STOCHASTIC CLAIMS RESERVING VIA A BAYESIAN SPLINE MODEL WITH RANDOM LOSS RATIO EFFECTS

BY

GUANGYUAN GAO AND SHENGWANG MENG

ABSTRACT

We propose a Bayesian spline model which uses a natural cubic *B*-spline basis with knots placed at every development period to estimate the unpaid claims. Analogous to the smoothing parameter in a smoothing spline, shrinkage priors are assumed for the coefficients of basis functions. The accident period effect is modeled as a random effect, which facilitate the prediction in a new accident period. For model inference, we use Stan to implement the no-U-turn sampler, an automatically tuned Hamiltonian Monte Carlo. The proposed model is applied to the workers' compensation insurance data in the United States. The lower triangle data is used to validate the model.

KEYWORDS

Stochastic claims reserving, tail factor, natural cubic spline, Hamiltonian Monte Carlo, Bayesian modeling.

1. INTRODUCTION

Modern enterprise risk management requires quantifying the uncertainty in crucial point estimates. This motivates the development of various stochastic models. In a general insurance company's balance sheet, the claims reserve is always the largest liability and it is important to estimate this liability accurately. An over-estimated reserve will increase the required capital and hence increase the capital costs to an insurer, while an under-estimated reserve can hide an insolvency problem.

Friedland (2010) discussed several deterministic claims reserving algorithms that return point estimates. Two monographs on stochastic claims reserving are Taylor (2000) and Wüthrich and Merz (2008). As one of the most widely used algorithms, the chain ladder method is studied frequently. Mack (1993, 1999) formulated the prediction uncertainty associated with the chain ladder method. In fact, many stochastic claims reserving models are based on the

Astin Bulletin 48(1), 55-88. doi: 10.1017/asb.2017.19

chain ladder method. This method is characterized by a multiplicative mean function with two factor covariates: accident period and development period. However, with factor covariates, it is impossible to (a) extrapolate the tail factor and (b) predict a new accident period. The main purpose of this paper is to address these two problems.

There are several extant papers that address the tail factor issue. Mack (1999) chose a tail factor by observing the trend of last few age-to-age factors in the traditional chain ladder method. Together with Mack (1993), Mack (1999) formulated the mean squared error of estimated unpaid claims. Wright (1990), Clark (2003) and Zhang *et al.* (2012) used parametric curves to extrapolate the tail factor as part of the model-fitting process. Wright (1990) derived the run-off pattern as a curve based on a risk theoretic model of the claims generating process, using the same model specification for claims reserving as for pricing: claim numbers are modeled by a Poisson distribution and claim severities are modeled by a gamma distribution. Zhang *et al.* (2012) modeled the cumulative claims rather than the incremental claims as in most other literature. Hence, Zhang *et al.* (2012) assumed an auto-regressive process error term, i.e., the cumulative claims are not independent. For model inference, Wright (1990) and Clark (2003) applied the maximum likelihood estimation in the generalized linear model (GLM) framework, while Zhang *et al.* (2012) applied Markov chain Monte Carlo method (MCMC) in a Bayesian framework. Zhang and Dukic (2013) and Antonio and Beirlant (2008) considered the equivalence between a penalized spline regression and a mixed effects model. They called the proposed models semi-parametric models since they allowed smoothing only of development periods. Both papers estimated the parameters in the models by MCMC method. Verrall (1996) applied the generalized additive model (GAM) in order to smooth over the accident period, using a locally weighted regression smoother. Verrall *et al.* (2012) is a compromise between a chain ladder model and a parametric curve model. It applied reversible jump MCMC (RJMCMC) to join the chain ladder run-off pattern and the parametric curve run-off pattern. The model parameters were estimated in WinBUGS with a RJMCMC add-in. The literature most relevant to our paper is England and Verrall (2001), in which the authors smoothed over both the accident periods and development periods using a smoothing spline in a GAM. We will discuss these models in detail in Section 2.

Essentially, there are three kinds of estimation for the uncertainty associated with estimated reserves. The first is the method of moments, which is used in Mack (1993) for the distribution-free model. The second is based on the resampling property of maximum likelihood estimation in GLMs and GAMs. Most statistical software can return the covariance matrix of linear predictors, from which the mean squared error of the total unpaid claims can be derived. The third is simulation-based estimation, including the bootstrap and MCMC methods. While the bootstrap method is always used to infer a GLM in the frequentist framework, MCMC is used to infer a Bayesian model. The first two estimation methods return an analytical expression of the uncertainty of the estimated reserve, while the third relies on the sample of simulated future claims to infer the reserve and the associated uncertainty.

Our approach to address the tail development issue is to expand the space of the development period covariate by using basis functions. The number of parameters associated with development periods equals the number of basis functions. The non-significant parameters are shrunk to zero by the assumed shrinkage priors. More specifically, we propose a Bayesian spline model, one of the most popular basis expansion models. There are several books covering the topic of spline models: Hastie and Tibshirani (1990), Ruppert *et al.* (2003), Wood (2006), Hastie *et al.* (2009), James *et al.* (2013) and Bishop (2006). Shrinkage priors are always chosen as Laplace distribution, Cauchy distribution, the generalized double Pareto distribution, etc. Park and Casella (2008) discussed inference using the Laplace prior distribution. Armagan *et al.* (2013) discussed generalized double Pareto shrinkage.

Among Bayesian computational tools, MCMC methods (Metropolis *et al.*, 1953; Hastings, 1970) are popular and widely used. In the context of a Bayesian model, MCMC methods can be used to generate a Markov chain whose stationary distribution is the posterior distribution of the quantities of interest. In this paper, we consider a variant of MCMC, Hamiltonian Monte Carlo (HMC). HMC was introduced to physics by Duane *et al.* (1987) and to statistical problems by Neal (1994, 2011). Compared with the random-walk Metropolis algorithm where the proposed value is not related to the target distribution, HMC proposes a value by computing a trajectory according to Hamiltonian mechanics to take account of the target distribution. The programming of HMC is more complicated than MCMC, so we use Sampling Through Adaptive Neighborhoods (Stan) (Stan Development Team, 2016) for model inference. Stan implements no-U-turn sampler (NUTS) (Homan and Gelman, 2014), an automatically tuned HMC.

The paper is structured as follows. In Section 2, we review several stochastic reserving models which can address the tail development. In Section 3, we propose a Bayesian spline model with random loss ratio effects for claims reserving. We also give an introduction to HMC and the implementation in Stan. Section 4 contains a workers' compensation insurance reserving example to illustrate the proposed method. Section 5 contains our conclusion.

2. Review of existing stochastic claims reserving models with a tail factor

The claims are usually cross aggregated by two factors: period of accident and period of development. We treat all claims with the same accident period as a cohort and track its development in the future. The accident periods are denoted by $i = 1, \ldots, I$, and the development periods by $j = 1, \ldots, I$. Here we consider the case when the number of accident periods is equal to the number of development periods. The unit can be a quarter year, half year or a full year, but the accident periods and development periods should use the same unit and the intervals should be equal. The experience periods (or calendar periods) are denoted by $i + j$, which contains a cross-section of experience from various accident periods lying on a diagonal line. The incremental claims of accident period i during the development period j are denoted by $y_{i,j}$. The cumulative claims for accident period i as of development period j are defined as $c_{i,j} := \sum_{l=1}^{j} y_{i,l}$, and the ultimate claims of accident period i as $c_{i,\infty}$ or u_i, which equals to $c_{i,I}$ when the claims are fully run-off by the development period I. The unpaid claims of accident period i are defined as $R_i := \sum_{j=I-i+2}^{\infty} y_{i,j}$. In the case of no development after I, $R_i = c_{i,I} - c_{i,I-i+1}$. The total unpaid claims are defined as $R := \sum_{i=1}^{I} R_i$. We define the upper triangle as $y^u := \{y_{i,j} : i+j \leq I+1\}$, the lower triangle as $y^l := \{y_{i,j} : i+j > I+1,\ j \leq I\}$. Assuming all the claims are paid until the development period $J(> I)$, the tail development is defined as $y^t := \{y_{i,j} : I < j \leq J\}$.

Table A2 shows a typical structure of an incremental claims run-off triangle, where the upper triangle y^u is available by the end of most recent accident year 1995. The earned premiums are measures of loss exposure of each accident period. The loss reserving problem is to predict the lower triangle y^l, and the possible tail development y^t given the upper triangle y^u. The final reserve set aside is not exactly equal to the summation of predicted lower triangle and tail development but depends on the uncertainty around them.

Except for Model (4), in which the response variable is cumulative claims, all the following models and the Bayesian spline model in Section 3 treat incremental claims as the response variable, and they have a common assumption that, given the mean parameters, incremental claims are independent of each other.

2.1. Mack's model

Mack (1993, 1999) established a distribution-free claims reserving model. He derived the mean squared error of both the individual future claims estimator and the aggregated future claims estimator. The age-to-age factors f_k and the individual age-to-age factor $F_{i,k}$ are defined by the following equations:

$$\mathbf{E}(c_{i,k+1}|c_{i,k}, \ldots, c_{i,1}) = c_{i,k} f_k,$$
$$F_{i,k} = \frac{c_{i,k+1}}{c_{i,k}}.$$

In Mack (1999), the tail factor estimate, denoted as $\hat{f}_{\text{ult}}$, is chosen by observing the trend of the last few estimated age-to-age factors, $\hat{f}_j$. For example, the R function MackChainLadder estimates the tail factor via a linear extrapolation of $\log(\hat{f}_j - 1)$. The standard errors of $\hat{f}_{\text{ult}}$ and $\hat{F}_{i,\text{ult}}$ are chosen to satisfy the following inequities:

$$\text{s.e.}(\hat{f}_{k-1}) > \text{s.e.}(\hat{f}_{\text{ult}}) > \text{s.e.}(\hat{f}_k),$$
$$\text{s.e.}(\hat{F}_{i,k-1}) > \text{s.e.}(\hat{F}_{i,\text{ult}}) > \text{s.e.}(\hat{F}_{i,k}), \quad (1)$$

where k is an index with

$$\hat{f}_{k-1} > \hat{f}_{\text{ult}} > \hat{f}_k.$$

The standard error estimate of the tail claims, denoted as $\hat{\sigma}_{\text{ult}}$, can be calculated as

$$\hat{\sigma}_{\text{ult}} = \text{s.e.}(\hat{F}_{i,\text{ult}})\sqrt{\hat{C}_{i,J}}.$$

Once the four quantities, $\hat{f}_{\text{ult}}$, s.e.$(\hat{f}_{\text{ult}})$, s.e.$(\hat{F}_{i,\text{ult}})$ and $\hat{\sigma}_{\text{ult}}$, are estimated, the mean squared error of aggregated future claims can be calculated using the formula in Mack (1999).

Mack's model plays an important role in the claims reserving practice due to its simplicity and ease of programming. In Section 4, we compare our results with those from Mack's model.

2.2. Parametric curve models

Rather than forming an additional stage in Mack's model, estimation of the tail factor is integrated into a parametric curve model fitting process. One of the problems associated with curve fitting is to choose a curve flexible enough to capture the systematic run-off pattern and still stable enough to discard the noise. A curve with too few parameters will under-fit the underlying run-off pattern, while one with too many parameters will be too unstable for prediction.

The Hoerl curve model. Wright (1990) derived the incremental run-off pattern as a Hoerl curve based on a risk theoretic model of the claims generating process. It used the same model specification in the claims reserving as in pricing: claim numbers are fitted by a Poisson distribution and claim severities are fitted by a gamma distribution. The simplified Hoerl curve model is as follows:

$$\begin{aligned} \frac{y_{ij}}{\phi} &\sim \text{Poi}\left(\frac{\mu_{ij}}{\phi}\right), \\ \mu_{ij} &= \exp(e_i + a_i + b_i \ln j + c_i j), \end{aligned} \tag{2}$$

where ϕ is the dispersion parameter, e_i is a measure of exposure in the ith accident period, a_i is the accident period effect, b_i and c_i are the development period effects in the ith accident period whose values can be estimated and smoothed over the accident periods using Kalman filter. A special case is created by letting $b_i = b$ and $c_i = c$ for all accident periods. Under this special case, the incremental run-off pattern is the same for all accident periods.

Clark's curve model. Clark's curve model (Clark, 2003) assumes the underlying cumulative run-off pattern is a loglogistic curve or a Weibull curve. Theoretically, any cumulative distribution function can be used to fit the underlying cumulative run-off pattern. Clark's curve model treats the incremental claims as

https://doi.org/10.1017/asb.2017.19 Published online by Cambridge University Press

a response variable and the model is specified as follows:

$$\frac{y_{ij}}{\phi} \sim \text{Poi}\left(\frac{\mu_{ij}}{\phi}\right),$$
$$\mu_{ij} = P_i \cdot LR \cdot [G(j|\omega, \theta) - G(j-1|\omega, \theta)], \tag{3}$$

where P_i is the earned premium in the accident period i, LR is the loss ratio and $G(\cdot|\omega, \theta)$ is a two-parameter growth curve. The parameters, LR, ω, θ and ϕ, are estimated via the maximum likelihood. The tail factor can be extrapolated from the estimated growth curve, i.e., $G(j|\hat{\omega}, \hat{\theta})$, $j > I$. In Section 4, we compare our results with those from Clark's curve model.

A non-linear curve model for cumulative claims. It is worth noting that Zhang *et al.* (2012) modeled the cumulative claims rather than the incremental claims as in most other literature. Since the cumulative claims are not independent, Zhang *et al.* (2012) assumed an autoregressive process error term. Zhang *et al.* (2012) also considered the dependence among insurance companies, i.e., dependence among several run-off triangles, so their model can borrow credibility from other companies for the parameter estimation. If only one insurance company is considered, the model in Zhang *et al.* (2012) can be simplified as follows:

$$\begin{aligned}
\log c_{ij} &= \log[e_i a_i G(j|\omega, \theta)] + \varepsilon_j, \\
\varepsilon_j &= \rho \varepsilon_{j-1} + \delta_j, \\
\delta_j &\overset{iid}{\sim} \text{N}(0, \sigma^2(1-\rho^2)), \\
\varepsilon_0 &\sim \text{N}(0, \sigma^2),
\end{aligned} \tag{4}$$

where e_i has the same interpretation as in Model (2), and $G(\cdot|\omega, \theta)$ has the same interpretation as in Model (3). Due to the dependent error structure, the model has a complicated likelihood function. Zhang *et al.* (2012) estimated the parameters in this model via MCMC in WinBUGS.

2.3. Smoothing models

Antonio and Beirlant (2008) and Zhang and Dukic (2013) used a penalized spline regression to smooth over the development periods. Verrall (1996) and England and Verrall (2001) used GAMs to smooth over the accident periods and both the accident and development periods, respectively.

Semi-parametric models. Antonio and Beirlant (2008) applied the statistical method proposed by Crainiceanu *et al.* (2005). They re-parameterized a penalized spline regression as a Bayesian mixed effects model, so that the penalized regression spline can be solved using MCMC. The same model was studied in Zhang and Dukic (2013). The Bayesian mixed effects model is specified as follows:

$$\frac{y_{ij}}{\phi} \sim \text{Poi}\left(\frac{\mu_{ij}}{\phi}\right),$$
$$\log \mu_{ij} = a_i + f(j),$$
$$f(j) = b_0 + b_1 j + [\mathbf{Z}c]_j, \tag{5}$$

where $c = (c_1, \ldots, c_K)$ is a K-vector with the prior $c_k \sim \text{N}(0, \sigma_c^2)$, and $\mathbf{Z} = \mathbf{Z}_1\mathbf{Z}_2^{-1/2}$ is an $I \times K$ matrix. Here, $\mathbf{Z}_1$ is a $I \times K$ matrix with the ith row of $(|i - \kappa_1|^3, \ldots, |i - \kappa_K|^3)$, and $\mathbf{Z}_2$ is a $K \times K$ matrix with the (s, t) entry of $|\kappa_s - \kappa_t|^3$. κ_k, $k = 1, \ldots, K$ are the K knots which are chosen to capture the shape of the run-off pattern. The non-informative priors are assumed for the other parameters, a_i, b_0 and b_1. The tail development from the development period $I + 1$ to J could be involved if we added $J - I$ rows below the matrix $\mathbf{Z}_1$ with the ith row equal to $(|i - \kappa_1|^3, \ldots, |i - \kappa_K|^3)$, $i = I + 1, \ldots, J$.

Generalized additive models. Verrall (1996) smoothed over the accident periods using a locally weighted smoothing regression. England and Verrall (2001) smoothed over both the accident periods and the development periods using a smoothing spline. Both models belong to GAMs. We present the model in England and Verrall (2001) as follows:

$$\mathbb{E}(y_{ij}) = \mu_{ij},$$
$$\text{Var}(y_{ij}) = \phi\mathbb{E}(y_{ij})^{\rho},$$
$$\log \mu_{ij} = e_i + a + s_{\theta_{\text{Acc}}}(i) + s_{\theta_{\text{Dev}}}(j) + s_{\theta_{\text{Dev}}}(\ln j), \tag{6}$$

where e_i is interpreted the same as in Model (2), $s_{\theta_{\text{Acc}}}(i)$ represents a smoothing of accident period i, obtained using a smoothing spline with smoothing parameter θ_{Acc}, $s_{\theta_{\text{Dev}}}(j)$ and $s_{\theta_{\text{Dev}}}(\ln j)$ represent smoothing splines specifying the shape of the run-off pattern, with smoothing parameter θ_{Dev} chosen to be the same for both functions. The value of ρ can be 0,1,2,3, which corresponds to normal, over-dispersed Poisson (ODP), gamma and inverse Gaussian distribution of y_{ij}, respectively, see Table A1.

2.4. A mixture model

Verrall and Wüthrich (2012) proposed a mixture model in the Bayesian framework. They assumed a log linear development curve after a particular development period. The model is a compromise between the chain ladder model and a curve model. The run-off pattern is divided into two parts: a chain ladder development pattern for the first few development periods, i.e., one parameter for one development period and a curve development pattern for the remaining development periods. RJMCMC is applied to perform parameter inference for this model, since the parameter space changes during the inferential process.

The incremental claims are assumed to follow an ODP distribution with the dispersion parameter fixed at the estimate from a frequentist GLM. The model can be written as follows:

$$\frac{y_{ij}}{\hat{\phi}} \sim \text{Poi}\left(\frac{\mu_{ij}}{\hat{\phi}}\right), \tag{7}$$

where

$$\log(\mu_{ij}) = \begin{cases} c_1 + a_i + b_j, \; j \leq k \\ c_2 + a_i + dj, \; j > k. \end{cases}$$

Non-informative priors are assumed for c_1, c_2, a_i, b_j and d. RJMCMC lets the data "determine" the location of the "change point" k, where the two run-off patterns are connected. However, RJMCMC is not easy to code and it involves between-model jumps and within-model jumps. While Verrall and Wüthrich (2012) wrote the specific RJMCMC, Verrall *et al.* (2012) used an add-in, named "Jump", in WinBUGS to implement RJMCMC.

2.5. Incurred-paid ratio method

Incurred-paid ratio method makes use of the additional information from the incurred claims data to calculate the tail factor. There is always payment delay in the paid claims run-off triangle. If we do not expect newly reported IBNR claims after the last development period, the incurred claims comprise all claims, and they are the estimated ultimate claims by claims adjusters. The incurred-paid ratio of accident period is defined as

$$Q_{i,j} := \frac{c^I_{i,j}}{c_{i,j}},$$

where $c^I_{i,j}$ denotes the cumulative incurred claims of the accident period i until the development period j. The tail factor is chosen by inspecting the upper right incurred-paid ratio triangle. For the prediction uncertainty, Mack's model can be applied. Several assumptions about the incurred claims must be fulfilled when using this method. For more details, please refer to Section 2.5 in Wüthrich and Merz (2015).

3. A Bayesian natural cubic spline model with random loss ratio effects model

In this paper, we assume the earned premiums are known, hence the coefficients of accident periods are interpreted as the loss ratios. Most literature assumes that the accident periods have fixed effects. Under a fixed effects model, we cannot predict the claims in a new accident period. Here we assume the accident periods have random effects to facilitate the prediction in a new accident period. The multiplication of a new accident period's loss ratio estimate and its earned premium is an estimate of the new accident period's ultimate loss. In the Bayesian framework, a random effect effectively means a two-level prior structure, which will be demonstrated in the model specification.

The incremental run-off pattern is smoothed by a natural cubic spline. Unlike the GAM in England and Verrall (2001) and the semi-parametric model in Antonio and Beirlant (2008), which maximized the penalized residual sum of squares (RSS) to estimate the coefficients of basis functions, in this paper, we assume shrinkage priors for the coefficients of basis functions. Gelman *et al.* (2015) discuss the Bayesian basis function models and shrinkage priors in detail (p.487).

Splines are a combination of polynomials and step functions. Polynomial models tend to capture the shape of the data as long as there are high-degree polynomials. A disadvantage of the polynomial models is global representation, which means all the data points can affect parameter estimation and every parameter can affect the shape of the polynomial. A step function model partitions the data into several parts and fits each part using a basis function whose value is zero for the remaining parts of the data. Step function models have the disadvantage of discontinuity at the boundaries of partition. Spline models combine the continuity property of polynomial models and the piecewise property of step function models. For example, a cubic spline is a series of piecewise-cubic polynomials joined continuously up to the second derivatives.

3.1. Model description

The response variable is the incremental claims, rather than the cumulative claims as in Zhang and Dukic (2013). We follow the generalized form used in England and Verrall (2001) and specify the model as follows:

$$
\begin{aligned}
\mathbb{E}(y_{ij}) &= \mu_{ij} = P_i \times LR_i \times G(j), \\
\mathrm{Var}(y_{ij}) &= \hat{\phi} V(\mu_{ij}) = \hat{\phi}\mu_{ij}^{\rho}, \\
G(j) &= \frac{F(j)}{\sum_{l=1}^{J} F(l)}, \; j = 1, \ldots, J, \\
\log F(j) &= \sum_{h=1}^{I} \beta_h b_h(j), \; j = 1, \ldots, J, \\
LR_i | \alpha_{LR}, \mu_{LR} &\overset{iid}{\sim} \mathrm{Gamma}(\alpha_{LR}, \alpha_{LR}/\mu_{LR}), \\
\beta_h | \sigma_\beta^2 &\overset{iid}{\sim} \mathrm{DoubleExp}(0, \sigma_\beta^2), \; h = 1, \ldots, I, \qquad (8)
\end{aligned}
$$

where P_i is the earned premium of the accident period i (assumed to be known), LR_i is the random loss ratio effect of accident period i, G is the "normalized" incremental claims run-off pattern, F is the "raw" incremental claims run-off pattern, $\hat{\phi}$ is the plug-in dispersion parameter (estimated via maximum quasilikelihood), V is the variance function, the interpretation of ρ is the same as in Model (6), b_h are basis functions, the hyper-parameter μ_{LR} is interpreted as the average loss ratio and the hyper-parameter $1/\alpha_{LR}$ measures the spread of loss ratios for different accident periods. The hyper-priors for the hyper-parameters will be discussed later. Note that the prior for LR_i applies for $i > I$, in which LR_i is the loss ratio for a new accident period i.

The coefficients of basis functions, β_h, follow a shrinkage prior. A shrinkage prior has high density at zero and heavy tails to avoid over-shrinking. It can be a t distribution with a small degree of freedom, or a double exponential distribution (Laplace distribution). The Laplace prior induces sparsity in the posterior mode, in that the posterior mode can be exactly equal to zero, which is related to the Lasso method (Park and Casella, 2008). The Laplace prior is the prior with the heaviest tails that still produces a computationally convenient unimodal posterior density. An alternative is to use a generalized double Pareto prior distribution (Gelman *et al.*, 2014), which resembles the double exponential near the origin while having arbitrarily heavy tails. Here we choose a Laplace distribution, where σ_β^2 can be given a non-informative hyper-prior or assumed to be constant. We consider two special cases. When $\sigma_\beta^2 \to 0$, there is no shrinkage on the coefficients, and the model approaches to the chain ladder model, i.e., one coefficient for one development period. When $\sigma_\beta^2 \to \infty$, the shrinkage force is infinitely strong, and the model approaches to the log-linear chain ladder model, i.e., the logarithm of run-off pattern is a straight line.

The "normalized" run-off pattern satisfies the desirable condition of summation equal to 1, which implies that $G(j)$ is the incremental proportion of ultimate claims. The transformation from F to G can be viewed as adding a constraint to the "raw" run-off pattern F. If we replaced G with F in the mean function, the predictive distribution of future claims would not be affected; however, the interpretation of LR_i, μ_{LR} and α_{LR} would not be straightforward. Another implication of using "normalized" run-off pattern is that the ultimate claims of accident period i depend on LR_i rather than β_h. This implication is crucial in predicting the ultimate claims amount for a new accident period.

In our model, the development period is treated as a continuous covariate and the basis functions expand the space of the development period covariate. A common choice for b_h is a polynomial. The mechanism of defining the set of $\{b_h : h = 1, \ldots, I\}$ determines the behaviour of $\log F$. We consider $\log F$ as full rank natural cubic splines, which use polynomials as basis functions with knots placed at every distinctive value, i.e., knots placed at every development period. An alternative is to use a fixed-knot[1] natural cubic spline basis, which leads to similar fit assuming the knots are chosen properly. The I knots implies a natural cubic spline with I degrees of freedom and hence I coefficients associated with I basis functions. The underlying logic is as follows. To join $I+1$ pieces of three-order polynomials (with intercept) continuously at the I knots up to the second derivatives, we have $4 \times (I+1) - 3 \times I = I+4$ degrees of freedom. Moreover, to obtain the "natural" property, we constrain the spline as a straight line beyond the two boundary knots, so four degrees of freedom are lost, implying I degrees of freedom left at the end.

We can rewrite $\log F$ as

$$\log F(j) = [\mathbf{B}\beta]_j,$$

where $\mathbf{B}$ is a $J \times I$ matrix, $\beta = (\beta_1, \ldots, \beta_I)$ is a I-vector and $[\mathbf{B}\beta]_j$ is the jth element of the J-vector $\mathbf{B}\beta$. The dimension of $\mathbf{B}$ is determined by the number of knots and the number of development periods including the tail development. The matrix $\mathbf{B}$ can be written as

$$\mathbf{B} = \begin{bmatrix} b_1(1) & b_2(1) & \cdots & b_I(1) \\ b_1(2) & b_2(2) & \cdots & b_I(2) \\ \cdots & \cdots & \cdots & \cdots \\ b_1(J) & b_2(J) & \cdots & b_I(J) \end{bmatrix}.$$

There are multiple bases $\{b_h : h = 1, \ldots, I\}$ which can span the same set of functions. The R function ns can return an orthogonal natural cubic spline basis, called the B-spline basis. The orthogonal matrix has computational efficiency. Our experience shows that the computing time of a spline using a B-spline basis is much less than that of a spline using a truncated basis (not an orthogonal basis). Under the definition of natural cubic spline, for any β, we have the following linear condition in the tail development period:

$$[\mathbf{B}\beta]_{I+1} - [\mathbf{B}\beta]_I = [\mathbf{B}\beta]_{I+2} - [\mathbf{B}\beta]_{I+1} = \ldots = [\mathbf{B}\beta]_J - [\mathbf{B}\beta]_{J-1}.$$

Corresponding to $\rho = 0, 1, 2, 3$, there are four candidates for the error distribution: normal, ODP, gamma and inverse Gaussian. As Table A1 shows, the variance of the incremental claims can be constant, proportional to the mean, proportional to the mean squared, or proportional to the mean cubic correspondingly. Typically, we choose a suitable distribution family by checking the diagnostic plot of residuals. However, due to the special data structure of run-off triangle, the pattern shown in the residual plot sometimes is misleading. We will illustrate this point in Section 4.

The non-informative hyper-priors are assumed for α_{LR}, μ_{LR}, and σ_β^2 as follows:

$$\alpha_{LR} \sim \text{Gamma}(0.0001, 0.0001),$$

$$\mu_{LR} \sim \text{Gamma}(0.0001, 0.0001),$$

$$\sigma_\beta^2 \sim \text{Gamma}(0.0001, 0.0001).$$

The two-level prior structure for LR_i essentially implies that LR_i is not independent unconditionally. The posterior of loss ratio LR_i is a trade-off between the data evidence and the prior. The precision parameter, α_{LR}, implies the degree of effect of the prior on the posterior. If $\alpha_{LR} \to 0$, the prior degenerates to a non-informative prior and the posterior of LR_i is determined by the data. In this case, the model is close to a chain ladder model. If $\alpha_{LR} \to \infty$, the prior poses a strong effect on the posterior of LR_i, and overrides the evidence from data. In this case, the model is close to a Bornhütter–Ferguson model (Bornhütter and Ferguson, 1972).

An obvious question is why we choose a natural cubic spline rather than a quadratic spline or a higher order spline. It turns out that the natural cubic spline is a solution to a fitting problem of finding a function to minimize the sum of the RSS and the integral of the squared second derivatives of this function. Remarkably, even without constraining this function as splines, it can be shown that this function is a natural cubic spline with knots placed at the unique values of the covariate (Hastie *et al.*, 2009).

3.2. Model inference

While we consider only one error distribution, the ODP distribution, the calculation in this section would also be applied for other error distributions. The ODP distribution implies that the variance of the incremental claims is proportional to their means, i.e., $\rho = 1$ in Model (8). The plug-in dispersion parameter, $\hat{\phi}$, is obtained by fitting a frequentist quasi-GLM with the accident periods and the development periods as the factor covariates and a logarithm link function (see Model (15)).

The main task in Bayesian model inference is to obtain the posterior distribution of parameters of interest. According to Bayes' theorem,

$$\begin{aligned}
&p\left(LR, \beta, \mu_{LR}, \alpha_{LR}, \sigma_\beta^2 \mid y^u\right) \\
&\times \propto p(y^u \mid LR, \beta) p(LR \mid \mu_{LR}, \alpha_{LR}) p\left(\beta | \sigma_\beta^2\right) p(\mu_{LR}) p(\alpha_{LR}) p\left(\sigma_\beta^2\right) \\
&= \prod_{y_{ij} \in y^u} \text{Poi}\left(\frac{y_{ij}}{\hat{\phi}} \,\middle|\, \frac{P_i LR_i \exp\left(\sum_{h=1}^{I} \beta_h b_h(j)\right)}{\hat{\phi} \sum_{l=1}^{J} \exp\left(\sum_{h=1}^{I} \beta_h b_h(l)\right)} \right) \\
&\times \prod_{i=1}^{I} \text{Gam}\left(LR_i | \alpha_{LR}, \alpha_{LR}/\mu_{LR}\right) \prod_{h=1}^{I} \text{DExp}\left(\beta_h \mid 0, \sigma_\beta^2\right) \\
&\times \text{Gam}(\mu_{LR} | 0.0001, 0.0001) \text{Gam}(\alpha_{LR} | 0.0001, 0.0001) \\
&\times \text{Gam}\left(\sigma_\beta^2 | 0.0001, 0.0001\right),
\end{aligned} \tag{9}$$

where $\text{Poi}(x|a)$ denotes the probability density function of a Poisson distribution with parameter a evaluated at x, as for the notation $\text{Gam}(x|a, b)$ and $\text{DExp}(x|a, b)$. The right-hand side of Equation (9) is not a commonly used distribution. We apply HMC to sample from the posterior distribution (9). The discussion of HMC is deferred to the next section.

Ultimately, we want to get the predictive distribution of the future claims. According to the relationship between joint distribution and marginal distribu-

https://doi.org/10.1017/asb.2017.19 Published online by Cambridge University Press

tion, the predictive distribution of future claims follows:

$$
\begin{aligned}
&p\left(y^l, y^t \mid y^u\right)\\
&= \int_{LR,\beta,\mu_{LR},\alpha_{LR},\sigma_\beta^2} p\left(y^l, y^t, LR, \beta, \mu_{LR}, \alpha_{LR}, \sigma_\beta^2 | y^u\right) dLR d\beta d\mu_{LR} d\alpha_{LR} d\sigma_\beta^2\\
&= \int_{LR,\beta,\mu_{LR},\alpha_{LR},\sigma_\beta^2} p(y^l, y^t, \mid LR, \beta) p\left(LR, \beta, \mu_{LR}, \alpha_{LR}, \sigma_\beta^2 \mid y^u\right)\\
&\times dLR d\beta d\mu_{LR} d\alpha_{LR} d\sigma_\beta^2, \qquad (10)
\end{aligned}
$$

where $p(y^l, y^t \mid LR, \beta)$ is the ODP density function and $p(LR, \beta, \mu_{LR}, \alpha_{LR}, \sigma_\beta^2 \mid y^u)$ is the posterior density function of parameters. From Equation (10), to get a sample of future claims, we can first simulate a sample of parameters from the posterior distribution (9), then simulate a sample of future claims from the ODP distribution given LR, β equal to the sampled parameters.

For a new accident period $i > I$, the loss ratio LR_i is not in the posterior distribution (9), i.e., $LR_i \notin LR$. So we generate LR_i from

$$
\begin{aligned}
p(LR_i|y_u) &= \int_{\mu_{LR},\alpha_{LR}} p(LR_i|\mu_{LR}, \alpha_{LR}) p(\mu_{LR}, \alpha_{LR}|y_u) d\mu_{LR} d\alpha_{LR}\\
&= \int_{\mu_{LR},\alpha_{LR}} p(LR_i|\mu_{LR}, \alpha_{LR})\\
&\quad \times \left[\int_{LR,\beta,\sigma_\beta^2} p\left(LR, \beta, \mu_{LR}, \alpha_{LR}, \sigma_\beta^2 \mid y^u\right) dLR d\beta d\sigma_\beta^2\right] d\mu_{LR} d\alpha_{LR}\\
&= \int_{LR,\beta,\mu_{LR},\alpha_{LR},\sigma_\beta^2} p(LR_i|\mu_{LR}, \alpha_{LR}) p\left(LR, \beta, \mu_{LR}, \alpha_{LR}, \sigma_\beta^2 \mid y^u\right)\\
&\quad \times dLR d\beta d\mu_{LR} d\alpha_{LR} d\sigma_\beta^2. \qquad (11)
\end{aligned}
$$

Suppose we get a T sample $\{(\mu_{LR}^t, \alpha_{LR}^t) : t = 1, \ldots, T\}$ from $p(LR, \beta, \mu_{LR}, \alpha_{LR}, \sigma_\beta^2 \mid y^u)$. For each sampled vector $(\mu_{LR}^t, \alpha_{LR}^t)$, we simulate an LR_i^t from $p(LR_i|\mu_{LR} = \mu_{LR}^t, \alpha_{LR} = \alpha_{LR}^t)$, a gamma distribution. Given LR_i^t, the ultimate claims of the new accident period i can be simulated from the ODP distribution as follows:

$$
\frac{u_i}{\hat{\phi}}|LR_i^t \sim \text{Poi}\left(\frac{P_i LR_i^t}{\hat{\phi}}\right).
$$

3.3. Hamiltonian Monte Carlo

Duane *et al.* (1987) first proposed HMC in physics, and Neal (1994, 2011) first introduced HMC to statistical problems. A good reference is Betancourt (2017). Unlike the random-walk Metropolis algorithm, where the proposed value is not

https://doi.org/10.1017/asb.2017.19 Published online by Cambridge University Press

related to the target distribution, HMC proposes a value by computing a trajectory according to Hamiltonian mechanics which takes account of the target distribution.

Our target distribution is $p(LR, \beta, \mu_{LR}, \alpha_{LR}, \sigma_\beta^2 \mid y^u)$. According to Equation (9), we know the kernel part of $p(LR, \beta, \mu_{LR}, \alpha_{LR}, \sigma_\beta^2 \mid y^u)$ except for a constant multiplied to it. For compactness and generality, we denote the target distribution (9) as $p(\theta)$. An equal length of $2I+3$ vector ψ is chosen as an auxiliary variable distributed as $p(\psi)$ with the covariance matrix Σ. We define the negative log probability density function of θ and ψ as follows:

$$U(\theta) := -\log p(\theta),\ K(\psi) := -\log p(\psi).$$

As an analogy to Hamiltonian mechanics, θ can be viewed as the position and ψ can be viewed as the momentum, which are both time-dependent parameters, i.e., $\theta = \theta(t)$ and $\psi = \psi(t)$. $U(\theta)$ is called the potential energy and $K(\psi)$ is called the kinetic energy. The joint density $p(\theta, \psi)$ defines a Hamiltonian function as follows:

$$H(\theta, \psi) := -\log p(\theta, \psi) = -\log p(\theta) - \log p(\psi) = U(\theta) + K(\psi). \tag{12}$$

The Hamiltonian mechanics evolve according to Hamilton's equations of motion:

$$\begin{aligned} \frac{d\theta_i}{dt} &= \frac{\partial H}{\partial \psi_i} = \frac{\partial K}{\partial \psi_i}, \\ \frac{d\psi_i}{dt} &= -\frac{\partial H}{\partial \theta_i} = -\frac{\partial U}{\partial \theta_i}. \end{aligned} \tag{13}$$

For computer implementation, Equation (13) must be approximated by discretizing time, using some small step size, δ. The Hamiltonian mechanics in the next infinitesimal time can be approximated by the leapfrog method as follows:

$$\begin{aligned} \psi_i(t+\delta/2) &= \psi_i(t) - (\delta/2)\frac{\partial U}{\partial \theta_i}(\theta_i(t)), \\ \theta_i(t+\delta) &= \theta_i(t) + \delta\frac{\partial K}{\partial \psi_i}(\psi_i(t+\delta/2)), \\ \psi_i(t+\delta) &= \psi_i(t+\delta/2) - (\delta/2)\frac{\partial U}{\partial \theta_i}(\theta_i(t+\delta)). \end{aligned} \tag{14}$$

This update would keep Hamiltonian function (12) constant if δ is small enough, i.e., $H(\theta(t), \psi(t)) \approx H(\theta(t+\delta), \psi(t+\delta))$. Note that for equations (13) and (14) only, t is interpreted as the continuous time variable. For the remainder, t is usually interpreted as the tth iteration.

The evolution in Hamiltonian mechanics implies a very specific proposal procedure: starting from the $t-1$th iterated values of $(\theta^{t-1}, \psi^{t-1})$, (θ^*, ψ^*) are proposed via leapfrog method (14). The detailed HMC algorithm in the tth iteration is as follows:

1. First, a value ψ^{t-1} for the momentum is drawn from a distribution $p(\psi)$.
2. Next, the Hamiltonian mechanics $(\theta^{t-1}, \psi^{t-1})$ evolves via the following leapfrog method for L steps to get the proposed value (θ^*, ψ^*):

$$\psi_i^{t-1+\delta/2} = \psi_i^{t-1} + (\delta/2)\frac{\partial \log p(\theta^{t-1})}{\partial \theta_i},$$
$$\theta_i^{t-1+\delta} = \theta_i^{t-1} - \delta\frac{\partial \log p(\psi^{t-1+\delta/2})}{\partial \psi_i},$$
$$\psi_i^{t-1+\delta} = \psi_i^{t-1+\delta/2} + (\delta/2)\frac{\partial \log p(\theta^{t-1+\delta})}{\partial \theta_i}.$$

   Note that $\theta^* = \theta^{t-1+\delta L}$, $\psi^* = \psi^{t-1+\delta L}$.
3. A Metropolis accept step is conducted with the acceptance rate as

$$\min\{1, \exp[H(\theta^{t-1}, \psi^{t-1}) - H(\theta^*, \psi^*)]\}.$$

4. $\theta^t = \theta^*$ if the proposed value is accepted and $\theta^t = \theta^{t-1}$ if not. Discard ψ^*.

If there were no approximation errors from the leapfrog discretization (i.e., $\delta \to 0$), the leapfrog trajectory would follow the exact trajectory and we would definitely accept (θ^*, ψ^*) since $\exp[H(\theta^{t-1}, \psi^{t-1}) - H(\theta^*, \psi^*)] = 1$. However, there are always errors given the non-zero step size δ. Neal (2011) suggested that HMC is optimally efficient when its acceptance rate is approximately 65%, while an optimal acceptance rate for the random walk Metropolis–Hastings (M–H) algorithm is around 23% (Roberts *et al.*, 1997).

The no-U-turn sampler (NUTS). In MCMC, all the tuning parameters should be fixed during the simulation that will be used for inference, otherwise the algorithm may converge to the wrong distribution. BUGS has an adaptive period during which suitable tuning parameters are selected. There are three tuning parameters in HMC: the covariance matrix Σ in $p(\psi)$, the step size δ and the number of steps L.

NUTS (Homan and Gelman, 2014) can dynamically adjust the number of leapfrog steps L at each iteration to send the trajectory as far as it can go during that iteration. If such a rule is applied in isolation, the simulation will not converge to the desired target distribution. The full NUTS is more complicated, going backward and forward along the trajectory in a way that satisfies detailed balance.

Stan. The programming of NUTS is much more complicated than an M–H algorithm and even more complicated than HMC. We use Stan to implement the NUTS inferential engine. Along with this algorithm, Stan can automatically optimize δ to match an acceptance rate target and estimate Σ based on warm-up iterations. Hence, we do not need to specify any tuning parameters in Stan. Besides NUTS, Stan can also approximate Bayesian inference using variational Bayes and perform penalized maximum likelihood estimation if we specify the priors as the penalized term.

The key steps in Stan include data and model input, computation of the log posterior density (up to an arbitrary constant such as in distribution (9)) and its gradients, a warm-up phase in which the tuning parameters, δ and Σ, are set, an implementation of NUTS to move through the parameter space, convergence monitoring and sample summary statistics at the end.

Compared with BUGS, Stan works seamlessly with R. Stan can analyze all the examples in the BUGS manual. It provides more instructive error messages than BUGS, which is particularly helpful when we work with a "black box" inferential engine. Furthermore, Stan can perform parameter inference for the multilevel models with unknown covariance matrices, which BUGS can not easily deal with, and it is easier to specify the support of parameters in Stan. On the other hand, Stan does not accommodate all the likelihood functions; in contrast, BUGS can accommodate any distribution via the zero trick[2]. Random quantities such as the future claims cannot be simulated in Stan, while BUGS can update the parameters and simulate the future claims at the same time.

4. Example: Workers' compensation insurance data

Workers' compensation insurance covers the liability associated with work-related injuries. The benefits of workers' compensation include salary replacement, doctor cost, hospital cost, rehabilitation, etc. The duration of a claim payment depends on the attributes of the claim. For example, a lump sum death benefit may be paid to a surviving dependant, or a life-time annuities to a permanently injured worker. In aggregate, workers' compensation insurance is a long-tailed insurance.

The dataset analyzed in this section is workers' compensation claims data for the United States, which is available from the Casualty Actuarial Society website. The claims data is extracted from Schedule P-Analysis of Losses and Loss Expenses in the National Association of Insurance Commissioners (NAIC) database. Meyers (2015) described the dataset in detail. This dataset was also studied in Zhang *et al.* (2012) and Zhang and Dukic (2013). Our analysis uses five self-explanatory variables: GRCODE, AccidentYear, DevelopmentLag, CumPaidLoss and EarnedPremNet. The company being studied in this section is California Casualty Group with the GRCODE code of 337.

The claims data for the GRCODE 337 insurer is a rectangle containing 10 accident years, from 1988 to 1997, and 10 development years; see Table A2. We extract the upper triangle from the first eight accident years (1988 to 1995) as

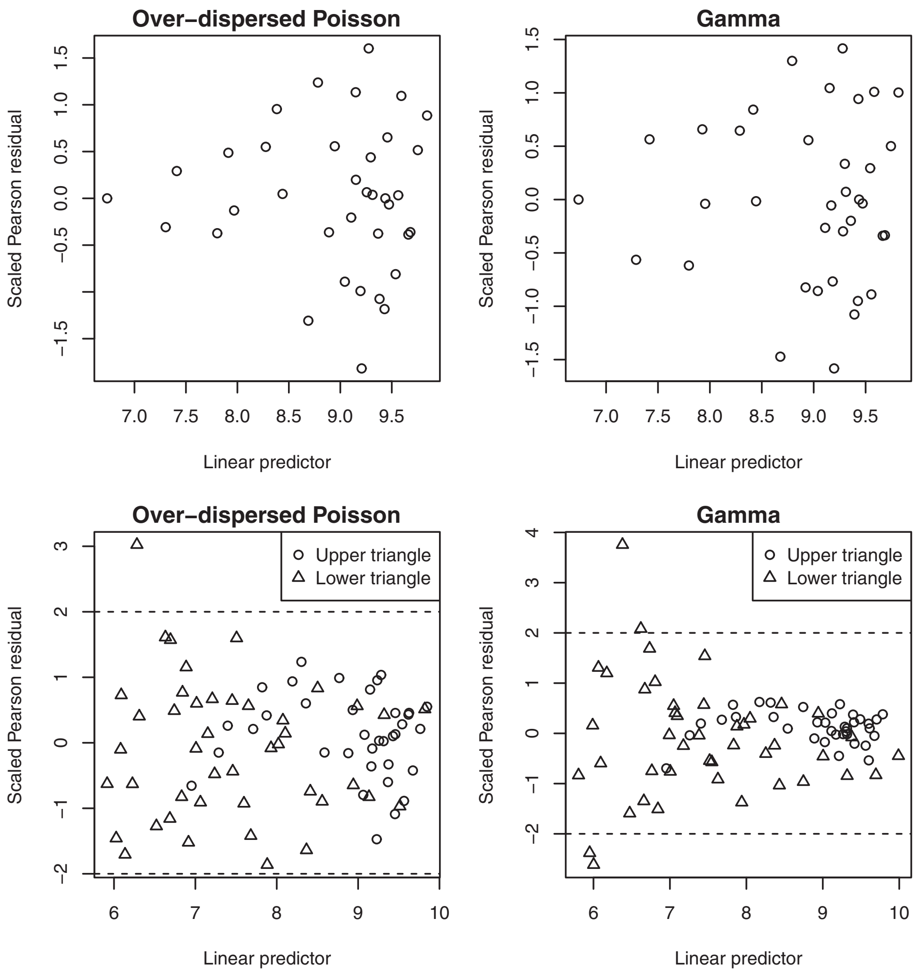

FIGURE 1: Top row: The scaled Pearson residual plots of frequentist GLMs fitted to the upper triangle (8×8) using an ODP distribution and a gamma distribution. Bottom row: The scaled Pearson residual plots of frequentist GLMs fitted to the rectangle (8×10) using an ODP distribution and a gamma distribution.

the training dataset and the remaining lower triangle and tail development as the out-of-sample validation dataset. The accident years 1996 and 1997 are used to illustrate how to predict the future claims for a new accident year. Obviously, there remain unpaid claims after 10 development years. For the purpose of out-of-sample model validation, we do not predict the claims after 10 development years. Hence, for this example, Model (8) has the indexes $I = 8$ and $J = 10$.

https://doi.org/10.1017/asb.2017.19 Published online by Cambridge University Press

4.1. Choice of error distribution

In comparison with a frequentist GLM, a disadvantage of Bayesian modeling is that we need to spend more time to simulate from the posterior distribution in order to assess the model. For a frequentist GLM, the diagnostic plots are available in seconds through most statistical softwares. To guide the choice of a suitable error distribution from Table A1, we do a preliminary analysis using a frequentist GLM. A frequentist quasi-GLM is fitted to the upper triangle (8×8) as follows:

$$\frac{y_{ij}}{\phi} \sim \text{Poi}\left(\frac{\mu_{ij}}{\phi}\right),$$
$$\log \mu_{ij} = \log P_i + a_i + b_j, \tag{15}$$

where the logarithm link function is assumed, $\log P_i$ is the offset term and the variance function equals to the mean.

The scaled Pearson residuals are defined as

$$e_{ij}^p = \frac{y_{ij} - \hat{y}_{ij}}{\sqrt{\hat{\phi}\hat{y}_{ij}}},$$

where $\hat{\phi} = 45.64$ (in thousands of dollars), and $\hat{y}_{ij}$ is the fitted value. If the error distribution is suitable, e_{ij}^p should display a constant spread over the fitted values. The residual plot is shown in the top-left of Figure 1. The funnel shape implies that the variance function should equal to the mean to a power greater than 1. Hence, another frequentist quasi-GLM with the same link function but with a variance function equal to mean squared is fitted. Surprisingly, there are no obvious changes in the new residual plot, as shown in the top-right of Figure 1. We then fit the full run-off rectangle (8×10) using the two error distributions: an ODP and a gamma. Two residual plots are shown at the bottom of Figure 1, from which it is now clear that the ODP distribution is more appropriate than the gamma distribution. In fact, the similar pattern of the top two residual plots in Figure 1 is due to the special triangle shape of the dataset. This also demonstrates the danger when we choose a proper error distribution according to the residual plot of the upper triangle.

4.2. Model inference

We use the plug-in estimate $\hat{\phi} = 45.64$ from the above frequentist GLM. In Stan, we simulate four chains, each with 10,000 iterations, which takes around 35 seconds.[3] The first half of these iterations are discarded as burn-in. For every 10 iterations in the second half, one is chosen as the inference sample. Hence, there are 2,000 iterations left, from which we make inference about the parameters and quantities of interest.

Table A3 in Appendix displays two statistics to judge the efficiency and convergence of HMC: the effective sample size (labelled as "n_eff") and the potential scale reduction factor (labelled as "Rhat"). The effective sample size is the adjusted sample size with regard to the autocorrelation among the sampled values. All the "n_eff"s are smaller than 2,000 and the deviation from 2,000 reflects the degree of autocorrelation among the sampled values. The potential scale reduction factor is derived by comparing the within-chain variability and the between-chain variability. A potential scale reduction factor close to 1 indicates the convergence. According to the last two columns in Table A3 in Appendix, HMC is converged after 5,000 iterations.

We would be confronted with no convergence of β and LR, if we replaced the normalized run-off pattern G with the raw run-off pattern F in Model (8). In this case, the posterior distribution of μ_{LR} and α_{LR} cannot be interpreted intuitively, so the prediction of ultimate claims in a new accident year would be problematic. Nevertheless, the means, μ_{ij}, would converge, and there would be no effect on the predictive distribution of future claims of old accident years.

Posterior distribution of parameters and quantities of interest. Table A3 in Appendix lists the posterior statistics of LR, β, μ_{LR}, α_{LR}, and σ_β^2. The "mean" column contains the estimated posterior means, calculated as follows:

$$\hat{\theta} := \hat{\mathbb{E}}(\theta|y^u) = \mathbb{E}_{P^*}(\theta) = \frac{\sum_{t=1}^{T}\theta^t}{T},$$

where a generic symbol θ is used to denote any parameter of interest, P^* denotes the empirical distribution of the inference sample, θ^t is the tth sampled value from the posterior distribution (9), and $T = 2,000$ is the inference sample size. The "se_mean" column contains the standard error of $\hat{\theta}$, calculated as follows:

$$\text{s.e.}(\hat{\theta}) = \frac{\text{s.e.}_{P^*}(\theta)}{\sqrt{T}} = \frac{\sum_{t=1}^{T}(\theta^t - \hat{\theta})^2}{\sqrt{T(T-1)}}.$$

The "sd" column contains the estimated standard error of parameters, calculated as follows:

$$\widehat{\text{s.e.}}(\theta|y^u) = \text{s.e.}_{P^*}(\theta) = \frac{\sum_{t=1}^{T}(\theta^t - \hat{\theta})^2}{\sqrt{T-1}}.$$

The next five columns contain the 2.5%, 25%, 50%, 75% and 97.5% quantiles of the inference sample. The 2.5% and 97.5% quantiles are the side points of the 95% central posterior density region (CPDR).

The posterior medians and the 50% CPDRs of loss ratios for each accident year are plotted in Figure 2. The realized loss ratios are all in the 50% CPDR except for the accident years 1994, 1996 and 1997. The spread becomes wider as accident years increase, since the evidence of data in recent accident years

https://doi.org/10.1017/asb.2017.19 Published online by Cambridge University Press

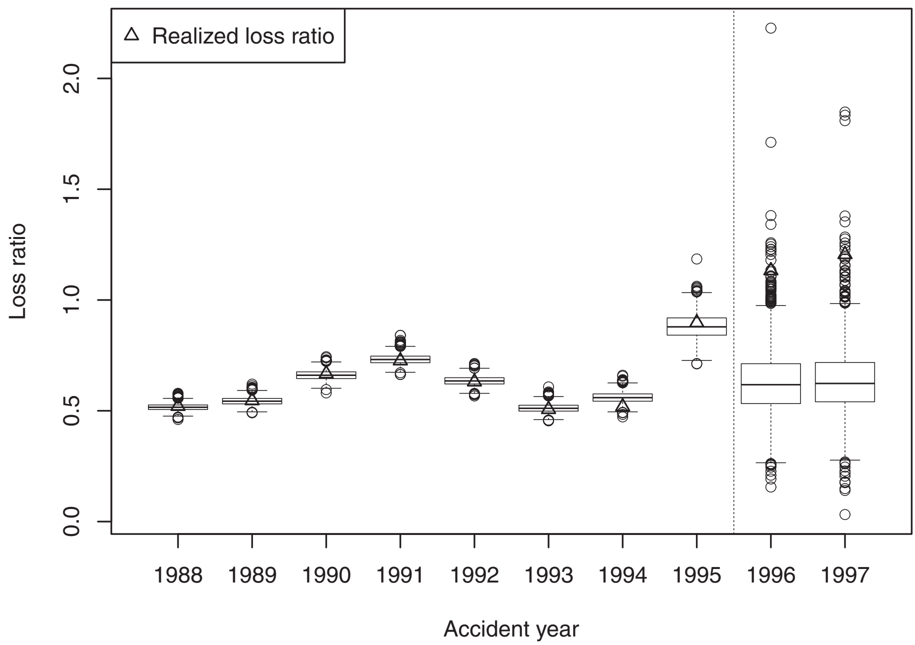

FIGURE 2: The box plot of posterior loss ratio for each accident year, compared with the realized loss ratios.

is weaker than that in early accident years. Note that there are no historical claims in the new accident years 1996 and 1997. The empirical distribution of new accident years loss ratio is obtained from the simulation (11).

The summary statistics for β do not tell us much about the development pattern. We are more interested in the shape of G. For each sampled value β^t in the inference sample, we can generate a set of $G^t(j)$, $j = 1, \dots, J$ as follows:

$$G^t(j) = \frac{\exp\left(\sum_{h=1}^{I=8} \beta_h^t b_h(j)\right)}{\sum_{l=1}^{J=10} \exp\left(\sum_{h=1}^{I=8} \beta_h^t b_h(l)\right)}.$$

Note that the generated $G^t(j)$ satisfies the normality condition, i.e., $\sum_{l=1}^{J=10} G^t(l) = 1$ for $t = 1, \dots, T$. Upon the generated $G^t(j)$, we can get the posterior mean and the 95% CPDR of G, which is shown in Table A4. The wider interval in the tail development is statistically desirable since there is no data to infer this period. However, from the perspective of actuaries, the variability in the tail development should decrease since there are fewer outstanding claims. We might assume a strong prior for the tail development by observing the trend of the run-off pattern.

Table A4 also lists the run-off patterns from Mack's model and Clark's curve model. Both models estimate an "aggregated" tail factor. We distribute the proportion of tail evenly to the development years 9 and 10. The Bayesian spline model and Mack's model have comparable run-off patterns, while Clark's curve model has a much larger tail factor. A Weibull growth curve is used in Clark's curve model, whose fit is much better than a log-logistic growth curve.

Predictive distribution of future claims of the accident years 1988 to 1995. For each sampled vector, (β^t, LR^t), in the inference sample, a set of means are calculated by the equation, $\mu_{ij}^t = P_i LR_i^t G^t(j)$. Then a set of future claims are simulated from $\hat{\phi}\text{Poi}(\mu_{ij}^t/\hat{\phi})$. As the support of a Poisson distribution is given by the set of non-negative integers, only multiplied values of $\hat{\phi}$ will be generated. Alternatively, we can use a gamma distribution with the same mean and variance as the ODP distribution to simulate the future claims:

$$y_{ij}^t \sim \text{Gamma}\left(\frac{P_i LR_i^t G^t(j)}{\hat{\phi}}, \frac{1}{\hat{\phi}}\right). \tag{16}$$

We will assess the accuracy of the gamma approximation of ODP distribution in Section 4.3.

Figure A1 in Appendix shows the posterior means and the 95% CPDRs of the future incremental claims. The 95% CPDRs cover the most realized points. Later, we will calculate the coverage rate of the 95% CPDRs as a measure of predictive accuracy. Table A5 compares the predicted unpaid claims from the Bayesian spline model, Mack's model and Clark's curve model. The posterior mean of total unpaid claims is estimated as 133,175 thousand dollars compared with the estimate of 134,752 thousand dollars in Mack's model (with tail factor) and 144,702 thousand dollars in Clark's model. The standard error of the total unpaid claims is estimated as 8,116 thousand dollars compared with 5,547 thousand dollars in Mack's model and 8,270 thousand dollars in Clark's model. The 95% CPDR of total unpaid claims is $(119,797,\ 151,610)$ covering the realized total unpaid claims, 129,374 thousand dollars.

In Mack's model, the standard errors for the predicted unpaid claims in the accident years 1988 to 1992 are significantly less than those estimated from the other two models. This is due to the deliberately chosen small standard error of the tail factor in Mack's model as shown in (1). In the other two models, the tail factor cannot be estimated accurately (i.e., the standard error of tail factor is large) from the upper triangle, which is remote from the tail development.

Predictive distribution of future claims of the accident years 1996 and 1997. The earned premiums of the accident years 1996 and 1997 are 60,244 and 45,933 thousand dollars. Applying the method discussed in Section 3.2, we get the estimated posterior mean of ultimate claims for the accident year 1996 as 37,980 thousand dollars with the standard error of 9,554 thousand dollars and the 95%CPDR of $(21,195,\ 58,181)$, and the estimated posterior mean of ultimate claims for the accident year 1997 as 29,325 thousand dollars with the standard error of 7,393 thousand dollars and the 95% CPDR of $(16,709,\ 45,351)$.

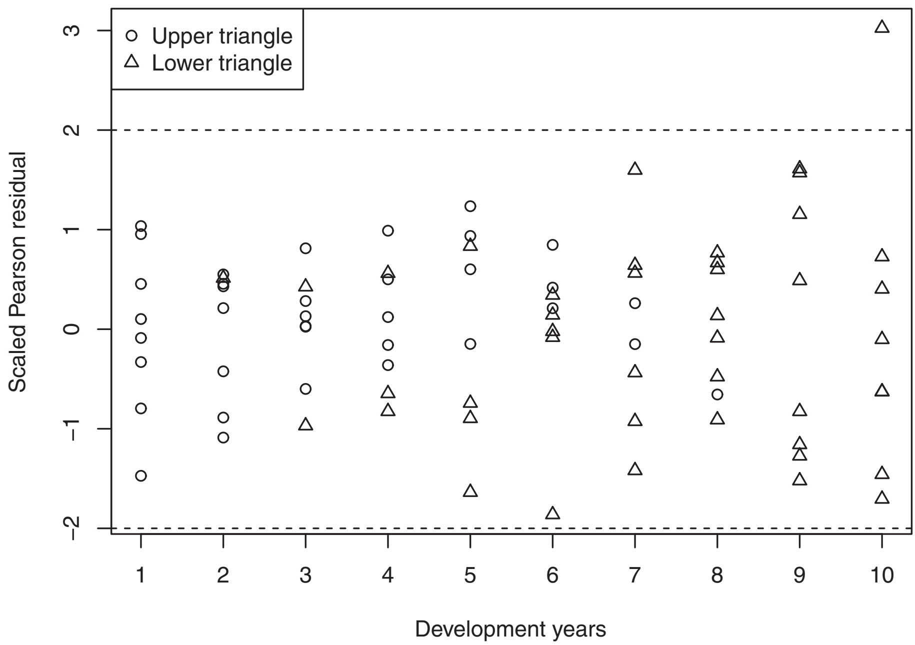

FIGURE 3: The scaled Pearson residual plots of a frequentist GLM fitted to the rectangle (8 × 10) using an ODP distribution.

The realized ultimate claims of the accident years 1996 and 1997 are 68,225 and 55,377 thousand dollars. The model underestimates the ultimate claims significantly, which is due to the unusually high loss ratios in the accident years 1996 and 1997. Special knowledge of the U.S. insurance market for that decade and actuarial judgment are required to get a reasonable informative prior for the loss ratios in the accident years 1996 and 1997.

4.3. Model assessment

In this section, we check three crucial assumptions made the constant dispersion $\hat{\phi}$, the ODP error distribution and the gamma approximation in (16).

Constant dispersion. It is worth checking the constant dispersion assumption as the true dispersion might vary across development years. We re-draw the third plot in Figure 1 with the development years as the x axis. The constant spread shown in Figure 3 indicates the validation of the constant dispersion assumption. Be aware that this plot may be distorted when a triangle dataset is used, as a triangle dataset has more data points in the earlier development years.

https://doi.org/10.1017/asb.2017.19 Published online by Cambridge University Press

Error distribution. The posterior means of scaled Pearson residuals are defined as

$$\bar{e}^p_{ij} = \frac{1}{T}\sum_{t=1}^{T} \frac{y_{ij} - \mu^t_{ij}}{\sqrt{\hat{\phi}\mu^t_{ij}}}.$$

The plot of the posterior means of scaled Pearson residuals against the logarithm of the posterior mean of fitted values is similar to the first plot in Figure 1. This shows an increasing spread pattern; as discussed above, this funnel shape is due to the triangle structure of the dataset and does not imply a higher order variance function. If we randomly selected cells from the rectangle to form a training set, the residual plot would not be misleading. In this example, the first two development years' claims amount to nearly 40% of the ultimate claims (see Table A4), so there tend to be a greater number of large response values in the upper triangle (corresponding to the earlier development years). The seemingly lesser spread in the area of small linear predictors is due to the smaller number of data points in this area (corresponding to the later development years).

The lower triangle and the tail development are also available, so we can compare the predicted claims with the realized claims to assess the predictive accuracy of the model. The coverage rate of the 95% CPDRs is defined as

$$\frac{1}{|y^l \cup y^t|} \sum_{y_{ij} \in y^l \cup y^t} I_{\mathrm{CPDR}_{ij}}(y_{ij}),$$

where $|A|$ is the cardinality of the set A, $I_\Omega(x) = 1$ if $x \in \Omega$ and $I_\Omega(x) = 0$ if $x \notin \Omega$, and CPDR_{ij} is the 95% CPDR for the (i, j)th cell. The coverage rate is calculated as 95.45% when the ODP distribution is assumed, and it reduces to 52.27% when the gamma distribution is assumed (discussed later).

Another measure of the model fitness on the validation set is the posterior empirical quantile values, defined as

$$q_{ij} = F_{P'_{ij}}(y_{ij}) = \Pr_{P'_{ij}}(Y_{ij} < y_{ij}),\ y_{ij} \in y^l \cup y^t,$$

where P'_{ij} denotes the empirical distribution of the future claim sample $\{y^t_{ij} : t = 1, \ldots, T\}$. If the specified model is correct, q_{ij} should be approximately uniformly distributed. We plot the histogram and the quantile-to-quantile plot of q_{ij} in the top row in Figure 4, neither of which indicate any deviation from the uniform distribution. The Kolmogorov–Smirnov (K–S) test of q_{ij} returns a p-value of 0.4629, revalidating the model assumption.

We compare the ODP error distribution with a gamma error distribution, i.e., $\rho = 2$ in Model (8). The dispersion parameter in the gamma distribution is fixed at $\hat{\phi} = 1/\hat{\alpha} = 0.005128$, which is estimated in a frequentist GLM similar to Model (15) but with the gamma error distribution. The coverage rate of the CPDRs is calculated as 52.27% and the p-value in the K–S test as 7×10^{-14},

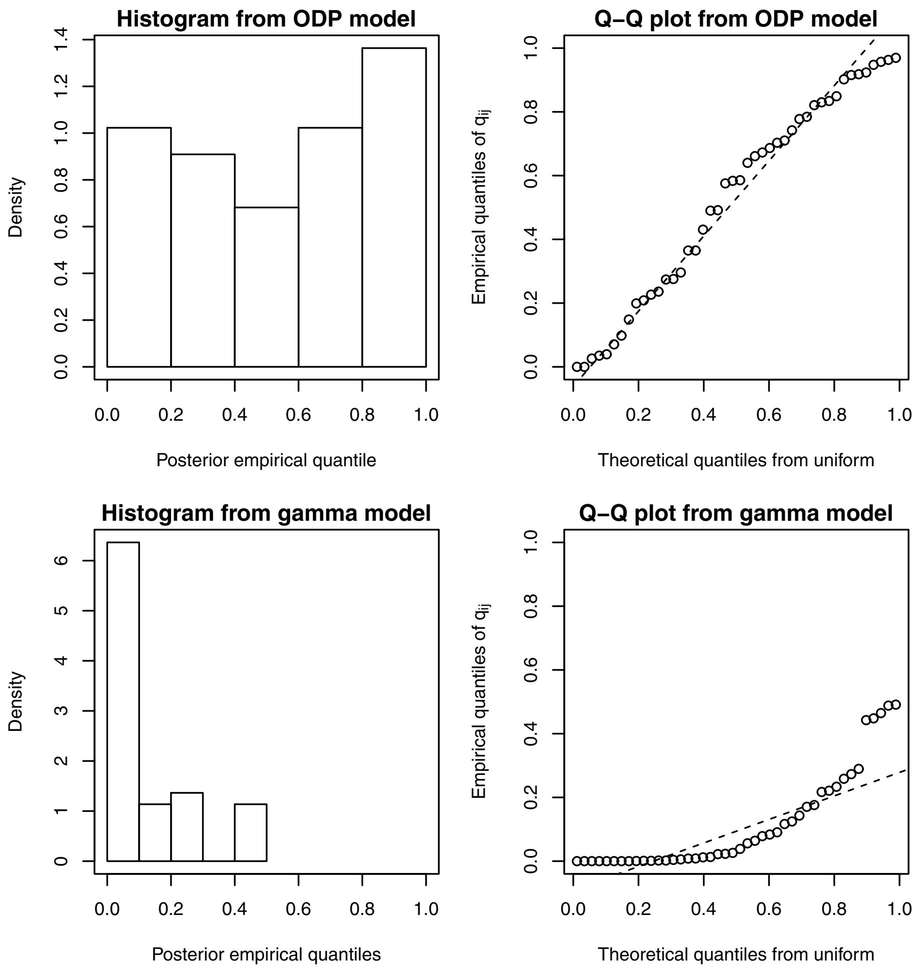

FIGURE 4: Top row: The histogram and q-q plot of the posterior empirical quantiles q_{ij}, obtained from the Bayesian spline model (8) with an ODP error distribution. Bottom row: The histogram and q-q plot of the posterior empirical quantiles q_{ij}, obtained from the Bayesian spline model (8) with a gamma error distribution.

both of which indicate that the gamma error distribution is not suitable for this dataset. The conclusion is reconfirmed in the bottom row in Figure 4 which displays the histogram and the q-q plot of q_{ij}. The values of q_{ij} are concentrated at zero, which implies that the model with the gamma error distribution tends to underestimate the future claims.

While the coverage rate and the posterior empirical quantiles discussed previously measure the predictive accuracy of the validation set, the information criteria can be viewed as the fitness of the model on the training set penalized by the model complexity. We calculate several information criteria commonly used in the Bayesian analysis, including leave-one-out information criterion (LOOIC), Watanabe–Akaike or widely available information criterion (WAIC) and deviance information criterion (DIC), all of which are in the form of a minus two log likelihood, plus double the number of effective parameters. Another predictive accuracy measure is the expected log predictive density for a new data point $\tilde{y}_{ij}$, defined as

$$\text{elpd} := \mathbb{E}_f\left(\log p(\tilde{y}_{ij}|y^u_{ij})\right) = \int \log p(\tilde{y}_{ij}|y^u_{ij}) f(\tilde{y}_{ij}) d\tilde{y}_{ij},$$

where $p(\tilde{y}_{ij}|y^u_{ij})$ is the predictive distribution (10), and f is the unknown data distribution. Applying the methods used in LOOIC, WAIC and DIC, we can estimate elpd as elpd_{loo}, $\text{elpd}_{\text{waic}}$ and elpd_{dic}.

The Bayesian model complexity is indexed by the number of effective parameters including p_{loo}, p_{waic} and p_{dic}, which correspond to LOOIC, WAIC and DIC, respectively. For more detailed discussion on the information criteria, the elpd, and the number of effective parameters, see Section 7.1 in Gelman *et al.* (2014). Table A6 compares the two models with the ODP error distribution and the gamma error distribution in terms of the predictive accuracy and the model complexity discussed above. While the complexities of the two models are comparable indicated by p_{loo}, p_{waic} and p_{dic}, all the information criteria and the estimated elpds prefer the model with the ODP error distribution.

Gamma approximation. Finally, we assess the approximation of an ODP distribution by a gamma distribution as in (16). The quantiles from an ODP distribution are compared with the corresponding quantiles from a gamma approximation as follows:

$$Q_{\text{ODP}(\hat{\phi},\mu)}(q) \text{ vs. } Q_{\text{Gamma}(\mu/\hat{\phi},1/\hat{\phi})}(q), \tag{17}$$

where $Q_D(q)$ denotes the q quantile from the distribution D and $\hat{\phi} = 45.64$. Considering the range of estimated mean parameters in this example, we choose $\mu = 1,000$ and 20, 000 for illustration. The quantile-to-quantile plot in Figure 5 compares 99 uniformly placed quantiles (i.e., 0.01, 0.02,..., 0.99 quantiles) from the two distributions. Both the plots show a roughly straight line, which indicates that the two distributions are close even in tails. Hence, the approximation of an ODP distribution by a gamma distribution is suitable for the purpose of mean, variance and tail-based risk measures. In fact, both distributions converge to normal distribution as $\mu \to \infty$.

5. Conclusions

One contribution of this paper is to introduce a Bayesian spline model to the claims reserving problem. The proposed model can address two problems that confront the chain ladder model: the inclusion of a tail factor and the prediction

https://doi.org/10.1017/asb.2017.19 Published online by Cambridge University Press

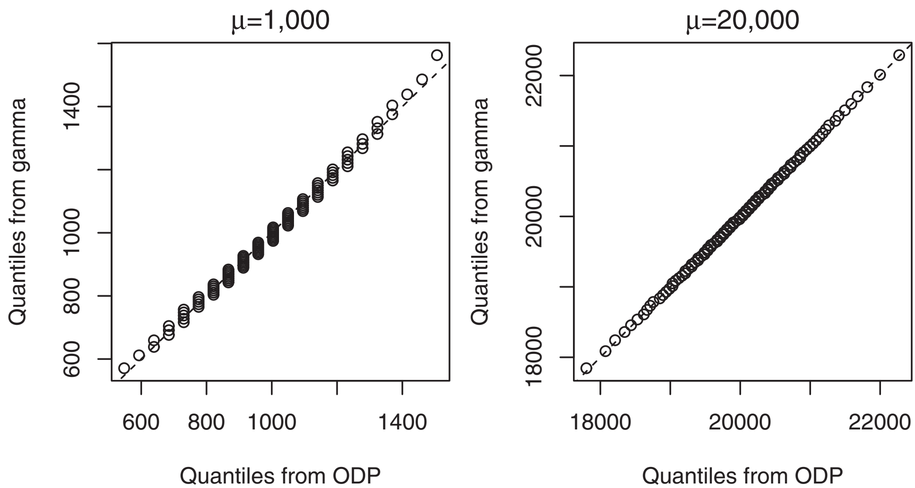

FIGURE 5: Left: The quantile-to-quantile plot comparing an ODP distribution with a gamma distribution. The parameters in (17) are $\hat{\phi} = 45.64$, $\mu = 1,000$. Right: The quantile-to-quantile plot comparing an ODP distribution with a gamma distribution. The parameters in (17) are $\hat{\phi} = 45.64$, $\mu = 20,000$.

for a new accident period. While Mack's model with tail factor performs equivalently for the total unpaid claims estimation in the analyzed example, our model has stronger predictive power in terms of quantifying the predictive distribution for the individual future claims. Due to the orthogonality of the *B*-spline basis, the time consumed in the Bayesian computational machinery for our model is much less than some Bayesian reserving models, such as the non-linear Bayesian model in Zhang *et al.* (2012) and the mixed Bayesian model in Verrall *et al.* (2012).

Another contribution of this paper is to apply HMC to make inferences about a Bayesian model. The programming of HMC involves the calculation of gradients of the posterior distribution, which is not necessary in MCMC. We use Stan to implement the NUTS, an automatically tuned HMC. Stan relieves the burden of programming HMC. However, when relying on Stan, we are restricted to its built-in functions and distributions. The ODP distribution is not a built-in distribution, so we work on the Poisson distribution assuming the dispersion parameter is fixed at a plug-in estimate. With the further development of Stan, we expect it will extend its range of distributions and functions.

The detailed example illustrates the danger of selecting an error distribution according to the residual plot, which is a routine model assessment practice. The residual plot is misleading due to the triangle structure of the dataset. From the prediction for new accident years, we can see that the actuarial judgment is important for getting a reasonable result since actuaries may have other important information not contained in the dataset.

Finally, this example shows a typical Bayesian claims reserving procedure:

1. Start from a simple model, such as a frequentist GLM, which is easy to fit. Select a suitable mean function and an appropriate error distribution according to the diagnostic plots and statistics.
2. Based on the simple model, construct the Bayesian model with particular focus on the parameterization.
3. Apply a Bayesian computational machinery to sample from the posterior distribution and simulate other quantities of interest.
4. Compare with alternative models in terms of the information criteria or the predictive accuracy on the out-of-sample dataset if available.

During this procedure, it may be necessary to go back to step 2 from time to time. For example, an absence of convergence in step 3 may be due to inadequate parameterization. Alternatively, a better model may be identified in step 4.

Acknowledgements

The authors appreciate the constructive comments made by two anonymous referees and Mario V. Wüthrich. Their comments helped improve the quality of this paper significantly. We acknowledge the financial support from National Social Science Fund of China (Grant No. 16ZDA052) and MOE National Key Research Bases for Humanities and Social Sciences (Grant No. 16JJD910001).

Notes

1. The knots are placed at selective development periods.
2. Based on a Poisson probability density evaluated at zero, see Lunn *et al.* (2000).
3. On a Mac of 4 GB 1,600 MHz DDR3 and 1.3 GHz Intel Core i5.

References

ANTONIO, K. and BEIRLANT, J. (2008) Issues in claims reserving and credibility: A semiparametric approach with mixed models. *Journal of Risk and Insurance*, **75**, 643–676.

ARMAGAN, A., DUNSON, D.B. and LEE, J. (2013) Generalized double Pareto shrinkage. *Statistica Sinica*, **23**, 119–143.

BETANCOURT, M. (2017) A conceptual introduction to Hamiltonian Monte Carlo. Preprint, arXiv:1701.02434.

BISHOP, C.M. (2006) *Pattern Recognition and Machine Learning*. New York: Springer.

BORNHÜTTER, R.L. and FERGUSON, R. (1972) The actuary and IBNR. *Proceedings of the Casualty Actuarial Society*, **59**, 181–195.

CLARK, D.R. (2003) LDF curve-fitting and stochastic reserving: A maximum likelihood approach. *Casualty Actuarial Society Forum* (Fall), 41–92.

CRAINICEANU, C.M., RUPPERT, D. and WAND, M.P. (2005) Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software*, **14**, 1–14.

DUANE, S., KENNEDY, A.D., PENDLETON, B.J. and ROWETH, D. (1987) Hybrid Monte Carlo. *Physics Letters B*, **195**, 216–222.

ENGLAND, P.D. and VERRALL, R.J. (2001) A flexible framework for stochastic claims reserving. *Proceedings of the Casualty Actuarial Society*, **88**, 1–38.

FRIEDLAND, J. (2010) Estimating unpaid claims using basic techniques. *Casualty Actuarial Society Study Notes*. Available at http://www.casact.org/library/studynotes/Friedland_estimating.pdf.
GELMAN, A., CARLIN, J.B., STERN, H.S. and RUBIN, D.B. (2014) *Bayesian Data Analysis*, 3rd edition. Boca Raton: Chapman & Hall.
GELMAN, A., LEE, D. and GUO, J. (2015) Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, **40**, 530–543.
HASTIE, T.J. and TIBSHIRANI, R.J. (1990) *Generalized Additive Models*. New York: Chapman & Hall.
HASTIE, T.J., TIBSHIRANI, R.J. and FRIEDMAN, J.H. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition. New York: Springer.
HASTINGS, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.
HOMAN, M.D. and GELMAN, A. (2014) The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, **15**, 1593–1623.
JAMES, G., WITTEN, D., HASTIE, T. and TIBSHIRANI, R. (2013) *An Introduction to Statistical Learning*. New York: Springer.
LUNN, D.J., THOMAS, A., BEST, N. and SPIEGELHALTER, D. (2000) WinBUGS — a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–337.
MACK, T. (1993) Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, **23**, 213–225.
MACK, T. (1999) The standard error of chain-ladder reserve estimates, recursive calculation and inclusion of a tail factor. *ASTIN Bulletin*, **29**, 361–366.
METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N., TELLER, A.H. and TELLER, E. (1953) Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1092.
MEYERS, G. (2015) Stochastic loss reserving using Bayesian MCMC models. *CAS Monograph Series*, **1**, 1–64.
NEAL, R.M. (1994) An improved acceptance procedure for the hybrid Monte Carlo algorithm. *Journal of Computational Physics*, **111**, 194–203.
NEAL, R.M. (2011) MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo* (eds. S. Brooks, A. Gelman, G. Jones, and X.L. Meng), pp. 113-160. Boca Raton, FL: CRC Press.
PARK, T. and CASELLA, G. (2008) The Bayesian lasso. *Journal of the American Statistical Association*, **103**, 681–686.
ROBERTS, G.O., GELMAN, A. and GILKS, W.R. (1997) Weak convergence and optimal scaling of random walk metropolis algorithms. *Annals of Applied Probability*, **7**, 110–120.
RUPPERT, D., WAND, M.P. and CARROLL, R.J. (2003) *Semiparametric Regression*. New York: Cambridge University Press.
Stan Development Team (2016) RStan: The R interface to Stan. http://mc-stan.org.
TAYLOR, G. (2000) *Loss Reserving: An Actuarial Perspective*. Huebner International Series on Risk, Insurance and Economic Security. Boston: Kluwer Academic Publishers.
VERRALL, R., HÖSSJER, O. and BJÖRKWALL, S. (2012) Modelling claims run-off with reversible jump Markov chain Monte Carlo methods. *ASTIN Bulletin*, **42**, 35–58.
VERRALL, R.J. (1996) Claims reserving and generalised additive models. *Insurance: Mathematics and Economics*, **19**, 31–43.
VERRALL, R.J. and WÜTHRICH, M.V. (2012) Reversible jump Markov chain Monte Carlo method for parameter reduction in claims reserving. *North American Actuarial Journal*, **16**, 240–259.
WOOD, S. (2006) *Generalized Additive Models: An Introduction with R*. New York: Chapman & Hall.
WRIGHT, T.S. (1990) A stochastic method for claims reserving in general insurance. *Journal of the Institute of Actuaries*, **117**, 677–731.
WÜTHRICH, M.V. and MERZ, M. (2008) *Stochastic Claims Reserving Methods in Insurance*. Chichester: John Wiley & Sons.

WÜTHRICH, M.V. and MERZ, M. (2015) Stochastic claims reserving manual: Advances in dynamic modeling. SSRN Manuscript ID 2649057.
ZHANG, Y., DUKIC, V. and GUSZCZA, J. (2012) A Bayesian non-linear model for forecasting insurance loss payments. *Journal of the Royal Statistical Society A*, **175**, 637–656.
ZHANG, Y.W. and DUKIC, V. (2013) Predicting multivariate insurance loss payments under the Bayesian copula framework. *Journal of Risk and Insurance*, **80**, 891–919.

GUANGYUAN GAO (Corresponding author)
Renmin University of China
Center for Applied Statistics and School of Statistics
Room 1006, Mingde Building
59 Zhongguancun Avenue, Haidian District
Beijing, 100872, China
E-Mail: guangyuan.gao@anu.edu.au

SHENGWANG MENG
Renmin University of China
Center for Applied Statistics and School of Statistics
59 Zhongguancun Avenue, Haidian District
Beijing, 100872, China
E-Mail: mengshw@ruc.edu.cn

APPENDIX

TABLE A1

FOUR ERROR DISTRIBUTIONS FOR THE INCREMENTAL CLAIMS: NORMAL, OVER-DISPERSED POISSON, GAMMA AND INVERSE GAUSSIAN. ρ IS THE POWER OF MEAN IN THE VARIANCE FUNCTION V. THE DISPERSION PARAMETER IS DENOTED BY ϕ.

ρ	Distribution	ϕ	$\mathrm{Var}(y_{ij}) = \phi V(\mu_{ij})$
0	$y_{ij} \sim \mathrm{N}(\mu_{ij}, \sigma^2)$	σ^2	σ^2
1	$y_{ij}/\phi \sim \mathrm{Poi}(\mu_{ij}/\phi)$	ϕ	$\phi\mu_{ij}$
2	$y_{ij} \sim \mathrm{Gamma}(\alpha, \alpha/\mu_{ij})$	$1/\alpha$	μ_{ij}^2/α
3	$y_{ij} \sim \mathrm{Inv\text{-}N}(\mu_{ij}, \lambda)$	$1/\lambda$	μ_{ij}^3/λ

https://doi.org/10.1017/asb.2017.19 Published online by Cambridge University Press

TABLE A2

THE INCREMENTAL CLAIMS FOR THE GRCODE 337 INSURER. THE UPPER TRIANGLE FROM ACCIDENT YEAR 1988 TO 1995 IS EXTRACTED AS THE TRAINING DATASET. THE DATA IN THE SHADOW IS USED FOR THE OUT-OF-SAMPLE VALIDATION. THE EARNED PREMIUMS ARE ASSUMED KNOWN. ALL THE NUMBERS ARE IN THOUSANDS OF DOLLARS.

Accident	Development Year										
Year	1	2	3	4	5	6	7	8	9	10	Premium
1988	9,558	13,220	10,520	7,050	4,798	2,902	1,734	841	1,189	127	99,779
1989	7,913	11,559	10,150	7,194	4,159	2,327	1,405	1,164	358	254	85,110
1990	8,744	15,558	11,104	8,006	4,645	2,840	1,982	1,077	484	417	82,187
1991	13,301	19,649	14,251	9,193	5,256	3,389	1,527	1,217	540	642	94,997
1992	11,424	17,662	12,948	8,876	5,496	3,031	1,592	1,325	683	369	100,508
1993	11,792	15,369	11,068	8,493	4,020	2,738	2,480	866	984	107	114,352
1994	11,194	15,699	11,595	7,092	3,256	1,723	1,560	1,307	1,240	589	106,540
1995	12,550	19,054	12,441	8,494	4,583	3,404	2,356	1,588	1,329	1,212	74,652
1996	13,194	18,280	12,596	7,623	5,427	3,333	3,046	2,706	1,218	802	60,244
1997	9,372	14,363	10,456	5,535	4,959	3,753	2,337	1,919	1,523	1,160	45,933

TABLE A3

INFERENCE FOR PARAMETERS IN THE BAYESIAN SPLINE MODEL 8: 4 CHAINS, EACH WITH 10,000 ITERATIONS; WARMUP=5,000; THIN=10; POST-WARMUP DRAWS PER CHAIN=500, TOTAL POST-WARMUP DRAWS=2,000.

Parameters	Mean	se_mean	sd	2.50%	25%	50%	75%	97.50%	n_eff	Rhat
β_1	1.153	0.007	0.306	0.477	0.974	1.182	1.342	1.710	1727	1.000
β_2	0.796	0.009	0.372	−0.019	0.580	0.830	1.026	1.484	1700	1.000
β_3	0.508	0.009	0.375	−0.306	0.287	0.537	0.750	1.206	1749	1.001
β_4	−0.064	0.009	0.378	−0.898	−0.284	−0.027	0.160	0.646	1713	1.000
β_5	−0.518	0.009	0.396	−1.377	−0.757	−0.488	−0.260	0.194	1767	1.001
β_6	−1.295	0.007	0.304	−1.931	−1.490	−1.286	−1.091	−0.701	1765	1.001
β_7	−0.408	0.020	0.818	−2.247	−0.877	−0.325	0.090	1.115	1715	1.000
β_8	−2.382	0.006	0.241	−2.877	−2.532	−2.375	−2.217	−1.920	1951	1.000
LR_1	0.517	0.000	0.016	0.486	0.506	0.517	0.528	0.550	2000	1.000
LR_2	0.545	0.000	0.019	0.508	0.532	0.545	0.557	0.585	1998	0.999
LR_3	0.662	0.001	0.022	0.621	0.647	0.661	0.676	0.708	1866	1.001
LR_4	0.733	0.001	0.023	0.690	0.716	0.732	0.747	0.778	1845	1.000
LR_5	0.635	0.001	0.022	0.595	0.620	0.634	0.649	0.680	1919	1.000
LR_6	0.513	0.000	0.020	0.476	0.499	0.512	0.525	0.553	2000	0.999
LR_7	0.561	0.001	0.026	0.513	0.542	0.561	0.578	0.612	2000	0.999
LR_8	0.879	0.001	0.059	0.767	0.838	0.877	0.918	0.999	2000	0.999
μ_{LR}	0.634	0.001	0.057	0.534	0.598	0.631	0.664	0.749	1900	1.000
α_{LR}	27.638	0.382	16.520	6.158	16.061	24.076	35.647	68.291	1874	1.003
σ_β	1.099	0.011	0.490	0.501	0.773	0.988	1.295	2.305	2000	1.000

TABLE A4

COMPARISON OF RUN-OFF PATTERNS FROM THE BAYESIAN SPLINE MODEL, MACK'S MODEL AND CLARK'S CURVE MODEL. FOR MACK'S MODEL AND CLARK'S CURVE MODEL, AN "AGGREGATED" TAIL FACTOR IS ESTIMATED, AND WE DISTRIBUTE THE PROPORTION OF TAIL TO THE DEVELOPMENT YEARS 9 AND 10.

Dev.	Bayesian Spline Model					
Year	Mean	sd	2.50%	97.50%	Mack Mean	Clark Mean
1	18.42%	0.50%	17.44%	19.38%	18.35%	17.21%
2	26.92%	0.62%	25.66%	28.09%	26.99%	27.65%
3	20.33%	0.55%	19.20%	21.35%	20.39%	18.91%
4	14.09%	0.48%	13.13%	15.02%	14.13%	12.61%
5	8.49%	0.43%	7.68%	9.32%	8.51%	8.31%
6	5.34%	0.39%	4.60%	6.12%	5.31%	5.43%
7	3.30%	0.41%	2.52%	4.17%	3.22%	3.53%
8	1.68%	0.36%	1.03%	2.44%	1.64%	2.28%
9(Tail)	0.90%	0.50%	0.25%	2.16%	0.73%	2.04%
10(Tail)	0.54%	0.54%	0.05%	2.00%	0.73%	2.04%
Sum	100.00%				100.00%	100.00%

TABLE A5

COMPARISON OF THE PREDICTED UNPAID CLAIMS FROM THE BAYESIAN SPLINE MODEL, MACK'S MODEL AND CLARK'S CURVE MODEL.

Acc.	Bayesian Spline Model				Mack's Model		Clark's Model		
Year	Mean	sd	2.50%	97.50%	Mean	sd	Mean	sd	Obs.
1988	751	585	112	2,278	750	74	1,273	368	1,316
1989	1,456	713	508	3,237	1,429	105	2,241	496	1,776
1990	3,503	836	2,238	5,487	3,432	199	4,643	757	3,960
1991	8,204	1,204	6,246	10,935	8,114	350	9,938	1,202	7,315
1992	12,916	1,299	10,580	15,760	12,839	778	14,637	1,486	12,496
1993	20,121	1,523	17,435	23,353	19,932	1,183	20,726	1,829	19,688
1994	32,627	2,135	28,679	37,096	32,421	1,953	32,064	2,567	28,362
1995	53,598	4,040	45,856	61,509	55,835	4,542	59,179	5,450	54,461
Sum	133,175	8,116	119,797	151,610	134,752	5,547	144,702	8,270	129,374

TABLE A6

COMPARISON OF TWO MODELS WITH AN OVER-DISPERSED POISSON ERROR DISTRIBUTION AND A GAMMA ERROR DISTRIBUTION. THE PREDICTIVE ACCURACY MEASURES INCLUDE COVERAGE RATE OF THE 95% CPDRS, p-VALUE IN THE K–S TEST, AND EXPECTED LOG PREDICTIVE DENSITY (ELPD). THE INFORMATION CRITERIA INCLUDE LEAVE-ONE-OUT INFORMATION CRITERION (LOOIC), WATANABE–AKAIKE OR WIDELY AVAILABLE INFORMATION CRITERION (WAIC), AND DEVIANCE INFORMATION CRITERION (DIC). THE MODEL COMPLEXITY IS INDEXED BY THE NUMBER OF EFFECTIVE PARAMETERS, p.

Measures	Poisson	Gamma
Coverage Rate	95.45%	52.27%
p-value	0.4629	$< 10^{-13}$
elpd_{loo}	−153.6	−288.1
$\text{elpd}_{\text{waic}}$	−151.0	−285.6
elpd_{dic}	−150.9	−285.6
LOOIC	307.2	576.2
WAIC	302.0	571.1
DIC	301.8	571.1
p_{loo}	14.1	14.1
p_{waic}	11.7	11.6
p_{dic}	14.9	14.7

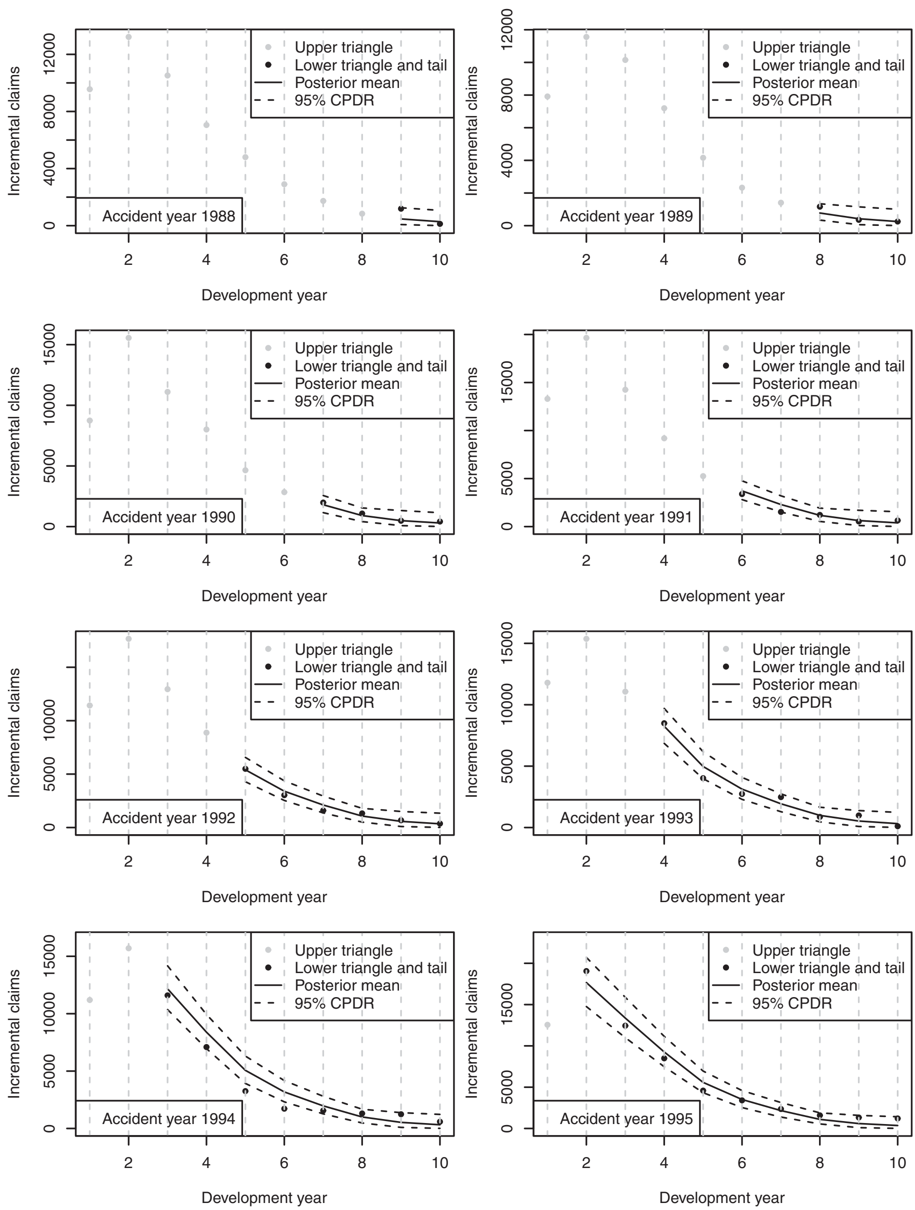

FIGURE A1: The posterior means and the 95% CPDRs of incremental claims in the lower triangle and the tail development. Each plot represents the incremental run-off pattern of an accident year.

https://doi.org/10.1017/asb.2017.19 Published online by Cambridge University Press