

A Bayesian Mixture Model Accounting for Zeros and Negatives in the Loss Triangle

Michelle Xia¹ & David P. M. Scollnik²

¹ Division of Statistics, Northern Illinois University, DeKalb, U.S.A.

² Department of Mathematics and Statistics, University of Calgary, Calgary, Canada

Correspondence: Michelle Xia, DuSable 366, 1425 West Lincoln Highway, DeKalb, IL 60115, U.S.A. Tel: 1-815-753-6795. E-mail: cxia@niu.edu

Received: December 30, 2014 Accepted: January 19, 2015 Online Published: March 11, 2015

doi:10.5539/ijsp.v4n2p10 URL: <http://dx.doi.org/10.5539/ijsp.v4n2p10>

Abstract

In insurance loss reserving, a large portion of zeros are expected at the later development periods of an incremental loss triangle. Negative losses occur frequently in the incremental loss triangle due to actuarial practices such as subrogation and salvation. The nature of the distributions assumed by most stochastic models, such as the lognormal and over-dispersed Poisson distributions, brings restrictions on the zeros and negatives appearing in the loss triangle. In this paper, we propose a Bayesian mixture model for stochastic reserving under the situation where there are both zeros and negatives in the incremental loss triangle. A multinomial regression model will be applied to model the sign of the loss data, while the lognormal distribution is assumed for the loss magnitudes of negatives and positives. Bayesian generalized linear models will be fitted for both the mixture and magnitude models. The model will be implemented using the Markov chain Monte Carlo (MCMC) techniques in BUGS. Our model provides a realistic tool for stochastic reserving in the cases of zeros and negatives.

Keywords: mixture models, Bayesian inference, Markov chain Monte Carlo

1. Introduction

Determining an appropriate amount of loss reserve is very important for the financial stability of an insurance company. Although the traditional methods such as the chain ladder and Bornhuetter-Ferguson methods are simple to implement, they do not thoroughly address the stochastic nature of the data. Recent researchers focus more on the stochastic loss reserving methods, in which the variability and tail values of the distribution of the reserve are studied.

For stochastic reserving (England and Verrall, 2002), specific distributions such as the lognormal (Kremer, 1982), over-dispersed Poisson (Renshaw and Verrall, 1998; England, Verrall and Wuthrich, 2012), negative binomial (Verrall, 2000) and gamma (de Alba and Nieto-Barajas, 2008) are assumed for the loss reserving data. For these models, classical generalized linear model (Nelder and Wedderburn, 1972) structures can be fitted to the mean or other parameters of the reserve distribution. The application of the generalized linear structures gives rise to the stochastic models reproducing the chain ladder reserves. Comparisons of the chain ladder model and the stochastic models reproducing the chain ladder reserves can be found in papers such as Kremer (1982), Verrall (2000) and Mack and Venter (2000).

With its capability of incorporating external information (Verrall and England, 2005), Bayesian inference is used frequently in stochastic reserving. In papers such as Scollnik (2002) and de Alba (2002, 2006), external information is incorporated into the stochastic reserving model by specifying prior distributions for the parameters. Bayesian models for the chain ladder and Bornhuetter-Ferguson methods were introduced in Scollnik (2004), Verrall (2004) and England et al. (2012) respectively. Antonio and Plat (2013) proposed a Bayesian stochastic reserving model under the individual claim level. Most of the above models can be implemented using the Markov Chain Monte Carlo (MCMC) simulation method in the Bayesian software package BUGS (Spiegelhalter et al., 2004). Reviews of the MCMC method, BUGS, and their application in actuarial science can be found in Scollnik (2001). See Makov et al. (1996) for a more general discussion of Bayesian methods in actuarial science.

In real data, a large portion of zeros are expected at the later development periods of an incremental loss triangle. Negative losses occur frequently in the incremental loss triangle due to actuarial practices such as subrogation, salvation, cancellation of a claim, initial over-estimation of the case reserve, consequences of judicial decisions, and errors. A large number of zeros and negatives occur in the loss triangle, which may make some of the models inappropriate or even undefined.

To cope with the problems caused by zero and negative values in stochastic loss reserving, some improved models have been proposed in the recent literature. An improved Bayesian lognormal model was introduced by de Alba (2002, 2006) to extend the lognormal model (Kremer, 1982) to situations where there are negative values in the loss triangle. Kunkler (2004, 2006) proposed a Bayesian binomial model to handle the situation where there are either zeros or negatives in the loss triangle. None of these models, however, handle the more realistic situation when there are notable numbers of both zeros and negatives in the loss triangle.

In this paper, we will propose a Bayesian mixture model for handling the situation when there are both zeros and negatives in the loss reserving data. A multinomial regression model will be applied to model the sign of the loss data, while the lognormal distribution is assumed for the loss magnitudes of negatives and positives. Bayesian generalized linear models will be constructed for both the mixture and magnitude models. A chain ladder type model structure derived from the structure in Zehnwirth (1994) is used for the magnitude of the positive losses. The model will be implemented using the Markov chain Monte Carlo (MCMC) techniques in BUGS. We will perform a case study using the adjusted loss triangle from the 'Historical Loss Development Study' (1991) by the Reinsurance Association of America.

The paper is organized as follows. Section 2 presents the mixture model in the context of Bayesian inference. In Section 3, the model will be implemented in BUGS with a well-studied loss triangle data. Section 4 concludes the paper.

2. A Bayesian Mixture Model

The earlier models such as de Alba (2002, 2006) and Kunkler (2004, 2006) are all aimed at solving the problem of loss reserving when either zeros or negatives appear in the loss triangle, but not both together. No model has been proposed for extending the stochastic reserving models to situations where there are both zeros and negatives existing in the loss triangle. A Bayesian mixture model will be proposed in this section to extend the conventional stochastic loss reserving model to a more general situation where there are a considerable number of both zero and negative values in the loss triangle.

2.1 Model for the Mixture Data

2.1.1 Distribution of Mixture Data

Here, we will use notation consistent with that in Kunkler (2004, 2006). Let y_{ij} denote the incremental losses in the loss triangle. Based on the sign of the data, the incremental loss triangle can be split into three subsets containing values of negatives, zeros, and positives, respectively. The three subsets are defined as $S^{(-)} = \{y_{ij} : y_{ij} < 0\}$, $S^{(0)} = \{y_{ij} : y_{ij} = 0\}$ and $S^{(+)} = \{y_{ij} : y_{ij} > 0\}$.

Denoting n_a as the number of accident periods included in the data, a mixture data triangle $\mathbf{z} = \{z_{ij} : i = 1, 2, \dots, n_a; j = 1, 2, \dots, n_a - i + 1\}$ can be defined for modelling the sign of the incremental loss triangle where

$$z_{ij} = \begin{cases} (1, 0, 0)^T & \text{if } y_{ij} < 0 \\ (0, 1, 0)^T & \text{if } y_{ij} = 0 \\ (0, 0, 1)^T & \text{if } y_{ij} > 0 \end{cases} .$$

The sum of the mixture data for each development year stands for the number of y_{ij} observations that are negative, zero, or positive from our definition. This can be written as

$$\mathbf{z}_j = \begin{pmatrix} z_{1j} \\ z_{2j} \\ z_{3j} \end{pmatrix} = \sum_{i=1}^{n_a} z_{ij} = \begin{pmatrix} \text{Number of } y_{ij} \text{ observations} < 0 \\ \text{Number of } y_{ij} \text{ observations} = 0 \\ \text{Number of } y_{ij} \text{ observations} > 0 \end{pmatrix} .$$

Denote $P(y_{ij} < 0) = \lambda_{1j}$, $P(y_{ij} = 0) = \lambda_{2j}$, then $P(y_{ij} > 0) = 1 - \lambda_{1j} - \lambda_{2j}$ as the mixture probabilities. Assuming conditional independence for losses from different accident years, the sum of the mixture data in each development year also follows a multinomial distribution with its probability distribution function given by

$$p(z_j|\lambda_j) = \binom{n_{a_j}}{z_{1j} z_{2j} z_{3j}} \lambda_{1j}^{z_{1j}} \lambda_{2j}^{z_{2j}} (1 - \lambda_{1j} - \lambda_{2j})^{n_{a_j} - z_{1j} - z_{2j}},$$

where

$$\lambda_j = \begin{pmatrix} \lambda_{1j} \\ \lambda_{2j} \\ \lambda_{3j} \end{pmatrix} = \begin{pmatrix} \lambda_{1j} \\ \lambda_{2j} \\ 1 - \lambda_{1j} - \lambda_{2j} \end{pmatrix} \quad \text{and} \quad \binom{n_{a_j}}{z_{1j} z_{2j} z_{3j}} = \frac{n_{a_j}!}{z_{1j}! z_{2j}! z_{3j}!}.$$

2.1.2 Generalized Linear Models

In the actual loss triangle, there tends to be more zeros and negatives in the later stage of development years. So we can assume that the proportion of zeros and negatives depends only on the development year as in Kunkler (2004, 2006). A Bayesian GLM for the multinomial probabilities on the development year j can be applied to model this structure. Two commonly used link functions for the multinomial distribution are the logistic and probit links (Dobson, 2002, Chapter 8, Page 135-148). Without loss of generality, only the logistic link is used for our analysis (i.e., a cumulative logistic regression model is assumed).

With a different factor δ_{ld} introduced for each development year in the cumulative logistic model, the piecewise linear relationship (Kunkler, 2004, 2006) gives a flexible structure. This model structure can be written as

$$\log\left(\frac{\lambda_{lj}}{\lambda_{3j}}\right) = \sum_{d=0}^{j-1} \delta_{ld}, \quad l = 1, 2.$$

With the above structure, some of the parameters can be set to zero or assigned equal values based on the size of the dataset available. As loss triangles are usually summarized by accident and development years, the datasets used for loss reserving are usually very small. For relatively small datasets, a simple linear regression on $j - 1$ can be used as a special case of the above model. That is,

$$\log\left(\frac{\lambda_{lj}}{\lambda_{3j}}\right) = \delta_{l0} + (j - 1)\delta_{l1}, \quad l = 1, 2. \tag{1}$$

2.2 Model for the Magnitude Data

2.2.1 Modelling Magnitude Data

To ease the analysis for this section, simplified notations for the distributions of the magnitudes of the positive and negative data can be introduced as

$$p(\mathbf{y}^-|\theta_1) = p(-y_{ij}|z_{ij} = -1, \theta_1)$$

$$p(\mathbf{y}^+|\theta_2) = p(y_{ij}|z_{ij} = 1, \theta_2),$$

where $\theta_1 = (\beta_1, \sigma_1^2)$ and $\theta_2 = (\beta_2, \sigma_2^2)$.

As discussed in Section 1, many distributions such as the lognormal (Kremer, 1982; de Alba, 2002, 2006; Kunkler, 2004, 2006), over-dispersed Poisson (Renshaw and Verrall, 1998; England et al., 2012), negative binomial (Verrall, 2000) and gamma (de Alba and Nieto-Barajas, 2008) can be assumed for the loss magnitude data. For demonstration purposes, lognormal sampling distributions are assumed for the loss magnitude data \mathbf{y}^- and \mathbf{y}^+ in our analysis. That is,

$$\log(\mathbf{y}^-)|\theta_1 \sim N(\mathbf{X}_1\beta_1, \sigma_1^2\mathbf{I}_1)$$

$$\log(\mathbf{y}^+)|\theta_2 \sim N(\mathbf{X}_2\beta_2, \sigma_2^2\mathbf{I}_2), \tag{2}$$

where \mathbf{I}_1 and \mathbf{I}_2 are identity matrices of $n_{y^+} \times n_{y^+}$, $n_{y^-} = \sum_{i=1}^{n_a} \sum_{j=1}^{n_a-i+1} \mathbf{I}_{(y_{ij}<0)}$, $n_{y^+} = \sum_{i=1}^{n_a} \sum_{j=1}^{n_a-i+1} \mathbf{I}_{(y_{ij}>0)}$, β_1 is a $k_{\beta_1} \times 1$ parameter vector, β_2 is a $k_{\beta_2} \times 1$ parameter vector, \mathbf{X}_1 is a design matrix of dimension $n_{y^-} \times k_{\beta_1}$, and \mathbf{X}_2 is a design matrix of dimension $n_{y^+} \times k_{\beta_2}$.

2.2.2 Generalized Linear Models

Zehnwirth (1994) put forward a flexible regression structure which can be applied to the loss magnitude data of \mathbf{y}^- and \mathbf{y}^+ . The model can be written as

$$(\mathbf{X}_{\beta_l}\beta_l)_{ij} = \alpha_{li} + \sum_{d=1}^{j-1} \gamma_{ld} + \sum_{t=1}^{i+j-2} \eta_{lt}, \quad l = 1, 2.$$

The α_{li} parameters are for modelling the effect of accident year, while the γ_{ld} and η_{lt} parameters are chosen to catch the effects of development year and calendar year respectively. Note that the observed data in the loss triangle provide no information concerning the η_{lt} parameters for future calendar years. Hence it is impossible to predict for future losses without making adjustments to the model and/or including prior information.

Setting zeros for all the η_{lt} parameters gives a structure comparable to that of the chain ladder model. The model under this structure reduces to

$$(\mathbf{X}_{\beta_l}\beta_l)_{ij} = \alpha_{li} + \sum_{d=1}^{j-1} \gamma_{ld}, \quad l = 1, 2, \tag{3}$$

where the transformed $e^{\gamma_{ld}}$ parameters are analogous to the development ratios in the chain ladder model.

Kunkler (2004, 2006) also introduced a simplified version of this model for relatively small datasets, in which

$$(\mathbf{X}_{\beta_l}\beta_l)_{ij} = \alpha_l + (j - 1)\gamma_l + (i + j - 2)\eta_l, \quad l = 1, 2.$$

There is an obvious limitation for this structure in that the trend may not be linear in either the development year or calendar year for real data.

When negatives appear in the loss triangle and the size of the data set is relatively small, some smoothing structures (see e.g., Zehnwirth, 1985) can be introduced in order to avoid possible over parameterization. One of the choices may be the Hoerl curve (Zehnwirth, 1985) given by

$$(\mathbf{X}_{\beta_l}\beta_l)_{ij} = c_l + a_{li} + b_{li} \log(j) + r_{li} j, \tag{4}$$

which provides a development pattern similar to those of the claim triangles.

A special case of the model in Equation (4) is when $b_i = b$ and $r_i = r$ for all i , assuming a common runoff pattern for all accident years. In this case, the model can be written as

$$(\mathbf{X}_{\beta_l}\beta_l)_{ij} = c_l + a_{li} + b_l \log(j) + r_l j.$$

2.3 Bayesian Inference

2.3.1 Overall Model Structure

In the framework of Bayesian inference, the outstanding liabilities will be estimated based on the posterior predictive distributions of the magnitude and mixture data for future incremental losses. In the analysis of this subsection, the future incremental data triangle and future mixture data triangle will be denoted as

$$\begin{aligned} \tilde{\mathbf{y}} &= \{\tilde{y}_{ij} : i = 2, \dots, n_a; j = n_a - i + 1, \dots, n_d\} \\ \tilde{\mathbf{z}} &= \{\tilde{z}_{ij} : i = 2, \dots, n_a; j = n_a - i + 1, \dots, n_d\}, \end{aligned}$$

where n_d denotes the number of development periods in the data.

Denoting $\Delta = \begin{pmatrix} \delta_{10} & \delta_{20} \\ \delta_{11} & \delta_{21} \end{pmatrix}$ as the parameters for cumulative logistic regression, the joint posterior distribution for θ and Δ can be written as

$$\pi(\theta, \Delta | \mathbf{y}, \mathbf{z}) \propto \pi(\theta)\pi(\Delta) p(\mathbf{y}, \mathbf{z} | \theta, \Delta),$$

where the joint sampling distribution for \mathbf{y} and \mathbf{z} in the above formulas can be obtained $p(y_{ij}, z_{ij}|\theta, \Delta) = p(z_{ij}|\Delta) p(y_{ij}|z_{ij}, \theta)$.

By averaging over z_{ij} , the marginal sampling distribution of y_{ij} gives us a form with which we can focus on the claim amount of our interest. Similar to Kunkler (2004, 2006), we obtain

$$\begin{aligned} p(y_{ij}|\theta, \Delta) &= P(z_{ij} = -1|\Delta) p(y_{ij}|z_{ij} = -1, \theta_1) + P(z_{ij} = 0|\Delta) p(y_{ij}|z_{ij} = 0) \\ &\quad + P(z_{ij} = 1|\Delta) p(y_{ij}|z_{ij} = 1, \theta_2) \\ &= \lambda_{1j} p(y_{ij}|z_{ij} = -1, \theta_1) + \lambda_{2j} p(y_{ij}|z_{ij} = 0) \\ &\quad + (1 - \lambda_{1j} - \lambda_{2j}) p(y_{ij}|z_{ij} = 1, \theta_2) . \end{aligned}$$

In the above formula, we can verify that $p(y_{ij}|z_{ij} = 0) = I_{(y_{ij}=0)}$.

To predict future losses, the joint posterior predictive distribution of \tilde{y}_{ij} and \tilde{z}_{ij} needs to be used with its formula given by

$$p(\tilde{y}_{ij}, \tilde{z}_{ij}|\mathbf{y}, \mathbf{z}) = p(\tilde{z}_{ij}|\mathbf{z}) p(\tilde{y}_{ij}|\tilde{z}_{ij}, \mathbf{y}) .$$

Treating the future mixture triangle $\tilde{\mathbf{z}}$ as nuisance parameters, the marginal posterior predictive distribution of the future incremental triangle $\tilde{\mathbf{y}}$ can be obtained as

$$\begin{aligned} p(\tilde{y}_{ij}|\mathbf{y}, \mathbf{z}) &= \lambda_{1j} p(\tilde{y}_{ij}|\tilde{z}_{ij} = -1, \mathbf{y}) + \lambda_{2j} p(\tilde{y}_{ij}|\tilde{z}_{ij} = 0, \mathbf{y}) \\ &\quad + (1 - \lambda_{1j} - \lambda_{2j}) p(\tilde{y}_{ij}|\tilde{z}_{ij} = 1, \mathbf{y}) . \end{aligned}$$

2.3.2 Posterior Analysis for Mixture Parameters

Under the GLM structure given in Subsection 2.1, the posterior distribution can be obtained for the cumulative logistic regression parameters Δ . That is

$$\pi(\Delta|\mathbf{z}) \propto \pi(\Delta) \prod_{j=1}^{n_d} p(\mathbf{z}_j|\Delta) ,$$

where $\pi(\Delta)$ is the prior distribution for Δ , and $p(\mathbf{z}_j|\Delta)$ is the multinomial sampling mixture distribution for \mathbf{z}_j given by

$$p(\mathbf{z}_j|\Delta) \propto \lambda_1(\Delta)^{z_{1j}} \lambda_2(\Delta)^{z_{2j}} (1 - \lambda_1(\Delta) - \lambda_2(\Delta))^{n_{aj} - z_{1j} - z_{2j}} .$$

With the logistic link function, the λ_l parameters can be expressed in terms of Δ as

$$\begin{aligned} \lambda_1(\Delta) &= \frac{\exp[\delta_{10} + (j-1)\delta_{11}]}{\exp[\delta_{10} + (j-1)\delta_{11}] + \exp[\delta_{20} + (j-1)\delta_{21}] + 1} \\ \lambda_2(\Delta) &= \frac{\exp[\delta_{20} + (j-1)\delta_{21}]}{\exp[\delta_{10} + (j-1)\delta_{11}] + \exp[\delta_{20} + (j-1)\delta_{21}] + 1} . \end{aligned}$$

This part of the model can be implemented using MCMC algorithms such as the Metropolis-Hastings algorithm.

2.3.3 Posterior Analysis for Sampling Distribution

Under the lognormal GLM model formulated in Equation (2), the joint posterior distribution for β_l and σ_l ($l = 1, 2$) can be factored into

$$\begin{aligned} \pi(\beta_1, \sigma_1^2|\mathbf{y}^-) &= \pi(\beta_1|\sigma_1^2, \mathbf{y}^-) \pi(\sigma_1^2|\mathbf{y}^-) \\ \pi(\beta_2, \sigma_2^2|\mathbf{y}^+) &= \pi(\beta_2|\sigma_2^2, \mathbf{y}^+) \pi(\sigma_2^2|\mathbf{y}^+) . \end{aligned}$$

Noninformative uniform prior on $(\beta_l, \log \sigma_l)$ may be assumed in order to make the normal regression model easier to implement. Then it is straightforward to verify that

$$\begin{aligned} \beta_1|\sigma_1^2, \mathbf{y}^- &\sim N(\hat{\beta}_1, (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \sigma_1^2) \\ \beta_2|\sigma_2^2, \mathbf{y}^+ &\sim N(\hat{\beta}_2, (\mathbf{X}_2^T \mathbf{X}_2)^{-1} \sigma_2^2) \\ \sigma_1^2|\mathbf{y}^- &\sim \text{Inv-}\chi^2(n_{y^-} - k_{\beta_1}, \hat{\sigma}_1^2) \\ \sigma_2^2|\mathbf{y}^+ &\sim \text{Inv-}\chi^2(n_{y^+} - k_{\beta_2}, \hat{\sigma}_2^2) , \end{aligned}$$

where β_l and σ_l^2 ($l = 1, 2$) can be estimated by

$$\begin{aligned}\hat{\beta}_1 &= (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \log(\mathbf{y}^-) \\ \hat{\beta}_2 &= (\mathbf{X}_2^T \mathbf{X}_2)^{-1} \mathbf{X}_2^T \log(\mathbf{y}^+) \\ \hat{\sigma}_1^2 &= \frac{1}{n_{y^-} - k_{\beta_1}} (\log(\mathbf{y}^-) - \mathbf{X}_1 \hat{\beta}_1)^T (\log(\mathbf{y}^-) - \mathbf{X}_1 \hat{\beta}_1) \\ \hat{\sigma}_2^2 &= \frac{1}{n_{y^+} - k_{\beta_2}} (\log(\mathbf{y}^+) - \mathbf{X}_2 \hat{\beta}_2)^T (\log(\mathbf{y}^+) - \mathbf{X}_2 \hat{\beta}_2).\end{aligned}$$

With the above analytical forms of the posterior distributions and posterior predictive distributions for all the unobserved variables (e.g., the regression coefficients and variance parameters), we will be able to implement our multinomial mixture model, using MCMC simulation methods such as the Metropolis-Hasting's algorithm. An alternative approach is to use a Bayesian software package such as BUGS that implements these algorithms after we specify the forms of the prior and sampling distributions. Due to its simple implementation, we will use BUGS to implement our model in the next section.

3. Model Implementation

The use of the specialized Bayesian software BUGS makes it relatively easy to implement our multinomial mixture model. Bayesian models including GLMs can be implemented in BUGS by specifying the sampling distributions, prior distributions and the regression functions. Hence, the posterior analysis given in Section 2 for our multinomial mixture model will not be needed for our model fitting in BUGS.

In this section, our multinomial mixture model will be fitted to the loss triangle adjusted from the original data at 'Historical Loss Development Study' (1991). Particularly for the positive magnitude where we have plenty of data, a GLM structure different from that in Kunkler (2004, 2006) will be constructed. The model is based on the three parameter lognormal model, with the interpretation of parameters more comparable to those from the chain ladder method. A calendar year trend parameter is introduced into this chain ladder type of structure.

The variances of the positive and negative magnitudes will be modelled using the same method as in the *ols.g()* function in MatLab (LeSage, 1999, page 176). The parameters for the mixture and magnitude models as well as the reserves will be estimated and compared to those from Kunkler (2006).

3.1 The Data

Since claim data from insurance companies are mostly confidential, it is generally hard to obtain a real dataset for a specific line of business that may contain zeros and negatives in the loss triangle. The open source datasets, on the other hand, are usually aggregate data from multiple insurance companies or multiple lines of business, that will generally not contain zeros or negatives due to the aggregation. For illustrative purposes, the original loss triangle from the 'Historical Loss Development Study' (1991) by the Reinsurance Association of America listed in Table 1 is adjusted so that it contains both values of zeros and negatives. The negative losses in our adjusted loss triangle are the same as those used by Kunkler (2006).

Table 1. Adjusted Incremental Loss Triangle with Zeros and Negatives.

Accident year	Development year									
	1	2	3	4	5	6	7	8	9	10
1	5012	3257	2638	-898	1734	2642	1828	599	-54	172
2	-106	4179	-1111	5270	3116	1817	-103	0	535	
3	3410	5582	4881	2268	2594	3479	0	603		
4	5655	5900	4211	5500	2159	2658	984			
5	1092	8473	6271	6333	3786	-225				
6	1513	4932	5257	1233	2917					
7	-557	3463	6926	1368						
8	1351	5596	6165							
9	3133	2262								
10	2063									

3.2 Model Construction

3.2.1 Modelling Mixture Data

In Section 2, we proposed using the multinomial distribution to model the mixture data, i.e.

$$z_{ij} \sim \text{multinomial}(\lambda_j, 1).$$

From the mixture data we can see that there tends to be more zeros and negatives during the later development years. Since we have the same negatives as those in Kunkler (2006), for the negative probability λ_{1j} we will use the same logit model structure given in Kunkler (2006). The model for the negative probability is given by

$$\log\left(\frac{\lambda_{1j}}{\lambda_{3j}}\right) = \delta_{10} + (j-5)\delta_{11}I_{(j>5)}, \quad j = 1, 2, \dots, n_a. \quad (5)$$

Since the zero values are only observed after the 6th development period, we assume the probability for zeros λ_{2j} differs only after the 6th development period. So we may use the following model structure for zeros:

$$\log\left(\frac{\lambda_{2j}}{\lambda_{3j}}\right) = \delta_{20} + (j-6)\delta_{21}I_{(j>6)}, \quad j = 1, 2, \dots, n_a. \quad (6)$$

Similar to Kunkler (2006), we assume the same diffuse priors for the δ parameters. That is,

$$\delta_{li} \sim N(0, 100), \quad l = 2, 3; i = 0, 1.$$

3.2.2 Modelling Magnitude Data

We assume the magnitudes of both positives and negatives follow lognormal distributions with different means and variances. That is,

$$|y_{ij}| \sim LN\left(\mu_{ij}^-, \frac{\sigma_{ij}^2}{\omega^-}\right) \quad \text{if } y_{ij} < 0 \quad (7)$$

$$|y_{ij}| \sim LN\left(\mu_{ij}^+, \frac{\sigma_{ij}^2}{\omega^+}\right) \quad \text{if } y_{ij} > 0. \quad (8)$$

Here, we use notations slightly different to those in the previous section in order to allow for a more flexible variance structure, as well revealing more detailed information on the expected loss amount for a specific combination of accident and development years.

Positive Magnitude

For the magnitude data of positives, we will model the mean of the lognormal distribution with the chain ladder type structure in Equation (3) with the same calendar trend factor ι from Kunkler (2006). The model structure is given by

$$\mu_{ij}^+ = \alpha_i^+ + \sum_{d=1}^{j-1} \gamma_d^+ + (i+j-2)\iota. \quad (9)$$

Diffuse priors of $N(0, 1000)$ are assumed for all of the α , γ and ι parameters in the above model. Since the loss triangle contains zeros, we are not able to use a common lognormal distribution for the loss triangle. The magnitudes of the positives and negatives have to be modelled using two lognormal distributions. Two vectors are used to store the values of i and j for the magnitude data.

Negative Magnitude

For the magnitude data of the negatives, a simplified model structure needs to be used to for relatively small datasets. Since we assume the same negative losses as in Kunkler (2006), the model in Kunkler (2006) can be used to model the mean of negative magnitude. The model structure for the negative magnitude is given by

$$\mu_{ij}^- = \alpha_1^- + [I_{(j \leq 3)}(j-1) + I_{(j > 3)}2]\gamma_1^- + I_{(j > 3)}(j-3)\gamma_2^- + (i+j-2)\iota. \quad (10)$$

Similar to the magnitude model for the positives, diffuse priors of $N(0, 1000)$ are assumed for all of the α , γ and ι parameters.

Modelling Variance

For the variance parameters σ_{ij}^2 in Equations (6) and (7), we will use a model specification similar to that in the *ols_g()* function in LeSage (1999, page 176). In our simulation, the prior density specification of σ_{ij}^2 is defined in this way:

$$\begin{aligned}\sigma_{ij}^2 &= sige \times v_{ij} \\ \frac{r}{v_{ij}} &\sim i.i.d. \frac{\chi^2(r)}{r} \\ sige &\sim U(0, 100),\end{aligned}\tag{11}$$

where we used the fixed value of $r = 100$ so as to be consistent with Kunkler (2006). For the prior distribution of *sige*, we used a diffuse uniform prior $U(0, L)$, with $L = 100$.

We estimate the parameters ω^+ and ω^- in Equations (7) and (8) based on the same analysis given in Kunkler (2006, pages 553-554). The estimates of ω^+ and ω^- are obtained by running the model using the variance construction in Equation (11). The estimates are given by $\omega^+ = 1/\text{var}(\mathbf{r}^+)$ and $\omega^- = 1/\text{var}(\mathbf{r}^-)$, where \mathbf{r}^+ and \mathbf{r}^- are the residual vectors for the positive and negative log magnitudes. That is,

$$\begin{aligned}\mathbf{r}^+ &= \log(\mathbf{y}^+) - \widehat{\log(\mathbf{y}^+)} \\ \mathbf{r}^- &= \log(\mathbf{y}^-) - \widehat{\log(\mathbf{y}^-)},\end{aligned}$$

where $\widehat{\log(\mathbf{y}^+)}$ and $\widehat{\log(\mathbf{y}^-)}$ are the predicted values of $\log(\mathbf{y}^+)$ and $\log(\mathbf{y}^-)$ in the upper loss triangle for the existing data. The values of $\widehat{\log(\mathbf{y}^+)}$ and $\widehat{\log(\mathbf{y}^-)}$ are calculated from Equations (9) and (10) using the posterior estimates of the α , γ and ι parameters.

For the variance calculation, we use the n normalized sample variance. That is, for the data $\mathbf{x} = (x_1, x_2, \dots, x_n)$ we let

$$\text{var}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n (\bar{x} - x_i)^2,$$

where \bar{x} is the sample mean given by $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

Using the data, we obtained the estimated values of ω^+ and ω^- as $\hat{\omega}^+ = 6.3880$ and $\hat{\omega}^- = 5.1547$. These values are used directly for our model implementation in BUGS.

3.3 Estimation and Prediction

3.3.1 Convergence of MCMC Simulation

Three chains with dispersed initial values are used for our simulation. With the uniform distribution $U(0, 100)$ assumed for the parameter *sige*, the simulation converges very quickly, i.e. before iteration 1,000. This can be observed from the history plots of the 3 chains for each parameter or quantity of interest. The three chains mix very quickly.

The refined potential scale reduction factors \hat{R}_c (Brooks and Gelman, 1998) are also monitored in order to diagnose convergence. In the case of all the parameters and quantities of interest, the corresponding refined potential scale reduction factor \hat{R}_c converged to approximately 1 within about 1,000 iterations. The values of the refined potential scale reduction factors \hat{R}_c are less than 1.01 for all the quantities after 2,000 iterations.

We will use the simulated values from iterations 2,001 to 32,000 from all three chains for our posterior analysis in the following. A total of 30,000 posterior samples will be used for our posterior analysis of parameters and reserves.

3.3.2 Mixture Model

For the multinomial mixture model in Subsection 3.2.1, the parameters for the logit models in Equations (5) and (6) are estimated by posterior simulation. The estimates of the parameters, their standard deviations and percentiles are listed in Table 2.

Table 2. Estimates for Multinomial Mixture Model

Parameter	mean	sd	MC error	2.50%	median	97.50%
δ_{10}	-2.176	0.503	0.001	-3.249	-2.144	-1.279
δ_{11}	0.261	0.324	0.001	-0.417	0.273	0.867
δ_{20}	-4.201	1.175	0.004	-6.984	-4.038	-2.388
δ_{21}	0.856	0.625	0.003	-0.430	0.865	2.074

From Table 2 we observe that the estimates of the δ_{1i} parameters from the multinomial model have very close absolute values with those of the δ_i parameters from the binomial model in Kunkler (2006) except for their signs. The closeness of the absolute values are due to the fact that we have the same negative values as those in Kunkler (2006). The difference in the sign is from the factor that the logit function we defined for the negatives is based on the ratio of negative probability over the positive, while the one in Kunkler (2006) is defined based on the opposite ratio.

The posterior mean for each future development year of each accident year in the lower part of the loss triangle is listed in Table 3 and 4 respectively for negatives and zeros. From the posterior mean of the negative probability we observe that the probability stays the same for the first 5 development years, and increases with development years from then on. Similarly, we observe that in the first 6 development years, the probability of zero is very small, and the probability increases significantly with every development year thereafter.

Table 3. Posterior Mean for Probability of Negatives, Multinomial Model

Accident year	Development year									
	1	2	3	4	5	6	7	8	9	10
1										
2										0.236
3									0.236	0.243
4								0.208	0.235	0.243
5							0.169	0.209	0.237	0.244
6						0.130	0.168	0.208	0.238	0.243
7					0.107	0.133	0.169	0.209	0.239	0.245
8				0.109	0.108	0.133	0.169	0.208	0.238	0.244
9			0.109	0.110	0.108	0.134	0.168	0.209	0.238	0.243
10		0.110	0.109	0.106	0.110	0.135	0.171	0.208	0.236	0.245

Table 4. Posterior Mean for Probability of Zeros, Multinomial Model

Accident year	Development year									
	1	2	3	4	5	6	7	8	9	10
1										
2										0.313
3									0.179	0.313
4								0.082	0.180	0.313
5							0.039	0.707	0.181	0.314
6						0.021	0.040	0.708	0.178	0.310
7					0.021	0.021	0.039	0.709	0.179	0.312
8				0.021	0.021	0.021	0.039	0.709	0.180	0.312
9			0.021	0.021	0.022	0.021	0.037	0.708	0.180	0.311
10		0.022	0.021	0.021	0.021	0.021	0.039	0.709	0.179	0.311

From Table 3 and 4, We observe that the probabilities of negatives and zeros increase with development years, which is consistent to what we observe from the data in Table 1. It is also consistent to the real practice that subrogations and salvations occur later in the development due to things such as time required for processing and investigation, and there will be no or little payment after many claims have been closed.

3.3.3 Magnitude Model

Positive Magnitude

For the lognormal positive magnitude model specified in Equations (9) and (10), we obtained the posterior predictive estimates of the α_i^+ , γ_d^+ and ι parameters. The estimates of the parameters, their standard deviations and percentiles are listed in Table 5.

Table 5. Parameter Estimation of Positive Magnitude, Multinomial Model

Parameter	mean	sd	MC error	2.50%	median	97.50%
α_1^+	7.831	0.299	0.003	7.235	7.831	8.416
α_2^+	7.813	0.390	0.006	7.042	7.813	8.580
α_3^+	7.635	0.387	0.009	6.872	7.634	8.402
α_4^+	7.561	0.479	0.013	6.611	7.561	8.516
α_5^+	7.470	0.595	0.018	6.286	7.471	8.646
α_6^+	6.848	0.704	0.022	5.455	6.851	8.258
α_7^+	6.669	0.856	0.026	4.970	6.670	8.375
α_8^+	6.676	0.953	0.031	4.772	6.679	8.592
α_9^+	6.335	1.090	0.035	4.167	6.335	8.500
α_{10}^+	6.218	1.258	0.039	3.705	6.218	8.725
γ_1^+	0.434	0.307	0.006	-0.169	0.433	1.045
γ_2^+	-0.082	0.309	0.005	-0.694	-0.081	0.524
γ_3^+	-0.702	0.339	0.005	-1.372	-0.701	-0.032
γ_4^+	-0.319	0.346	0.006	-0.998	-0.319	0.363
γ_5^+	-0.201	0.379	0.005	-0.949	-0.201	0.551
γ_6^+	-0.807	0.501	0.006	-1.796	-0.808	0.184
γ_7^+	-0.910	0.581	0.007	-2.066	-0.909	0.238
γ_8^+	-0.356	0.757	0.008	-1.852	-0.353	1.132
γ_9^+	-1.149	0.857	0.007	-2.827	-1.152	0.556

We observe from Table 5 that the both accident year effects (i.e., α^+ parameters) decreases with the accident year, and the development year effects (i.e., γ^+ parameters) also decreases as with the payoff of more claims there expect to be lower payments in the later years. The estimates and MC error are comparable to those from Kunkler (2004, 2006) as we are using very similar data and model structures, despite the difference in dealing with zeros and negatives all at once.

Negative Magnitude

Similar to the previous subsection, the posterior predictive means for the α^- and γ_i^- parameters in Equation (6) and (10) can be estimated using the simulated samples from BUGS. The estimates of the parameters, their standard deviations and percentiles are listed in Table 6. We observe that the estimates for the parameters are very close to those from Kunkler (2006), as we use the same model and same negative data for our simulation.

Table 6. Parameter Estimation of Negative Magnitude, Multinomial Model

Parameter	mean	sd	MC error	2.50%	median	97.50%
α^-	4.948	0.614	0.016	3.726	4.948	6.154
γ_1^-	0.819	0.311	0.003	0.207	0.817	1.438
γ_2^-	-0.698	0.175	0.004	-1.046	-0.698	-0.352

From the above table, we observe that the estimates and MC errors for the accident year and development year parameters for our negative model are very close to those from Kunkler (2006), as we have the same negatives values in the loss triangle and we are using the same model structure. Since the negative values we added are arbitrarily chosen, the estimates would not possess a meaningful interpretation for real practice.

Common Parameters and Ultimate Losses

We also estimated the common parameters of the positive and negative magnitudes models, i.e. the calendar trend factor ι and the τ parameters for the inverse variances (i.e., precisions). The posterior mean, standard deviation and percentiles of the calendar parameters are listed in Table 7. Note that the parameter for the cumulative logistic regression model does not have a meaningful interpretation.

Table 7. Parameter Estimation of Calendar Trend Factor, Multinomial Model

Parameter	mean	sd	MC error	2.50%	median	97.50%
ι	0.157	0.126	0.004	-0.095	0.156	0.407

The posterior means of the precision parameters are listed in Table 8.

Table 8. Posterior Mean for Precision Parameters, Multinomial Model

Acc year	Development year									
	1	2	3	4	5	6	7	8	9	10
1	0.572	0.580	0.574	0.578	0.580	0.582	0.577	0.581	0.578	0.578
2	0.576	0.581	0.581	0.577	0.582	0.578	0.581	0.578	0.578	0.579
3	0.581	0.582	0.582	0.580	0.582	0.580	0.579	0.580	0.578	0.579
4	0.575	0.583	0.579	0.577	0.580	0.582	0.577	0.578	0.579	0.579
5	0.556	0.579	0.583	0.576	0.582	0.577	0.578	0.578	0.579	0.579
6	0.581	0.580	0.581	0.573	0.579	0.579	0.579	0.579	0.579	0.578
7	0.577	0.582	0.575	0.576	0.578	0.579	0.578	0.578	0.579	0.579
8	0.574	0.580	0.581	0.579	0.579	0.579	0.578	0.578	0.579	0.578
9	0.576	0.576	0.579	0.578	0.578	0.578	0.579	0.579	0.579	0.579
10	0.578	0.579	0.579	0.579	0.578	0.578	0.579	0.578	0.578	0.578

3.3.4 Reserves

The posterior mean reserve for each accident year and development year in the lower part of the loss triangle is estimated. The triangle reserve estimates are listed in Table 9.

Table 9. Mean Reserve by Accident & Development Years, Multinomial Model

Accident year	Development year									
	1	2	3	4	5	6	7	8	9	10
1										
2										117
3									428	111
4								608	457	119
5							1585	650	490	127
6						1933	963	387	293	70
7					1990	1916	957	380	287	64
8				2637	2324	2248	1110	440	336	73
9			3637	2098	1913	1846	916	342	261	50
10		4238	3988	2353	2110	2079	1006	371	286	48

The estimated reserve amounts from our model are lower than those from Kunkler (2004, 2006), as we have added both zeros and negatives in our loss triangle.

The posterior means, standard deviations and percentiles for the total reserve (i.e. over all accident years) and reserves by accident year are listed in Table 10.

Table 10. Mean, STD and Percentiles of Reserve Estimates, Multinomial Model

Acc year	mean	sd	MC error	2.50%	5.00%	median	95.00%	97.50%
1	0	0	0	0	0	0	0	0
2	117	254	1.0	-75	-52	0	547	754
3	539	848	3.7	-158	-109	332	1878	2498
4	1184	1269	4.9	-265	-160	954	3310	4154
5	2852	2452	10	-309	-10	2410	7086	8656
6	3647	2713	12	-301	268	3225	8350	10130
7	5594	4111	19	-229	705	4944	12710	15290
8	9168	6471	37	-799	1183	8334	20190	24020
9	11060	10590	79	-5618	-1009	10160	26940	32640
10	16480	19860	149	-8143	-1432	13610	45710	57330
Total	50640	26900	261	9887	19400	48420	91340	105500

From the table, we observe that the reserves (i.e., the mean) for our multinomial mixture model are smaller than those for the binomial model in Kunkler (2006). It is what we would expect to see, as we have two extra zeros included in the loss triangle we used, while all the other loss data are the same. However, the high quantiles from our model such as the 95% and 97.5% percentiles are higher (i.e., higher variance in the estimation), due to the uncertainty introduced by the extra parameters.

4. Concluding Remarks

Papers such as de Alba (2002, 2006) and Kunkler (2004, 2006) have put forward two types of models to deal with either zeros or negatives in the loss triangle. No model has been introduced for a situation with notable numbers of both zeros and negative. Inspired by the models of Kunkler (2004, 2006) and other previous work, we proposed a Bayesian multinomial mixture model for a more general situation when there are both zeros and negatives in the loss triangle.

The Bayesian multinomial mixture model we proposed in Section 2 seems to work very well in dealing with zeros and negatives in stochastic loss reserving. From the simulation results in Section 3, we observe that the estimates of parameters and reserves look reasonable according to the data we are using. With the multinomial mixture model for modelling the sign of the data, the model is able to deal with situations where there are large portions of zeros and negatives. The number of zeros or negatives that can be handled is never restricted. The generalized linear modelling structure gives the flexibility of innovation as well as replicating various existing models, such as the chain ladder model. With a Bayesian implementation, external information can be incorporated by specifying specific prior distributions for the parameters or quantities of interest.

The Bayesian mixture model implemented in Section 3 is only an example of the Bayesian mixture models that can be used for dealing with zeros and negatives in the loss triangle. For different loss triangle data, different generalized linear models can be fitted for both the mixture and magnitude models. Other link functions such as the probit and log-log link functions can be used, while a nonlinear regression equation can be fitted for the multinomial mixture model. Instead of the lognormal model, other models such as the over-dispersed Poisson model can be chosen for the magnitude models of the negatives and positives.

Another possible application of the Bayesian mixture model in loss reserving is the situation where the losses are from several notably different distributions (e.g., large losses vs. small losses, losses from different territory, gender). To better reflect the actual distributions of the different groups, a Bayesian binomial or multinomial mixture can be applied to model the probabilities of losses from different groups, while different distributions or models can be fitted for losses from each group.

Acknowledgment

The authors would like to thank the editor and anonymous referees for their valuable comments and suggestions that improved the quality of the paper. This research was supported by the Natural Sciences and Engineering Research Council of Canada.

References

- Antonio, K., & Plat, H. (2014). Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, 7, 649-669. <http://dx.doi.org/10.1080/03461238.2012.755938>
- Brooks, S., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434-455.
- de Alba, E. (2002). Claims reserving when there are negative values in the runoff triangle. 37th Actuarial Research Conference, ARCH 2003 Proceedings.
- de Alba, E. (2006). Claims reserving when there are negative values in the runoff triangle: Bayesian analysis using the three-parameter log-normal distribution. *North American Actuarial Journal*, 10(3), 45-59. <http://dx.doi.org/10.1080/10920277.2006.10597402>
- de Alba, E., & Nieto-Barajas, L. E. (2008). Claims reserving: A correlated Bayesian model. *Insurance: Mathematics and Economics*, 43, 368-376. <http://dx.doi.org/10.1016/j.insmatheco.2008.05.007>
- Dobson, A. J. (2002). *An Introduction to Generalized Linear Models*, Second ed. Chapman & Hall/CRC, Boca Raton, FL.
- England, P. D., Verrall, R., & Wuthrich, M. (2012). Bayesian over-dispersed Poisson model and the Bornhuetter & Ferguson claims reserving method. *Annals of Actuarial Science*, 6(2), 258-283. <http://dx.doi.org/10.1017/S1748499512000012>
- England, P. D., & Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3), 519-544. <http://dx.doi.org/10.1017/S1357321700003809>
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis*, Second ed. Chapman & Hall/CRC, Boca Raton, FL.
- Kremer, E. (1982). IBNR claims and the two way model of ANOVA. *Scandinavian Actuarial Journal*, 47-55. <http://dx.doi.org/10.1080/03461238.1982.10405432>
- Kunkler, M. (2004). Modelling zeros in stochastic reserving models. *Insurance: Mathematics and Economics*, 34(1), 23-35. <http://dx.doi.org/10.1016/j.insmatheco.2003.09.011>
- Kunkler, M. (2006). Modelling negatives in stochastic reserving models. *Insurance: Mathematics and Economics*, 38(3), 540-555. <http://dx.doi.org/10.1016/j.insmatheco.2005.11.006>
- LeSage, J. P. (1999). *Applied econometrics using MATLAB*. Available at <http://www.spatial-econometrics.com/>
- Lunn, D., Thomas, A., Best, N. G., & Spiegelhalter, D. J. (2000). Winbugs A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10, 325-337. <http://dx.doi.org/10.1023/A:1008929526011>
- Mack, T., & Venter, G. (2000). A comparison of stochastic models that reproduce chain ladder reserve estimates. *Insurance: Mathematics and Economics*, 26(1), 101-107. [http://dx.doi.org/10.1016/S0167-6687\(99\)00039-6](http://dx.doi.org/10.1016/S0167-6687(99)00039-6)
- Makov, U. E., Smith, A. F. M., & Liu, Y.-H. (1996). Bayesian methods in actuarial science. *The Statistician*, 45, 503-15. <http://dx.doi.org/10.2307/2988548>
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, 135(3), 370-384. <http://dx.doi.org/10.2307/2344614>
- Renshaw, A. E., & Verrall, R. J. (1998). A stochastic model underlying the chainladder techniques. *British Actuarial Journal*, 4(4), 903-923. <http://dx.doi.org/10.1017/S1357321700000222>
- Scollnik, D. P. M. (2001). Actuarial modeling with MCMC and BUGS. *North American Actuarial Journal*, 5(2), 96-125. <http://dx.doi.org/10.1080/10920277.2001.10595987>
- Scollnik, D. P. M. (2002). Implementation of four models for outstanding liabilities in WinBUGS: A discussion of a paper by Ntzoufras and Dellaportas. *North American Actuarial Journal*, 6(1), 128-136. <http://dx.doi.org/10.1080/10920277.2002.10596033>
- Scollnik, D. P. M. (2004). Bayesian reserving models inspired by chain ladder methods and implemented using WinBUGS. *Actuarial Research Clearing House*, 2(1), 1-43.

- Verrall, R. J. (2000). An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: Mathematics and Economics*, 26(1), 91-99. [http://dx.doi.org/10.1016/S0167-6687\(99\)00038-4](http://dx.doi.org/10.1016/S0167-6687(99)00038-4)
- Verrall, R. J. (2004). A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving. *North American Actuarial Journal*, 8(3), 67-89. <http://dx.doi.org/10.1080/10920277.2004.10596152>
- Verrall, R. J., & England, P. D. (2005). Incorporating expert opinion into a stochastic model for the chain-ladder technique. *Insurance: Mathematics and Economics*, 37(2), 355-370. <http://dx.doi.org/10.1016/j.insmatheco.2005.04.005>
- Zehnwirth, B. (1994). Probabilistic development factor models with applications to loss reserve variability, prediction intervals and risk based capital. *Casualty Actuarial Society Forum*, 2, 447-606.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).