for rigorous derivations we refer to Jørgensen and de Souza (1994), Smyth and Jørgensen (2002) and Wüthrich (2003).

Model Assumptions 2.1 (1st Representation). *We assume that $Y_{i,j}$ are independent for $i, j \in \{0, \ldots, I\}$ and have a compound Poisson distribution*

$$Y_{i,j} = 1_{\{N_{i,j}>0\}} \sum_{k=1}^{N_{i,j}} X_{i,j}^{(k)}, \tag{2.1}$$

in which (a) $N_{i,j}$ and $X_{i,j}^{(k)}$ are independent for all k, (b) $N_{i,j}$ is Poisson distributed with parameter $\lambda_{i,j}$; (c) $X_{i,j}^{(k)}$ are independent gamma severities with the mean

$\tau_{i,j} > 0$ and the shape parameter $\gamma > 0$. Hereafter, we denote $1_{\{\}}$ as an indicator function.

2nd Representation. The random variable $Y_{i,j}$ given in (2.1) belongs to the family of Tweedie's compound Poisson models, see Tweedie (1984). The distribution of $Y_{i,j}$ can be reparameterized in such a way that it takes a form of the exponential dispersion family, see e.g. formula (3.5) and Appendix A in Wüthrich (2003):
$Y_{i,j}$ has a probability weight at 0 given by

$$P[Y_{i,j} = 0] = P[N_{i,j} = 0] = \exp\{-\phi_{i,j}^{-1} \kappa_p(\theta_{i,j})\} \tag{2.2}$$

and for $y > 0$ the random variable $Y_{i,j}$ has continuous density

$$f_{\theta_{i,j}}(y; \phi_{i,j}, p) = c(y; \phi_{i,j}, p) \exp\left\{ \frac{y\,\theta_{i,j} - \kappa_p\left(\theta_{i,j}\right)}{\phi_{i,j}} \right\}. \tag{2.3}$$

Here $\theta_{i,j} < 0$, $\phi_{i,j} > 0$, the normalizing constant is given by

$$c(y; \phi, p) = \sum_{r \geq 1} \left(\frac{(1/\phi)^{\gamma+1} y^{\gamma}}{(p-1)^{\gamma}(2-p)} \right)^{r} \frac{1}{r!\Gamma(r\gamma) y} \tag{2.4}$$

and the cummulant generating function $\kappa_p(.)$ is given by

$$\kappa_p(\theta) \overset{def.}{=} \frac{1}{2-p} \left[(1-p)\,\theta \right]^{\gamma}, \tag{2.5}$$

where $p \in (1,2)$ and $\gamma = (2-p)/(1-p)$.

The parameters, in terms of the 1st representation quantities, are:

$$p = p(\gamma) = \frac{\gamma + 2}{\gamma + 1} \in (1,2), \tag{2.6}$$

$$\phi_{i,j} = \frac{\lambda_{i,j}^{1-p} \tau_{i,j}^{2-p}}{2-p} > 0, \tag{2.7}$$

$$\theta_{i,j} = \left(\frac{1}{1-p} \right) \left(\mu_{i,j} \right)^{(1-p)} < 0, \tag{2.8}$$

$$\mu_{i,j} = \lambda_{i,j} \tau_{i,j} > 0. \tag{2.9}$$

Then the mean and variance of $Y_{i,j}$ are given by

$$E[Y_{i,j}] = \frac{\partial}{\partial \theta_{i,j}} \kappa_p(\theta_{i,j}) = \kappa_p'(\theta_{i,j}) = [(1-p)\theta_{i,j}]^{1/(1-p)} = \mu_{i,j}, \tag{2.10}$$

$$\mathrm{Var}(Y_{i,j}) = \phi_{i,j} \kappa_p''(\theta_{i,j}) = \phi_{i,j}\, \mu_{i,j}^p. \tag{2.11}$$

That is, $Y_{i,j}$ has the mean $\mu_{i,j}$, dispersion $\phi_{i,j}$ and variance function with the variance parameter p. The extreme cases $p \to 1$ and $p \to 2$ correspond to the overdispersed Poisson and the gamma models, respectively. Hence, in this spirit, Tweedie's compound Poisson model with $p \in (1,2)$ closes the gap between the Poisson and the gamma models. Often in practice, p is assumed to be known and fixed by the modeller. The aim of this paper is to study *Model Uncertainty*, that is, we would like to study the sensitivity of the claims reserves within this subfamily, i.e. Tweedie's compound Poisson models (which are now parameterized through p). This answers model uncertainty questions within the family of Tweedie's compound Poisson models. In this paper the restriction on $p \in (1,2)$ is taken in the context of practical application of these models to claims reserving, Wüthrich (2003) comments that the majority of claims reserving problems will be captured under this assumption. However, in general, in the exponential dispersion family p can be outside of the (1,2) range, e.g. $p = 0$ produces a Gaussian density and $p = 3$ leads to an inverse Gaussian model.

3rd Representation. Utilizing the above definitions, the distribution of $Y_{i,j}$ can be rewritten in terms of $\mu_{i,j}$, p and $\phi_{i,j}$ as

$$P\left[Y_{i,j} = 0\right] = P\left[N_{i,j} = 0\right] = \exp\left\{-\phi_{i,j}^{-1} \frac{\mu_{i,j}^{2-p}}{2-p}\right\} \tag{2.12}$$

and for $y > 0$

$$f_{\mu_{i,j}}\left(y; \phi_{i,j}, p\right) = c\left(y; \phi_{i,j}, p\right) \exp\left\{\phi_{i,j}^{-1}\left[y \frac{\mu_{i,j}^{1-p}}{1-p} - \frac{\mu_{i,j}^{2-p}}{2-p}\right]\right\}. \tag{2.13}$$

3. Parameter estimation

Our goal is to estimate the parameters $\mu_{i,j}$, p and $\phi_{i,j}$ based on the observations $\mathcal{D}_I$. In order to estimate these parameters we need to introduce additional structure in the form of a multiplicative model.

Model Assumptions 3.1. *Assume that there exist exposures* $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_I)$ *and a development pattern* $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_I)$ *such that we have for all* $i, j \in \{0, \ldots, I\}$

$$\mu_{i,j} = \alpha_i \beta_j. \tag{3.1}$$

Moreover, assume that $\phi_{i,j} = \phi$ *and* $\alpha_i > 0$, $\beta_j > 0$.

In addition, we impose the normalizing condition $\alpha_0 = 1$, so that the estimation problem is well-defined. That is we have $(2I + 3)$ unknown parameters p, ϕ, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ that have to be estimated from the data $\mathcal{D}_I$. Next we present the likelihood function for this model and then develop the methodology for parameter estimation using the maximum likelihood and Bayesian inference methods.

3.1. Likelihood function

Define the parameter vector $\boldsymbol{\theta} = (p, \phi, \boldsymbol{\alpha}, \boldsymbol{\beta})$. Then the likelihood function for $Y_{i,j}$, $i + j \leq I$, is given by

$$L_{\mathcal{D}_I}(\boldsymbol{\theta}) = \prod_{i+j \leq I} c\left(Y_{i,j}; \phi, p\right) \exp\left\{\phi^{-1}\left[Y_{i,j}\frac{\left(\alpha_i \beta_j\right)^{1-p}}{1-p} - \frac{\left(\alpha_i \beta_j\right)^{2-p}}{2-p}\right]\right\}, \tag{3.2}$$

where we set $c(0; \phi, p) = 1$ for $Y_{i,j} = 0$. The difficulty in the evaluation of the likelihood function is the calculation of $c(y; \phi, p)$ which contains an infinite sum

$$c(y; \phi, p) = \sum_{r \geq 1}\left(\frac{(1/\phi)^{\gamma+1} y^{\gamma}}{(p-1)^{\gamma}(2-p)}\right)^r \frac{1}{r!\Gamma(r\gamma)y} = \frac{1}{y}\sum_{r \geq 1} W_r, \tag{3.3}$$

where $\gamma = \gamma(p) = (2-p)/(1-p)$. Tweedie (1984) identified this summation as Wright's (1935) generalized Bessel function, which can not be expressed in terms of more common Bessel functions. To evaluate this summation we follow the approach of Dunn and Smyth (2005) which directly sums the infinite series, including only terms which significantly contribute to the summation. Consider the term

$$\log W_r = r \log z - \log\Gamma(1+r) - \log\Gamma(\gamma r),$$

where

$$z = \frac{(1/\phi)^{\gamma+1} y^{\gamma}}{(p-1)^{\gamma}(2-p)}.$$

Replacing the gamma functions using Stirling's approximation and approximating γr by $\gamma r + 1$ we get

$$\log W_r \approx r\{\log z + (1+\gamma) - \gamma \log\gamma - (1+\gamma)\log r\} - \log(2\pi) - \frac{1}{2}\log\gamma - \log r,$$

which is also a reasonable approximation for small r. Treating r as continuous and taking the partial derivative w.r.t. r gives

$$\frac{\partial \log W_r}{\partial r} \approx \log z - \log r - \gamma \log(\gamma r).$$

Hence, the sequence W_r is unimodal in r. Solving $\partial W_r / \partial r = 0$, to find (approximately) the maximum of W_r, results in the approximate maximum lying close to

$$R_0 = R_0(\phi, p) = \frac{y^{2-p}}{(2-p)\phi}. \tag{3.4}$$

This gives a surprisingly accurate approximation to the true maximum of W_r, $r \in \mathbb{N}$. Finally, the aim is to find $R_L < R_0 < R_U$ such that the following approximation is sufficiently accurate for the use in the evaluation of the likelihood terms,

$$c(y; \phi, p) \approx \tilde{c}(y; \phi, p) = \frac{1}{y} \sum_{r=R_L}^{R_U} W_r. \tag{3.5}$$

The fact that $\partial \log W_r / \partial r$ is monotonic and decreasing implies that $\log W_r$ is strictly convex in r and hence the terms in W_r decay at a faster rate than geometric on either side of R_0. Dunn and Smyth (2005) derive the following bounds,

$$c(y; \phi, p) - \tilde{c}(y; \phi, p) < W_{R_L - 1} \frac{1 - q_L^{R_L - 1}}{1 - q_L} + W_{R_U + 1} \frac{1}{1 - q_U} \tag{3.6}$$

with

$$q_L = \exp\left(\frac{\partial \log W_r}{\partial r} \right) \Bigg|_{r = R_L - 1}, \quad q_U = \exp\left(\frac{\partial \log W_r}{\partial r} \right) \Bigg|_{r = R_U + 1}. \tag{3.7}$$

These bounds are typically too conservative since the decay is much faster than geometric. In practice, an adaptive approach balancing accuracy and efficiency is to continue adding terms either side of the maximum until the lower and upper terms satisfy the double precision constraints $W_{R_L} \leqslant e^{-37} W_{R_0}$ (or $R_L = 1$) and $W_{R_U} \leqslant e^{-37} W_{R_0}$. When evaluating the summation for $\tilde{c}(y; \phi, p)$, it was important to utilize the following identity to perform the summation in the log scale to avoid numerical overflow problems,

$$\log \tilde{c}(y; \phi, p) = -\log y + \log W_{R_0} + \log \left(\sum_{r=R_L}^{R_U} \exp\big(\log(W_R) - \log\big(W_{R_0}\big) \big) \right).$$

We made an additional observation when analyzing this model. For our data set, as p approaches 1 (i.e. when the distribution approaches the overdispersed Poisson model) the likelihood may become multimodal. Therefore, to avoid numerical complications in actual calculations, we restrict to $p \geqslant 1.1$. At the other extreme, when $p = 2$ the number of terms required to evaluate $c(y; \phi, p)$ may become very large, hence to manage the computation burden, we restrict $p \leqslant 1.95$. These limitations are also discussed in Dunn and Smyth (2005). For our data set, we checked that this restriction did not have a material impact on the results.

3.2. Maximum likelihood estimation

The maximum likelihood estimator (MLE) for the parameters is given by maximizing $L_{\mathcal{D}_I}(\boldsymbol{\theta})$ in $\boldsymbol{\theta} = (p, \phi, \boldsymbol{\alpha}, \boldsymbol{\beta})$ under the constraints $\alpha_i > 0$, $\beta_j > 0$, $\phi > 0$ and $p \in (1,2)$. This leads to the MLEs $\widehat{\boldsymbol{\theta}}^{\mathrm{MLE}} = (\hat{p}^{\mathrm{MLE}}, \hat{\phi}^{\mathrm{MLE}}, \widehat{\boldsymbol{\alpha}}^{\mathrm{MLE}}, \widehat{\boldsymbol{\beta}}^{\mathrm{MLE}})$ and to the best estimate reserves for R, given $\mathcal{D}_I$,

$$\widehat{R}^{\mathrm{MLE}} = \sum_{i+j>I} \widehat{\alpha}_i^{\mathrm{MLE}} \widehat{\beta}_j^{\mathrm{MLE}}. \tag{3.8}$$

A convenient practical approach to obtain the MLEs is to use the fact that at the maximum of the likelihood, $\boldsymbol{\beta}$ are expressed through $\boldsymbol{\alpha}$ and p according to the following set of equations, $p \in (1,2)$:

$$\beta_k = \frac{\sum_{i=0}^{I-k} Y_{i,k} \alpha_i^{1-p}}{\sum_{i=0}^{I-k} \alpha_i^{2-p}}, \quad k = 0, \ldots, I, \tag{3.9}$$

obtained by setting partial derivatives

$$\begin{aligned} \frac{\partial \ln L_{\mathcal{D}_I}(\boldsymbol{\theta})}{\partial \beta_k} &= \frac{\partial}{\partial \beta_k} \sum_{j=0}^{I} \sum_{i=0}^{I-j} \phi^{-1} \left(Y_{i,j} \frac{(\alpha_i \beta_j)^{1-p}}{1-p} - \frac{(\alpha_i \beta_j)^{2-p}}{2-p} \right) \\ &= \sum_{i=0}^{I-k} \phi^{-1} \left(Y_{i,k} \alpha_i^{1-p} \beta_k^{-p} - \alpha_i^{2-p} \beta_k^{1-p} \right) \end{aligned} \tag{3.10}$$

equal to zero. Hence, after maximizing the likelihood in $\boldsymbol{\alpha}, p, \phi$ one then calculates the set of equations (3.9) for the remaining parameters utilizing the normalization condition $\alpha_0 = 1$.

Under an asymptotic Gaussian approximation, the distribution of the MLEs is Gaussian with the covariance matrix elements

$$\operatorname{cov}(\widehat{\theta}_i^{\mathrm{MLE}}, \widehat{\theta}_j^{\mathrm{MLE}}) \approx (\boldsymbol{I}^{-1})_{i,j}, \tag{3.11}$$

where $\boldsymbol{I}$ is Fisher's information matrix that can be estimated by the observed information matrix

$$(\boldsymbol{I})_{i,j} \approx \left. \frac{\partial^2 \ln L_{\mathcal{D}_I}(\boldsymbol{\theta})}{\partial\theta_i \, \partial\theta_j} \right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{\mathrm{MLE}}}. \tag{3.12}$$

It is interesting to note that, $\widehat{\beta}_I^{\mathrm{MLE}} = Y_{0,I}$. Also, it is easy to show (using (3.10) and (3.11)) that $\widehat{\beta}_I^{\mathrm{MLE}}$ is orthogonal to all other parameters, i.e.

$$\operatorname{cov}\big(\widehat{\beta}_I^{\mathrm{MLE}}, \widehat{\theta}_i^{\mathrm{MLE}}\big) = 0, \quad \widehat{\theta}_i^{\mathrm{MLE}} \neq \widehat{\beta}_I^{\mathrm{MLE}}. \tag{3.13}$$

The next step is to estimate the parameter estimation error in the reserve as a function of the parameter uncertainty. We do this via propagation of error by forming a Taylor expansion around the MLEs, see England and Verrall (2002) formulae (7.6)-(7.8) and Wüthrich (2003) formulae (5.1)-(5.2),

$$\operatorname{stdev}\Big(\widehat{R}^{\mathrm{MLE}}\Big) = \sqrt{\operatorname{Var}\Big(\widehat{R}^{\mathrm{MLE}}\Big)} \tag{3.14}$$

$$\begin{aligned}\widehat{\operatorname{Var}}\Big(\widehat{R}^{\mathrm{MLE}}\Big) = {} & \sum_{i_1+j_1>I} \sum_{i_2+j_2>I} \hat{a}_{i_1}^{\mathrm{MLE}} \hat{a}_{i_2}^{\mathrm{MLE}} \operatorname{cov}\Big(\hat{\beta}_{j_1}^{\mathrm{MLE}}, \hat{\beta}_{j_2}^{\mathrm{MLE}}\Big) \\ & + \sum_{i_1+j_1>I} \sum_{i_2+j_2>I} \hat{\beta}_{j_1}^{\mathrm{MLE}} \hat{\beta}_{j_2}^{\mathrm{MLE}} \operatorname{cov}\Big(\hat{a}_{i_1}^{\mathrm{MLE}}, \hat{a}_{i_2}^{\mathrm{MLE}}\Big) \\ & + 2 \sum_{i_1+j_1>I} \sum_{i_2+j_2>I} \hat{a}_{i_1}^{\mathrm{MLE}} \hat{\beta}_{j_2}^{\mathrm{MLE}} \operatorname{cov}\Big(\hat{a}_{i_2}^{\mathrm{MLE}}, \hat{\beta}_{j_1}^{\mathrm{MLE}}\Big).\end{aligned} \tag{3.15}$$

Additionally, using the independence assumption on $Y_{i,j}$ and (2.11), the process variance is estimated as

$$\widehat{\operatorname{Var}}(R) = \sum_{i+j>I} \Big(\hat{a}_i^{\mathrm{MLE}} \hat{\beta}_j^{\mathrm{MLE}}\Big)^{\hat{p}^{\mathrm{MLE}}} \hat{\phi}^{\mathrm{MLE}}. \tag{3.16}$$

Then the conditional MSEP (1.6) is estimated by

$$\begin{aligned}\widehat{\operatorname{msep}}_{R|\mathcal{D}_i}\Big(\widehat{R}^{\mathrm{MLE}}\Big) &= \widehat{\operatorname{Var}}(R) + \widehat{\operatorname{Var}}\Big(\widehat{R}^{\mathrm{MLE}}\Big) \\ &= \text{MLE process variance + MLE estimation error.}\end{aligned} \tag{3.17}$$

Note that, in practice, typically MLE is done for a fixed p (expert choice) and hence model selection questions are neglected. In our context it means that the expert chooses p and then estimates $\widehat{\boldsymbol{\alpha}}^{\mathrm{MLE}}$, $\widehat{\boldsymbol{\beta}}^{\mathrm{MLE}}$ and $\widehat{\phi}^{\mathrm{MLE}}$ (see also Wüthrich (2003), section 4.1). The case $p = 1$ corresponds to the overdispersed Poisson model and provides the chain-ladder estimate for the claims reserves (see Wüthrich and Merz (2008), section 2.4). It is important to note that, often the dispersion parameter ϕ is estimated using Pearson's residuals as

$$\widehat{\phi}^{\mathrm{P}} = \frac{1}{N-k} \sum_{i+j \le I} \frac{\left(Y_{i,j} - \hat{a}_i^{\mathrm{MLE}} \hat{\beta}_j^{\mathrm{MLE}}\right)^2}{\left(\hat{a}_i^{\mathrm{MLE}} \hat{\beta}_j^{\mathrm{MLE}}\right)^p}, \tag{3.18}$$

where N is the number of observations $Y_{i,j}$ in $\mathcal{D}_I$ and k is the number of estimated parameters α_i, β_j (see e.g. Wüthrich and Merz (2008), formula (6.58)). Also note that for a given p, $\widehat{R}^{\mathrm{MLE}}$ given by (3.8) does not depend on ϕ and the estimators for the process variance (3.16) and estimation error (3.15) are proportional to ϕ. Next we present the Bayesian model which provides the posterior distribution of the parameters given the data. This will be used to analyze the model uncertainty within Tweedie's compound Poisson models.

3.3. Bayesian inference

In a Bayesian context all parameters, $p, \phi, \alpha_i > 0$ and $\beta_j > 0$, are treated as random. Using Bayesian inference we adjust our *a priori* beliefs about the parameters of the model utilizing the information from the observations. Through the Bayesian paradigm we are able to learn more about the distribution of $p, \phi, \boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ after having observed $\mathcal{D}_I$.

Our *a priori* beliefs about the parameters of the model are encoded in the form of a prior distribution on the parameters $\pi(\boldsymbol{\theta})$. Then the joint density of $\mathcal{D}_I = \{Y_{i,j} > 0;\ i + j \le I\}$ and $\boldsymbol{\theta} = (p, \phi, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is given by

$$L_{\mathcal{D}_I}(\boldsymbol{\theta})\, \pi(\boldsymbol{\theta}). \tag{3.19}$$

Now applying Bayes' law, the posterior distribution of the model parameters, given the data $\mathcal{D}_I$, is

$$\pi(\boldsymbol{\theta} \,|\, \mathcal{D}_I) \propto L_{\mathcal{D}_I}(\boldsymbol{\theta})\, \pi(\boldsymbol{\theta}). \tag{3.20}$$

Usually, there are two problems that arise in this context, the normalizing constant of this posterior is not known in closed form. Additionally, generating samples from this posterior is typically not possible using simple inversion or rejection sampling approaches. In such cases it is usual to adopt techniques such as Markov chain Monte Carlo (MCMC) methods, see for example Gilks

et al. (1996) and Robert and Casella (2004) for detailed expositions of such approaches.

The Bayesian estimators typically considered are the Maximum a Postiori (MAP) estimator and the Minimum Mean Square Estimator (MMSE), that is the mode and mean of the posterior, defined as follows:

$$MAP: \quad \widehat{\boldsymbol{\theta}}^{MAP} = \arg\max_{\boldsymbol{\theta}} [\pi(\boldsymbol{\theta} \,|\, \mathcal{D}_I)], \tag{3.21}$$

$$MMSE: \quad \widehat{\boldsymbol{\theta}}^{MMSE} = E[\boldsymbol{\theta} \,|\, \mathcal{D}_I]. \tag{3.22}$$

We mention here that if the prior $\pi(\boldsymbol{\theta})$ is constant and the parameter range includes the MLE, then the MAP of the posterior is the same as the MLE. Additionally, one can approximate the posterior using a second order Taylor series expansion around the MAP estimate as

$$\begin{aligned} \ln \pi(\boldsymbol{\theta} \,|\, \mathcal{D}_I) \approx{}& \ln \pi\left(\widehat{\boldsymbol{\theta}}^{MAP} \,\middle|\, \mathcal{D}_I\right) \\ &+ \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial \theta_i \, \partial \theta_j} \ln \pi(\boldsymbol{\theta} \,|\, \mathcal{D}_I) \bigg|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{MAP}} \left(\theta_i - \hat{\theta}_i^{MAP}\right)\left(\theta_j - \hat{\theta}_j^{MAP}\right). \end{aligned} \tag{3.23}$$

This corresponds to $\pi(\boldsymbol{\theta} \,|\, \mathcal{D}_I)$ approximated by the Gaussian distribution with the mean $\widehat{\boldsymbol{\theta}}^{MAP}$ and covariance matrix calculated as the inverse of the matrix

$$(\tilde{\boldsymbol{I}})_{i,j} = - \frac{\partial^2}{\partial \theta_i \, \partial \theta_j} \ln \pi(\boldsymbol{\theta} \,|\, \mathcal{D}_I) \bigg|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{MAP}}, \tag{3.24}$$

which in the case of diffuse priors (or constant priors defined on a large range) compares with the Gaussian approximation for the MLEs (3.11)-(3.12).

In the Bayesian context, the conditionally expected future payment, for Model Assumptions 3.1, is given by

$$E[R \,|\, \mathcal{D}_I] = \sum_{i+j>I} E\left[\alpha_i \, \beta_j \,\middle|\, \mathcal{D}_I\right]. \tag{3.25}$$

Denote the expected reserves, given the parameters $\boldsymbol{\theta}$, by

$$\tilde{R} = E[R \,|\, \boldsymbol{\theta}] = \sum_{i+j>I} \alpha_i \, \beta_j. \tag{3.26}$$

Then, the best consistent estimate of reserves (ER) is given by

$$\widehat{R}^{\mathrm{B}} = E\left[\tilde{R} \,\middle|\, \mathcal{D}_I\right] = \sum_{i+j>I} E\left[\alpha_i \, \beta_j \,\middle|\, \mathcal{D}_I\right] = E[R \,|\, \mathcal{D}_I], \tag{3.27}$$

which is, of course, a $\mathcal{D}_I$-measurable predictor. Hence, the conditional MSEP is simply

$$\operatorname{msep}_{R\,|\,\mathcal{D}_I}\left(\widehat{R}^{\mathrm{B}}\right) = E\left[\left(R-\widehat{R}^{\mathrm{B}}\right)^2\,\middle|\,\mathcal{D}_I\right] = \operatorname{Var}(R\,|\,\mathcal{D}_I). \tag{3.28}$$

This term, in the Bayesian approach for Tweedie's compound Poisson model, is decomposed as, see also (1.7),

$$\operatorname{Var}(R\,|\,\mathcal{D}_I) = \operatorname{Var}\left(\sum_{i+j>I} Y_{i,j}\,\middle|\,\mathcal{D}_I\right) = \sum_{i+j>I} E\left[\left(\alpha_i\,\beta_j\right)^p \phi\,\middle|\,\mathcal{D}_I\right] + \operatorname{Var}\left(\tilde{R}\,\middle|\,\mathcal{D}_I\right). \tag{3.29}$$

Hence, we obtain the familiar decoupling into average process variance and estimation error. However, in addition we incorporate model uncertainty within Tweedie's compound Poisson model, which enters the calculation by the averaging over all possible values of the variance parameter p.

3.4. Random walk Metropolis Hastings-algorithm within Gibbs

In this section we describe an MCMC method to be used to sample from the posterior distribution (3.20). The following notations are used: $\boldsymbol{\theta} = (p, \phi, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is the vector of parameters; $U(a,b)$ is the uniform distribution on the interval (a, b); $f_N(x;\mu,\sigma)$ and $F_N(x;\mu,\sigma)$ are the Gaussian density and distribution correspondingly with the mean $\mu \in R$ and standard deviation $\sigma > 0$ at position $x \in R$.

TABLE 2

DATA – ANNUAL CLAIMS PAYMENTS $Y_{i,j}$ FOR EACH ACCIDENT YEAR i AND DEVELOPMENT YEAR j, $i + j \leq 9$.

Year	0	1	2	3	4	5	6	7	8	9
0	594.6975	372.1236	89.5717	20.7760	20.6704	6.2124	6.5813	1.4850	1.1130	1.5813
1	634.6756	324.6406	72.3222	15.1797	6.7824	3.6603	5.2752	1.1186	1.1646	
2	626.9090	297.6223	84.7053	26.2768	15.2703	6.5444	5.3545	0.8924		
3	586.3015	268.3224	72.2532	19.0653	13.2976	8.8340	4.3329			
4	577.8885	274.5229	65.3894	27.3395	23.0288	10.5224				
5	618.4793	282.8338	57.2765	24.4899	10.4957					
6	560.0184	289.3207	56.3114	22.5517						
7	528.8066	244.0103	52.8043							
8	529.0793	235.7936								
9	567.5568									

Prior Structure: We assume that all parameters are independent under the prior distribution $\pi(\boldsymbol{\theta})$ and all distributed uniformly with $\theta_i \sim U(a_i, b_i)$. The prior domains we used for our analysis were $p \in (1.1, 1.95)$, $\phi \in (0.01, 100)$, $\alpha_i \in (0.01, 100)$ and $\beta_j \in (0.01, 10^4)$. These are reasonable ranges for the priors in view of our data in Table 2 and corresponding to the MLEs in Table 3. Other priors such as diffuse priors can be applied with no additional difficulty. The choice of very wide prior supports was made with the aim of performing inference in the setting where the posterior is largely implied by the data. Subsequently, we checked that making the ranges wider does not affect the results.

Next we outline a random walk Metropolis-Hastings (RW-MH) within Gibbs algorithm. This creates a reversible Markov chain with the stationary distribution corresponding to our target posterior distribution (3.20). That is, we will run the chain until it has sufficiently converged to the stationary distribution (=posterior distribution) and in doing so we obtain samples from that posterior

TABLE 3

MLE AND BAYESIAN ESTIMATORS. σ_{RW} IS THE PROPOSAL STANDARD DEVIATION IN THE MCMC ALGORITHM AND $[Q_{0.05}; Q_{0.95}]$ IS THE PREDICTIVE INTERVAL, WHERE Q_α IS THE QUANTILE OF THE POSTERIOR DISTRIBUTION AT LEVEL α. THE NUMERICAL STANDARD ERROR, IN BAYESIAN ESTIMATORS DUE TO FINITE NUMBER OF MCMC ITERATIONS, IS INCLUDED IN BRACKETS NEXT TO ESTIMATES.

	MLE	**MLE stdev**	**Bayesian posterior**			σ_{RW}
			MMSE	**stdev**	$[Q_{0.05}; Q_{0.95}]$	
p	1.259	0.149	1.332 (0.007)	0.143 (0.004)	[1.127;1.590]	1.61
ϕ	0.351	0.201	0.533 (0.013)	0.289 (0.005)	[0.174;1.119]	1.94
α_1	0.918	0.056	0.901 (0.004)	0.074 (0.001)	[0.778;1.022]	0.842
α_2	0.946	0.051	0.946 (0.003)	0.073 (0.001)	[0.833;1.072]	0.907
α_3	0.861	0.048	0.861 (0.003)	0.068 (0.001)	[0.756;0.977]	0.849
α_4	0.891	0.049	0.902 (0.003)	0.072 (0.002)	[0.794;1.027]	0.893
α_5	0.879	0.051	0.876 (0.003)	0.070 (0.001)	[0.768;0.994]	0.932
α_6	0.842	0.048	0.843 (0.002)	0.069 (0.001)	[0.736;0.958]	0.751
α_7	0.762	0.046	0.762 (0.003)	0.066 (0.001)	[0.660;0.876]	0.888
α_8	0.763	0.047	0.765 (0.003)	0.067 (0.001)	[0.661;0.874]	0.897
α_9	0.848	0.059	0.856 (0.003)	0.090 (0.002)	[0.716;1.009]	1.276
β_0	669.1	27.7	672.7 (2.1)	39.7 (0.7)	[610.0;740.0]	296
β_1	329.0	14.4	331.1 (1.0)	20.6 (0.4)	[298.1;365.9]	190
β_2	77.43	4.38	78.06 (0.24)	6.10 (0.06)	[68.58;88.29]	75.4
β_3	24.59	1.96	24.95 (0.08)	2.64 (0.03)	[20.89;29.64]	40.9
β_4	16.28	1.55	16.65 (0.05)	2.09 (0.03)	[13.44;20.30]	40.6
β_5	7.773	1.028	8.068 (0.024)	1.356 (0.020)	[6.064;10.473]	26.0
β_6	5.776	0.937	6.115 (0.022)	1.261 (0.016)	[4.246;8.347]	24.1
β_7	1.219	0.396	1.494 (0.006)	0.609 (0.013)	[0.739;2.609]	13.1
β_8	1.188	0.476	1.622 (0.008)	0.802 (0.016)	[0.674;3.070]	15.1
β_9	1.581	0.790	2.439 (0.021)	1.496 (0.026)	[0.829;5.250]	32.1

distribution. It should be noted that the Gibbs sampler creates a Markov chain in which each iteration of the chain involves scanning either deterministically or randomly over the variables that comprise the target stationary distribution of the chain. This process involves sampling each proposed parameter update from the corresponding full conditional posterior distribution. The algorithm we present generates a Markov chain that will explore the parameter space of the model in accordance with the posterior mass in that region of the parameter space. The state of the chain at iteration t will be denoted by $\boldsymbol{\theta}^t$ and the chain will be run for a length of T iterations. The manner in which MCMC samplers proceed is by proposing to move the ith parameter from state θ_i^{t-1} to a new proposed state θ_i^*. The latter will be sampled from an MCMC proposal transition kernel (3.30). Then the proposed move is accepted according to a rejection rule which is derived from a reversibility condition. This makes the acceptance probability a function of the transition kernel and the posterior distribution as shown in (3.31). If under the rejection rule one accepts the move then the new state of the ith parameter at iteration t is given by $\theta_i^t = \theta_i^*$, otherwise the parameter remains in the current state $\theta_i^t = \theta_i^{t-1}$ and an attempt to move that parameter is repeated at the next iteration. In following this procedure, one builds a set of correlated samples from the target posterior distribution which have several asymptotic properties. One of the most useful of these properties is the convergence of ergodic averages constructed using the Markov chain samples to the averages obtained under the posterior distribution.

Next we present the algorithm and then some references that will guide further investigation into this class of simulation methodology. Properties of this algorithm, including convergence results can be found in the following references Casella and George (1992), Robert and Casella (2004), Gelman *et al.* (1995), Gilks *et al.* (1996) and Smith and Roberts (1993).

Random Walk Metropolis Hastings (RW-MH) within Gibbs algorithm.

1. Initialize randomly or deterministically for $t = 0$ the parameter vector $\boldsymbol{\theta}^0$ (e.g. MLEs).
2. For $t = 1, \ldots, T$
   a) Set $\boldsymbol{\theta}^t = \boldsymbol{\theta}^{t-1}$
   b) For $i = 1, \ldots, 2I + 3$

Sample proposal θ_i^* from Gaussian distribution whose density is truncated below a_i and above b_i and given by

$$f_N^T\left(\theta_i^*; \theta_i^t, \sigma_{RWi}\right) = \frac{f_N\left(\theta_i^*; \theta_i^t, \sigma_{RWi}\right)}{F_N\left(b_i; \theta_i^t, \sigma_{RWi}\right) - F_N\left(a_i; \theta_i^t, \sigma_{RWi}\right)} \tag{3.30}$$

to obtain $\boldsymbol{\theta}^* = (\theta_1^t, \ldots, \theta_{i-1}^t, \theta_i^*, \theta_{i+1}^{t-1}, \ldots)$.

Accept proposal with acceptance probability

$$\alpha\left(\boldsymbol{\theta}^t, \boldsymbol{\theta}^*\right) \min\left\{1, \frac{\pi\left(\boldsymbol{\theta}^* \middle| \mathcal{D}_I\right) f_N^T\left(\theta_i^t; \theta_i^*, \sigma_{RWi}\right)}{\pi\left(\boldsymbol{\theta}^t \middle| \mathcal{D}_I\right) f_N^T\left(\theta_i^*; \theta_i^t, \sigma_{RWi}\right)}\right\}, \tag{3.31}$$

where $\pi(\boldsymbol{\theta}^* | \mathcal{D}_I)$ is given by (3.20). That is, simulate $U \sim U(0,1)$ and set $\theta_i^t = \theta_i^*$ if $U < \alpha(\boldsymbol{\theta}^t, \boldsymbol{\theta}^*)$.

$\Rightarrow$ Note that in (3.31) the normalizing constant of the posterior $\pi(\boldsymbol{\theta} | \mathcal{D}_I)$ from (3.20) is not needed.

Remark. The RW-MH algorithm is simple in nature and easily implemented. However, if one does not choose the proposal distribution carefully, then the algorithm only gives a very slow convergence to the stationary distribution. There have been several studies regarding the optimal scaling of proposal distributions to ensure optimal convergence rates. Gelman *et al.* (1997), Bedard and Rosenthal (2007) and Roberts and Rosenthal (2001) were the first authors to publish theoretical results for the optimal scaling problem in RW-MH algorithms with Gaussian proposals. For d-dimensional target distributions with i.i.d. components, the asymptotic acceptance rate optimizing the efficiency of the process is 0.234 independent of the target density. In this case we recommend that the selection of σ_{RWi} are chosen to ensure that the acceptance probability is roughly close to 0.234. This number is the acceptance probability obtained for asymptotically optimal acceptance rates for RW-MH algorithms when applied to multidimensional target distributions with scaling terms possibly depending on the dimension. To obtain this acceptance rate, one is required to perform some tuning of the proposal variance prior to final simulations. An alternative approach is to utilize a new class of Adaptive MCMC algorithms recently proposed in the literature, see Atchade and Rosenthal (2005) and Rosenthal (2007), but these are beyond the scope of this paper.

3.5. Markov chain results and analysis

This section presents the results comparing both MLE and Bayesian estimates for the parameters of Tweedie's compound Poisson model. It is also demonstrated how additional information in a Bayesian framework can be obtained through the complete knowledge of the target posterior distribution obtained from the MCMC algorithm described above. In this regard we demonstrate how this additional information can be exploited in the claims reserving setting to provide alternative statistical analysis not obtainable if one just considers point estimators. We also analyze model averaging solutions in section 5. These can be obtained by forming estimates using the information given by the full posterior distribution $\pi(\boldsymbol{\theta} | \mathcal{D}_I)$ that we find empirically from the MCMC samples.

The maximum likelihood and MCMC algorithms were implemented in Fortran. The maximization routine for the MLEs utilizes the direct search algorithm DBCPOL (that requires function evaluation only) from the IMSL numerical library. Note that, gradient based optimization routines such as the BFGS algorithm can be more efficient, but the direct search algorithm we used was sufficient for our problem in terms of computing time ($\approx$ 4 seconds on a typical desktop PC[1]).

The algorithm was analyzed on synthetic data and found to provide correct estimates. In particular with uniform priors the MAP estimates of the parameters are the same as the MLEs, up to numerical errors. This was confirmed for different sized claims triangles. The actual data set studied in this paper is presented in Table 2. The data we study is the standard data set used in Wüthrich and Merz (2008) scaled by 10,000.

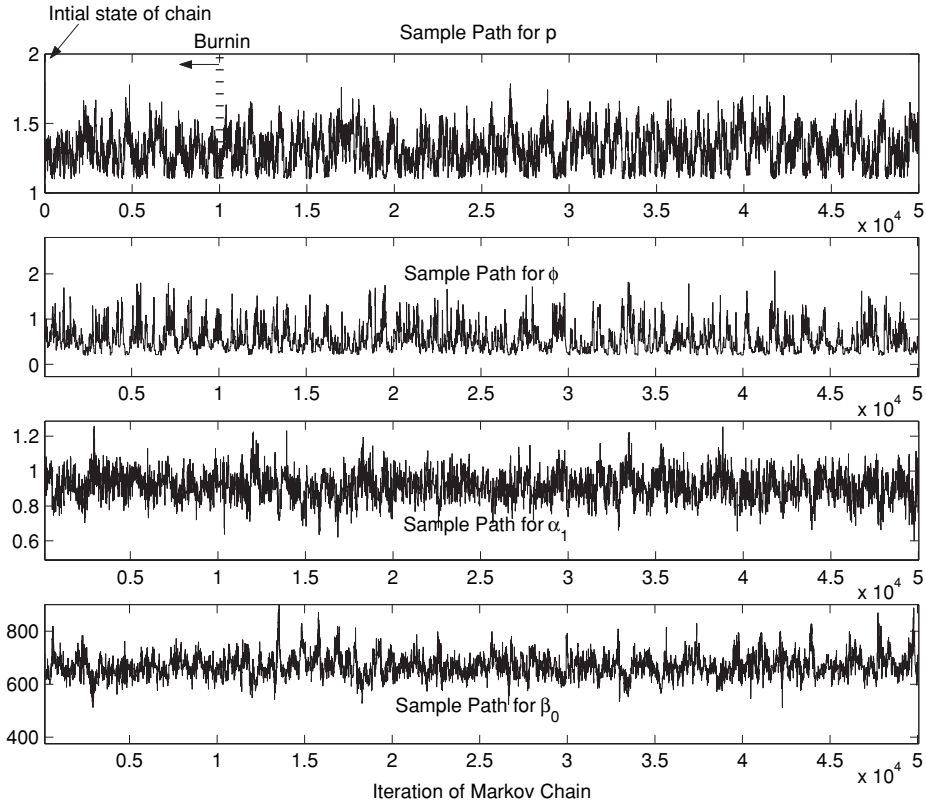


FIGURE 1: Markov chain sample paths ($p, \phi, \alpha_1, \beta_0$).

[1] Intel® Core™ 2 Duo, 2.13GHz processor.

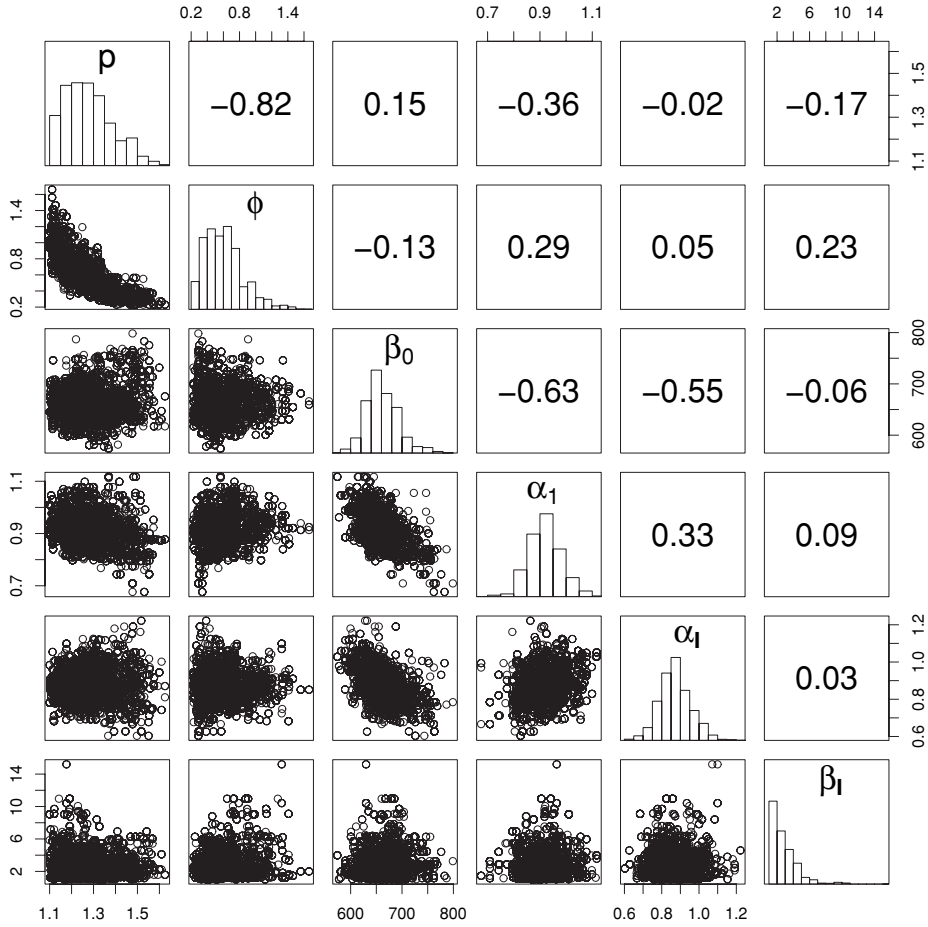


FIGURE 2: Posterior scatter plots, marginal posterior histograms and linear correlations for $(p, \phi, \alpha_1, \beta_0, \alpha_I, \beta_I)$.

The results presented for the Bayesian approach were obtained after pretuning the Markov chain random walk standard deviations, σ_{RW_i}, to produce average acceptance probabilities of 0.234. Then the final simulation was for 10^5 iterations from a Markov chain (≈ 17 min[1]) in which the first 10^4 iterations were discarded as burnin when forming the estimates.

The pretuned proposal standard deviations σ_{RW_i} are presented in Table 3. The first set of results in Table 3 demonstrates the MLE versus the Bayesian posterior estimator MMSE for all model parameters. Included are the [5%, 95%] predictive intervals for the Bayesian posterior distribution. The MLE standard deviations are calculated using (3.11). The numerical standard errors (due to a finite number of MCMC iterations) in the Bayesian estimates are obtained by blocking the MCMC samples post burnin into blocks of length 5000

and using the estimates on each block to form the standard error (these are given in brackets next to the estimates).

The next set of analysis demonstrates the performance of the MCMC approach in converging to the stationary distribution given by the target posterior $\pi(\boldsymbol{\theta}|\mathcal{D}_I)$. To analyze this, in Figure 1, we present the trace plots for the Markov chain for the parameters, $(p, \phi, \alpha_1, \beta_0)$. Also, in Figure 2, we demonstrate the marginal posterior distribution histograms and pair-wise posterior scatter plots for $(p, \phi, \alpha_1, \beta_0, \alpha_I, \beta_I)$. The lower panels in Figure 2 are the scatter plots for the pair-wise marginal posteriors, the diagonal contains the marginal posteriors and the upper panels contains the correlations between parameters. These plots demonstrate strong linear correlations between several parameters. Some of these correlations are similar to MLE correlations calculated using (3.11). For example, we found that under the posterior distribution $\rho(p, \phi) \approx -0.82$ and $\rho(\beta_0, \alpha_1) \approx -0.63$, see Figure 2, are similar to $\rho(\hat{p}^{\text{MLE}}, \hat{\phi}^{\text{MLE}}) \approx -0.94$ and $\rho(\hat{\beta}_0^{\text{MLE}}, \hat{\alpha}_1^{\text{MLE}}) \approx -0.68$ correspondingly. However, we also observed that under the posterior distribution $\rho(p, \beta_I) \approx -0.17$ and $\rho(\phi, \beta_I) \approx 0.23$, see Figure 2, while corresponding MLE correlations are zero, see (3.13).

4. Variable selection via posterior model probabilities

In the development so far it has been assumed that variable selection is not being performed, that is we are assuming that the model is known and we require parameter estimates for this model. This is equivalent to specifying that the number of α and β parameters is fixed and known in advance. We now relax this assumption and will demonstrate how the variable selection problem can be incorporated into our framework. The procedure we utilize for the variable selection is based on recent work of Congdon (2006) and specifies the joint support of the posterior distribution for the models and parameters under the product space formulation of Carlin and Chib (1995).

In this section we consider the subset of nested models which create homogenous blocks in the claims reserving triangle ($I = 9$) for the data set in Table 2.

- $M_0 : \boldsymbol{\theta}_{[0]} = (p, \phi, \tilde{\alpha}_0 = \alpha_0, \ldots, \tilde{\alpha}_I = \alpha_I, \tilde{\beta}_0 = \beta_0, \ldots, \tilde{\beta}_I = \beta_I)$ – **saturated model**.
- $M_1 : \boldsymbol{\theta}_{[1]} = (p, \phi, \tilde{\beta}_0)$ with $(\tilde{\beta}_0 = \beta_0 = \ldots = \beta_I)$, $(\alpha_0 = \ldots = \alpha_I = 1)$.
- $M_2 : \boldsymbol{\theta}_{[2]} = (p, \phi, \tilde{\alpha}_1, \tilde{\beta}_0, \tilde{\beta}_1)$ with $(\alpha_0 = \ldots = \alpha_4 = 1)$, $(\tilde{\alpha}_1 = \alpha_5 = \ldots = \alpha_I)$, $(\tilde{\beta}_0 = \beta_0 = \ldots = \beta_4)$, $(\tilde{\beta}_1 = \beta_5 = \ldots = \beta_I)$.
- $M_3 : \boldsymbol{\theta}_{[3]} = (p, \phi, \tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2)$ with $(\alpha_0 = \alpha_1 = 1)$, $(\tilde{\alpha}_1 = \alpha_2 = \ldots = \alpha_5)$ $(\tilde{\alpha}_2 = \alpha_6 = \ldots = \alpha_I)$, $(\tilde{\beta}_0 = \beta_0 = \beta_1)$, $(\tilde{\beta}_1 = \beta_2 = \ldots = \beta_5)$, $(\tilde{\beta}_2 = \beta_6 = \ldots = \beta_I)$
- $M_4 : \boldsymbol{\theta}_{[4]} = (p, \phi, \tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3, \tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3)$ with $(\alpha_0 = \alpha_1 = 1)$, $(\tilde{\alpha}_1 = \alpha_2 = \alpha_3)$, $(\tilde{\alpha}_2 = \alpha_4 = \alpha_5 = \alpha_6)$, $(\tilde{\alpha}_3 = \alpha_7 = \alpha_8 = \alpha_I)$, $(\tilde{\beta}_0 = \beta_0 = \beta_1)$, $(\tilde{\beta}_1 = \beta_2 = \beta_3)$, $(\tilde{\beta}_2 = \beta_4 = \beta_5 = \beta_6)$, $(\tilde{\beta}_3 = \beta_7 = \beta_8 = \beta_I)$.

- $M_5 : \boldsymbol{\theta}_{[5]} = (p, \phi, \widetilde{\alpha}_1, \widetilde{\alpha}_2, \widetilde{\alpha}_3, \widetilde{\alpha}_4, \widetilde{\beta}_0, \widetilde{\beta}_1, \widetilde{\beta}_2, \widetilde{\beta}_3, \widetilde{\beta}_4)$ with $(\alpha_0 = \alpha_1 = 1)$, $(\widetilde{\alpha}_1 = \alpha_2 = \alpha_3)$, $(\widetilde{\alpha}_2 = \alpha_4 = \alpha_5)$, $(\widetilde{\alpha}_3 = \alpha_6 = \alpha_7)$, $(\widetilde{\alpha}_4 = \alpha_8 = \alpha_I)$, $(\widetilde{\beta}_0 = \beta_0 = \beta_1)$, $(\widetilde{\beta}_1 = \beta_2 = \beta_3)$, $(\widetilde{\beta}_2 = \beta_4 = \beta_5)$, $(\widetilde{\beta}_3 = \beta_6 = \beta_7)$, $(\widetilde{\beta}_4 = \beta_8 = \beta_I)$.
- $M_6 : \boldsymbol{\theta}_{[6]} = (p, \phi, \alpha_0, \widetilde{\alpha}_1, \beta_0, \beta_1, \ldots, \beta_I)$ with $(\widetilde{\alpha}_1 = \alpha_1 = \ldots = \alpha_I)$.

Now, to determine the optimal model, we first consider the joint posterior distribution for the model probability and the model parameters denoted $\pi(M_k, \boldsymbol{\theta}_{[k]} | \mathcal{D}_I)$, where $\boldsymbol{\theta}_{[k]} = (\widetilde{\theta}_{1,[k]}, \widetilde{\theta}_{2,[k]}, \ldots, \widetilde{\theta}_{N[k],[k]})$ is the parameter vector for model $[k]$. Additionally we denote the prior bounds for $\widetilde{\theta}_{i,[k]}$ as $[a_{\widetilde{\theta}_{i,[k]}}, b_{\widetilde{\theta}_{i,[k]}}]$. We assume a prior distribution $\pi(M_k)$ for the model selection and a prior for the parameters conditional on the model $\pi(\boldsymbol{\theta}_{[k]} | M_k)$. It is no longer possible to run the standard MCMC procedure we described in section 3.4 for this variable selection setting. This is because the posterior is now defined on either a support consisting of disjoint unions of subspaces or a product space of all such subspaces, one for each model considered. A popular approach to run Markov chains in such a situation is to develop a more advanced sampler than that presented above, typically in the disjoint union setting. This involves developing a Reversible Jump RJ-MCMC framework, see Green (1995) and the references therein. This type of Markov chain sampler is complicated to develop and analyze. Hence, we propose as an alternative in this paper to utilize a recent procedure that will allow us to use the above MCMC sampler we have already developed for a model M_k. The process we must follow involves first running the sampler in the simulation technique described in section 3.4 for each model considered. Then the calculation of the posterior model probabilities $\pi(M_k | \mathcal{D}_I)$ is performed using the samples from the Markov chain in each model to estimate (4.3).

Furthermore, our approach here removes the assumption on the priors across models, made by Congdon (2006), p. 348,

$$\pi(\boldsymbol{\theta}_{[m]} | M_k) = 1, m \neq k \tag{4.1}$$

and instead we work with the prior

$$\pi(\boldsymbol{\theta}_{[m]} | M_k) = \prod_{i=1}^{N_{[m]}} \left[b_{\widetilde{\theta}_{i,[m]}} - a_{\widetilde{\theta}_{i,[m]}} \right]^{-1}, \ m \neq k. \tag{4.2}$$

That is, instead we use a class of priors where specification of priors for a model M_k automatically specifies priors for any other model. This is a sensible set of priors to consider given our product space formulation and it has a clear interpretation in our setting where we specify our models through a series of constraints, relative to each other. In doing this we also achieve our goal of having posterior model selection insensitive to the choice of the prior and being data driven. The modified version of Congdon's (2006), formula A.3, we obtain after relaxing Congdon's assumption, allows the calculation of the posterior

model probabilities $\pi(M_k|\mathcal{D}_I)$ using the samples from the Markov chain in each model to estimate

$$\begin{aligned}\pi(M_k|\mathcal{D}_I) &= \int \pi\left(M_k, \boldsymbol{\theta}_{[k]}\middle|\mathcal{D}_I\right) d\boldsymbol{\theta}_{[k]} = \int \pi\left(M_k\middle|\boldsymbol{\theta}_{[k]}, \mathcal{D}_I\right)\pi\left(\boldsymbol{\theta}_{[k]}\middle|\mathcal{D}_I\right) d\boldsymbol{\theta}_{[k]} \\ &\approx \frac{1}{T-T_b}\sum_{j=T_b+1}^{T}\pi\left(M_k\middle|\mathcal{D}_I, \boldsymbol{\theta}_{j,[k]}\right) \\ &= \frac{1}{T-T_b}\sum_{j=T_b+1}^{T}\frac{L_{\mathcal{D}_I}\left(M_k, \boldsymbol{\theta}_{j,[k]}\right)\prod_{k=0}^{K}\pi\left(\boldsymbol{\theta}_{j,[k]}\middle|M_k\right)\pi(M_k)}{\sum_{m=0}^{K}L_{\mathcal{D}_I}\left(m_m, \boldsymbol{\theta}_{j,[m]}\right)\prod_{k=0}^{K}\pi\left(\boldsymbol{\theta}_{j,[k]}\middle|M_m\right)\pi(M_m)} \\ &= \frac{1}{T-T_b}\sum_{j=T_b+1}^{T}\frac{L_{\mathcal{D}_I}\left(M_k, \boldsymbol{\theta}_{j,[k]}\right)}{\sum_{m=0}^{K}L_{\mathcal{D}_I}\left(m_m, \boldsymbol{\theta}_{j,[m]}\right)}.\end{aligned} \tag{4.3}$$

Here $K = 6$, and for a proof, see Congdon (2006), formula A.3. Note that, the prior of parameters (given model) contributes in the above implicitly as $\boldsymbol{\theta}_{j,[k]}$ are MCMC samples from the k^{th} models posterior distribution. In the actual implementation we used $T = 100{,}000$ and the burnin period $T_b = 10{,}000$. Note, the prior probabilities for each model are considered diffuse and are set such that all models *a priori* are equiprobable, hence $\pi(M_k) = 1/(K+1)$ and $\pi(\boldsymbol{\theta}_{j,[k]}|M_k)$ is the prior for model M_k's parameters evaluated at the j^{th} Markov chain iteration. Once we have the posterior model probabilities we can then take the MAP estimate for the optimal model (variable selection) for the given data set. In this paper we do not consider the notion of model averaging over different parameterized models in the variable selection context. Instead we simply utilize these results for optimal variable selection from a MAP perspective for the marginal posterior $\pi(M_k|\mathcal{D}_I)$.

In addition to this model selection criterion we also consider in the Bayesian framework the Deviance Information Criterion (DIC), see Bernardo and Smith (1994). From a classical maximum likelihood perspective we present the likelihood ratio (LHR) p-values.

Application of this technique to the simulated MCMC samples for each of the considered models produced the posterior model probabilities given in Table 4. This suggests that within this subset of models considered, the saturated model M_0 was the optimal model to utilize in the analysis of the claims reserving problem, $\pi(M_0|\mathcal{D}_I) \approx 0.7$. It is followed by model M_6 with $\pi(M_0|\mathcal{D}_I) \approx 0.3$. Additionally, the choice of M_0 was also supported by the other criteria we considered: DIC and LHR.

In future research it would be interesting to extend to the full model space which considers all models in the power set $|\boldsymbol{\theta}_{[0]}|$. This is a large set of models including all combinatorial combinations of model parameters for α's and β's. In such cases it is no longer feasible to run standard MCMC algorithms in each

TABLE 4

POSTERIOR MODEL PROBABILITIES $\pi(M_k|\mathcal{D}_I)$, DEVIANCE INFORMATION CRITERION (DIC) FOR VARIABLE SELECTION MODELS $M_0, \ldots, M_6$ AND LIKELIHOOD RATIO (LHR) p-VALUES (COMPARING M_0 TO $M_1, \ldots, M_6$).

	M_0	M_1	M_2	M_3	M_4	M_5	M_6
$\pi(M_k\|\mathcal{D}_I)$	0.71	4.19E-54	3.04E-43	1.03E-28	6.71E-20	2.17E-21	0.29
DIC	399	649	600	535	498	507	398
LHR *p-value*	1	2.76E-50	1.67E-40	3.53E-28	5.78E-21	3.03E-23	0.043

model since this will involve an impractical number of simulations. Hence, more sophisticated model exploration techniques will be required such as RJ-MCMC, see Green (1995) or the product space samplers of Carlin and Chib (1995).

We note here that we do not claim M_0 is the optimal model in all possible models, only in the subset we consider in this section. In saying this we acknowledge that we aim to work in the saturated model but consider it important to illustrate how variable selection can be performed in this class of models and also raise awareness that this will impact the model uncertainty analysis subsequently performed.

Hence, using these findings and the analysis of the MCMC results for model M_0 provided above, we may now proceed to analyze the claims reserving problem. Of interest to the aim of this paper is the sensitivity of the model choice parameter p to the parameterization of the claims reserving triangle. This is particularly evident when one considers the MMSE estimate of the model specification parameter p estimated under each model. In the most parsimonious, yet inflexible model M_1 the estimate obtained was $MMSE(p) \approx 1.9$, a very similar estimate was obtained in models M_2, M_3, M_4 and M_5, however, interestingly in the saturated model the estimate was $MMSE(p) \approx 1.3$ which is almost at the other extreme of the considered range for which the parameter p is defined.

5. CALCULATION OF THE CLAIMS RESERVES

We now demonstrate the results for several quantities in the claims reserving setting, utilizing the MCMC simulation results we obtained for the Bayesian posterior distribution under the variable selection model M_0 (saturated model). In particular, we start by noting that we use uniform prior distributions with a very wide ranges to perform inference implied by the data only. In this case, theoretically, the Bayesian MAP (the posterior mode) and MLEs for the parameters should be identical up to numerical error due to the finite number of MCMC iterations. A large number of MCMC iterations was performed so that the numerical error is not material. In general, the use of more informative

priors will lead to the differences between the MAP and MLE. Some of the MMSE estimates (the posterior mean) were close to the MAP estimates, indicating that the marginal posterior distributions are close to symmetric. When the posterior is not symmetric, MMSE and MAP can be very different. Also, note that the uncertainties in the parameter MLEs are estimated using the asymptotic Gaussian approximation (3.11)-(3.12). In the case of constant priors, this should lead to the same inferences as corresponding Bayesian estimators if the posterior distributions are close to the Gaussian approximation, see (3.23)-(3.24). In addition, the MLEs for the reserves, estimation error and process variance, see section 3.2, are based on a Taylor expansion around parameter MLEs assuming small errors. In many cases the posterior is materially different from the Gaussian distribution, has significant skewness and large standard deviation leading to the differences between the MLEs and corresponding Bayesian estimators. Having mentioned this, we now focus on the main point of this paper which involves analysis of the quantities in Table 5 related to the model uncertainty within Tweedie's compound Poisson models (introduced by fixing model parameter p) in a Bayesian setting.

TABLE 5

QUANTITIES USED FOR ANALYSIS OF THE CLAIMS RESERVING PROBLEM UNDER MODEL AVERAGING AND MODEL SELECTION IN RESPECT TO p.

	Model Averaging	**Model Selection for p**
Estimated Reserves	$ER = \widehat{R}^{\mathrm{B}} = E[\widehat{R} \mid \mathcal{D}_I]$	$ERp = E[\tilde{R} \mid \mathcal{D}_I, p]$
Process Variance	$PV = E[\sum \phi(\alpha_i \beta_j)^p \mid \mathcal{D}_I]$	$PVp = E[\sum \phi(\alpha_i \beta_j)^p \mid \mathcal{D}_I, p]$
Estimation Error	$EE = \mathrm{Var}(\widehat{R} \mid \mathcal{D}_I)$	$EEp = \mathrm{Var}(\widehat{R} \mid \mathcal{D}_I, p)$

It is worth noting that point estimates of model parameters are either in the frequentists approach MLEs or in a Bayesian approach the MAP or MMSE estimates. These are under the auspice that we wish to perform model selection (i.e. selection of p). The focus of this paper is to demonstrate the difference in results obtained for reserve estimates that can arise by performing model averaging instead of the typical approach of model selection, using a *priori* chosen p. In this regard we perform estimation utilizing the full posterior distribution of the parameters and not just point estimators. This allows us to capture the influence of the model uncertainty (uncertainty in p), since in a Bayesian setting we can account for this uncertainty using the posterior distribution. In particular, the Bayesian analysis specifies the optimal p (either in the MAP or the MMSE context) and it also provides a confidence interval for the choice of p (see Figure 7), which corresponds to the choice of the optimal model within Tweedie's compound Poisson models. Moreover, we demonstrate the impact on the claims reserve by varying p from 1.1 to 1.9 (i.e. for a fixed model choice).

5.1. Results: average over p

Initially it is worth considering the predicted reserve distribution for the estimator $\tilde{R}$. This is obtained by taking the samples $t = 10{,}001$ to $100{,}000$ from the MCMC simulation $\{p^t, \phi^t, \boldsymbol{\alpha}^t, \boldsymbol{\beta}^t\}$ and calculating $\{\tilde{R}^t\}$ via (3.26). The histogram estimate is presented in Figure 3. In the same manner, we also estimate the distributions of $\tilde{R}_{i,j} = \alpha_i \beta_j$ for the individual cells of the $I \times I$ claims matrix, presented as subplots in Figure 4. Note that the total observed loss in the upper triangle (≈ 9274) is consistent with $E\left[\sum_{i+j\leq I} \alpha_i \beta_j\right]$ and $\left[\mathrm{Var}(\sum_{i+j\leq I} \alpha_i \beta_j)\right]^{1/2}$ estimated using the MCMC samples as (≈ 9311) and (≈ 190) respectively. The maximum likelihood approach results in $\sum_{i+j\leq I} \widehat{\alpha}_i^{MLE} \widehat{\beta}_j^{MLE} \approx 9275$ with standard deviation ≈ 124 also conforming with the observed total loss.

Now we focus on quantities associated with the estimated distribution for $\tilde{R}$ to calculate the results, see Table 5, which can only be estimated once the entire posterior distribution is considered. These quantities are the key focus of this paper since they allow assessment of the conditional MSEP as specified in (3.28). In particular, we may now easily use the posterior probability samples obtained from the MCMC algorithm to evaluate the estimated reserve (ER), the process variance (PV) and the estimation error (EE) in the conditional MSEP. This provides an understanding and analysis of the behaviour of the

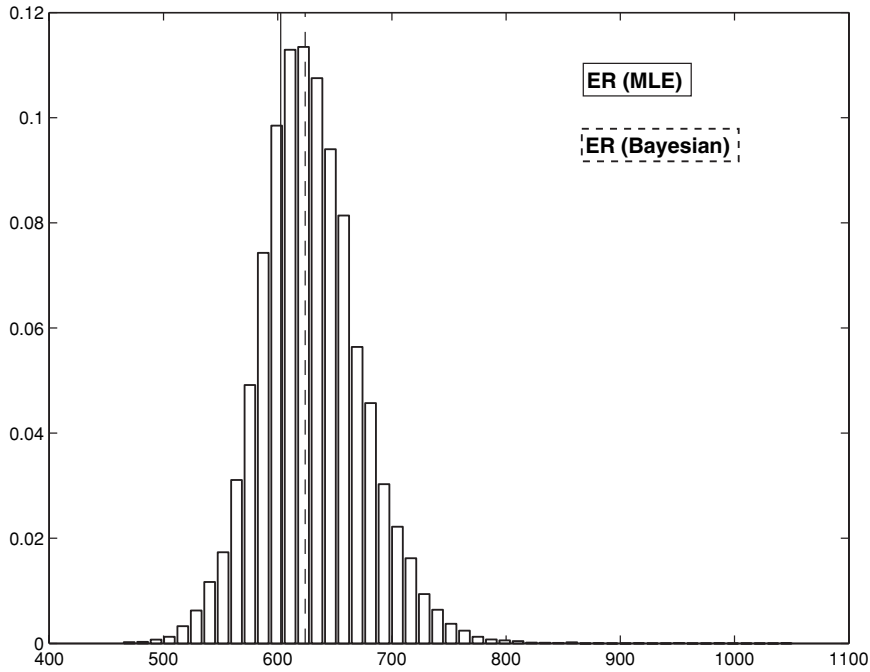


FIGURE 3: Predicted distribution of reserves, $\tilde{R} = \sum_{i+j>I} \alpha_i \beta_j$.

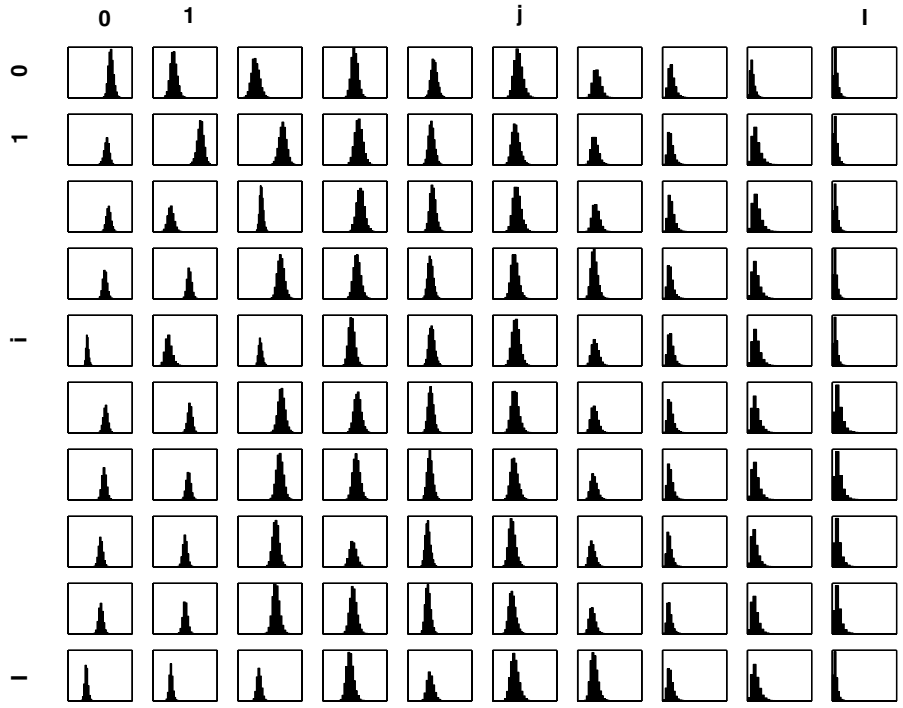


FIGURE 4: Posterior distributions for $\tilde{R}_{i,j} = \alpha_i \beta_j$ estimated using MCMC.

proposed model in both the model averaging and model selection (i.e. selection of p) contexts whilst considering the issue of model uncertainty, the goal of this paper. The Bayesian estimates for ER, PV, EE and MSEP are presented in Table 6. The corresponding MLEs were calculated using (3.8), (3.16), (3.15)

TABLE 6

MODEL AVERAGED ESTIMATES OF CLAIM RESERVE, PROCESS VARIANCE AND ESTIMATION ERROR. NUMERICAL ERROR IN BAYESIAN ESTIMATES IS REPORTED IN BRACKETS. SEE TABLE 5 FOR DEFINITIONS OF ER, PV, EE AND $MSEP = EE + PV$.

Statistic	Model Averaging	
	Bayesian Estimate	**MLE Estimate**
ER	624.1 (0.7)	602.630
$\sqrt{PV}$	37.3 (0.2)	25.937
$\sqrt{EE}$	44.8 (0.5)	28.336
$\sqrt{MSEP}$	58.3 (0.5)	38.414

and (3.17) respectively and presented in Table 6 for comparison. The results demonstrate the following:

- Claims reserves MLE, $\widehat{R}^{\mathrm{MLE}}$, is less than Bayesian estimate $\widehat{R}^{\mathrm{B}}$ by approximately 3%, which is the estimation bias of the claims reserve MLE (see also Wüthrich and Merz (2008), Remarks 6.15.
- $\sqrt{EE}$ and $\sqrt{PV}$ are of the same magnitude, approximately 6-7% of the total claims reserves.
- MLEs for $\sqrt{EE}$ and $\sqrt{PV}$ are less than corresponding Bayesian estimates by approximately 37% and 30%, respectively.
- The difference between $\widehat{R}^{\mathrm{MLE}}$ and $\widehat{R}^{\mathrm{B}}$ is of the same order of magnitude as $\sqrt{EE}$ and $\sqrt{PV}$ and thus is significant.

Note that we use constant priors with very wide ranges, the MLE uncertainties are calculated using an asymptotic Gaussian approximation and numerical error due to the finite number of MCMC iterations is not material (also see the 1st paragraph, section 5). The observed significant differences between the MLEs and corresponding Bayesian estimators suggest that our posterior distributions are skewed and materially different from the Gaussian distribution.

We conclude this section with the distribution of R, the total outstanding claims payment, see Figure 5. This is obtained from the MCMC samples of

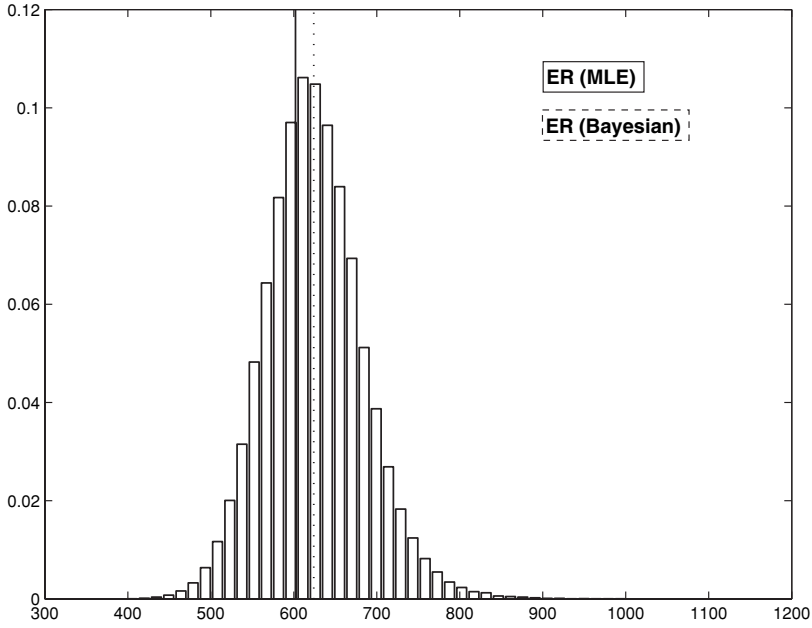


FIGURE 5: Distribution of total outstanding claims payment $R = \sum_{i+j>I} Y_{i,j}$, accounting for all process, estimation and model uncertainties.

TABLE 7

BAYESIAN MODEL AVERAGED ESTIMATES OF VALUE AT RISK FOR OUTSTANDING CLAIMS PAYMENT R AND CLAIM RESERVES $\tilde{R}$.

VaR_q	Model Averaging	
	R	$\tilde{R}$
$\text{VaR}_{75\%}$	659.8 (0.9	650.6 (1.0)
$\text{VaR}_{90\%}$	698.4 (1.2)	680.4 (1.3)
$\text{VaR}_{95\%}$	724.0 (1.5)	701.7 (1.6)

the parameters $(p, \phi, \boldsymbol{\alpha}, \boldsymbol{\beta})$ which we then transform to parameters $(\lambda, \gamma, \boldsymbol{\tau})$ from model representation 1, section 2, and simulate annual losses in $i + j > I$. That is, these samples of R are obtained from the full predictive distribution $f(R|\mathcal{D}_I) = \int g(R|\boldsymbol{\theta})\, \pi(\boldsymbol{\theta}|\mathcal{D}_I)\, d\boldsymbol{\theta}$, where $g(R|\boldsymbol{\theta})$ is the distribution of R given by (1.3) and (2.1). It takes into account both process uncertainty and parameter uncertainty. We note that while reserving by some measure of centrality such as $\widehat{R}^{\text{B}}$ may be robust, it will not take into account the distributional shape of R. A viable alternative may be Value-at-Risk (VaR) or a coherent risk measure such as Expected Shortfall. In Table 7 we demonstrate estimates of the VaR for $\tilde{R}$ and R at the 75%, 90% and 95% quantiles.

5.2. Results: conditioning on p

As part of the model uncertainty analysis, it is useful to present plots of the relevant quantities in the model selection (selection of p) settings, see Figure 6, where we present $ER_p = E[\tilde{R}|\mathcal{D}_I, p]$, $PV_p = \sum_{i+j>I} E[\phi(\alpha_i\beta_j)^p|\mathcal{D}_I, p]$ and $EE_p = \text{Var}(\tilde{R}|\mathcal{D}_I, p)$ as a function of p. Figure 6 shows:

- MLE of ER_p is almost constant, varying approximately from a maximum of 603.96 ($p = 1.1$) to a minimum of 595.78 ($p = 1.9$) while the MLE for ER was 602.63.
- The Bayesian estimates for ER_p change as a function of p. Approximately, it ranged from a maximum of 646.4 ($p = 1.9$) to a minimum of 621.1 ($p = 1.5$) while the Bayesian estimator for ER was 624.1. Hence, the difference (estimation bias) within this possible model range is ≈ 25 which is of a similar order as the process uncertainty and the estimation error.
- Bayesian estimators for $\sqrt{PV_p}$ and $\sqrt{EE_p}$ increase as p increases approximately from 33.1 to 68.5 and from 37.4 to 102.0 respectively, while the Bayesian estimators for $\sqrt{PV}$ and $\sqrt{EE}$ are 37.3 and 44.8 correspondingly. Hence, the resulting risk measure strongly varies in p which has a large influence on quantitative solvency requirements. The MLEs for PV_p and EE_p are

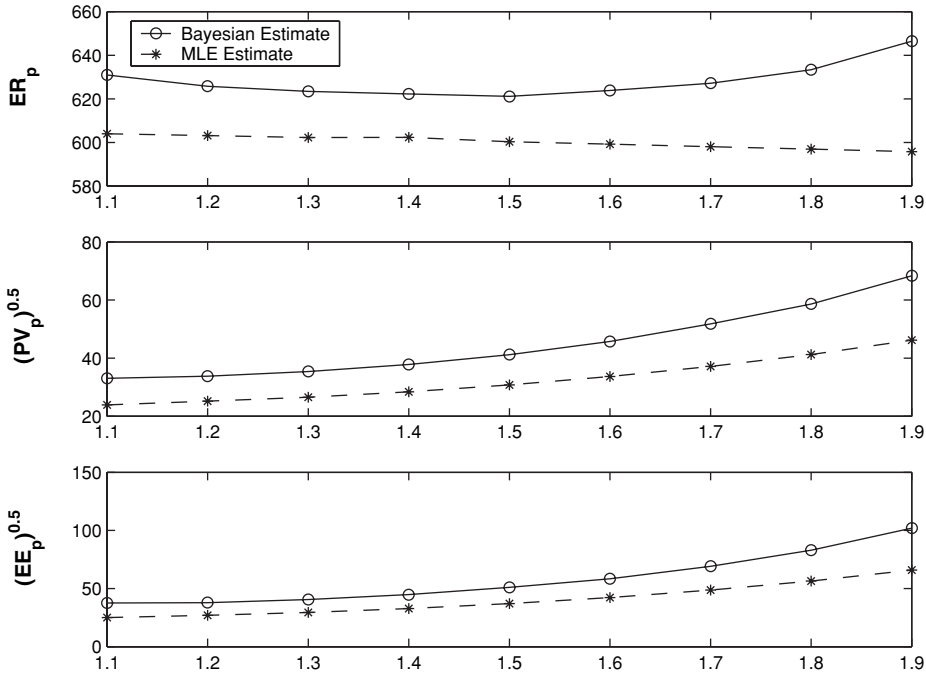


FIGURE 6: Estimates of quantities from Table 5 conditional on p. Note, numerical standard errors are not included as they are negligible and are less than the size of the symbols.

significantly less than the corresponding Bayesian estimators. Also, the difference between the MLE and the Bayesian estimators increases as p increases.

For interpretation purposes of the above results it is helpful to use the following relations between model averaging and model selection quantities (easily derived from their definitions in Table 5):

$$ER = E[ER_p|\mathcal{D}_I], \tag{5.1}$$

$$PV = E[PV_p|\mathcal{D}_I], \tag{5.2}$$

$$EE = E[EE_p|\mathcal{D}_I] + \mathrm{Var}(ER_p|\mathcal{D}_I). \tag{5.3}$$

Here, the expectations are calculated with respect to the posterior distribution of p. The histogram estimate of the later is presented in Figure 7 and highlights significant uncertainty in p (model uncertainty within Tweedie's compound Poisson model).

We also provide Figure 8 demonstrating a Box and Whisker summary of the distributions of $\tilde{R}|p$ for a range of values of p. This plot provides the first, second and third quartiles as the box. The notch represents uncertainty in the

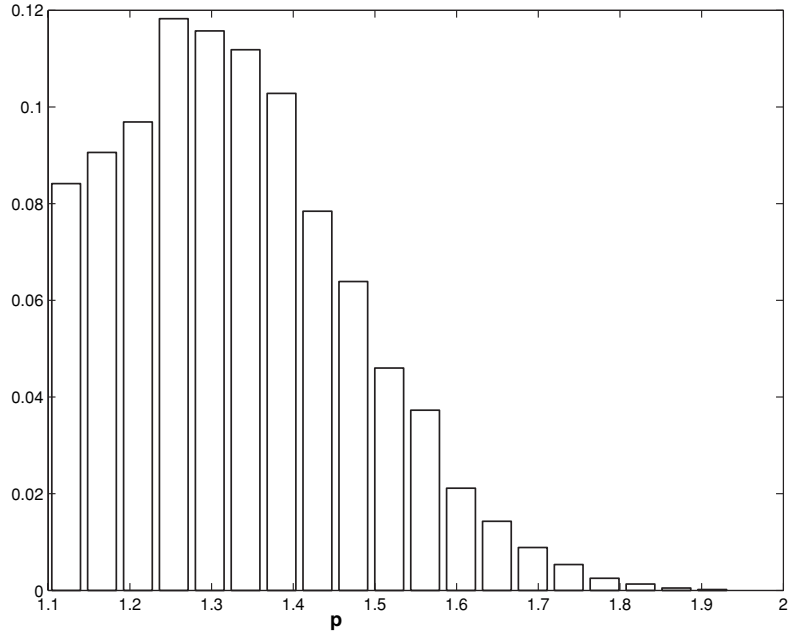


FIGURE 7: Posterior distribution of the model parameter p.

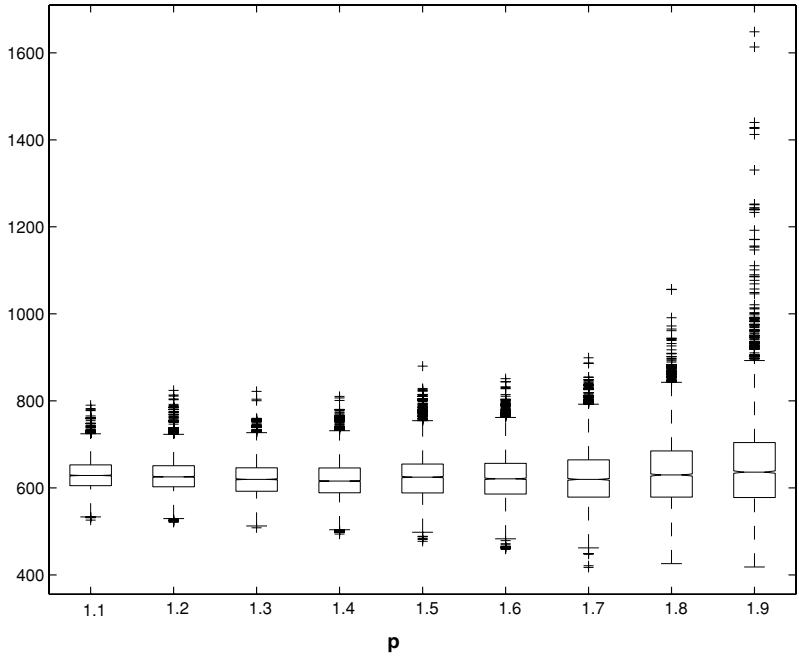


FIGURE 8: Predicted claim reserves $\tilde{R}$ distributional summaries conditional on model parameter p.

median estimate for model comparison, across values of p, and the whiskers demonstrate the smallest and largest data points not considered as outliers. The outliers are included as crosses and the decision rule to determine if a point is an outlier was taken as the default procedure from the statistical software package R.

The conclusion from this section is that if model selection is performed (i.e. p is fixed by the modeller), the conditional MSEP will increase significantly if a poor choice of the model parameter p is made. In particular, though the median is fairly constant for the entire range of $p \in (1,2)$ the shape of the distribution of $\widetilde{R} \mid p$ is clearly becoming more diffuse as $p \to 2$. This will lead to significantly larger variance in the reserve estimate. If risk measures such as Value-at-Risk are used in place of the mean, it will result in reserves which are too conservative (if a poor choice of p is made). Also, using the maximum likelihood approach may significantly underestimate the claims reserves and associated uncertainties.

5.3. Overdispersed Poisson and Gamma models

There are several popular claims reserving models, however we restrict our comparison to the overdispersed Poisson and gamma models since they fit into Tweedie's compound Poisson framework when $p = 1$ and $p = 2$ respectively. Note that the overdispersed Poisson model and several other stochastic models lead to the same reserves as the chain ladder method but different in higher moments. The detailed treatment of these models can be found in e.g. England and Verrall (2002) or Wüthrich and Merz (2008), section 3.2.

The MLEs for the reserves and associated uncertainties within the overdispersed Poisson and gamma models are provided in Table 8. These results are obtained when the dispersion ϕ is estimated by $\widehat{\phi}^P$ using Pearson's residuals (3.18) and when ϕ is estimated by $\widehat{\phi}^{\text{MLE}}$ obtained from the maximization of the likelihood. The results for the first case are also presented in Wüthrich and Merz (2008), Table 6.4. Firstly note that, the values of $\widehat{\phi}^P$ and $\widehat{\phi}^{\text{MLE}}$ are significantly different both for the overdispersed Poisson and gamma models. As we mentioned in section 3.2, for a fixed p, the MLE for the reserves does not depend on ϕ while the estimation error, process variance and MSEP are proportional to ϕ. As one can see from Table 8, different estimators for the dispersion ϕ lead to the same estimators for the reserves but very different estimators for the uncertainties. Also note that, our MLE calculations for Tweedie's distribution conditional on p, i.e. Figure 6, are obtained using $\widehat{\phi}^{\text{MLE}}$ and are consistent with the corresponding results for the overdispersed Poisson and Gamma models when $p \to 1$ and $p \to 2$ respectively. Though, in the case of the overdispersed Poisson we had to use an extended quasi-likelihood to estimate $\widehat{\phi}^{\text{MLE}}$. In Figure 6, we do not show the results based on $\widehat{\phi}^P$ but would like to mention that these are always above the MLEs and below the Bayesian estimators for the process variance and estimation error and are consistent with corresponding

TABLE 8

THE MLES FOR THE OVERDISPERSED POISSON ($p = 1$) AND GAMMA ($p = 2$) MODELS, WHEN THE DISPERSION ϕ IS ESTIMATED AS $\widehat{\phi}^{\mathrm{P}}$ USING PEARSON'S RESIDUALS (3.18) OR $\widehat{\phi}^{\mathrm{MLE}}$.

Statistic	Overdispersed Poisson		Gamma model	
	$\widehat{\boldsymbol{\phi}}^{P} \approx 1.471$	$\widehat{\boldsymbol{\phi}}^{MLE} \approx 0.954$	$\widehat{\boldsymbol{\phi}}^{P} \approx 0.045$	$\widehat{\boldsymbol{\phi}}^{MLE} \approx 0.031$
ER_p	604.706	604.706	594.705	594.705
$\sqrt{PV_p}$	29.829	24.017	62.481	52.162
$\sqrt{EE_p}$	30.956	24.925	92.826	77.496
$\sqrt{MSEP_p}$	42.989	34.613	111.895	93.415

overdispersed Poisson and gamma model limits. Interestingly, the ratio $\widehat{\phi}^{P}/\widehat{\phi}^{\mathrm{MLE}}$ is approximately 1.4-1.5 for all considered cases of p within a range [1,2].

The MLEs obtained using both $\widehat{\phi}^{\mathrm{MLE}}$ and $\widehat{\phi}^{\mathrm{P}}$ underestimate the uncertainties compared to the Bayesian analysis. Note that, while the MLEs for the uncertainties are proportional to the dispersion estimator, the corresponding Bayesian estimators are averages over all possible values of ϕ according to its posterior distribution. The uncertainty in the estimate for the dispersion is large which is also highlighted by a bootstrap analysis in Wüthrich and Merz (2008), section 7.3. This indicates that ϕ should also depend on the individual cells (i, j). However, in this case overparameterization needs to be considered with care and Bayesian framework should be preferred.

6. DISCUSSION

The results demonstrate the development of a Bayesian model for the claims reserving problem when considering Tweedie's compound Poisson model. The sampling methodology of a Gibbs sampler is applied to the problem to study the model sensitivity for a real data set. The problem of variable selection is addressed in a manner commensurate with the MCMC sampling procedure developed in this paper and the most probable model under the posterior marginal model probability is then considered in further analysis. Under this model we then consider two aspects, model selection and model averaging with respect to model parameter p. The outcomes from these comparisons demonstrate that the model uncertainty due to fixing p plays a significant role in the evaluation of the claims reserves and its conditional MSEP. It is clear that whilst the frequentist MLE approach is not sensitive to a poor model selection, the Bayesian estimates demonstrate more dependence on poor model choice, with respect to model parameter p. We use constant priors with very wide ranges to perform inference in the setting where the posterior is largely implied by data only. Also, we run a large number of MCMC iterations so that numerical error

in the Bayesian estimators is very small. In the case of the data we studied, the MLEs for the claims reserve, process variance and estimation error were all significantly different (less) than corresponding Bayesian estimators. This is due to the fact that the posterior distribution implied by the data and estimated using MCMC is materially different from Gaussian, i.e. more skewed.

Future research will examine variable selection aspects of this model in a Bayesian context considering the entire set of possible parameterizations. This requires development of advanced approaches such as Reversible Jump MCMC and variable selection stochastic optimization methodology to determine if a more parsimonious model can be selected under assumptions of homogeneity in adjacent columns/rows in the claims triangle.

Acknowledgements

The first author is thankful to the Department of Mathematics and Statistics at the University of NSW for support through an Australian Postgraduate Award and to CSIRO for support through a postgraduate research top up scholarship. Thank you also goes to Robert Kohn for discussions.

References

Atchade Y. and Rosenthal, J. (2005) On adaptive Markov chain Monte Carlo algorithms. *Bernoulli* **11(5)**, 815-828.

Bedard, M. and Rosenthal, J.S. (2008) Optimal scaling of Metropolis algorithms: heading towards general target distributions. *The Canadian Journal of Statistics* **36(4)**, 483-503.

Bernardo, J.M. and Smith, A.F.M. (1994) *Bayesian Theory*. John Wiley and Sons, NY.

Cairns, A.J.G. (2000) A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics* **27**, 313-330.

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

Casella, G. and George, E.I. (1992) Explaining the Gibbs Sampler. *The American Statistician* **46(3)**, 167-174.

Congdon P. (2006) Bayesian model choice based on Monte Carlo estimates of posterior model probabilities. *Computational Statistics and Data Analysis* **50(2)**, 346-357.

Dunn, P.K. and Smyth, G.K. (2005) Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing* **15**, 267-280.

England, P.D. and Verrall, R.J. (2002) Stochastic claims reserving in general insurance. *British Actuarial Journal* **8(3)**, 443-510.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995) *Bayesian Data Analysis*. Chapman and Hall / CRC Texts in Statistical Science Series, **60**.

Gelman, A., Gilks, W.R. and Roberts, G.O. (1997) Weak convergence and optimal scaling of random walks metropolis algorithm. *Annals of Applied Probability* **7**, 110-120.

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

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

Jørgensen, B. and de Souza, M.C.P. (1994) Fitting Tweedie's compound Poisson model to insurance claims data. *Scandinavian Actuarial Journal*, 69-93.

Robert, C.P. and Casella, G. (2004) *Monte Carlo Statistical Methods*, 2nd Edition Springer Texts

in Statistics.
ROBERTS, G.O. and ROSENTHAL, J.S. (2001) Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science* **16**, 351-367.
ROSENTHAL, J.S. (2007) AMCMC: An R interface for adaptive MCMC. *Computational Statistics and Data Analysis* **51(12)**, 5467-5470.
SMITH, A.F.M. and ROBERTS, G.O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of Royal Statistical Society Series B* **55(1)**, 3-23.
SMYTH, G.K. and JØRGENSEN, B. (2002) Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. *Astin Bulletin* **32**, 143-157.
TWEEDIE, M.C.K. (1984) An index which distinguishes between some important exponential families. In Statistics: Applications in new directions. *Proceeding of the Indian Statistical Institute Golden Jubilee International Conference*, J.K. Ghosh and J. Roy (eds.), 579-604, Indian Statistical Institute Canada.
WRIGHT, E.M. (1935) On asymptotic expansions of generalized Bessel functions. *Proceedings of London Mathematical Society* **38**, 257-270.
WÜTHRICH, M.V. (2003) Claims reserving using Tweedie's compound Poisson model. *Astin Bulletin* **33**, 331-346.
WÜTHRICH, M.V. and MERZ, M. (2008) *Stochastic Claims Reserving Methods in Insurance*, Wiley Finance.

GARETH W. PETERS
CSIRO Mathematical and Information Sciences,
Sydney, Locked Bag 17,
North Ryde, NSW, 1670, Australia
and
UNSW Mathematics and Statistics Department,
Sydney, 2052, Australia.
E-Mail: peterga@maths.unsw.edu.au

PAVEL V. SHEVCHENKO (corresponding author)
CSIRO Mathematical and Information Sciences,
Sydney, Locked Bag 17,
North Ryde, NSW, 1670, Australia.
E-Mail: Pavel.Shevchenko@csiro.au

MARIO V. WÜTHRICH
ETH Zurich, Department of Mathematics,
CH-8092 Zurich,
Switzerland.
E-Mail: wueth@math.ethz.ch