A GAMMA MOVING AVERAGE PROCESS FOR MODELLING DEPENDENCE ACROSS DEVELOPMENT YEARS IN RUN-OFF TRIANGLES

BY

LUIS E. NIETO-BARAJAS AND RODRIGO S. TARGINO

ABSTRACT

We propose a stochastic model for claims reserving that captures dependence along development years within a single triangle. This dependence is based on a gamma process with a moving average form of order $p \geq 0$ which is achieved through the use of poisson latent variables. We carry out Bayesian inference on model parameters and borrow strength across several triangles, coming from different lines of businesses or companies, through the use of hierarchical priors. We carry out a simulation study as well as a real data analysis. Results show that reserve estimates, for the real data set studied, are more accurate with our gamma dependence model as compared to the benchmark over-dispersed poisson that assumes independence.

KEYWORDS

Claims reserving, gamma process, Incurred But Not Reported, latent variables, moving average processes.

JEL codes: C110, C130, C320, C530

1. INTRODUCTION

One of the largest liabilities of an insurance company is its future claims; therefore, estimation of adequate reserves for outstanding claims is one of the main activities of actuaries in insurance and a major topic in actuarial science. The need to estimate future claims, given available information about the past, has led to the development of many loss reserving models.

In this paper, we study the problem of claims reserving using several run-off triangles, each of them coming from different lines of business or from different companies. For each triangle, we propose a stochastic dependence model of moving average form of order $p \geq 0$. Furthermore, we pull strength across lines

Astin Bulletin 51(1), 245-266. doi:10.1017/asb.2020.36

of business/companies by considering a hierarchical prior under a Bayesian inferential approach.

In the sequel, we discuss a strand of the claims reserving literature related to this work. For an in-depth introduction to the topic, the reader is referred to Taylor (2012) or Wuthrich (2019).

The oldest and most widely used technique for reserving is the "chain-ladder" (CL), which, due to its widespread utilisation and ease of implementation, is frequently taken as a benchmark. Although the CL has been originally derived as a purely deterministic algorithm, several stochastic models provide the same predictions. For instance, Mack (1993) provides a simple "distribution free" CL model, using just moments assumptions. One of the first models with distributional assumptions on the claims payments was the over-dispersed poisson (ODP) model, proposed by Renshaw and Verrall (1998) and popularised by England and Verrall (2002). Among several other techniques, the ODP model is implemented in the `R` package `ChainLadder` (Gesmann *et al.*, 2018). For a recent formulation of the stochastic CL, see Sriram and Shi (to appear).

Early Bayesian stochastic formulations of the reserving problem for one line of business can be found in Verrall (1991), de Alba (2002) and Ntzoufras and Dellaportas (2002). A Bayesian formulation of the CL model is provided in Gisler and Wüthrich (2008). More recent examples include Antonio and Beirlant (2008), de Alba and Nieto-Barajas (2008), Peters *et al.* (2009), Meyers (2009), the monograph Meyers (2015), the survey paper Taylor (2015), Gao and Meng (2018), and the recent book Gao (2018). Going down a different route and interpreting the run-off triangles as a spatially organised data set, Lally and Hartman (2018) use Gaussian process regression techniques to estimate the reserves.

On the other hand, Pinheiro *et al.* (2003) showed that in some cases the gamma model presents a better fit to the observed values than other models that are used more frequently. Under the Bayesian paradigm, the assumption of conditionally gamma-distributed claims is made by several authors. For example, de Alba and Nieto-Barajas (2008), Gisler (2006) and Merz and Wüthrich (2015) propose similar Bayesian gamma models, differing mainly by their choices of priors.

Over the past few years, several authors proposed Bayesian models for multivariate loss reserving. One of the first contributions to this literature is Merz and Wüthrich (2010), where the authors develop a log-normal paid-incurred chain model. This model is further extended in Peters *et al.* (2014). Shi *et al.* (2012) use a multivariate log-normal model for incremental claims to examine calendar year effects when multiple triangles are available, while Merz *et al.* (2013) assume a log-normal model for the log-link ratios. Zhang and Dukic (2013) perform full Bayesian inference for models defined through several combinations of copulas and marginal distributions and Avanzi *et al.* (2016) explore a multivariate Tweedie family of models.

Correlation in triangles has been studied by several authors and in different ways. First, we discuss the dependence among accidental years and/or

https://doi.org/10.1017/asb.2020.36 Published online by Cambridge University Press

TABLE 1

RUN-OFF TRIANGLE OF AVAILABLE DATA FOR BUSINESS $k = 1, \ldots, K$.

	Business k						
Year of	Development year						
origin	1	2	$\cdots$	j	$\cdots$	$n-1$	n
1	$X_{1,1,k}$	$X_{1,2,k}$	$\cdots$	$X_{1,j,k}$		$X_{1,n-1,k}$	$X_{1,n,k}$
2	$X_{2,1,k}$	$X_{2,2,k}$	$\cdots$	$X_{2,j,k}$		$X_{2,n-1,k}$	
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$			
i	$X_{i,1,k}$	$X_{i,2,k}$	$\cdots$	$X_{i,n+1-i,k}$			
$\vdots$	$\vdots$	$\vdots$					
$n-1$	$X_{n-1,1,k}$	$X_{n-1,2,k}$					
n	$X_{n,1,k}$						

calendar years. For instance, Avanzi *et al.* (2020) use generalised linear models; Wüthrich (2010) proposes a Bayesian CL and Guszcza (2008), Zhang *et al.* (2012), Shi and Hartman (2016) and Shi (2017) use hierarchical models. Next, we highlight some of the articles that introduce dependence among development years. For example, Kremer (2005) proposes a first-order autoregressive model along development years and uses least squares to pool information across multiple triangles; De Jong (2006) considers lognormal cumulative claims ratios together with an integrated moving average model; de Alba and Nieto-Barajas (2008) take gamma distributions for the incremental claims and introduce order one dependence through a set of latent variables inducing a Markov process for the claims. Our proposed model, on the other hand, assumes gamma distributions for the incremental claims and introduces an order-p dependence through poisson latent variables. We further pull strength across several triangles through the use of hierarchical priors on the model parameters.

To formally state the problem, we denote by $Y_{i,j,k}$ the incremental claim amount arising from year of origin i and paid in development year j; $P_{i,k}$ is the net premium received at year i. Index k can refer to either different lines of business within the same company, or the same line of business across different companies. Furthermore, we define $X_{i,j,k} = g(Y_{i,j,k})$, that is a function of the incremental payments, where $g(\cdot)$ could be a rescaling function or the loss ratio $Y_{i,j,k}/P_{i,k}$. Let us assume that for each k we are in calendar year n and the available information is given by $\mathcal{D}_n = \bigcup_{k=1}^{K} \mathcal{D}_n^k$, where

$$\mathcal{D}_n^k = \left\{X_{i,j,k} \,:\, i = 1, \ldots, n,\; j = 1, \ldots, n - i + 1\right\}.$$

These available data can be represented in terms of run-off triangles as the one given in Table 1.

The problem consists in predicting the quantities $X_{i,j,k}$ (and $Y_{i,j,k}$), for $i = 2, 3, \ldots, n$ and $j = n + 2 - i, n + 3 - i, \ldots, n$, which correspond to the right-lower triangle in Table 1. In particular, more than predicting individual values $Y_{i,j,k}$, we are interested in the prediction of outstanding claims and thus having the necessary information to constitute adequate reserves. Let $R_{i,k}=\sum_{j=n+2-i}^{n} Y_{i,j,k}$ the total aggregate outstanding claims for each year of origin $i=2,\ldots,n$ and business k. Moreover, the total outstanding claims for business k, considering all years, is $R_k=\sum_{i=2}^{n} R_{i,k}$, and the grand total becomes $R=\sum_{k=1}^{K} R_k$.

Let us assume that the probabilistic model for $X_{i,j,k}$ is described conditional on some parameter vector $\boldsymbol{\theta}$. Then, for business k, all information on the outstanding reserves at time n is given by the distribution of $R_{i,k}\mid\mathcal{D}_n$:

$$p(R_{i,k}\mid\mathcal{D}_n)=\int p(R_{i,k}\mid\boldsymbol{\theta},\mathcal{D}_n)p(\boldsymbol{\theta}\mid\mathcal{D}_n)d\boldsymbol{\theta},$$

which is written as an average of the model $p(R_{i,k}\mid\boldsymbol{\theta},\mathcal{D}_n)$ over all possible parameters, weighted by their posterior probability, $p(\boldsymbol{\theta}\mid\mathcal{D}_n)$.

Before proceeding, we introduce some notation: $\mathrm{Ga}(\alpha,\beta)$ denotes a gamma density with mean α/β and variance α/β^2; $\mathrm{Po}(\gamma)$ denotes a poisson density with mean and variance γ.

2. Dependent gamma model

Let $\mathbf{X}=\{X_{i,j,k}\}$ be the set of variables of interest, for origin year $i=1,\ldots,n$, development year $j=1,\ldots,n$ and business $k=1,\ldots,K$. For a particular claim $X_{i,j,k}$, we propose a dependence model of moving average nature for the claims made in the previous p development years, say $j-1,\ldots,j-p$. To achieve our objective, we define a set of latent variables $\mathbf{Z}=\{Z_{i,j,k}\}$, one for each (i,j,k).

Therefore, our model is defined through a two-level hierarchical specification of the form:

$$X_{i,j,k}\mid\mathbf{Z}\sim\mathrm{Ga}\left(\alpha_{i,k}+\sum_{l=0}^{p} Z_{i,j-l,k}\,,\ \beta_{j,k}+\sum_{l=0}^{p}\gamma_{j-l,k}\right),$$
$$Z_{i,j,k}\sim\mathrm{Po}\big(\alpha_{i,k}\gamma_{j,k}\big) \tag{2.1}$$

independently for $i,j=1,\ldots,n$ and $k=1,\ldots,K$, where $\{\alpha_{i,k}\}$, $\{\beta_{j,k}\}$ and $\{\gamma_{j,k}\}$ are all nonnegative parameters, and $p\geq 0$ is the order of dependence across development years. We define $Z_{i,j,k}=0$ with probability one (w.p.1), and $\gamma_{j,k}\equiv 0$, for $j\leq 0$. We will refer to (2.1) as dependent gamma model (DGM).

The role of the latent variables $\mathbf{Z}$ is to introduce dependence across development years. The marginal distribution for the set $\mathbf{X}$ is an infinite (discrete) mixture of gamma distributions with poisson weights, that is:

$$f(\mathbf{x})=\sum_{z_{1,1,1}=0}^{\infty}\cdots\sum_{z_{n,n,K}=0}^{\infty}\prod_{i,j,k} f(x_{i,j,k}\mid\mathbf{z})f(z_{i,j,k})$$

where $f(x_{i,j,k}\mid\mathbf{z})$ and $f(z_{i,j,k})$ are the gamma and poisson distributions given in (2.1). The discreteness of the poisson random variables does not affect the

https://doi.org/10.1017/asb.2020.36 Published online by Cambridge University Press

model performance. If $\gamma_{j,k}=0$, then $Z_{i,j,k}=0$ w.p.1 so the influence (dependence) of that specific development year j with future years is null. If $\gamma_{j,k}=0$ for all j and k, the $X_{i,j,k}$ become independent. Alternatively, independence across all development years can be achieved by taking $p=0$.

The choice of the poisson distributions for the latent variables is convenient to obtain an appealing interpretation of the model. By taking iterative expectations in model (2.1), the marginal expected value of each $X_{i,j,k}$ becomes

$$\mu_{i,j,k}=\mathrm{E}\big(X_{i,j,k}\big)=\frac{\alpha_{i,k}\big(1+\sum_{l=0}^{p}\gamma_{j-l,k}\big)}{\beta_{j,k}+\sum_{r=0}^{p}\gamma_{j-r,k}}=\alpha_{i,k}\pi_{j,k}, \tag{2.2}$$

where $\pi_{j,k}=\big\{1+\sum_{l=0}^{p}\gamma_{j-l,k}\big\}\,/\,\big\{\beta_{j,k}+\sum_{r=0}^{p}\gamma_{j-r,k}\big\}$ are development year specific weights. When we take $\gamma_{j,k}=0$ for all j and k, these weights become $\pi_{j,k}=1/\beta_{j,k}$ so $\mu_{i,j,k}=\alpha_{i,k}/\beta_{j,k}$.

Since $\sum_{j=1}^{n}\pi_{j,k}\neq 1$, we propose the following transformations to define interpretable quantities:

$$\alpha^*_{i,k}=\alpha_{i,k}\sum_{j=1}^{n}\pi_{j,k}\quad\text{and}\quad\pi^*_{j,k}=\frac{\pi_{j,k}}{\sum_{l=1}^{n}\pi_{l,k}} \tag{2.3}$$

so that Equation (2.2) can be written as $\mu_{i,j,k}=\alpha^*_{i,k}\pi^*_{j,k}$, where $\sum_{j=1}^{n}\pi^*_{j,k}=1$. In this case, $\alpha^*_{i,k}=\sum_{j=1}^{n}\mu_{i,j,k}$ can be interpreted as the ultimate total amount for business k at origination year i, and $\pi^*_{j,k}$ can be interpreted as the proportion of $\alpha^*_{i,k}$ corresponding to development year j in business k.

The marginal variance for each $X_{i,j,k}$ can also be computed using iterative variance results. This has the form:

$$\mathrm{Var}\big(X_{i,j,k}\big)=\frac{\alpha_{i,k}\big(1+2\sum_{l=0}^{p}\gamma_{j-l,k}\big)}{\big(\beta_{j,k}+\sum_{r=0}^{p}\gamma_{j-r,k}\big)^2}. \tag{2.4}$$

Additionally, as a measure of the dependence induced by our model, we can compute in closed form the covariance between any two claims for development years j and $j+s$, with $1\le s\le p$, for the same origin year i and the same business k. This becomes

$$\mathrm{Cov}(X_{i,j,k},X_{i,j+s,k})=\frac{\alpha_{i,k}\sum_{l=0}^{p-s}\gamma_{j-l,k}}{\big(\beta_{j,k}+\sum_{l=0}^{p}\gamma_{j-l,k}\big)\big(\beta_{j+s,k}+\sum_{l=0}^{p}\gamma_{j+s-l,k}\big)} \tag{2.5}$$

and takes the value of zero for $s>p$ or if the two claims come from different origin years ($i's$) or different business ($k's$).

https://doi.org/10.1017/asb.2020.36 Published online by Cambridge University Press

Finally, with expressions (2.4) and (2.5), we can easily compute the correlation between $X_{i,j,k}$ and $X_{i,j+s,k}$, for $1 \leq s \leq p$, which has the form:

$$\text{Corr}(X_{i,j,k}, X_{i,j+s,k}) = \frac{\sum_{l=0}^{p-s} \gamma_{j-l,k}}{\sqrt{1 + 2\sum_{l=0}^{p} \gamma_{j-l,k}}\sqrt{1 + 2\sum_{l=0}^{p} \gamma_{j+s-l,k}}}. \tag{2.6}$$

We note that expression (2.6) does not depend on the parameters $\alpha_{i,k}$ and $\beta_{j,k}$, it is only a function of parameters $\{\gamma_{j,k}\}$. Since expression (2.6) does not depend on the origin year i, we will denote the correlation as $\rho_{j,j+s,k}$ for $s \leq p$. Specifically, the numerator is a function of the parameters shared by development years j and $j+s$, that is, $\gamma_{l,j}$ for $l = j+s-p, j+s-p+1, \ldots, j$. Larger/smaller values of $\gamma_{l,k}$ induce a larger/smaller correlation.

Therefore, the set $\{\gamma_{j,k}\}$ controls the degree of dependence across developments years in the business-specific triangle k, for instance, if $\gamma_{j,k} = 0$ for all development years j, the correlation becomes zero between any two claims for that specific business k. In general, for values of $\gamma_{j,k} > 0$, the dependence (in terms of correlation), induced by our model (2.1), is positive and takes values in the whole range, that is $\rho_{j,j+s,k} \in [0, 1]$, which is useful in the modelling of trends along development years in a run-off triangle.

3. Bayesian inference

Recall that available data consist of run-off triangles as those depicted in Table 1, that is, $\mathbf{X} = \{X_{i,j,k}\}$ for $i = 1, \ldots, n$, $j = 1, \ldots, n+1-i$ and $k = 1, \ldots, K$. To carry out a full Bayesian analysis of the model, we rely on data augmentation techniques (e.g., Tanner, 1991) to deal with the unobserved latent variables $\mathbf{Z}$. If we denote by $\boldsymbol{\theta} = \{\alpha_{i,k}, \beta_{j,k}, \gamma_{j,k}\}$, the set of all model parameters, the likelihood, assuming that we have observed the latent variables, is simply the joint distribution of the observed data as well as the latent variables, that is,

$$f(\mathbf{x}, \mathbf{z} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{j=1}^{n+1-i} \prod_{k=1}^{K} \left\{ \text{Ga}\left(x_{i,j,k} \mid \alpha_{i,k} + \sum_{l=0}^{p} z_{i,j-l,k}\,,\ \beta_{j,k} + \sum_{l=0}^{p} \gamma_{j-l,k}\right) \times \text{Po}\left(z_{i,j,k} \mid \alpha_{i,k}\gamma_{j,k}\right) \right\}.$$

To borrow strength across different triangles in the estimation procedure, we propose a hierarchical prior of the form:

$$\begin{aligned} &\alpha_{i,k} \mid a_{\alpha i}, b_{\alpha i} \sim \text{Ga}(a_{\alpha i}, b_{\alpha i}) \text{ with } a_{\alpha i} \sim \text{Ga}(a_{\alpha 0}, b_{\alpha 0}) \text{ and } b_{\alpha i} \sim \text{Ga}(a_{\alpha 0}, b_{\alpha 0}) \\ &\beta_{j,k} \mid a_{\beta j}, b_{\beta j} \sim \text{Ga}(a_{\beta j}, b_{\beta j}) \text{ with } a_{\beta j} \sim \text{Ga}(a_{\beta 0}, b_{\beta 0}) \text{ and } b_{\beta j} \sim \text{Ga}(a_{\beta 0}, b_{\beta 0}) \\ &\gamma_{j,k} \mid a_{\gamma j}, b_{\gamma j} \sim \text{Ga}(a_{\gamma j}, b_{\gamma j}) \text{ with } a_{\gamma j} \sim \text{Ga}(a_{\gamma 0}, b_{\gamma 0}) \text{ and } b_{\gamma j} \sim \text{Ga}(a_{\gamma 0}, b_{\gamma 0}), \end{aligned} \tag{3.1}$$

https://doi.org/10.1017/asb.2020.36 Published online by Cambridge University Press

conditionally independent for $k = 1, \ldots, K$, with suitable values $a_{\alpha 0}$, $b_{\alpha 0}$, $a_{\beta 0}$, $b_{\beta 0}$, $a_{\gamma 0}$ and $b_{\gamma 0}$. Smaller/larger prior variance on the hyper-parameters allows for a lower/higher borrowing of strength across k.

Posterior distribution is characterised through the full conditional distributions of all model parameters $\boldsymbol{\theta}$ plus the conditional distributions for the latent variables $\mathbf{Z}$, as well as the conditional distributions of the hyper-parameters of the hierarchical prior specification, which are given in the Appendix. Since these distributions do not have standard form, posterior inference would require the implementation of a Gibbs sampler (Smith and Roberts, 1993) with Metropolis-Hastings steps (Tierney, 1994). Alternatively, since all model and prior distributions involved are of standard form, Markov Chain Monte Carlo (MCMC) procedures can also be implemented in the `R` package `rjags` (Plummer, 2018).

4. Results

4.1. Simulated data

Before presenting a detailed application of the proposed DGM on a real data set, we briefly discuss its performance on a simulation study.

For this example, we fix the number of lines of business $K = 2$ and assume the claims are fully developed after $n = 4$ years. We then generate a data set from the model described in (2.1), with the following specifications: $p = 1$; $\alpha_{i,1} = 1$ and $\alpha_{i,2} = 2$ for $i = 1, \ldots, 4$; $\beta_{j,k} = 1$ $\forall j, k$; and $\{\gamma_{j,k}, j = 1, \ldots, 4\} = \{1, 4, 6, 2\}$, for $k = 1, 2$. In this case, $\alpha^*_{i,1} = 4$ and $\alpha^*_{i,2} = 8$ for $i = 1, \ldots, 4$; and $\pi^*_{j,k} = 1/4$ $\forall j, k$. Additionally, $\rho_{1,2,k} = 0.174$, $\rho_{2,3,k} = 0.263$ and $\rho_{3,4,k} = 0.317$, for $k = 1, 2$. Since we are only changing the value of $\alpha_{i,k}$ for $k = 1, 2$, the proportions and correlations across development years remain the same in both triangles.

To perform inference on this model, we run two independent Markov chains, both with priors as discussed in Section 3 and hyper-parameters $a_{\alpha 0} = 2$, $b_{\alpha 0} = 1$, $a_{\beta 0} = 2$, $b_{\beta 0} = 2$, and $a_{\gamma 0} = 3$, $b_{\gamma 0} = 1$. The burn-in was set to 10, 000 samples and 10, 000 samples were kept for each chain.

The assessment of the convergence of the Markov chains is made visually, through their trace plots, as show in Figure 1. Mixing of the chains is appropriate and shows no trend and the estimated densities are smooth. The other parameters perform similarly.

The 90% high posterior density (HPD) intervals using the samples from both chains are shown in Figure 2. The plot is divided in three panels, where each point in the horizontal axis denote one of the unknown (identifiable) parameters: $\alpha^*_{i,k}$, $\rho_{j,j+s,k}$ and $\pi^*_{j,k}$, respectively.

From Figure 2, we can see that almost all parameters (apart from $\alpha^*_{2,2}$) lie within their respective 90% HPD interval (red bar) and most of the point estimates (red dots) are very close to the real values (black dots). The reason why $\alpha^*_{2,2}$ lies away from its HPD interval is because we are using a single replicate of the data.

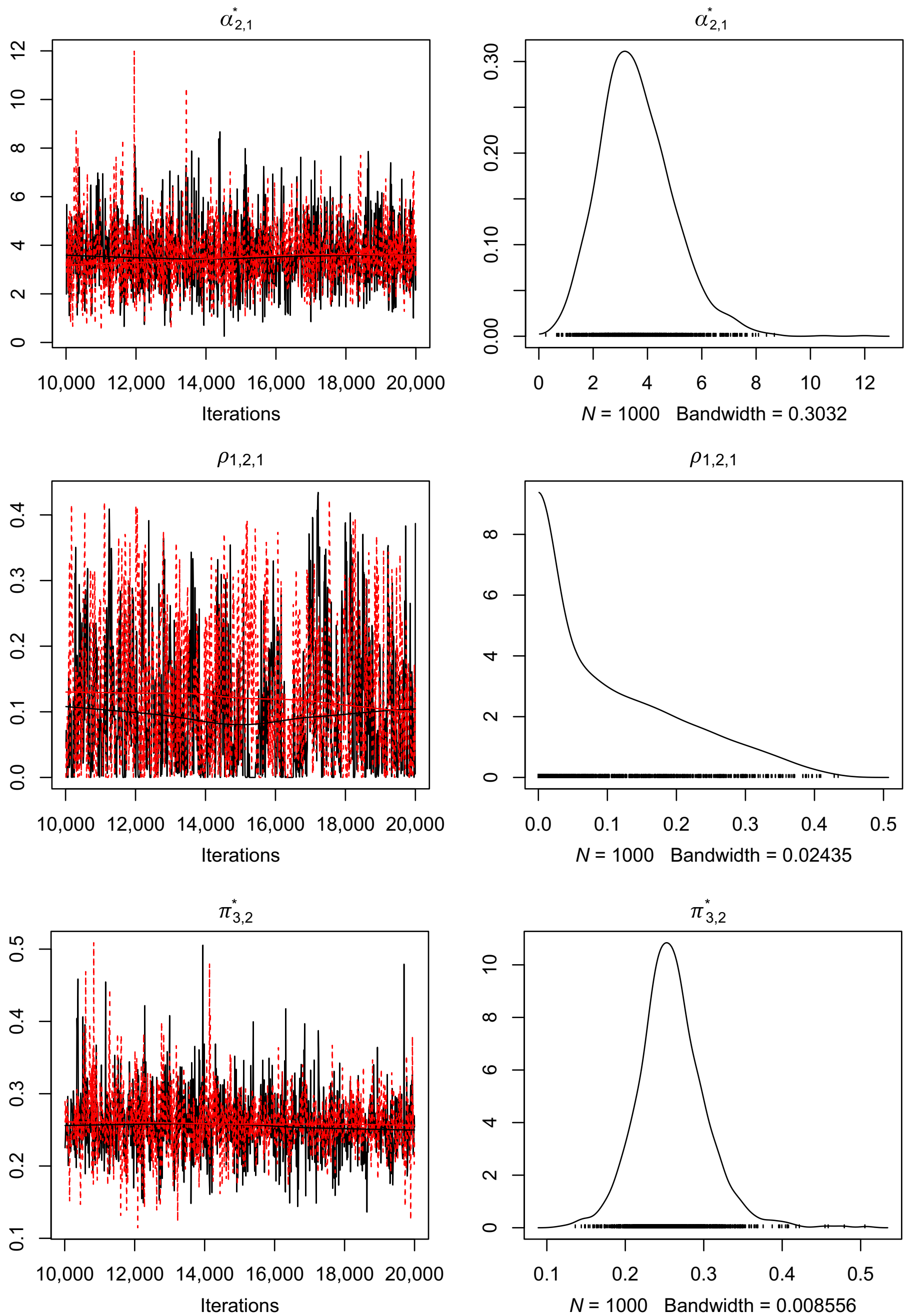

FIGURE 1: Trace-plots of some parameters for the Markov Chains (left) and posterior density (right) for the simulated data.

https://doi.org/10.1017/asb.2020.36 Published online by Cambridge University Press

TABLE 2

GROUP CODES AND NAMES FOR THE 10 COMPANIES ANALYSED. THESE CORRESPOND TO THE LARGEST COMPANIES BASED ON THEIR POSTED RESERVES IN 1997.

k	Group codes	Group name
1	1767	State Farm Mut Grp
2	2003	United Services Automobile Asn Grp
3	7080	New Jersey Manufacturers Grp
4	4839	FL Farm Bureau Grp
5	388	Federal Ins Co Grp
6	1090	Kentucky Farm Bureau Mut Ins Grp
7	3240	NC Farm Bureau Ins Grp
8	6947	Tenn Farmers Mut
9	620	Employers Mut Co Of Des Moines
10	692	Wawanesa Ins Grp

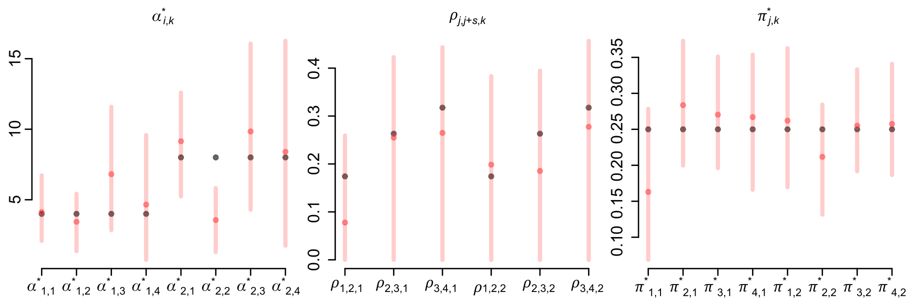

FIGURE 2: Parameter estimates for simulated data. ■ 90% HPD intervals, ● posterior median and ● true parameters.

4.2. Real data

The data used in this section consist of incremental paid losses for personal auto insurance in the US, a data set compiled by Meyers and Shi (2011). In order to fit the proposed DGM, we only used information up to 1997, which leaves 10 years for testing. Based on the subset of insurers selected (Meyers, 2015, Appendix A), we selected the 10 largest insurers (by posted reserves in 1997) to perform our analysis on. The group codes for these insures are presented in the first column of Table 2 (in decreasing order of their 1997 reserves).

Before fitting any model, we compute the average (over accident years) sample auto-correlation function (ACF) of the incremental payments across development years. More precisely, for each company $k = 1, \ldots, K$ and for each accident year $i = 1, \ldots, I-1$, we computed the ACF across development years, that is, the ACF of $\{\{Y_{i,j,k}\}\}_{j=1}^{J}$. Then, for each lag of the ACF, we took the average over possible accident years. Note that for the ℓ-th lag in the ACF, we computed $J - \ell$ values, one for each accident year. In Figure 3, we plot

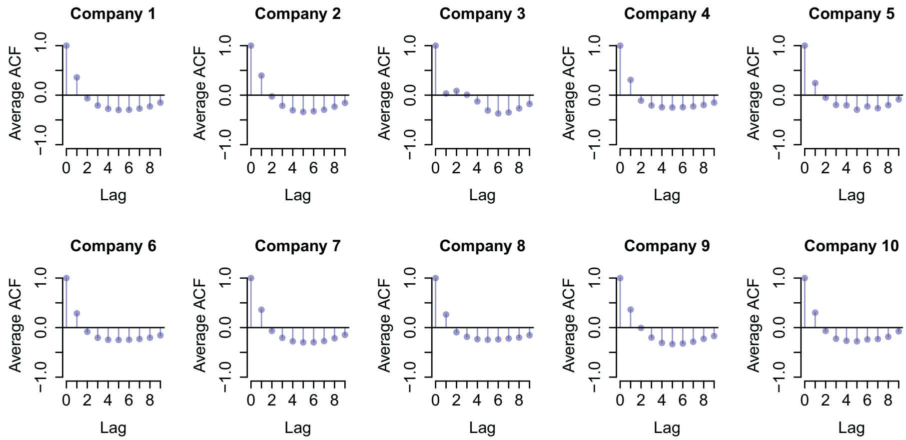

FIGURE 3: Average auto-correlation function for upper triangle of incremental payments with lags on the development year. Average across accident years.

these average ACFs for each one of the selected companies. From the figure, we can see a clear pattern: the average ACF is positive and assumes its largest value for lag 1. For all other lags, the average ACF is negative with comparably smaller absolute values.

To avoid numerical problems we rescaled the data, first dividing all claims by 1000 and then taking its square root, that is, $X_{i,j,k} = g(Y_{i,j,k}) = \sqrt{Y_{i,j,k}/1000}$. Then played with different specifications of the prior distributions. Specifically, we took $p \in \{0, \ldots, 5\}$, $a_{\alpha 0} = b_{\alpha 0} \in \{1, 10\}$, $a_{\beta 0} = b_{\beta 0} \in \{1, 10\}$ and $a_{\gamma 0} = b_{\gamma 0} = 10$. The combinations of these values lead to 24 models which are summarised in Table 3. Note that $a_{\beta 0} = b_{\beta 0} = 1$ (or 10) induce a larger (or smaller) prior variance on $a_{\beta j}$ and $b_{\beta j}$ which in turn implies a higher (or lower) borrowing of strength across β_{jk}. The same applies to α_{ik}. For γ_{jk}, our hierarchical prior choice allows for a low borrowing since the dependence among development years might be different across triangles.

To show the proposed hierarchical priors (3.1) do have an effect of borrowing strength across triangles, we also fitted our model with independent priors of the form:

$$\alpha_{i,k} \sim \text{Ga}(a_\alpha, b_\alpha) \quad \beta_{j,k} \sim \text{Ga}(a_\beta, b_\beta) \quad \text{and} \quad \gamma_{j,k} \sim \text{Ga}(a_\gamma, b_\gamma) \tag{4.1}$$

with $a_\alpha = b_\alpha = a_\beta = b_\beta = a_\gamma = b_\gamma = 1$. The hyper-parameters here are set as their expected values of the hierarchical prior (3.1) for all models of Table 3.

For all models, we ran Gibbs samplers with two parallel chains, each one with 10,000 iterations after a burn-in of 100,000 iterations and keeping one of every 10th to compute posterior estimates. Trace plots for the model parameters with hierarchical priors behave similarly to those in Figure 1, which show a satisfactory convergence.

https://doi.org/10.1017/asb.2020.36 Published online by Cambridge University Press

TABLE 3

PRIOR SPECIFICATION FOR THE 24 MODELS. THE TABLE IS SORTED (IN ASCENDING ORDER) BY p, THEN $a_{\alpha 0}$ AND FINALLY BY $b_{\beta 0}$.

Model	p	$a_{\alpha 0} = b_{\alpha 0}$	$a_{\beta 0} = b_{\beta 0}$
1	0	1	1
2	1	1	1
⋮	⋮	⋮	⋮
6	5	1	1
7	0	10	1
⋮	⋮	⋮	⋮
12	5	10	1
13	0	1	10
⋮	⋮	⋮	⋮
18	5	1	10
19	0	10	10
⋮	⋮	⋮	⋮
24	5	10	10

The best model was selected based on two goodness-of-fit measures: the deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002), which is a model selection criterion that penalises for model complexity; and the L-measure (Ibrahim and Laud, 1994) based on the posterior predictive density and defined as:

$$L(\nu) = \frac{1}{M} \sum_{k=1}^{K} \sum_{i=2}^{n} \sum_{j=n-i+2}^{n} \mathrm{Var}(X_{i,j,k}^{F} \mid \mathbf{x}) + \frac{\nu}{M} \sum_{k=1}^{K} \sum_{i=2}^{n} \sum_{j=n-i+2}^{n} \{\mathrm{E}(X_{i,j,k}^{F} \mid \mathbf{x}) - x_{i,j,k}\}^2,$$

where $X_{i,j,k}^{F}$ and $x_{i,j,k}$ are the predictive and observed values of $X_{i,j,k}$, respectively, $\nu \in [0, 1]$ is a weighting term which determines a trade-off between variance and bias, and $M = Kn(n-1)/2$ is the number of unknowns.

Figure 4 presents the DIC values obtained for each of the 24 models. Values are shown in four blocks of six models, each block has the same prior specifications but with varying p. In the four blocks, the best fitting is obtained with $p = 1$, with a shorter difference in the third block for $p = 2, 3$. Across blocks, it seems that there is an increasing trend in the DIC values. Overall the best fitting is achieved by model 2, which corresponds to $p = 1$, $a_{\alpha 0} = b_{\alpha 0} = 1$ and $a_{\beta 0} = b_{\beta 0} = 1$.

On the other hand, Figure 5 includes the L-measure with $\nu = 1/2$ for in-sample data (left panel) and out-of-sample data (right panel). Again results are reported in four blocks of six models as in Figure 4. The qualitative behaviour of the measures in- and out-of-sample is very similar and also similar to what is

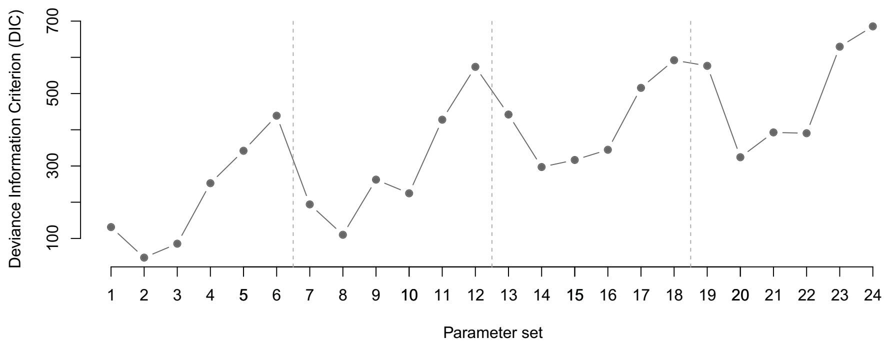

FIGURE 4: Deviance Information Criterion (DIC) for the 24 models of Table 3. Vertical dotted lines divide the models in four blocks of six models.

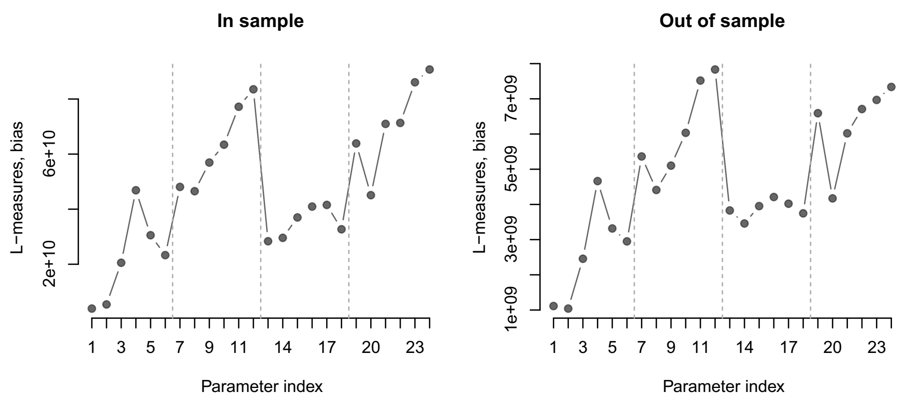

FIGURE 5: L-measure with $\nu = 1/2$ for the 24 models of Table 3. In sample measures (left) and out of sample measures (right). Vertical dotted lines divide the models in four blocks of six models.

observed with the DIC in Figure 4. However, the increasing trend of the DIC is not shown with the L-measures. The third block has lower values than the second and fourth blocks. Again, the best fit is achieved by model 2.

Moreover, the goodness-of-fit measures, DIC and L-measure, for the six model specifications with independence priors (4.1) and for $p \in \{0, \ldots, 5\}$ are reported in Table 4. These values are a lot higher than then ones we obtain with the four different specifications of the hierarchical prior.

Therefore, results from the winning model 2 of Table 3 will be discussed in the sequel. Figure 6 shows the point estimates, based on the 50% quantile, for $\alpha^*_{i,k}$ (left panel) and $\pi^*_{j,k}$ (right panel). One can clearly see that the values of $\alpha^*_{i,k}$ (the ultimate total amount for business k with origin at year i) decrease along the columns, that is, when k increases. This is expected, as the businesses are ordered in a decreasing way based on their size (posted reserves in 1997). Although less pronounced, it is also possible to see the origin year effect, represented by the fact that $\alpha^*_{i,k}$ is (slightly) increasing in i (across rows). On the other hand, a clear pattern also emerges from the estimates of $\pi^*_{j,k}$, where one

https://doi.org/10.1017/asb.2020.36 Published online by Cambridge University Press

TABLE 4

GOODNESS-OF-FIT MEASURES WITH INDEPENDENCE PRIORS (4.1) FOR $p \in \{0, 1, 2, 3, 4, 5\}$. DIC AND L-MEASURE (IN AND OUT OF SAMPLE) WITH $\nu = 1/2$.

Measure/p	0	1	2	3	4	5
DIC	1.89e+03	1.72e+03	1.65e+03	1.63e+03	1.67e+03	1.71e+03
L-measure (in)	9.83e+11	9.19e+11	9.22e+11	9.02e+11	1.00e+12	1.02e+12
L-measure (out)	4.01e+10	3.14e+10	3.44e+10	3.33e+10	2.98e+10	3.04e+10

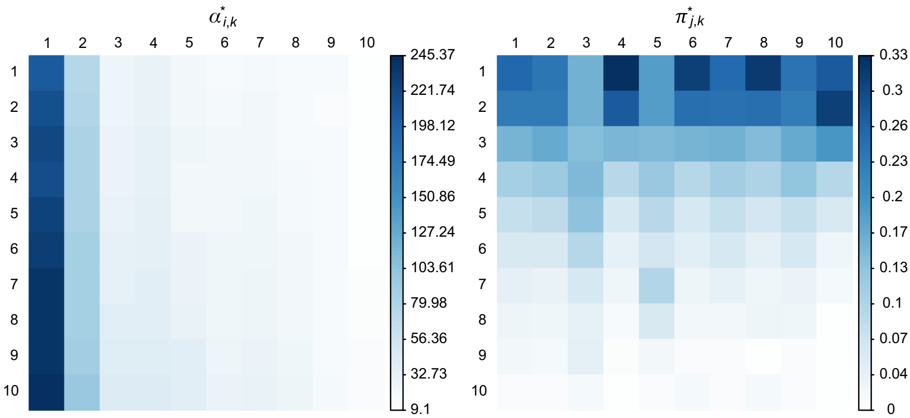

FIGURE 6: Posterior estimates (medians) for $\alpha^*_{i,k}$ (left) and $\pi^*_{j,k}$ (right) for the real data obtained with the best fitting model with hierarchical priors.

can observe a decrease in their values when the development year j increases (across rows). The intuition for this result is that the further we are from the first development years j, the smaller is the proportion of the ultimate total claim amount $\alpha^*_{i,k}$ expected in development year j.

Since for the best fitting model $p = 1$, there are only nine correlation coefficients $\rho_{j,j+1,k}$, as in (2.6), for $j = 1, \ldots, 9$ and for each company $k = 1, \ldots, 10$. Posterior densities of these coefficients are presented in Figure 7. From these plots we can see that, for most development years j, all companies have similar correlation distributions. One particularly different case is for development year $j = 5$ (centre panel), where almost all posterior distributions are left-skewed, apart from two, which are right-skewed and concentrated on smaller values. A similar situation occurs for development year $j = 7$, when only one company's posterior distribution stands out as left-skewed. In general, all correlations are likely to be smaller than 0.5. An interesting fact is that correlations tend to be larger (around 0.5) and with low dispersion for odd years $j = 1, j = 3$ and $j = 5$; be very small (less than 0.2) for years $j = 2$ and $j = 4$; and with a range between 0 and 0.4 for years $j = 6, 7, 8, 9$. This could provide a degree of importance for each development year in the whole triangle.

https://doi.org/10.1017/asb.2020.36 Published online by Cambridge University Press

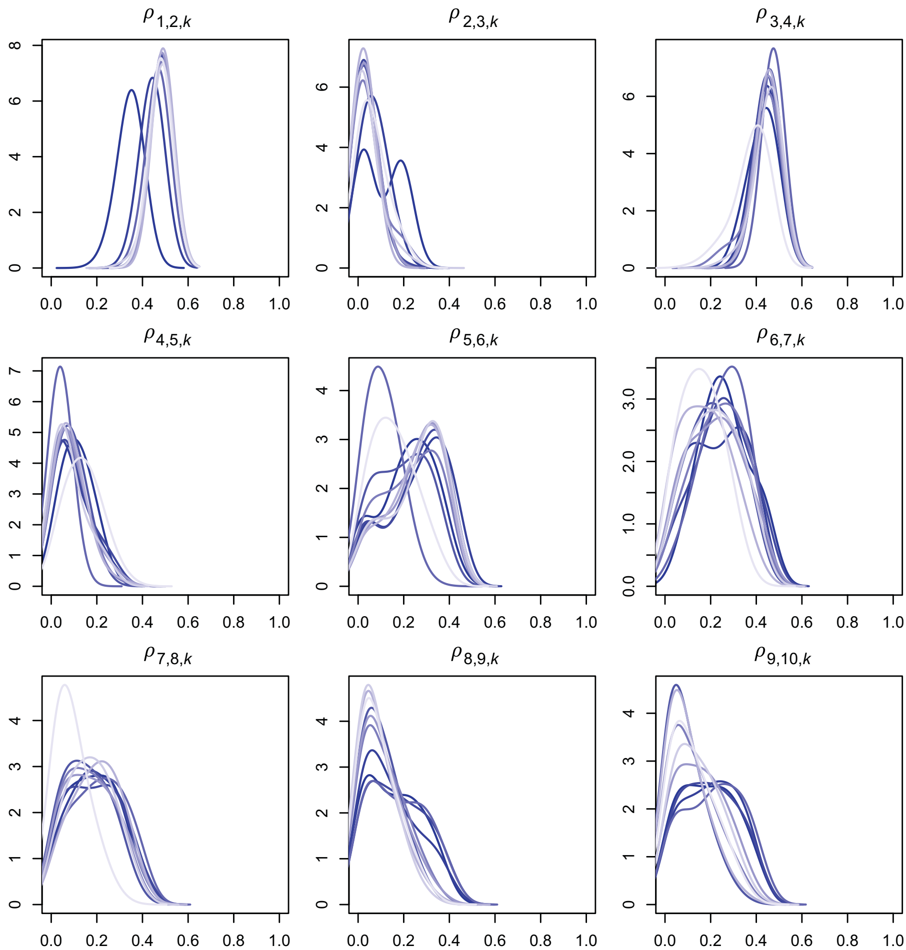

FIGURE 7: Posterior densities of the correlation coefficients $\rho_{j,j+1,k}$, $j = 1, \ldots, 9$ obtained by the best fitting model with hierarchical priors. Darker colours denote larger companies while lighter colours denote smaller companies.

We compute posterior predictive distributions for $Y_{i,j,k}$ in the lower-right triangle, that is for $i = 2, \ldots, n$, $j = n + 2 - i, \ldots, n$ and $n = 10$. We rescaled back the values with the inverse function $g^{-1}(\cdot)$ and present the median (dark red) together with the 95% credibility intervals (light red), for two companies: $k = 1$, which corresponds to the largest company, State Farm Mut Grp (code 1767), in Figure 8; and $k = 10$, which corresponds to the smallest company, Wawanesa Ins Grp (code 692), in Figure 9. For each accident year, the light grey dots represent the observed loss data (upper triangle) and the dark grey dots are observations not provided to the model (lower triangle).

For the largest company (Figure 8), the model is able to fit the in-sample data perfectly well, and it produces very precise predictions. For the smallest

https://doi.org/10.1017/asb.2020.36 Published online by Cambridge University Press

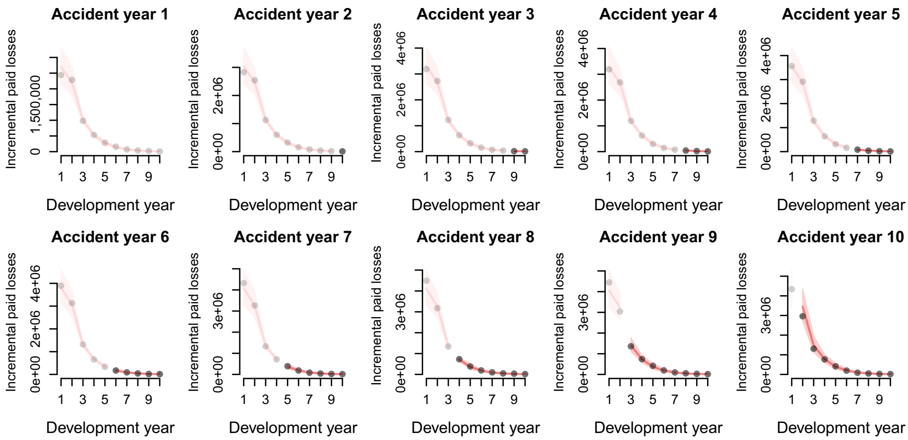

FIGURE 8: Largest company in the sample (code 1767). Posterior predictions of incremental paid losses per development year i, for all accident years j obtained with the best fitting model with hierarchical priors. • Observed claims (upper triangle), • Non-observed claims (lower triangle), ■ 95% credibility interval, — median.

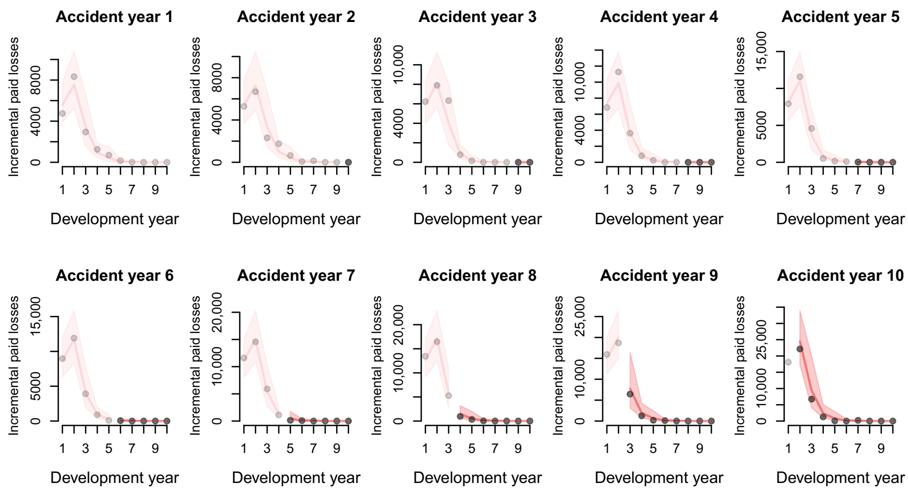

FIGURE 9: Smallest company in the sample (code 692). Posterior predictions of incremental paid losses per development year i, for all accident years j obtained with the best fitting model with hierarchical priors. • Observed claims (upper triangle), • non-observed claims (lower triangle), ■ 95% credibility interval, — median.

company (Figure 9), the model also performs well, even being able to fit the hump at development year 2 in all accident years. The credible intervals are, perhaps, a little wider than those for the largest company ($k = 1$).

To see the advantage of our modelling with respect to standard methods, we provide a comparison between the proposed DGM, which considers dependence across development years and pulls strength across the 10 companies, with the ODP. We fitted 10 independent ODP models using the R-package

https://doi.org/10.1017/asb.2020.36 Published online by Cambridge University Press

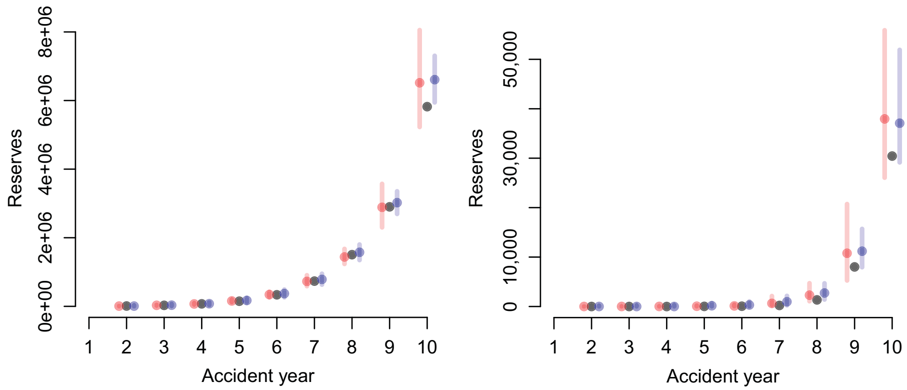

FIGURE 10: Reserves for each accident year obtained with the best fitting model with hierarchical priors. (Left) largest company (code 1767) and (right) smallest company (code 692). • median and 95% CI for DGM, • true claims and • median and 95% CI for ODP.

ChainLadder (Gesmann *et al.*, 2018). For the comparison, we show the reserves per accident year i as well as the aggregated reserves for the 10 companies.

Figure 10 presents the reserves estimates from the DGM (red) and the ODP model (blue) for each accident year $i = 1, \ldots, 10$, for the largest company $k = 1$ (code 1767) and the smallest company $k = 10$ (code 692). For the DGM, posterior predictive estimates are given by the median and the quantiles to form a 95% CI. For the ODP, predictions are based on maximum likelihood estimates and bootstrap 95% CI. For comparison, we also include the true observed claims (grey). For both companies, estimates produced by both models, DGM and ODP, are very precise for the first seven accident years ($i = 1, \ldots, 7$), but for the last three accident years ($i = 8, 9, 10$) interval estimates of DGM are wider than those of the ODP showing more dispersion. However, for these last 3 years, DGM point estimates show less bias than those from ODP model.

To further analyse the prediction performance, we aggregate the reserves for all 10 companies and compute the predictive distribution of the aggregated reserve, which is shown in Figure 11 as a probability histogram. As postulated, the DGM (red) has larger variance and smaller bias than the ODP (blue). The pink area corresponds to the intersection of both histograms. Additionally, we can see that the predictive distribution of the ODP lies away from the true observed reserve (grey vertical line), whereas the predictive distribution of the DGM captures well the true reserve.

As expected from the goodness-of-fit measures presented, the reserves computed based on the model with independent priors is far from the results achieved by the DGM with hierarchical priors and the ODP and for this reason are not included in the plots. For the sake of comparison, for the best performing model with independent priors, $p = 3$, the best estimate for the reserves is around 1.0×10^7. Another expected result from the goodness-of-fit measures is the fact that the models with $p = 0$ and $p = 1$ would have similar predictive

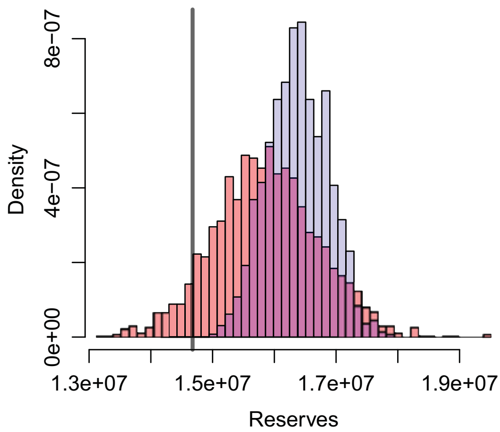

FIGURE 11: Predictive distributions of the aggregated reserves for all 10 companies obtained with the best fitting model with hierarchical priors. – True claims; DGM; ODP (Bootstrap).

power, since both their DICs and L-measures are similar. This is, indeed, the case with some advantage for the model with $p = 1$ over the model with $p = 0$.

It is important to notice that this exercise was performed to showcase the ability of the proposed DGM to perform accurate forecasts of future claims payments. For this task, point estimates were chosen as the 50% quantile (median), and the 95% credibility intervals were determined by the 2.5% and 97.5% quantiles of the posterior distributions. In practice, modern solvency regulations, such as Solvency II (ECB, 2009) and the Swiss Solvency Test (FINMA, 2007), require reserves to be computed in a much more conservative fashion. For example, the first requires reserves to be computed based on the 99.5%-quantile or value at risk (VaR) of the distribution of the losses, while the latter uses the expected shortfall (also called conditional VaR) at the level 99% of the same distribution. Both quantities, and their related credibility intervals, can be easily computed from the model's outputs, as the Markov Chain Monte Carlo algorithm returns samples of the required posterior (predictive) distributions.

Lastly, as an alternative procedure, we also adjusted our model to the loss ratios $X_{i,j,k} = g(Y_{i,j,k}) = Y_{i,j,k}/P_{i,k}$. We fitted the 24 model specifications of Table 3. The qualitative behaviour of the DIC and L-measures was roughly the same as when using the scaling transformation. In both set-ups, the best model is found to be model 2, with $p = 1$, and the posterior predictive distributions for the reserves are very similar. Nevertheless, it is worth mentioning that the reserves for the other 23 models are more precise when using the scaling transformation squared root of incremental payment divided by 1000.

Additionally, given that the size of the 10 companies is different, we test the need of hierarchical priors for the accident year effects $\alpha_{i,k}$. For that we adjusted a model with independent priors for the α parameters and hierarchical priors for β and γ. The hyper-parameters used were the same as above. Here,

TABLE 5

POSTERIOR QUANTILES FOR THE RESERVES UNDER DIFFERENT MODELS. H: HIERARCHICAL PRIORS FOR ALL PARAMETERS, I: INDEPENDENT PRIORS FOR ALL PARAMETERS, HI: INDEPENDENT PRIORS FOR α'S AND HIERARCHICAL PRIORS FOR β'S AND γ'S, LR: LOSS RATIO WITH HIERARCHICAL PRIORS FOR ALL PARAMETERS, T: TRUE RESERVES.

	$q_{2.5\%}$	$q_{50\%}$	$q_{97.5\%}$
H	14,215,629	15,861,776	17,547,465
I	6,514,440	10,839,046	23,304,484
HI	4,193,141	7,373,883	16,627,489
LR	14,309,830	16,008,333	18,104,569
T		14,676,308	

the results for the best model ($p = 1$ with $a_{\beta 0} = b_{\beta 0} = 1$) are as bad as in the independent case (see Table 4).

We also summarise the performance of the different analysis, in terms of the predicted reserves, in Table 5. The best performance is when we use hierarchical priors for all parameters, with similar numbers for the analysis made using the loss ratio.

5. CONCLUDING REMARKS

We presented an easily interpretable stochastic model for claims reserving, the DGM, that captures dependence across development years within a single triangle and combines information from multiple triangles through a judicious choice of prior distributions. Dependence is of any order $p \geq 0$ of past developments years, with a very appealing parametrisation that can be easily interpreted.

Posterior inference of our model can easily be obtained through a MCMC procedure implemented in `rjags`. Code is available upon request from the second author.

Our examples, using the National Association of Insurance Commissioners data set, show that our method works well for both small and large companies, with more accurate predictions than the benchmark model (ODP). Moreover, dependence is summarised through correlation coefficients across different development years within the same company ($\rho_{j,j+s,k}$) and the other interpretable parameters ($\alpha^*_{i,k}$ and $\pi^*_{j,k}$) provide some insight of the ultimate total claims per accident year i and how it is divided into the development years j, respectively, for each triangle k.

Due to the simple hierarchical structure of the DGM, it may be possible (at least numerically) to study further quantities, such as the claims development result – the difference between (a) the reserves predicted at time t and (b) the reserves predicted at time $t + 1$ plus the payments made at time $t + 1$. For more information on this "one-year risk," see Ohlsson and Lauzeningks (2009) or Merz and Wüthrich (2015).

Our DGM construction (2.1) is very flexible and can also be adapted to seasonal dependence. That is, if the seasonality of the data is s, then the model would be $X_{i,j,k} \mid \mathbf{Z} \sim \text{Ga}(\alpha_{i,k} + \sum_{l=0}^{p} Z_{i,j-sl,k},\ \beta_{j,k} + \sum_{l=0}^{p} \gamma_{j-sl,k})$. For instance, if the data were available in a monthly basis, we can assume a seasonality of order $s = 12$ so that the incremental claims made in January of the current year can depend on the January claims of the previous p years. On the other hand, our DGM does not currently consider negative dependence. These and other generalisations are worth studying in the future.

Acknowledgements

The authors are grateful to the insightful comments of two anonymous referees. This research was supported by the School of Applied Mathematics (EMAp) from Fundação Getulio Vargas (FGV), Brazil. The first author acknowledges support from *Asociación Mexicana de Cultura A.C.*–Mexico.

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**(3), 643–676.

AVANZI, B., TAYLOR, G., VU, P.A. and WONG, B. (2016) Stochastic loss reserving with dependence: A flexible multivariate Tweedie approach. *Insurance: Mathematics and Economics*, **71**, 63–78.

AVANZI, B., TAYLOR, G., VU, P.A. and WONG, B. (2020) A multivariate evolutionary generalised linear model framework with adaptive estimation for claims reserving. *Insurance: Mathematics and Economics*, **93**, 50–71.

DE ALBA, E. (2002) Bayesian estimation of outstanding claims reserves. *North American Actuarial Journal*, **6**(4), 1–20.

DE ALBA, E. and NIETO-BARAJAS, L.E. (2008) Claims reserving: A correlated Bayesian model. *Insurance: Mathematics and Economics*, **43**(3), 368–376.

DE JONG, P. (2006) Forecasting runoff triangles. *North American Actuarial Journal*, **10**(2), 28–38.

ECB. (2009) *Directive 2009/138/EC of the European parliament and of the council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II)*. Technical report.

ENGLAND, P.D. and VERRALL, R.J. (2002) Stochastic claims reserving in general insurance. *British Actuarial Journal*, **8**(3), 443–518.

FINMA. (2007) *Technical document on the Swiss Solvency Test*. Technical report.

GAO, G. (2018) Bayesian chain ladder models. In *Bayesian Claims Reserving Methods in Non-life Insurance with Stan*, pp. 73–115. Springer.

GAO, G. and MENG, S. (2018) Stochastic claims reserving via a bayesian spline model with random loss ratio effects. *ASTIN Bulletin*, **48**(1), 55–88.

GESMANN, M., MURPHY, D., ZHANG, Y.(W.), CARRATO, A., WUTHRICH, M., CONCINA, F. and DAL MORO, E. (2018) *ChainLadder: Statistical Methods and Models for Claims Reserving in General Insurance*. R package version 0.2.7.

GISLER, A. (2006) The estimation error in the chain-ladder reserving method: A Bayesian approach. *Astin Bulletin*, **36**(02), 554–565.

GISLER, A. and WÜTHRICH, M.V. (2008) Credibility for the chain ladder reserving method. *Astin Bulletin*, **38**(02), 565–600.

GUSZCZA, J. (2008) Hierarchical growth curve models for loss reserving. In *CAS Forum*, pp. 146–173.

https://doi.org/10.1017/asb.2020.36 Published online by Cambridge University Press

IBRAHIM, J.G. and LAUD, P.W. (1994) A predictive approach to the analysis of designed experiments. *Journal of the American Statistical Association*, **89**, 309–319.

KREMER, E. (2005). The correlated chain-ladder method for reserving in case of correlated claims developments. *Blätter DGVFM*, **27**, 315–322. https://doi.org/10.1007/BF02808313

LALLY, N. and HARTMAN, B. (2018) Estimating loss reserves using hierarchical Bayesian Gaussian process regression with input warping. *Insurance: Mathematics and Economics*, **82**, 124–140.

MACK, T. (1993) Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bulletin*, **23**(02), 213–225.

MERZ, M. and WÜTHRICH, M.V. (2010) Paid–incurred chain claims reserving method. *Insurance: Mathematics and Economics*, **46**(3), 568–579.

MERZ, M. and WÜTHRICH, M.V. (2015) Claims run-off uncertainty: The full picture. Available at SSRN 2524352, version of 3/Jul/2015.

MERZ, M., WÜTHRICH, M.V. and HASHORVA, E. (2013) Dependence modelling in multivariate claims run-off triangles. *Annals of Actuarial Science*, **7**(1), 3–25.

MEYERS, G. (2009) Stochastic loss reserving with the collective risk model. *Variance*, **3**(2), 239–269.

MEYERS, G. (2015) *Stochastic Loss Reserving Using Bayesian MCMC Models*. New York: Casualty Actuarial Society.

MEYERS, G.G. and SHI, P. (2011) The retrospective testing of stochastic loss reserve models. In *Casualty Actuarial Society E-Forum, Summer*.

NTZOUFRAS, I. and DELLAPORTAS, P. (2002) Bayesian modeling of outstanding liabilities incorporating claim count uncertainty. *North American Actuarial Journal*, **6**(1), 113–136.

OHLSSON, E. and LAUZENINGKS, J. (2009) The one-year non-life insurance risk. *Insurance: Mathematics and Economics*, **45**(2), 203–208.

PETERS, G.W., DONG, A.X.D. and KOHN, R. (2014) A copula based Bayesian approach for paid–incurred claims models for non-life insurance reserving. *Insurance: Mathematics and Economics*, **59**, 258–278.

PETERS, G.W., SHEVCHENKO, P.V. and WÜTHRICH, M.V. (2009) Model uncertainty in claims reserving within Tweedie's compound Poisson models. *ASTIN Bulletin*, **39**(1), 1–33.

PINHEIRO, P.J.R., ANDRADE E SILVA, J.M. and CENTENO, M.L. (2003) Bootstrap methodology in claim reserving. *The Journal of Risk and Insurance*, **70**, 701–714.

PLUMMER, M. (2018) *rjags: Bayesian Graphical Models using MCMC*. R package version 4-8.

RENSHAW, A.E. and VERRALL, R.J. (1998) A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, **4**(4), 903–923.

SHI, P. (2017) A multivariate analysis of intercompany loss triangles. *Journal of Risk and Insurance*, **84**(2), 717–737.

SHI, P., BASU, S. and MEYERS, G.G. (2012) A Bayesian log-normal model for multivariate loss reserving. *North American Actuarial Journal*, **16**(1), 29–51.

SHI, P. and HARTMAN, B.M. (2016) Credibility in loss reserving. *North American Actuarial Journal*, **20**(2), 114–132.

SMITH, A.F.M. and ROBERTS, G.O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*, **55**(1), 3–23.

SPIEGELHALTER, D.J., BEST, N.G., CARLIN, B.P. and VAN DER LINDE, A. (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.

SRIRAM, K. and SHI, P. (to appear) Stochastic loss reserving: A new perspective from a Dirichlet model. *Journal of Risk and Insurance*. https://doi.org/10.1111/jori.12311

TANNER, M.A. (1991) *Tools for Statistical Inference: Observed Data and Data Augmentation Methods*. New York: Springer.

TAYLOR, G. (2015) Bayesian chain ladder models. *Astin Bulletin*, **45**(01), 75–99.

TAYLOR, G. (2012) *Loss Reserving: An Actuarial Perspective*. Vol. 21. New York: Springer Science & Business Media.

TIERNEY, L. (1994) Markov chains for exploring posterior distributions. *Annals of Statistics*, **22**, 1701–1728.

VERRALL, R.J. (1991) On the estimation of reserves from loglinear models. *Insurance: Mathematics and Economics*, **10**, 75–80.
WÜTHRICH, M.V. (2010) Accounting year effects modeling in the stochastic Chain Ladder reserving method. *North American Actuarial Journal*, **14**, 235–255.
WUTHRICH, M.V. (2019) Non-life insurance: mathematics & statistics. Available at SSRN 2319328.
ZHANG, Y. and DUKIC, V. (2013) Predicting multivariate insurance loss payments under the bayesian copula framework. *Journal of Risk and Insurance*, **80**(4), 891–919.
ZHANG, Y., DUKIC, V. and GUSZCZA, J. (2012) A Bayesian non-linear model for forecasting insurance loss payments. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **175**(2), 637–656.

LUIS E. NIETO-BARAJAS (Corresponding author)
Department of Statistics
Instituto Tecnológico Autónomo de México (ITAM)
Mexico City, Mexico
E-Mail: lnieto@itam.mx

RODRIGO S. TARGINO
School of Applied Mathematics
Fundação Getulio Vargas (FGV)
Rio de Janeiro, Brazil
E-Mail: rodrigo.targino@fgv.br

APPENDIX

Posterior conditional distributions of model parameters $\boldsymbol{\theta}$ and the hyperparameters, for $i, j = 1, \ldots, n$, $k = 1, \ldots, K$, are

(i) Full conditional of $\alpha_{i,k}$

$$f(\alpha_{i,k} \mid \text{rest}) \propto \left\{ e^{-(b_{\alpha i} + \sum_{j=1}^{n+1-i} \gamma_{j,k})} \prod_{j=1}^{n+i-1} (\beta_{j,k} + \sum_{l=0}^{p} \gamma_{j-l,k}) x_{i,j,k} \right\}^{\alpha_{k,i}} \alpha_{i,k}^{a_{\alpha i} + \sum_{j=1}^{n+1-i} z_{i,j,k} - 1}$$

$$\times \frac{1}{\prod_{j=1}^{n+1-i} \Gamma(\alpha_{i,k} + \sum_{l=0}^{p} z_{i,j-l,k})} I(\alpha_{i,k} > 0)$$

(ii) Full conditional of $\beta_{j,k}$

$$f(\beta_{j,k} \mid \text{rest}) \propto \left(\beta_{j,k} + \sum_{l=0}^{p} \gamma_{j-l,k} \right)^{\sum_{i=1}^{n+1-j} (\alpha_{i,k} + \sum_{l=0}^{p} z_{i,j-l,k})} e^{-\beta_{j,k}(b_{\beta j} + \sum_{i=1}^{n+1-j} x_{i,j,k})}$$

$$\times \beta_{j,k}^{a_{\beta j} - 1} I(\beta_{j,k} > 0)$$

(iii) Full conditional of $\gamma_{j,k}$

$$f(\gamma_{j,k} \mid \text{rest}) \propto \left\{ \prod_{r=0}^{p} \left(\beta_{j+r,k} + \sum_{l=0}^{p} \gamma_{j+r-l,k} \right)^{\sum_{i=1}^{n+1-j} (\alpha_{i,k} + \sum_{l=0}^{p} z_{i,j+r-l,k})} \right\}$$
$$\times\, e^{-\gamma_{j,k}\left(b_{\gamma j} + \sum_{r=0}^{p} \sum_{i=1}^{n+1-j} x_{i,j+r,k} + \sum_{i=1}^{n+1-j} \alpha_{i,k}\right)} \gamma_{j,k}^{a_{\gamma j} + \sum_{i=1}^{n+1-j} z_{i,j,k} - 1} I(\gamma_{j,k} > 0)$$

(iv) Full conditional of $z_{i,j,k}$

$$f(z_{i,j,k} \mid \text{rest}) \propto \left\{ \alpha_{i,k} \gamma_{j,k} \prod_{r=0}^{p} (\beta_{j+r,k} + \sum_{l=0}^{p} \gamma_{j+r-l,k}) x_{i,j+r,k} \right\}^{z_{i,j,k}}$$
$$\times\, \Gamma^{-1}(z_{i,j,k} + 1) \left\{ \prod_{r=0}^{p} \Gamma^{-1}(\alpha_{i,k} + \sum_{l=0}^{p} z_{i,j+r-l,k}) \right\} I_{\{0,1,\ldots\}}(z_{i,j,k})$$

(v) Full conditionals of $a_{\alpha i}$ and $b_{\alpha i}$

$$f(a_{\alpha i} \mid \text{rest}) \propto \frac{(bv)^{K a_{\alpha i}}}{\Gamma^{K}(a_{\alpha i})} \left(\prod_{k=1}^{K} \alpha_{i,k} \right)^{a_{\alpha i}} \text{Ga}(a_{\alpha i} \mid a_{\alpha 0}, b_{\alpha 0}) \quad \text{and}$$
$$f(b_{\alpha i} \mid \text{rest}) \propto (b_{\alpha i})^{K a_{\alpha i}} e^{-b_{\alpha i} \sum_{k=1}^{K} \alpha_{i,k}} \text{Ga}(b_{\alpha i} \mid a_{\alpha 0}, b_{\alpha 0})$$

(vi) Full conditionals of $a_{\beta j}$ and $b_{\beta j}$

$$f(a_{\beta j} \mid \text{rest}) \propto \frac{(b_{\beta j})^{K a_{\beta j}}}{\Gamma^{K}(a_{\beta j})} \left(\prod_{k=1}^{K} \beta_{j,k} \right)^{a_{\beta j}} \text{Ga}(a_{\beta j} \mid a_{\beta 0}, b_{\beta 0}) \quad \text{and}$$
$$f(b_{\beta j} \mid \text{rest}) \propto (b_{\beta j})^{K a_{\beta j}} e^{-b_{\beta j} \sum_{k=1}^{K} \beta_{j,k}} \text{Ga}(b_{\beta j} \mid a_{\beta 0}, b_{\beta 0})$$

(vii) Full conditionals of $a_{\gamma j}$ and $b_{\gamma j}$

$$f(a_{\gamma j} \mid \text{rest}) \propto \frac{(b_{\gamma j})^{K a_{\gamma j}}}{\Gamma^{K}(a_{\gamma j})} \left(\prod_{k=1}^{K} \gamma_{j,k} \right)^{a_{\gamma j}} \text{Ga}(a_{\gamma j} \mid a_{\gamma 0}, b_{\gamma 0}) \quad \text{and}$$
$$f(b_{\gamma j} \mid \text{rest}) \propto (b_{\gamma j})^{K a_{\gamma j}} e^{-b_{\gamma j} \sum_{k=1}^{K} \gamma_{j,k}} \text{Ga}(b_{\gamma j} \mid a_{\gamma 0}, b_{\gamma 0})$$