THE UNIVERSITY OF CALGARY

A Bayesian Mixture Model for Zeros and Negatives in Stochastic Reserving

by

Chaoxiong (Michelle) Xia

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

August, 2007

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a Dissertation entitled "A Bayesian Mixture Model for Zeros and Negatives in Stochastic Reserving" submitted by Chaoxiong (Michelle) Xia in partial fulfillment of the requirements for the degree of Master of Science.

Supervisor, Dr. David P. M. Scollnik
Department of Mathematics and Statistics

Dr. Ka Chun Cheung
Department of Mathematics and Statistics

Dr. Ryan Lee
Department of Finance

August 23, 2007

Date

Abstract

In 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 thesis, the existing stochastic reserving models will be introduced and compared, particularly those dealing with zeros and negatives in the loss triangle. The specialized Bayesian software BUGS (Spiegelhalter et al., 1996) will be used to implement the model introduced by Kunkler (2006) for the situation where there are negatives in the loss triangle. Logit model and prior specifications different to those in Kunkler (2006) will be considered. We will compare the results from BUGS to those Kunkler (2006) obtained with the Econometrics Toolbox (Lesage, 1999) in MatLab (developed by the MathWorks, Inc.).

Inspired by the work of Kunkler (2004, 2006), we will propose a Bayesian mixture model to extend the stochastic reserving models to a situation where there are both zeros and negatives in the incremental loss triangle. A multinomial mixture 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.

Acknowledgements

First of all, I am full of gratitude to my supervisor, Dr. David P. M. Scollnik, for his persistent guidance, inspiration and support during my research. His professionalism and passion of science have encouraged me to stick to high standards and keep improving. I am sure his values will highly benefit my future research and work.

I would like to express my sincere thanks to the examiners of my thesis, Dr. R. Lee and Dr. K. C. Cheung for their time and pertinent comments. Many thanks to Ms. J. Mellard for the considerate assistance and support she provided for my defense.

Finally, I would like to thank my husband, Lei Hua, for his love, encouragement and support in every moment of my life.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Loss Reserving

Loss reserving is the process of estimating the amount of money (i.e., reserve) an insurance company needs to set aside for future losses, based on the loss data from the same group of policies over a past period of time. Determining an appropriate amount of loss reserve is very important for the financial stability of an insurance company. An inadequate reserve amount may lead to insolvency, while an overestimation of the reserve will reduce the periodic income of the insurance company.

The loss reserving data are typically listed as a loss triangle by the accident year (i.e., the year when the accident occurs) and development year (i.e., number of years between the accident year and actual payment of the loss). Details as well as an example will be given in Section 2.1 of this thesis for the loss triangle.

Traditionally, two of the most popular loss reserving methods used by insurance companies are the chain ladder method (Harnek, 1966) and the Bornhuetter-Ferguson method (Bornhuetter and Ferguson, 1972). This is partly due to their simplicity of implementation. The chain ladder method assumes the same ratio (i.e., development ratio) for losses from the same adjacent development years. The development ratio can be estimated from previous data by some mean measures such

as the arithmetic mean or geometric mean. Bornhuetter and Ferguson (1972) introduced an external estimate of ultimate loss into the chain ladder method for each accident year, which solved the problem of instability in the chain ladder method. This method is named the Bornhuetter-Ferguson method, and is very popular in loss reserving practice. In Section 2.1 of this thesis we will give the details and examples of these two methods.

1.2 Stochastic Loss Reserving Models

Although the traditional methods such as the chain ladder and Bornhuetter-Ferguson methods are simple to implement, they do not consider 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.

1.2.1 Classical Stochastic Reserving

In stochastic loss reserving, specific distributions such as the lognormal (Kremer, 1982), over-dispersed Poisson (Renshaw and Verrall, 1998), and negative binomial (Verrall, 2000) are assumed for the loss reserving data, with which the risk of an underestimation or overestimation can be quantified. 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.

An introduction of the lognormal, over-dispersed Poisson, and negative binomial models will be given in Section 2.2 of this thesis. Comparisons of the chain ladder model and the stochastic models reproducing the chain ladder reserves can be found in papers such as Kremer (1982), Renshaw and Verrall (1998), Mack (1994), Verrall (2000), Mack and Venter (2000), and Verrall and England (2000).

1.2.2 Bayesian Models

In Bayesian statistics, the parameters of a distribution are assumed to be random variables instead of definite values. External information or expert opinion can be incorporated into the model via the distribution (i.e., prior distribution) specified for the parameters. Inferences can be made for the mean or variance of the parameters or quantities of interest based on Bayes' Theorem (Bayes, 1763). By assuming certain prior distributions for the parameters, traditional generalized linear models can also be implemented in the framework of Bayesian statistics.

With its capability of incorporating external information, Bayesian method is used frequently in stochastic reserving. In papers such as Scollnik (2002a), de Alba (2002a, 2002b, 2006), and Ntzoufras and Dellaportas (2002), 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 by Scollnik (2004) and Verrall (2004) respectively. Most of the above models are implemented using the Markov Chain Monte Carlo (MCMC) simulation method in the Bayesian software package BUGS (Spiegelhalter et al., 1996). Reviews of the MCMC method, BUGS, and their application in actuarial

science can be found in Scollnik (1996, 2001). In Chapter Two of this thesis, we will give a detailed review of the Bayesian methods and their application in loss reserving and actuarial science.

1.3 Zeros and Negatives

1.3.1 The Problem

Due to the nature of the distributions assumed, the stochastic models reviewed in the previous section have their limitation in handling zero and negative values in the loss triangle. For example, the lognormal model works only in the cases of positive losses. Using the quasi-likelihood approach (McCullagh and Nelder, 1989, Chapter 9, pages 323-356), the over-dispersed Poisson and negative binomial models can be implemented even when there are non-integer or negative values in the data. However, due to the parameterization of the variance, the sum of the incremental claims in each development year has to be positive. Hence the quasi-likelihood approach is also restricted in the number of zeros and negatives it can handle.

In loss reserving, however, 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 a loss, consequences of judicial decisions, and errors. A large number of zeros and negatives occur in the loss triangle may make some of the models inappropriate or even undefined.

1.3.2 Previous Work

Although various techniques have been proposed in the recent actuarial literature (e.g., de Alba, 2002a, 2006; Kunkler, 2004, 2006) in order to deal with the problem of zeros and negatives, none of them can handle the situation when there are notable number of zeros and negatives in the loss triangle.

A threshold parameter is introduced by de Alba (2002a, 2006) into the lognormal model to handle the negatives, while the number of zeros is highly restricted. The improved lognormal model assumes a minimum value of the negative losses (i.e., the negative of the threshold parameter) which is not consistent with the true nature of the loss data. The improved lognormal model in de Alba (2006) will be introduced in more detail in Subsection 4.2.1 of this thesis.

Kunkler (2004) proposed a binomial mixture model to handle the situation where there are zeros in the loss triangle. A binomial mixture model is used to model the probabilities of zeros and positives, while a lognormal model is used for the magnitude of the positive losses. A similar model is introduced by Kunkler (2006) to deal with the negatives in the loss triangle. The same binomial mixture model is used for the probabilities of negatives and positives, while two different lognormal model structures are used for the magnitudes of negatives and positives. Bayesian generalized linear models are fitted for both the mixture and magnitude models. We will give detailed introductions of these two models in Subsection 4.2.2 and 4.2.3 of this thesis.

In Chapter Five, we will implement the binomial mixture model of Kunkler (2006) in the Bayesian software package BUGS (Spiegelhalter et al., 1996) in order to reproduce and verify the results Kunkler obtained using the Econometrics Toolbox developed by LeSage (LeSage, 1999) for MatLab (developed by the MathWorks, Inc.). Due to the vagueness of model specification in Kunkler (2006), we will try different prior specifications as well as different implementations of the binomial logit model.

1.4 A Multinomial Mixture Model

Inspired by the work of Kunkler (2004, 2006), we will propose a Bayesian mixture model in Chapter Six to extend the stochastic reserving models to a situation where there are both zeros and negatives in the incremental loss triangle. A multinomial mixture 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.

In Chapter Seven, the model will be implemented using the Markov chain Monte Carlo (MCMC) techniques in BUGS. For the sake of comparison, a loss triangle similar as to that in Kunkler (2006) is used for our model implementation. The loss triangle is adjusted from the 'Historical Loss Development Study' (1991) by the Reinsurance Association of America, keeping the same negative values as in Kunkler (2006). The same model structure as the one in Kunkler (2006) is chosen for the

negative magnitude model. Prior distributions similar to those in Kunkler (2006) are specified to make the results more comparable. A chain ladder type model structure derived from the structure in Zehnwirth (1994) is used for the magnitude of the positive losses. The model implementation, BUGS codes, as well as the results for the estimation of parameters and reserves are given in this chapter.

Chapter 2

Stochastic Loss Reserving

2.1 Loss Reserving

To meet the future claims on the policies currently in force, an insurance company needs to set aside an amount of money named the reserve. Loss reserving or claims reserving is the process of estimating the amount of reserve the insurance company needs to hold, based on the losses to the specific group of policies over past periods.

A typical data format used to tabulate the loss data for loss reserving is the loss triangle or claim triangle (Scollnik, 2004). In a loss triangle the loss data are listed by the accident year (period) and development year (period). The accident year refers to the year when the accident occurs, while the development year is the number of years between the accident year and the year the insurance company actually pays for the loss. The data set looks like a triangle at the time when the outstanding claim reserve needs to be estimated.

A loss triangle frequently referenced in the loss reserving literature is from the 'Historical Loss Development Study' (1991) by the Reinsurance Association of America. This data set was analyzed by Mack (1994), Renshaw and Verrall (1998) and Kunkler (2004, 2006). Please refer to Table 2.1 and Table 2.2 for the loss triangles in which the incremental losses and cumulative losses in units of $1000 are listed by

the accident year and development year.

Table 2.1: Incremental Loss Triangle, Historical Loss Development Study (1991)

	Development year									
Accident 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	673	535	
3	3410	5582	4881	2268	2594	3479	649	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									

Table 2.2: Cumulative Loss Triangle, Historical Loss Development Study (1991)

Development year									
1	2	3	4	5	6	7	8	9	10
5012	8269	10907	11805	13539	16181	18009	18608	18662	18834
106	4285	5396	10666	13782	15599	15496	16169	16704	
3410	8992	13873	16141	18735	22214	22863	23466		
5655	11555	15766	21266	23425	26083	27067			
1092	9565	15836	22169	25955	26180				
1513	6445	11702	12935	15852					
557	4020	10946	12314						
1351	6947	13112							
3133	5395								
2063									

2.1.1 Chain Ladder Method

One of the simplest and most popular models for loss reserving is the chain ladder method, in which a fixed development ratio λ_j is assumed for cumulative losses from

two adjacent development years, i.e. from development year $j-1$ to j. The introduction of this model can be dated back to Harnek (1966).

For an incremental loss triangle over n consecutive accident years, $P_{i,j}$ $(i = 1,\ldots,n;\ j = 1,\ldots,n-i+1)$, a cumulative loss triangle is required for applying the chain ladder method. The following formula can be used to obtain the corresponding cumulative loss triangle $C_{i,j}$ $(i = 1,\ldots,n;\ j = 1,\ldots,n-i+1)$:

$$C_{i,j} = \sum_{k=1}^{j} P_{i,k}\ , \qquad i = 1,\ldots,n;\ j = 1,\ldots,n-i+1.$$

The development ratio for each development year can be estimated based on the cumulative loss triangle data as in Verrall (1989).

Let

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} C_{i,j}}{\sum_{i=1}^{n-j+1} C_{i,j-1}}\ , \qquad j = 2,\ldots,n\,.$$

With the estimated development ratios $\hat{\lambda}_j$ $(j = 2,\ldots,n)$, the losses for future years can be obtained based on past loss data. That is,

$$\hat{C}_{i,n-i+2} = C_{i,n-i+1} \times \hat{\lambda}_{n-i+2}\ , \qquad i = 1,\ldots,n$$

$$\hat{C}_{i,j} = \hat{C}_{i,j-1} \times \hat{\lambda}_j\ , \qquad i = 1,\ldots,n;\ j = n-i+3,\ldots,n\,.$$

The total loss reserve R is the sum of all the estimated future losses, which is equal to the difference between the estimated final losses $\hat{F}$ and loss to date P. Hence, the estimated loss reserve is given by

$$\hat{R} = \hat{F} - P\ ,$$

where $\hat{F} = C_{1,n} + \sum_{i=2}^{n} \hat{C}_{i,n}$ and $P = \sum_{i=1}^{n} C_{i,n-i+1}$.

Applying the chain ladder method, we can now estimate the loss development ratios, future losses, and reserve for the cumulative loss triangle in Table 2.2. The resulting estimated future losses are listed in Table 2.3.

Table 2.3: Estimated Future Cumulative Losses, Chain Ladder Method

Accident	Development year									
year	1	2	3	4	5	6	7	8	9	10
1										
2										16871
3									23935	24174
4								27879	28437	28721
5							27227	28044	28605	28891
6						17596	18300	18849	19226	19418
7					14407	15992	16632	17131	17474	17649
8				16652	19483	21626	22491	23166	23629	23865
9			8740	11100	12987	14416	14993	15443	15752	15910
10		6189	10026	12733	14898	16537	17198	17714	18068	18249
Dev ratio	-	3.00	1.62	1.27	1.17	1.11	1.04	1.03	1.02	1.01

The total loss reserve based on the above estimated cumulative losses is

$$\begin{aligned}
\hat{R} &= \hat{F} - P \\
&= \left(C_{1,n} + \sum_{i=2}^{n} \hat{C}_{i,n} \right) - \left(\sum_{i=1}^{n} C_{i,n-i+1} \right) \\
&= 212582 - 160987 \\
&= 51595 .
\end{aligned}$$

2.1.2 Bornhuetter-Ferguson Method

For the chain ladder method, the estimate of final reserve can be affected dramatically by the most recent losses, except for the first accident year there is no reserve. This is easy to see when we write the estimate of outstanding claims for each accident year in the following form, as in England and Verrall (2002):

$$\hat{R}_i = C_{i,n-i+1}\left(\hat{\lambda}_{n-i+2}\hat{\lambda}_{n-i+3}\ldots\hat{\lambda}_n - 1\right), \qquad i = 2,\ldots,n\,. \tag{2.1}$$

Now, observe that the estimated total losses for accident year i from the chain ladder method can be calculated as

$$\hat{U}_i^{(CL)} = C_{i,n-i+1}\hat{\lambda}_{n-i+2}\hat{\lambda}_{n-i+3}\ldots\hat{\lambda}_n\,, \qquad i = 2,\ldots,n\,.$$

Substituting this result into Equation (2.1) gives another expression for the estimated final reserve for each accident year, namely

$$\hat{R}_i = \hat{U}_i^{(CL)}\frac{1}{\hat{\lambda}_{n-i+2}\hat{\lambda}_{n-i+3}\ldots\hat{\lambda}_n}\left(\hat{\lambda}_{n-i+2}\hat{\lambda}_{n-i+3}\ldots\hat{\lambda}_n - 1\right), \qquad i = 2,\ldots,n\,.$$

For the Bornhuetter-Ferguson method (Bornhuetter and Ferguson, 1972), an outside estimate of total loss $\hat{U}_i^{(BF)}$ is introduced based on past experience and company practices. The Bornhuetter-Ferguson estimate of outstanding claims tends to be more stable than the estimate given by the chain ladder method, as the formula incorporates some external information. The estimated reserve for each accident year under the Bornhuetter-Ferguson method is given by

$$\hat{R}_i = \hat{U}_i^{(BF)}\frac{1}{\hat{\lambda}_{n-i+2}\hat{\lambda}_{n-i+3}\ldots\hat{\lambda}_n}\left(\hat{\lambda}_{n-i+2}\hat{\lambda}_{n-i+3}\ldots\hat{\lambda}_n - 1\right), \qquad i = 2,\ldots,n\,,$$

where $\hat{\lambda}_j$ $(j = 2, \ldots, n)$ are the development factors calculated using the chain ladder method.

Now based on the results in Table 2.3 from the chain ladder method, we can calculate the reserves using the Bornhuetter-Ferguson method. The Bornhuetter-Ferguson reserve estimates for our example in Table 2.1 using different assumptions of $\hat{U}_i^{(BF)}$ $(i = 2, \ldots, n)$ are listed in Table 2.4.

Table 2.4: Reserves vs Estimated Ultimate Losses, Bornhuetter-Ferguson Method

Accident	50% CL		Chain Ladder		150% CL		200% CL	
year	$\hat{U}_i^{BF}$	$\hat{R}_i$	$\hat{U}_i^{BF}$	$\hat{R}_i$	$\hat{U}_i^{BF}$	$\hat{R}_i$	$\hat{U}_i^{BF}$	$\hat{R}_i$
1	9417	0	18834	0	28251	0	37668	0
2	8436	84	16871	167	25307	251	33742	334
3	12087	354	24174	709	36261	1063	48348	1417
4	14361	827	28721	1654	43082	2481	57442	3308
5	14446	1356	28891	2711	43337	4066	57782	5422
6	9709	1783	19418	3566	29127	5349	38836	7132
7	8825	2667	17649	5334	26474	8002	35298	10669
8	11933	5377	23865	10753	35798	16130	47730	21507
9	7955	5257	15910	10514	23865	15771	31820	21029
10	9125	8093	18249	16186	27374	24279	36498	32372
Total		25798		51594		77392		103190

From the above table we can see that the final reserve is proportional to the estimated ultimate losses we assume. We get the same reserve estimates when assuming the same ultimate losses as those calculated from the chain ladder method. The difference of 1 is due to the rounding process in calculation. For the years when the total losses are extremely small or large, it will be effective to stabilize the result by applying an ultimate loss estimate based on long term experiences.

2.2 Stochastic Models for Loss Reserving

The traditional loss reserving methods such as the chain ladder and Bornhuetter-Ferguson methods are the most popular reserving models in practice, as they are simple to model and also give good estimates for outstanding reserves. However, in recent years attention has focussed on the variability and tail values of the distribution of the reserve, which brings the necessity of investigating the stochastic nature of the data.

By specifying the distribution or certain statistical measures for the loss data, stochastic loss reserving models have become a popular tool with which to estimate measures such as the percentiles or prediction error of the outstanding reserve. Stochastically based chain ladder models define the first two moments (Mack, 1993), or assume specific distributions such as lognormal (Kremer, 1982), over-dispersed Poisson (Renshaw and Verrall, 1998), or negative binomial (Verrall, 2000) for the loss reserving data. Many researchers (e.g., Kremer, 1982; Renshaw and Verrall, 1998; Mack, 1994; Verrall, 2000; Mack and Venter, 2000; Verrall and England, 2000) have focused on the comparison of the chain ladder method and the stochastic models reproducing the chain ladder reserves. Among these models, the distribution based models are the most popular stochastic models in recent study.

2.2.1 Lognormal Model

In the lognormal model, the incremental loss values P_{ij} are assumed to follow lognormal distributions, that is

$$P_{ij} \sim LN(\mu_{ij},\ \sigma^2)\ , \qquad i = 1, \ldots, n;\ j = 1, \ldots, n\ .$$

The three-parameter ANOVA structure is very popular in which the mean is modelled by

$$\mu_{ij} = \mu + \alpha_i + \beta_j\ , \qquad i = 1, \ldots, n;\ j = 1, \ldots, n\ .$$

This model first appeared in Kremer (1982) to reproduce the results of the traditional chain ladder method. But due to the stochastic nature of the model, the results may still differ from those of the chain ladder method. Before the logarithmic transformation, the structure of the mean is multiplicative in this model, which is similar to the chain ladder method. A limitation of this model is that the incremental loss data used for this model must be positive so as to ensure that the logarithm is defined. A wide range of reserving literatures investigated the implementation of the model using various statistical techniques such as generalized additive models (Verrall, 1996), Bayesian inference (Scollnik, 2004; de Alba, 2006), and mixed models (Antonio et al., 2006).

2.2.2 Over-Dispersed Poisson Model

The Poisson distribution has a mean equal to the variance, which is usually not true for the actual incremental loss data. The over-dispersed Poisson distribution is

more flexible by adding an over-dispersion parameter to allow for the variance to be proportional to the mean. Renshaw and Verrall (1998) introduced an over-dispersed Poisson model which reproduces the simple chain ladder reserves. In their model the incremental loss P_{ij} is assumed to follow an over-dispersed Poisson distribution given by

$$P_{ij} \sim \textit{over-dispersed Poisson}(m_{ij},\ \phi)\ ,$$

where

$$\mathrm{E}[P_{ij}] = m_{ij} = x_i\, y_j\ , \qquad i = 1, \ldots, n;\ j = 1, \ldots, n$$

$$\mathrm{Var}[P_{ij}] = \phi\, x_i\, y_j\ , \qquad i = 1, \ldots, n;\ j = 1, \ldots, n\ ,$$

with

$$\sum_{j=1}^{n} y_j = 1\ .$$

In the over-dispersed Poisson model, y_j is the proportion of ultimate losses occurring in development year j, while x_i is the expected ultimate loss for accident year i. The product of $x_i y_j$ is then the expected loss for development year j of accident year i. This mean structure is multiplicative, which allows it to reproduce the results of the chain ladder method. The over-dispersion parameter ϕ relaxes the restriction of equality for the mean and variance.

For modelling the mean of the over-dispersed Poisson model, a log linear model structure can be used to facilitate estimation. In Renshaw and Verrall (1998), the same model structure as that for the lognormal model (Kremer, 1982) was proposed for the over-dispersed Poinsson distribution. That is,

$$\log(m_{ij}) = c + \alpha_i + \beta_j\ , \qquad i = 1, \ldots, n;\ j = 1, \ldots, n\ .$$

This ANOVA model structure is equivalent to the multiplicative mean structure $m_{ij} = x_i y_j$, but is more convenient for estimation. Constraints such as the corner constraints can be used to the sets of parameters in this model.

Using the quasi-likelihood approach (McCullagh and Nelder, 1989, Chapter 9, pages 323-356), the over-dispersed Poisson model can be implemented even when there are some non-integer or negative values in the data. Hence, it is applicable to loss triangles where there are non-integer and negative values. Details about the over-dispersed Poisson model for loss reserving can be found in Renshaw and Verrall (1998). The earliest ideas of linking the chain ladder method and the Poisson distribution can be dated back to Wright (1990) and Mack (1991).

2.2.3 Negative Binomial Model

The negative binomial model (Verrall, 2000) looks more similar to the chain ladder method and gives similar results with the over-dispersed Poisson model. The model is derived from the over-dispersed Poisson model. It has parameters $\lambda_j \, (j = 1, \ldots, n)$ which are analogous to the development ratios in the chain ladder method. An over-dispersion parameter is contained in the model to allow over-dispersion of the variance. In the negative binomial model, the incremental losses P_{ij} are assumed to follow over-dispersed negative binomial distributions. That is,

$$P_{ij} \sim \textit{over-dispersed negative binomial} \, ,$$

with mean and variance given by

$$\begin{aligned} \mathrm{E}[P_{ij}] &= (\lambda_j - 1)\, C_{i,j-1}\,, & i = 1,\ldots,n;\ j = 1,\ldots,n \\ \mathrm{Var}[P_{ij}] &= \phi\, \lambda_j\, (\lambda_j - 1)\, C_{i,j-1}\,, & i = 1,\ldots,n;\ j = 1,\ldots,n\,, \end{aligned}$$

where $C_{ij} = C_{i,j-1} + P_{ij}$ is the cumulative loss. In this recursive model, an estimate of $C_{i,j-1}$ needs to be obtained before modelling P_{ij}.

The model can also be written in terms of cumulative losses C_{ij}. It is easy to verify that C_{ij} also follows an over-dispersed negative binomial distribution with mean and variance given by

$$\begin{aligned} \mathrm{E}[C_{ij}] &= \lambda_j C_{i,j-1}\,, & i = 1,\ldots,n;\ j = 1,\ldots,n \\ \mathrm{Var}[C_{ij}] &= \phi\, \lambda_j\, (\lambda_j - 1)\, C_{i,j-1}\,, & i = 1,\ldots,n;\ j = 1,\ldots,n\,. \end{aligned}$$

Details of this model can be found in Verrall (2000), and England and Verrall (2002).

2.2.4 Other Models

There are many other stochastic reserving models that are popular in the reserving literature. A comprehensive review of the stochastic reserving models is given in England and Verrall (2002). Besides the distributions listed above, the gamma distribution can be used for the claim amounts (Mack, 1991). A normal approximation can be used for the negative binomial model (McCullagh and Nelder, 1989, Chapter 4, pages 103-107). Mack (1993) brought forward another recursive stochastic model by only specifying the first two moments of the claim distribution which also pro-

duces the chain ladder reserves.

To avoid over-parameterization of the model, other generalized linear model (GLM) structures or parametric curves have been introduced by early researchers. One of the most popular structures is the Hoerl curve or gamma curve (Wright, 1990; Renshaw, 1994a) which can be applied to the lognormal model or gamma model with a log link function. Wright (1990) was the first to model the claim frequency and severity separately with GLM structures for the incremental losses in stochastic reserving. Non-parametric smoothing techniques can also be applied, an example of which is the introduction of generalized additive models (GAM) in stochastic reserving by Verrall (1996).

The Bornhuetter-Ferguson method shows that the use of external information is a great help for stabilizing the reserve estimates. Bayesian inference provides an effective and flexible way of implementing external information into the model. A wide range of Bayesian models such as the Bayesian hierarchical models (Scollnik, 2002a), Bayesian Bornhuetter-Ferguson method (Scollnik, 2004; Verrall, 2004), Bayesian mixture model (Kunkler, 2004, 2006), Bayesian GLM (Verrall, 2004) and general Bayesian techniques (Ntzoufras and Dellaportas, 2002; de Alba, 2002b, 2006) can be applied for stochastic reserving.

Chapter 3

Bayesian Methods

3.1 Bayesian Inference

In classical statistics, the parameters of a distribution are assumed to be definite values. Based on Bayes' Theorem (Bayes, 1763), Bayesian statistics, details of which can be found in Gelman, Carlin, Stern and Rubin (2005), releases this restriction and catches the uncertainty in the parameters with the use of random variables.

Bayesian inference is the process of statistical modelling with statistical probability distributions fitted for the observed data set as well for the unknown parameters and unobserved future observations. Bayes' Theorem can be applied to new observations by treating the former posterior distribution as the new prior distribution, which automatically updates the model.

The process of a Bayesian analysis can be broken down into three steps (Gelman et al., 2005, Chapter 1):

1. Specifying a joint probability distribution for the data as well as for the unobservable quantities such as the parameters, which determines a full probability model for the specific problem under consideration.

2. Obtaining an appropriate posterior distribution for the parameters. The pos-

terior distribution is the conditional probability of the parameters of interest given the observed data.

3. Evaluating the model fit and interpreting the posterior distributions obtained.

3.1.1 Bayesian Statistics

Posterior Inference

In Bayesian statistics, the parameters of a probability distribution are assumed to be random variables. We assume that y is a vector $y = (y_1, y_2, \ldots, y_n)$ with n observations and θ is a vector of parameters in the sampling distribution. The probability distribution of the observations y can be written as a conditional distribution $p(y|\theta)$, which is called the sampling distribution. This sampling distribution is conditional on the model parameters which are denoted by θ. A probability distribution $\pi(\theta)$, known as the prior distribution in Bayesian statistics, is specified for θ. The resulting full probability model is given by

$$p(\theta, y) = \pi(\theta)\, p(y|\theta) \,.$$

Using Bayes' Theorem (Bayes, 1763), the posterior distribution for θ, i.e. the conditional distribution of θ given y, can be obtained as

$$\pi(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{\pi(\theta)\, p(y|\theta)}{p(y)} \,.$$

Here, $p(y) = \sum_\theta \pi(\theta)\, p(y|\theta)$ in the discrete case or $p(y) = \int \pi(\theta)\, p(y|\theta)\, \mathrm{d}\theta$ in the continuous case. Since $p(y)$ is independent of θ and can be considered as a constant, an unnormalized posterior probability can be simply obtained as

$$\pi(\theta|y) \propto \pi(\theta)\, p(y|\theta) \,.$$

Predictive Inference

With the observed data of y in hand, predictions may be made for future observations $\tilde{y}$ using Bayesian analysis. The posterior predictive distribution, or the conditional distribution of $\tilde{y}$ given y, needed for predicting future observations of our interest, is defined by

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y},\, \theta|y)\, \mathrm{d}\theta \\ &= \int p(\tilde{y}|\theta,\, y)\, \pi(\theta|y)\, \mathrm{d}\theta\,. \end{aligned}$$

If we assume that $\tilde{y}$ and y are conditionally independent given θ, then

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)\, \pi(\theta|y)\, \mathrm{d}\theta\,.$$

Details of the analysis can be found in Gelman et al. (2005, Chapter 2, pages 6-9).

A Parametric Example

The earliest parametric example of Bayesian inference is the binomial model which can be dated back to Bayes (1763). In the binomial model, the sampling distribution of y is assumed to be binomial, with number of trials denoted by n and probability of success denoted by θ. That is,

$$y \sim \mathit{binomial}(n,\, \theta)\,,$$

for which the sampling distribution's probability function is given by

$$p(y|\, n,\, \theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}\,, \qquad y = 0, 1, \ldots, n\,.$$

A beta prior can be assumed for θ, i.e.

$$\theta \sim beta(a,\ b)\ .$$

Then the prior density function for θ is

$$\begin{aligned}\pi(\theta) &= \frac{1}{\beta(a,b)}\,\theta^{a-1}(1-\theta)^{b-1} \\ &\propto \theta^{a-1}(1-\theta)^{b-1}\ ,\end{aligned}$$

where $\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, $0 < \theta < 1$, $a > 0$ and $b > 0$. From the earlier results, the posterior probability distribution for θ can be obtained by

$$\begin{aligned}\pi(\theta|y) &\propto \pi(\theta)\,p(y|\theta) \\ &\propto \theta^{a-1}(1-\theta)^{b-1}\theta^{y}(1-\theta)^{n-y} \\ &= \theta^{a+y-1}(1-\theta)^{b+n-y-1} \\ &= \theta^{a^*-1}(1-\theta)^{b^*-1}\ .\end{aligned}$$

The form of the posterior distribution in this example can be recognized as a beta distribution with parameters $a^* = a + y$ and $b^* = b + n - y$. Here the beta distribution is said to be a conjugate prior for the binomial distribution, as the posterior distribution is of the same form as the prior.

Predictive analysis can now proceed on the basis of the posterior predictive dis-

tribution given by

$$\begin{aligned} p(\tilde{y}|y) &= \int_0^1 p(\tilde{y}|\theta)\,\pi(\theta|y)\,\mathrm{d}\theta \\ &= \int_0^1 \binom{n}{\tilde{y}} \theta^{\tilde{y}}(1-\theta)^{n-\tilde{y}}\,\frac{1}{\beta(a^*,\,b^*)}\theta^{a^*-1}(1-\theta)^{b^*-1}\,\mathrm{d}\theta \\ &= \frac{\binom{n}{\tilde{y}}}{\beta(a^*,\,b^*)} \int_0^1 \theta^{\tilde{y}+a^*-1}(1-\theta)^{n-\tilde{y}+b^*-1}\,\mathrm{d}\theta \\ &= \frac{\binom{n}{\tilde{y}}\beta(\tilde{y}+a^*,\,n-\tilde{y}+b^*)}{\beta(a^*,\,b^*)}\,, \qquad \tilde{y}=0,1,\ldots,n\,, \end{aligned}$$

since $\tilde{y}$ has the same binomial distribution as y, i.e. $\tilde{y} \sim \mathit{binomial}(n,\,\theta)$, and $\theta|y \sim \mathit{beta}(a^*,\,b^*)$ as verified before.

Details of this model can be found in text books such as Leonard and Hsu (1999, Chapter 3, pages 108-114), and Gelman et al. (2005, Chapter 2, pages 31-46).

For the above example, an informative prior was used with definite information specified for the prior distribution. There are also other forms of priors such as the noninformative prior and improper prior that can be used for Bayesian analysis. A noninformative prior is a prior that gives vague information about the prior distribution, an example of which can be a distribution with an extremely large variance. An improper prior with a sum or integral of the prior density larger than 1 or infinite can be assumed, so long as sensible answers for the posterior probabilities exist.

3.1.2 Regression Models

Classical Regression Models

Classical regression models are widely used for exploring the relationship between

a dependent variable (response) and some independent variables (predictors). The regression equation is the mathematical formulation of the relationship between the response and predictors. In general linear regression models, the responses are assumed to follow independent normal distributions and the regression equation is assumed to be linear.

The introduction of the generalized linear models (Nelder and Wedderburn, 1972) relaxed the restrictions of the independence, normality, and linearity assumptions. A generalized linear model (GLM) has three components (Dobson, 2002, Chapter 3, pages 49-53):

1. Response variables $Y_1, \ldots, Y_n$ from the same distribution of the exponential distribution family, such as the normal, Poisson, or binomial distribution;

2. A parameter vector $\beta\,(p \times 1)$ and a predictor vector

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} = \begin{pmatrix} x_{11} & \ldots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & & x_{np} \end{pmatrix};$$

3. A monotone link function l relaxing the linear restriction, such as the log link function given by $l(\mu) = \log(\mu)$.

With these components a GLM can be formulated as

$$l(\mu_i) = \mathbf{x}_i^T \beta\,, \qquad i = 1, \ldots, n\,,$$

where

$$\mu_i = \mathrm{E}(Y_i)\ , \qquad i = 1, \ldots, n\ .$$

For the classical GLM, the parameters β can be estimated by the method of maximum likelihood estimation, instead of the least squares estimation method routinely used for the linear regression models. However, in the case of normally distributed responses, the results given by the maximum likelihood estimation and least squares estimation methods are equivalent. Further details concerning GLMs can be found in Nelder and Wedderburn (1972), and Dobson (2002).

GLMs can be fitted for the stochastic reserving models introduced in Chapter Two.

Bayesian Regression Models

An alternative to the maximum likelihood estimation and least squares estimation approaches to regression model fitting is given by the Bayesian analysis. In the Bayesian regression model, the coefficients of regression β are treated as random variables. The estimation of parameters and future predictions are accomplished using posterior and predictive distributions. Bayesian regression can be extended to the generalized linear model (GLM), which gives the Bayesian generalized linear model. Reviews of Bayesian regression models and Bayesian GLMs can be found in Gelman et al. (2005, Chapters 14 and 16).

A binomial GLM will be used to illustrate the Bayesian analysis of a GLM. In the model developed by Kunkler (2004, 2006) for stochastic loss reserving, the response

variables $Y_1, Y_2, \ldots, Y_n$ are assumed to follow binomial distributions with different parameters, i.e.

$$Y_i \sim \textit{binomial}(n, \theta_i) , \qquad i = 1, 2, \ldots, n .$$

Two commonly used link functions for binomial responses (Dobson, 2002, Chapter 7, pages 116-124) are the logistic link and probit link functions given by

$$\text{logit}(\theta) = \log\left(\frac{\theta}{1-\theta}\right)$$
$$\text{probit}(\theta) = \Phi^{-1}(\theta) ,$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution.

Given a parameter vector $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ and a predictor vector

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} = \begin{pmatrix} x_{11} & \ldots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & & x_{np} \end{pmatrix} ,$$

the logit GLM for the parameter vector $\theta_1, \theta_2, \ldots, \theta_n$ of the binomial model can be written as

$$\text{logit}(\theta_i) = \mathbf{x}_i^T \beta , \qquad i = 1, \ldots, n .$$

In the Bayesian GLM framework, prior distributions for the parameters need to be specified. In this way, external information is introduced into the model. In our example, we may assume that each β_i follows a normal distribution with different means and variances specified, i.e.

$$\beta_i \sim N(\mu_i, \sigma_i^2) , \qquad i = 1, 2, \ldots, n .$$

If the response follows a distribution with several parameters, then a Bayesian GLM structure can be fitted to some or all of the parameters.

3.1.3 Mixture Models

Mixture models can be used for modelling the data from a population which has several subpopulations with different distributions, where data from each subpopulation share the same distribution. The following discussion is based on the review of mixture models given in Gelman et al. (2005, Chapter 18).

For each observation y from the mixture distribution, let ζ denote a vector of indicator variables identifying the subpopulation from which the observation is drawn. Given the value of this indicator data, the distribution of the observation y is determined. So in the mixture model, the distribution of y is specified conditionally on both its parameters θ and the mixture data ζ. Implementation of this model can be performed under the framework of a Bayesian analysis by specifying the prior distributions for θ and ζ.

Assume the observed data are from a population with M subpopulations, with the distribution of the mth subpopulation given by $f_m(y|\theta_m)$ with a parameter vector θ_m. The proportion of the population from component m can be denoted as λ_m, with $\sum_{m=1}^{M} \lambda_m = 1$. Let $\theta = (\theta_1, \theta_2, \ldots, \theta_M)$ and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_M)$. Then the sampling distribution of the ith observation y_i is given by

$$p(y_i|\theta, \lambda) = \lambda_1 f(y_i|\theta_1) + \lambda_2 f(y_i|\theta_2) + \cdots + \lambda_M f(y_i|\theta_M) \, .$$

In this case, the indicator vector for the ith observation can be defined as

$$\zeta_i = (\zeta_{i1}, \zeta_{i2}, \ldots, \zeta_{iM}) ,$$

where

$$\zeta_{im} = \begin{cases} 1 & \text{if the } i\text{th observation is from the } m\text{th subpopulation} \\ 0 & \text{otherwise .} \end{cases}$$

It is easy to see that the distribution of each indicator vector ζ_i, given λ, is given by

$$\zeta_i \sim multinomial(\lambda, 1) .$$

The joint sampling distribution of y and ζ can be written as

$$p(y, \zeta|\theta, \lambda) = p(\zeta|\lambda)\, p(y|\zeta, \theta) = \prod_{i=1}^{n} \prod_{m=1}^{M} (\lambda_m f(y_i|\theta_m))^{\zeta_{im}} .$$

Prior distributions for λ and θ must now be specified in order for the Bayesian posterior analysis to proceed.

Note, the discussion above assumed that the number of subpopulations, M, was fixed and known. Gelman et al. (2005, page 466) discuss the more general case when there is uncertainty concerning the value of M.

3.2 Model Implementation

Gelman et al. (2005, Chapters 10-13) give a detailed review of various topics concerning Bayesian model implementation for complex problems. In particular, they

emphasize posterior simulation and approximation. Crude estimation is usually a starting point for a more accurate and complicated posterior sampling analysis. For example, for the cases when there are missing data, crude estimation can be performed by simply ignoring all the missing data. Due to its roughness, crude inference can only serve as a starting point and reference for later analysis.

For complex problems, the Bayesian analysis often proceeds on the basis of simulation from the posterior distribution of the parameters. The shape of the posterior distribution can be described on the basis of the mean, variance, percentiles, and various plots of the simulated parameter values. Another use of the posterior simulation is to make inferences about the predictive distributions. With the simulated value of the parameter θ^l, it is now possible to simulate a predictive value $\tilde{y}^l$ from the predictive distribution $p(\tilde{y}^l|\tilde{y})$ by making a conditional draw from the conditional distribution $p(\tilde{y}^l|\theta^l)$.

However, direct simulation from the posterior distribution is possible only for simple Bayesian models, such as ones for which a conjugate prior is assumed. The numerical method known as Markov chain Monte Carlo (MCMC) is an important and useful tool for more complicated full Bayesian analyses.

3.2.1 Markov Chain Monte Carlo

The Main Idea

Scollnik (2001) describes the main ideas behind MCMC. In a MCMC simulation, sample data are simulated from a Markov chain which has a stationary distribution

the same as the posterior distribution $p(\theta|y)$. From the key property of a Markov chain, the distribution of any sampled draw θ^t depends only on the last simulated value θ^{t-1}. Under some regularity conditions, these dependent draws $\theta^1, \theta^2, \ldots$ can be shown to satisfy the statements

$$\theta^t \xrightarrow{d} p(\theta|y), \qquad \text{as} \quad t \to \infty ,$$

and

$$\frac{1}{n}\sum_{t=1}^{n} h(\theta^t) \to \mathrm{E}[h(\theta)] \quad a.s. , \quad \text{as} \quad n \to \infty ,$$

where $h(\cdot)$ is an integrable function.

For a MCMC algorithm, it is always necessary to check the convergence of the sequence, to ensure that the distribution of the random draw is close enough to its actual distribution.

MCMC methods have been widely used in the actuarial literature, examples of which can be found in Scollnik (1993, 1996, 2001, 2002a), Haastrup and Arjas (1996), Ntzoufras and Dellaportas (2002), Verrall (2004), de Alba (2006), and Ntzoufras, Katsis and Karlis (2005). A detailed review of actuarial modelling with MCMC can be found in Scollnik (1996, 2001).

The Gibbs Sampler

There are many different algorithms that can be used in the construction of an MCMC simulation. Many of these have been described in the statistics literature (e.g., Gelfand and Smith, 1990; Smith and Roberts, 1993; and Tierney, 1994). Two

of the most popular algorithms are the Metropolis-Hastings algorithm (Metropolis and Ulam, 1949; Metropolis et al., 1953; and Hastings, 1970) and Gibbs sampler (Geman and Geman, 1984). As a special case of the Metropolis-Hastings algorithm, the Gibbs sampler or alternating conditional sampling is one of the simplest and most useful methods for MCMC.

A review of the Gibbs sampler is given in Gelman et al. (2005, Chapter 11, pages 287-289). Gelfand (2000) reviews the origins of the Gibbs sampler and assesses its impact on the research community. For the algorithm of Gibbs sampler, the parameter vector θ is divided into several subvectors, i.e. $\theta = (\theta_1, \theta_2, \ldots, \theta_d)$ for some integer d. During each iteration, a sample of each subvector is drawn conditional on the simulated values of the rest of subvectors. That is, at iteration t, θ_j^t ($j \in \{1, 2, \ldots, d\}$) is drawn from the conditional distribution

$$p(\theta_j | \theta_{-j}^{t-1}) ,$$

where

$$\theta_{-j}^{t-1} = (\theta_1^t, \ldots, \theta_{j-1}^t, \theta_{j+1}^{t-1}, \ldots, \theta_d^{t-1}) ,$$

For example, at the first iteration (i.e. for $t = 1$), we have

$$\begin{aligned}
\theta_{-1}^0 &= (\theta_2^0, \theta_3^0, \ldots, \theta_d^0) \\
\theta_{-2}^0 &= (\theta_1^1, \theta_3^0, \ldots, \theta_d^0) \\
&\cdots \\
\theta_{-d}^0 &= (\theta_1^1, \theta_2^1, \ldots, \theta_{d-1}^1) .
\end{aligned}$$

In this manner, the values of each θ_j will be updated in each iteration. Under appropriate conditions, the distribution of the simulated values of θ^t will get closer to the posterior distribution of θ when t gets larger. Depending on the problem, this convergence may occur immediately (or almost immediately), or it may require many (or even tens of thousands of) iterations. Posterior inference can be conducted based on the portion of samples drawn from the iterations after convergence.

Details about other MCMC methods such as the Metropolis and Metropolis-Hastings algorithms can be found in Gelman et al. (2005, Chapter 11).

Regression Models

With noninformative priors utilized for all the parameters, Bayesian regression models (including Bayesian GLMs) give estimates and standard errors equivalent to those from classical regression models. In this case, the difference is that we may still use posterior simulations as an effective tool for implementing predictive inference and model checking in the Bayesian setting.

For Bayesian regression models, external information can be incorporated by specifying informative priors for the parameters. Conjugate priors may be assumed in order to obtain the exact form of the posterior distributions, which may make the model easier to implement. Nonconjugate priors on the regression parameters may also be used. Refer to Gelman et al. (2005, Chapters 14 and 16) for additional details.

3.2.2 Computation in BUGS

BUGS (Spiegelhalter et al., 1996) is a specialized program for MCMC based analysis. BUGS stands for Bayesian inference Using Gibbs Sampling. Several versions of BUGS have been developed for different computer platforms. WinBUGS exists for Windows. OpenBUGS is a version that can run on Windows and Linux, and within the statistical package R. GeoBUGS and PkBUGS are add-ons to WinBUGS that can fit spatial and pharmacokinetic models, respectively. Bayesian full probability models based on MCMC can be implemented in BUGS very conveniently. Examples can be found in Scollnik (2001, 2002a, 2002b, 2004), Verrall (2005), and Gelman et al. (2005, Appendix C). The various BUGS packages can be obtained from these websites:

www.mrc-bsu.cam.ac.uk/bugs

mathstat.helsinki.fi/openbugs.

Simple Bayesian Models

Suppose we have a data set $y = (y_1, y_2, \ldots, y_k)$ from the binomial model specified in Subsection 3.1.1. That is,

$$y_i|n, \theta \sim binomial(n, \theta) , \qquad i = 1, 2, \ldots, k$$

and

$$\theta \sim beta(a, b) .$$

This model can be defined in BUGS with these lines of code:

```
model {
```

```
  for (i in 1:k){
    y[i] ~ dbin(theta, n)
  }

  theta ~ dbeta(a, b)
  a ~ dlnorm(0, 1.0E-6)
  b ~ dlnorm(0, 1.0E-6)
}
```

In the illustration above, noninformative lognormal priors with large variances were specified for the parameters $a > 0$ and $b > 0$ of the beta distribution. Note, for the normal and lognormal distributions in BUGS, the first parameter is the mean and the second parameter is the inverse of the variance (also known as the precision).

Predictive inference can be implemented for the y variables by adding an additional variable y_{k+1} with the same distribution, i.e. simply by adding one more term in the first loop.

```
model {
 for (i in 1:k+1){
   y[i] ~ dbin(theta, n)
 }

 theta ~ dbeta(a, b)
```

```
 a ~ dlnorm(0, 1.0E-6)
 b ~ dlnorm(0, 1.0E-6)
}
```

Once a full probability model is properly defined and coded in BUGS, and the data loaded, BUGS will compile and then run an implementation of a MCMC simulation for the model.

Regression Models

It is also very convenient to define a regression model such as a GLM in BUGS (e.g., see Scollnik, 2002b). For example, we can define a GLM for the binomial data above. The model is

$$y_i|n, \theta_i \sim binomial(n, \theta_i) , \qquad i = 1, 2, \ldots, k ,$$

and

$$\text{logit}(\theta_i) = \beta_1 + \beta_2 i , \qquad i = 1, 2, \ldots, k ,$$

with normal noninformative priors for β_1 and β_2.

The regression model is defined in BUGS by specifying the regression equation with a left arrow $< -$ composed of $<$ and $-$. The above model can be defined in BUGS with these lines of code:

```
model {
 for (i in 1:k){
   y[i] ~ dbin(theta[i], n)
```

```
    theta[i] <- beta1 + beta2 * i
  }

  beta1 ~ dnorm(0, 1.0E-6)
  beta2 ~ dnorm(0, 1.0E-6)
}
```

After loading the data and initial values for the unknown parameters, posterior simulation and inference can be performed in BUGS via a menu-driven interface.

Monitoring Convergence

When the number of iterations is not large enough, the distribution of the simulated values may not be close enough to the target distribution. Therefore, convergence needs to be checked before the simulated samples can be used for posterior analysis. A common method is to monitor convergence by simulating multiple sequences with different starting points. Convergence needs to be monitored for the entire distribution including all the parameters and quantities of interest.

A useful tool for checking the convergence was first introduced by Gelman and Rubin (1992). It is an estimator of a potential scale reduction factor R that is defined based on the between-sequence variance B and within-sequence variance W. Suppose ϕ is a parameter or quantity of interest in the model. With m parallel

sequences each with length n, the estimates of B and W for ϕ are given by

$$B = \frac{n}{m-1}\sum_{j=1}^{m}\left(\bar{\phi}_{.j} - \bar{\phi}_{..}\right)^2$$

$$W = \frac{1}{m}\sum_{j=1}^{m} s_j^2 \,,$$

where

$$\bar{\phi}_{.j} = \frac{1}{n}\sum_{i=1}^{n}\phi_{ij}$$

$$\bar{\phi}_{..} = \frac{1}{m}\sum_{j=1}^{m}\bar{\phi}_{.j}$$

$$s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\phi_{ij} - \bar{\phi}_{.j}\right)^2 \,,$$

and ϕ_{ij} is the ith draw from the jth sequence.

Gelman and Rubin (1992) defined an estimator of the potential scale reduction factor by

$$\hat{R} = \frac{\hat{V}}{W} = \frac{m+1}{m}\frac{\widehat{\mathrm{var}}^{+}(\phi|y)}{W} - \frac{n-1}{mn} \,,$$

where

$$\hat{V} = \widehat{\mathrm{var}}^{+}(\phi|y) + \frac{B}{mn}$$

is a pooled posterior variance estimate taking into account the sampling variability in the estimation of the mean μ, and

$$\widehat{\mathrm{var}}^{+}(\phi|y) = \frac{n-1}{n}W + \frac{1}{n}B$$

is an estimator of the marginal posterior variance of ϕ. $\widehat{\mathrm{var}}^{+}(\phi|y)$ is an unbiased estimator for $\mathrm{var}(\phi|y)$, if the simulation has a starting distribution that is identical

to the target distribution, but overestimates $\text{var}(\phi|y)$ when the initial distribution is overdispersed. The within-sequence variance W is an underestimate of $\text{var}(\phi|y)$ with its expectation increases to $\text{var}(\phi|y)$ as n goes to infinity. Hence, the estimate $\hat{R}$ will decline to 1 as $n \to \infty$. The potential scale reduction factor $\hat{R}$ can be used as an indicator for convergence. When $\hat{R}$ is close to 1, we may consider the m sequences of the n simulated values to be converged.

Brooks and Gelman (1998) further refined the potential scale reduction factor by incorporating a correction factor accounting for the sampling variability of the variance estimates. The correction factor is determined using the method of Fisher (1935). Their refined potential scale reduction factor is given by

$$\hat{R}_c = \frac{d+3}{d+1}\hat{R} = \frac{d+3}{d+1}\frac{\hat{V}}{W}, \tag{3.1}$$

where $d \approx 2\hat{V}/\widehat{\text{var}}(\hat{V})$. In this refined potential scale reduction factor, the estimator $\hat{V}$ is corrected for the degrees of freedom on which it is based. See Brooks and Gelman (1998, pages 437-438) for details of this derivation.

BUGS calculates the refined potential scale reduction factor $\hat{R}_c$ (Brooks and Gelman, 1998) automatically for use of monitoring the convergence of the simulation. A graphical approach is used which makes the monitoring of convergence easier to conduct.

Note, as in Gelman et al. (2005, Chapter 11, pages 294-299), the potential scale reduction factor can also be defined as

$$\hat{R} = \sqrt{\frac{\widehat{\text{var}}^{+}(\phi|y)}{W}},$$

which will also decline to 1 as $n \to \infty$.

To monitor the convergence of the entire distribution, we need to estimate the potential scale reduction factors for all the parameters and quantities of interest. The simulation needs to be run until every parameter's $\hat{R}_c$ is close to 1. Usually, $\hat{R}_c$ values below 1.1 will be acceptable. We may combine the second halves of all the sequences to use as our sample for posterior reference. Details can be found in Gelman and Rubin (1992), Brooks and Gelman (1998), Gelman et al. (2005, Chapter 11, pages 294-299), or other books with topics in MCMC simulations.

Chapter 4

Zeros and Negatives

4.1 The Problem

At this point, let us return to the loss reserving context introduced in Chapter Two. In practice, it is frequently the case that zero losses appear in some of the cells making up the incremental loss triangle. This is especially true at the later stage of the development years, as most outstanding claims will have been settled by that time (Kunkler, 2004, Abstract).

We may even expect negative values in the loss triangle due to various reasons arising from insurance practices (Kunkler, 2004, 2006; de Alba, 2002a, 2006). By subrogation, the insurance company can obtain the right to claim from a third party for paid losses. So after paying a claim the insurance company may recover an amount of money at a future date from that party which will appear as a negative loss. Similarly, in marine insurance, it is a customary practice that the insurance company pay the full amount of goods to the insured party and get the residual value, which is called salvage. In this case, a negative loss arises when the insurance company sells the goods and recover some amount of money back. There are other reasons for which negative losses may occur, such as the cancellation of a claim, initial over-estimation of a loss, consequences of judicial decisions, and errors.

However, most of the stochastic models presented in Chapter Two assume positive values for the claim data. For the lognormal model (Kremer, 1982) in which lognormal distributions are specified for the loss data, zero and negative losses will make the model undefined. A similar problem arises with the over-dispersed Poisson (Renshaw and Verrall, 1998) and negative binomial (Verrall, 2000) loss reserving models. Although the quasi-likelihood method can be used for these models so as to accommodate non-integer, zero and negative values, too many zero or negative values may result in negatives in some columns, which will make the model inappropriate.

4.2 Previous Work

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 (2002a, 2006) to extend the lognormal model (Kremer, 1982) to situations where there are negative values in the loss triangle. Kunkler (2004) put forward a Bayesian binomial mixture model for the situation when there are zeros in the loss triangle data. Kunkler (2006) proposed a similar model for a loss triangle with values composed of positives and negatives.

4.2.1 An Improved Lognormal Model

For the incremental loss triangle the notation y_{ij} is used instead of the P_{ij} used in the previous chapters so as to be consistent with the commonly used notation for probabilistic models. In the improved lognormal model proposed by de Alba (2002a,

2006), a threshold parameter $\delta > 0$ is introduced so that the lognormal model can be applied to adjusted loss data $y_{ij} + \delta > 0$. That is,

$$y_{ij} + \delta \sim LN(\mu_{ij},\ \sigma^2)\ , \qquad i = 1, \ldots, n;\ j = 1, \ldots, n\ .$$

The three-parameter ANOVA structure is used to model the mean by

$$\mu_{ij} = \mu + \alpha_i + \beta_j\ , \qquad i = 1, \ldots, n;\ j = 1, \ldots, n\ .$$

Corner constraints are assumed whereby $\alpha_1 = \beta_1 = 0$.

For a Bayesian analysis to proceed, the prior distributions must be specified for the model parameters. For example, de Alba (2002a, 2006) assumes the prior distributions

$$\begin{aligned}
\mu &\sim N(0,\ \sigma_\mu^2) \\
\alpha_i &\sim N(0,\ \sigma_{\alpha_i}^2)\ , \quad i = 1, 2, \ldots, n \\
\beta_j &\sim N(0,\ \sigma_{\beta_j}^2)\ , \quad j = 1, 2, \ldots, n \\
\sigma^2 &\sim IG(\nu,\ \lambda) \\
\delta &\sim N(\mu_\delta,\ \sigma_\delta^2)\ ,
\end{aligned}$$

where $IG(\nu,\ \lambda)$ stands for an inverse Gamma distributions with parameters ν and λ. The parameters (hyperparameters) in these prior distributions must now be specified themselves. If precise values for the hyperparameters are unavailable, then a Bayesian hierarchical modelling approach can be adopted, and the hyperparameters can be assigned prior distributions reflecting a lack of information. de Alba (2006,

page 56) took this approach and adopted the hyperprior distributions

$$\begin{aligned}
\sigma_{\mu}^2 &\sim IG(0.1,\ 0.1) \\
\sigma_{\alpha_i}^2 &\sim IG(0.001,\ 0.001)\ , \quad i = 1, 2, \ldots, n \\
\sigma_{\beta_j}^2 &\sim IG(0.001,\ 0.001)\ , \quad j = 1, 2, \ldots, n \\
\nu &\sim G(2.5,\ 0.1) \\
\lambda &\sim G(2,\ 0.1) \\
\mu_{\delta} &\sim N(200,\ 10000) \\
\sigma_{\delta}^2 &\sim IG(0.0001,\ 0.1)\ ,
\end{aligned}$$

where $G(a,\ b)$ stands for a Gamma distributions with parameters a and b.

In the manner described above, de Alba (2002a, 2006) improved the lognormal model so that the model can be applied even if there are negative values in the loss triangle. One limitation of this model is that it assumes a minimum value $-\delta$ for the loss data, which is not true in reality. And it may also not be appropriate where there are zeros in the loss triangle, especially when the proportion of zeros in the loss triangle is large.

4.2.2 A Binomial Mixture Model for Zeros

Modelling Mixture Data

A binomial mixture model is applied by Kunkler (2004) for modelling zeros in the lognormal model. An indicator or mixture data triangle $\mathrm{z} = \{z_{ij} :\ i = 1, 2, \ldots, n_a;\ j =$

$1, 2, \ldots, n_a - i + 1\}$ is introduced to model the sign of the data, where

$$z_{ij} = \begin{cases} 0 & \text{if } y_{ij} = 0 \\ 1 & \text{if } y_{ij} > 0 \end{cases} .$$

The model assumes that the probability of getting positive losses depends only on development year j. This is consistent with the empirical observation that there tends to be more zeros at the later development years. Denoting $\mathrm{P}(z_{ij} = 1) = \lambda_j$, it is easy to see z_{ij} follows a Bernoulli distribution with its sampling probability given by

$$p(z_{ij}|\lambda_j) = \lambda_j^{z_{ij}} (1 - \lambda_j)^{1 - z_{ij}} .$$

Since the Bernoulli distribution is the special case of a binomial distribution with a single trial (i.e., $n = 1$), link functions such as the logistic, probit and complementary log-log link functions can be used for the mixture data. Putting $l(\lambda_j)$ as the link function for λ_j, a piecewise linear relationship is proposed in the form of

$$l(\lambda_j) = \sum_{d=0}^{j-1} \delta_d , \tag{4.1}$$

a special case of which is

$$l(\lambda_j) = \delta_0 + (j - 1)\delta_1 , \tag{4.2}$$

when $\delta_1, \ldots, \delta_d$ are assumed to be equal.

Modelling Magnitude Data

For the magnitude data of the positive losses, the author uses the lognormal distribution as the sampling distribution. Denoting $p(\mathrm{y}^+|\theta) = p(y_{ij}|z_{ij} = 1, \theta)$, the sampling distribution of loss magnitudes is given by

$$\log(\mathrm{y}^+)|\theta \sim N(\mathrm{X}_\beta\beta, \sigma^2 I) ,$$

where $\theta = (\beta, \sigma^2)$, and I is an identity matrix with dimension equal to the number of positive values. Kunkler (2004) proposed the model structure of Zehnwirth (1994) for the mean of the lognormal model. In this model the form of $\mathrm{X}_\beta\beta$ is given by

$$(\mathrm{X}_\beta\beta)_{ij} = \alpha_i + \sum_{d=1}^{j-1}\gamma_d + \sum_{t=1}^{i+j-2}\iota_t \,. \tag{4.3}$$

Bayesian posterior analysis can be performed when prior information is specified for the variance parameter and the parameters for the regression models in Equations (4.1), (4.2), and (4.3).

A drawback of the structure in (4.3) is that for loss triangle data, we are usually not able to get information about losses for $i + j > n_a + 1$, the lower triangle. So it will not be possible to get estimates of ι_t for $t > n_a - 1$, which will complicate the predictive analysis. And the model only copes with the situation where there are zeros, probably due to the computational difficulty.

4.2.3 A Binomial Mixture Model for Negatives

Modelling Mixture Data

Kunkler (2006) proposed another binomial mixture model for extending the lognormal model to situations where there are negative values in the loss triangle. The key difference between this model and the one in Kunkler (2004) is that it defines a different indicator or mixture data triangle $\mathrm{z} = \{z_{ij} : \; i = 1, 2, \ldots, n_a; \; j = 1, 2, \ldots, n_a - i + 1\}$ as

$$z_{ij} = \begin{cases} -1 & \text{if } y_{ij} < 0 \\ 1 & \text{if } y_{ij} > 0 \end{cases} .$$

Similar to Kunkler (2004), it is easy to see that $z'_{ij} = (z_{ij} + 1)/2$ follows a Bernoulli distribution with its sampling probability given by

$$p(z'_{ij}|\lambda_j) = \lambda_j^{z'_{ij}} (1 - \lambda_j)^{1-z'_{ij}} \ ,$$

where $\lambda_j = \mathrm{P}(z_{ij} = 1)$.

The same link functions proposed in Kunkler (2004), such as the logistic, probit and complementary log-log link functions, can be used with this model. letting $l(\lambda_j)$ denote the link function for λ_j, the same piecewise linear relationship as in Kunkler (2004) is proposed as

$$l(\lambda_j) = \sum_{d=0}^{j-1} \delta_d \ , \tag{4.4}$$

with a special case of this being

$$l(\lambda_j) = \delta_0 + (j-1)\delta_1 \ . \tag{4.5}$$

Modelling Magnitude Data

The lognormal distribution is assumed as the sampling distribution for the magnitude data of both the negative and positive losses. That is,

$$\begin{aligned} |y_{ij}| &\sim LN\left(\mu_{ij}, \frac{\sigma^2}{\omega^-}\right) \quad \text{if } y_{ij} < 0 \\ |y_{ij}| &\sim LN\left(\mu_{ij}, \frac{\sigma^2}{\omega^+}\right) \quad \text{if } y_{ij} > 0 \ . \end{aligned}$$

Two model structures are proposed by Kunkler (2006) to use for the model for the magnitude data, each corresponds to an assumption for the form of the parameters

μ_{ij}.

The first one is derived from the three parameter ANOVA structure introduced by Kremer (1982). Based on this, Kunkler (2006) suggested

$$\mu_{ij} = \mu + \left(\alpha_i^+ + \gamma_j^+\right) I_{(z_{ij}=1)} + \left(\alpha_i^- + \gamma_j^-\right) I_{(z_{ij}=-1)} \,, \tag{4.6}$$

where a common parameter of μ is assumed for both the positive and negative magnitude, I_A is the indicator function with a value of 1 when A is true and 0 otherwise, and α_i^+, γ_j^+, α_i^-, γ_j^- are the row and column parameters for the positive and negative magnitudes.

The second structure is obtained from the probabilistic trend family of models described by Zehnwirth (1994). Based on this Kunkler (2006) proposed

$$\mu_{ij} = \left(\alpha_i^+ + \sum_{d=1}^{j-1} \gamma_d^+\right) I_{(z_{ij}=1)} + \left(\alpha_i^- + \sum_{d=1}^{j-1} \gamma_d^-\right) I_{(z_{ij}=-1)} + \sum_{t=1}^{i+j-2} \iota_t \,, \tag{4.7}$$

where common calendar year trend factors of ι_t $(t = 1, 2, \ldots, 2n_a - 2)$ but different parameters of α_i^+, γ_d^+, α_i^-, γ_d^- $(i = 1, 2, \ldots, n_a,\ d = 1, 2, \ldots, n_a - 1)$ are assumed for positive and negative magnitudes.

An Example

Kunkler (2006) performed a Bayesian analysis of his binomial mixture model for an adjusted loss triangle from the 'Historical Loss Development Study' (1991) published by the Reinsurance Association of America. The adjusted losses considered by Kunkler (2006) is given in Table 4.1. Observe that this loss triangle contains both

Table 4.1: Adjusted Loss Triangle with Negatives

	Development year									
Accident 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	673	535	
3	3410	5582	4881	2268	2594	3479	649	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									

positive and negative losses.

For the mixture data, Kunkler (2006) adopted a model structure different from (4.4) and (4.5) for the binomial GLM. Based on a preliminary data analysis described in Kunkler (2004), he chose a logit model

$$\text{logit}(\lambda_j) = \delta_0 + (j-5)\,\delta_1 I_{(j>5)}\,,$$

and assumed noninformative priors for the parameters δ_0 and δ_1.

For the loss magnitude data, Kunkler (2006) used a model structure which is different from (4.6) and (4.7). This was a simplified model, with linear relationships assumed for both the development year and calendar year parameters so as to reduce the number of parameters. The simplified model is given by

$$\mu_{ij} = (\alpha^+ + (j-1)\gamma^+)I_{(z_{ij}=1)} + (\alpha^- + (j-1)\gamma^-)I_{(z_{ij}=-1)} + (i+j-2)\iota\,, \quad (4.8)$$

where α^+, γ^+, α^- and γ^- are base magnitudes and development year parameters

which are assumed to be different for positives and negatives, and ι is the common calendar year parameter.

On the basis of a residual analysis, Kunkler (2006, page 547) found that this simplified model failed to capture certain major trends in the development period directions for both the negative and positive data, and also failed to capture a level change between accident periods five and six for the positive data. Accordingly, Kunkler revised the model to

$$\begin{aligned}\mu_{ij} = &\left(I_{(i<=5)}\alpha_1^+ + I_{(i>5)}\alpha_2^+\right) I_{(z_{ij}=1)} + \alpha_1^- I_{(z_{ij}=-1)} \\ &+ \left\{I_{(j>1)}\gamma_1^+ + I_{(j>2)}\gamma_2^+ + I_{(j>3)}\gamma_3^+ + \left[I_{(4\leq j\leq 6)}(j-4) + I_{(j>6)}2\right]\gamma_4^+ \right. \\ &\quad \left. + I_{(j>6)}(j-6)\gamma_5^+\right\} I_{(z_{ij}=1)} \\ &+ \left\{\left[I_{(j\leq 3)}(j-1) + I_{(j>3)}2\right]\gamma_1^- + I_{(j>3)}(j-3)\gamma_2^-\right\} I_{(z_{ij}=-1)} \\ &+ (i+j-2)\iota\,.\end{aligned}$$

Kunkler (2006) found that this revised model appeared to capture the major levels and trends in the data. Further details of his analysis are provided in Kunkler (2006). In the next chapter, we will show how the model in Kunkler (2006) can be implemented in BUGS.

Chapter 5

Implementing Kunkler's Model in BUGS

5.1 Coding the Mixture Model

In the previous chapter, we reviewed several loss reserving models that have been proposed for use in special situations when zero and negative values appear in the loss triangle. In this chapter we will implement one of these, the model by Kunkler (2006), in BUGS. The model in Kunkler (2006) was originally implemented using MatLab (developed by the MathWorks, Inc.), along with the Econometrics Toolbox of econometric functions for use in MatLab developed by LeSage (LeSage, 1999).

For the binomial mixture model, we denote $\frac{z_{ij}+1}{2}$ in Subsection 4.2.3 as z_{ij} for simplicity. We first code the binomial mixture model

$$z_{ij} \sim Bernoulli(\lambda_j), \qquad i = 1, 2, \ldots, n_a;\ j = 1, 2, \ldots, n_a\,,$$

where the Bernoulli probability λ_j is modelled with a logit structure

$$\text{logit}(\lambda_j) = \delta_0 + (j-5)\,\delta_1 I_{(j>5)}\ .$$

Mildly informative priors are assumed for the parameters δ_0 and δ_1. That is,

$$\begin{aligned} \delta_0 &\sim N(0,\ 100) \\ \delta_1 &\sim N(0,\ 100)\ . \end{aligned}$$

The BUGS codes for this model is

```
for (j in 1:10) {
  for (i in 1:10) {
    z[i, j] ~ dbern(p[j])
    }
      }

for (j in 1:10) {
  # Logit model
  logit(p[j]) <- delta[1]+ (j-5)*step(j-6)* delta[2]
     }

for (j in 1:2) {
  # Prior distribution
  delta[j] ~ dnorm(0, 0.01)
     }
```

5.2 Coding the Magnitude Model

The magnitude data for both the negatives and positives are assumed to be lognormal. That is,

$$|y_{ij}| \sim LN\left(\mu_{ij}, \frac{\sigma_{ij}^2}{\omega^-}\right) \quad \text{if } y_{ij} < 0 \tag{5.1}$$

$$|y_{ij}| \sim LN\left(\mu_{ij}, \frac{\sigma_{ij}^2}{\omega^+}\right) \quad \text{if } y_{ij} > 0\,, \tag{5.2}$$

where the means of these lognormal distributions are modelled with a GLM structure chosen by Kunkler (2006). The model is given by

$$\begin{aligned}
\mu_{ij} = &\left(I_{(i<=5)}\alpha_1^+ + I_{(i>5)}\alpha_2^+\right) I_{(z_{ij}=1)} + \alpha_1^- I_{(z_{ij}=-1)} \\
&+ \left\{I_{(j>1)}\gamma_1^+ + I_{(j>2)}\gamma_2^+ + I_{(j>3)}\gamma_3^+ + \left[I_{(4\le j\le 6)}(j-4) + I_{(j>6)}2\right]\gamma_4^+\right. \\
&\quad \left. + I_{(j>6)}(j-6)\gamma_5^+\right\} I_{(z_{ij}=1)} \\
&+ \left\{\left[I_{(j\le 3)}(j-1) + I_{(j>3)}2\right]\gamma_1^- + I_{(j>3)}(j-3)\gamma_2^-\right\} I_{(z_{ij}=-1)} \\
&+ (i+j-2)\iota\,.
\end{aligned}$$

Kunkler (2006) assumed noninformative (actually, mildly informative) priors of $N(0, 1000)$ for all of the α, γ and ι parameters appearing in the above model.

The parameters ω^+ and ω^- in Equations (5.1) and (5.2) were estimated as 5.9511 and 6.1646, respectively, by Kunkler (2006) on the basis of a preliminary data analysis. See Kunkler (2006, pages 550, 553-554) for details. We use these same estimated values of ω^- and ω^+. Kunkler (2006) used the *ols_g*() function (LeSage, 1999, Chapter 6, pages 175-178) in MatLab to implement Gibbs sampling for this GLM model.

This function defines the prior density specification of σ_{ij}^2 in this way:

$$\begin{aligned}
\sigma_{ij}^2 &= sige \times v_{ij} \\
\frac{r}{v_{ij}} &\sim i.i.d. \ \frac{\chi^2(r)}{r} \\
sige &\sim gamma(nu, d0) \\
r &\sim gamma(nr, kr)\,.
\end{aligned}$$

Kunkler (2006) used a fixed value of r, i.e. $r = 100$. We assume Kunkler (2006) used the values $nu = 0$ and $d0 = 0$, which are the default values used by $ols\_g()$ function as described in LeSage (1999, page 176). This results in a diffuse or noninformative $gamma(0, 0)$ prior for *sige*.

We code the same model for the magnitudes data in BUGS in the following manner.

```
for (i in 1:10) {
   for (j in 1:10) {
     y[i, j] <- yp[i, j]*(2*z[i, j]-1)
     yp[i, j] ~ dlnorm(mu[i, j],  tao[i, j])

     # Modelling the mean
     mu[i, j] <-a[i, j]+gp[i, j]+gn[i, j]+(i+j-2)*iota
     a[i, j] <- (step(5-i)*alphap[1]+step(i-6)*alphap[2])*z[i, j]
                +alphan*(1-z[i, j])
     gp[i, j] <- (step(j-2)*gammap[1]+step(j-3)*gammap[2]
```

```
                +step(j-4)*(gammap[3]+step(6-j)*(j-4)*gammap[4])
                +step(j-7)*(2*gammap[4]+(j-6)*gammap[5]))*z[i, j]
    gn[i, j] <- (step(3-j)*(j-1)*gamman[1]+step(j-4)*(2*gamman[1]
                +(j-3)*gamman[2]))*(1-z[i, j])
       }
          }

# Modelling the inverse-variance (precision)
for (i in 1:10) {
  for (j in 1:10) {
    tao[i, j] <- tau[i,j]*(6.1646*z[i, j]+5.9511*(1-z[i, j]))
    tau[i,j] <- 1/(sige*v[i,j])
    v[i,j] <- r*r/c[i,j]
    c[i,j] ~ dchisqr(r)
       }
          }

# Priors for the parameters
sige ~ dgamma(0, 0)

alphan ~ dnorm(0, 1.0E-3)
iota ~ dnorm(0, 1.0E-3)

for (i in 1:2) {
```

```
    alphap[i] ~ dnorm(0, 1.0E-3)
    gammap[i] ~ dnorm(0, 1.0E-3)
    gamman[i] ~ dnorm(0, 1.0E-3)
     }

  for (i in 3:5) {
    gammap[i] ~ dnorm(0, 1.0E-3)
     }
```

5.3 Comparison of Results

We ran three chains for the model with dispersed initial values, and monitored all of the parameters for convergence. Consider the inverse variance (i.e., precision) parameters $tau[i,j] = \frac{1}{\sigma_{ij}^2}$. From the history plots of these parameters in BUGS, we observe that it only takes 100 iterations before the three paths start to mix for every one of the $tau[i,j]$ parameters. Figure 5.1 is the history plot for the sampled values of the precision $tau[1,1]$ for the first magnitude entry in the loss triangle.

In order to diagnose convergence, we also monitor the refined potential scale reduction factors $\hat{R}_c$ (Brooks and Gelman, 1998), previously defined in Equation (3.1). In the case of the precision parameters $tau[i,j]$, the corresponding refined potential scale reduction factor $\hat{R}_c$ converged to approximately 1 within about 1000 iterations (e.g., see Table 5.1).

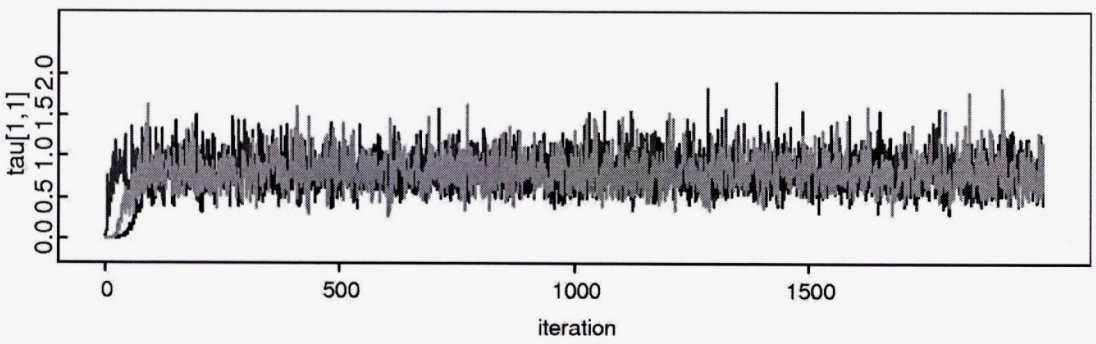

Figure 5.1: History Plot of the Precision Parameter $tau[1, 1]$

Table 5.1: Refined Potential Scale Reduction of the Precision Parameter $tau[1, 1]$

Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$
550	1.016	1050	1.007	1550	1.008	2050	1.003
600	1.033	1100	1.002	1600	1.004	2100	1.003
650	1.001	1150	0.995	1650	1.001	2150	1.003
700	0.986	1200	0.996	1700	1.009	2200	1.001
750	0.995	1250	0.998	1750	1.006	2250	1.003
800	0.985	1300	1.001	1800	1.003	2300	1.003
850	1.000	1350	1.002	1850	1.004	2350	1.004
900	0.997	1400	1.006	1900	1.003	2400	1.002
950	1.006	1450	0.999	1950	1.001	2450	1.002
1000	1.010	1500	1.006	2000	1.003	2500	1.001

On the basis of history plots and examination of $\hat{R}_c$ for the other parameters in the model, we determine that all of these other parameters converge before the 2000th iteration. So we can use a sequence with the same length as in Kunkler (2006) to estimate the posterior distribution of the parameters and the reserves. We use the simulated values from iterations 2001 to 12000 from all three chains in the simulation, a total of 30000 posterior samples, for our analysis. The estimates of our parameters are very close to those from Kunkler (2006). Please refer to Table 5.2

below for details.

Table 5.2: Estimates of Parameters for the Magnitude Model

	BUGS results		Kunkler 2006	
Parameter	mean	STD	mean	STD
α_1^+	7.797	0.220	7.800	0.226
α_2^+	7.187	0.356	7.195	0.368
γ_1^+	0.557	0.222	0.552	0.229
γ_2^+	0.020	0.234	0.022	0.236
γ_3^+	-0.624	0.250	-0.614	0.25
γ_4^+	-0.219	0.142	-0.225	0.142
γ_5^+	-0.681	0.105	-0.677	0.106
α_1^-	5.253	0.372	5.251	0.376
γ_1^-	0.809	0.244	0.816	0.242
γ_2^-	-0.611	0.107	-0.614	0.109
ι	0.068	0.045	0.068	0.047

The percentiles of the reserve estimates from Kunkler (2006) are given in Table 5.3. The reserve estimates, their corresponding standard deviations, and some of their percentiles on the basis of the BUGS output are listed in Table 5.4.

From the above results, we can see that the mean and percentiles from MatLab and BUGS are very close. Some of the standard deviations for the reserve estimates are quite large, which may be the reason for any differences in the mean and percentile estimates.

Table 5.3: Percentiles of the Reserve Estimates from Kunkler (2006)

Year	1	2.5	5	50	Mean	95	97.5	99
1	0	0	0	0	0	0	0	0
2	-88	-65	-50	141	147	437	539	691
3	-169	-124	-88	444	464	1107	1316	1602
4	-261	-169	-49	1104	1144	2445	2796	3352
5	-317	19	367	2466	2551	5007	5753	6545
6	-304	115	519	2834	2956	5757	6530	7471
7	-13	699	1385	4838	4987	9058	10092	11456
8	198	1387	2407	7493	7670	13450	15029	17085
9	163	2175	3913	12229	12469	21696	24144	27331
10	2382	4801	6701	17144	17724	30478	34025	39022
Total	22624	26940	30157	49098	50112	73325	79139	87377

5.4 Differing Priors Density Specifications

As noted earlier, the $gamma(0,0)$ prior presumably adopted by Kunkler (2006) for the parameter *sige* is described as a diffuse prior by LeSage (1999, page 176). However, we are not sure whether in BUGS $gamma(0,0)$ is necessarily defined in quite the same way. In this section, we will code the model using several different diffuse prior density specifications for *sige*.

5.4.1 Types of Diffuse Priors

Proportional Density

The gamma distribution $gamma(\alpha,\beta)$ is normally only defined for $\alpha > 0$ and $\beta > 0$. Kunkler (2006) used a diffuse prior of $gamma(0,0)$. From the definition of the gamma density function, assuming $gamma(0,0)$ is defined in the obvious way, its density function would have the form of $p(x) = \frac{0^0}{\Gamma(0)}x^{-1}, \quad x > 0$. From this form, a

Table 5.4: Mean, STD and Percentiles of Reserve Estimates from BUGS

Year	mean	sd	MC error	2.50%	5.00%	median	95.00%	97.50%
1	0	0	0	0	0	0	0	0
2	153	168	1.11	-66	-50	146	444	539
3	486	375	2.74	-116	-81	466	1130	1316
4	1192	740	6.09	-147	29	1156	2475	2826
5	2692	1420	13.65	221	576	2589	5138	5856
6	3104	1575	11.51	334	765	2980	5822	6587
7	5272	2356	15.54	1129	1788	5082	9325	10490
8	8023	3337	24.61	2007	2993	7828	13750	15270
9	13100	5426	47.00	3104	4760	12780	22270	24850
10	18590	7176	84.20	5923	7881	18030	31030	34200
Total	52620	13260	156.4	29760	33070	51460	75970	81890

reasonable guess of the diffuse prior in MatLab for *sige* would be

$$p(sige) \propto \frac{1}{sige}. \tag{5.3}$$

This form of prior is discussed in Gelman et al. (2005, Chapter 2, pages 61-65).

This prior can be defined in BUGS using the "ones trick" (Spiegelhalter, Thomas, Best, and Lunn, 2004), where an imaginary observation with value 1 is used to obtain the desired prior. The 1 is assumed to be an observation from a Bernoulli distribution with probability p. Keeping in mind that the contribution made to the likelihood by a Bernoulli observation with a value of 1 is given by p, we see that the correct prior density contribution results if p is set equal to a term proportional to the desired prior density.

The diffuse prior in Equation (5.3) can be defined in BUGS using the following code.

```
one <- 1
c <- 1000          # a large number to ensure that p<1
sige ~ dflat()
p <- (1/sige)/c    # expression for desired prior of sige
one ~ dbern(p)
```

Uniform Prior

Another type of diffuse prior is a uniform prior on an interval. For the prior of the parameter *sige* in Kunkler (2006), a uniform prior on an interval $(0, L)$ can be assumed. That is,

$$sige \sim U(0, L)\ . \tag{5.4}$$

For the interval upper bound L, we can choose big values such as 10 or 100 in order to make the prior information vague. This prior can be easily defined in BUGS as follows:

```
L <- 100
sige ~ dunif(0, L)
```

Flat Prior

BUGS has a standard function $dflat()$ which represents a useful form of diffuse prior. The flat prior function $dflat()$ is an improper (flat) prior which assumes equal probability for each value on the whole real line. It is very straightforward to specify a flat prior for *sige* in BUGS. That is,

```
sige ~ dflat().
```

5.4.2 Comparison of Results

Time for Convergence

Using the proportional prior defined in Equation (5.3), it took much longer for the sample paths of the precision parameter $tau[i, j]$ to converge. The simulated values from the three chains did not mix until after about 20,000 iterations. This is shown in the history plot of $tau[1, 1]$ in Figure 5.2. The history plots for all of the other $tau[i, j]$ parameters are similar.

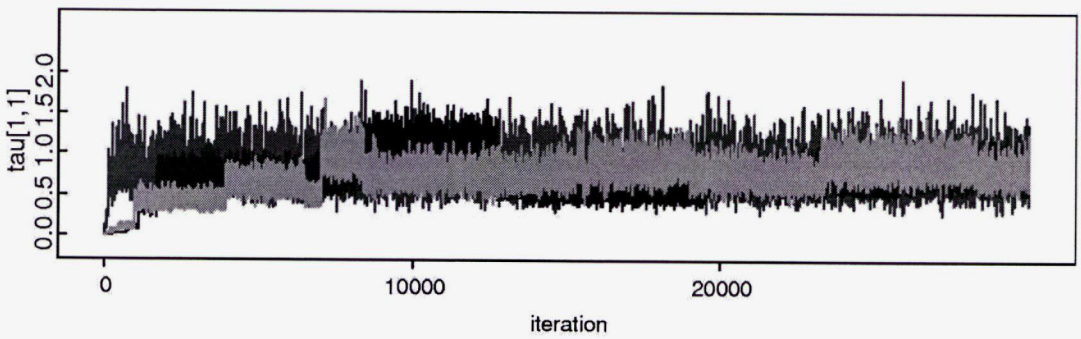

Figure 5.2: History Plot of $tau[1, 1]$ with Proportional Prior

From the refined potential scale reduction factors, it is easy to see that after 20,000 iterations, the potential scale reduction is very close to 1. The values of the potential scale reduction for iterations from 550 to 21,500 are listed in Table 5.5 as an example. For the iterations 22,001 to 30,000, the values are similar to those from iterations 20,050 to 21,500. The refined potential scale reduction factors for all the

other parameters stay close to 1 after about 20,000 iterations.

Table 5.5: Refined Potential Scale Reduction of $tau[1,1]$ with Proportional prior

Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$
550	5.366	5050	1.213	20050	1.129	20800	1.039
600	5.002	5100	1.038	20100	1.062	20850	1.038
650	5.604	5150	0.982	20150	1.073	20900	1.027
700	4.632	5200	0.923	20200	1.100	20950	1.033
750	4.498	5250	0.975	20250	1.106	21000	1.029
800	4.685	5300	1.056	20300	1.100	21050	1.032
850	4.809	5350	1.013	20350	1.062	21100	1.030
900	4.545	5400	1.008	20400	1.028	21150	1.024
950	4.666	5450	0.997	20450	1.056	21200	1.023
1000	3.404	5500	0.998	20500	1.060	21250	1.021
1050	2.838	5550	1.017	20550	1.024	21300	1.018
1100	2.650	5600	0.989	20600	1.026	21350	1.022
1150	2.183	5650	0.992	20650	1.034	21400	1.021
1200	2.091	5700	1.005	20700	1.039	21450	1.017
⋮	⋮	⋮	⋮	20750	1.041	21500	1.017

For the uniform prior of *sige* defined in Equation (5.4), we choose $L = 100$. The convergence of the simulation is very quick and similar to the simulation when using $gamma(0,0)$ as the prior. From the history plot of $tau[1,1]$ in Figure 5.3, it is apparent that the three chains began to mix after about a hundred iterations. The history plots for all of the other $tau[i,j]$ parameters look similar. From the refined potential scale reduction factors, it is easy to see that after 1000 iterations, the potential scale reduction is very close to 1. The sample values of the potential scale reduction for iterations from 550 to 2500 are listed in Table 5.6. The values of the potential scale reduction for the iterations after 2500 look similar. All the other parameters seem to converge after about 1000 iterations.

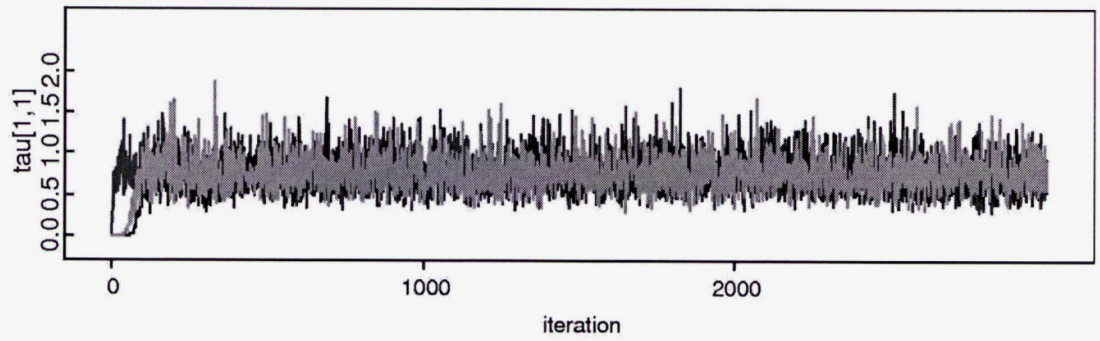

Figure 5.3: History Plot of $tau[1,1]$ with Uniform Prior

Table 5.6: Refined Potential Scale Reduction of $tau[1,1]$ with Uniform prior

Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$
550	1.023	1050	0.996	1550	1.010	2050	0.994
600	1.046	1100	0.996	1600	1.009	2100	0.994
650	0.992	1150	1.011	1650	1.006	2150	0.995
700	1.002	1200	1.008	1700	1.006	2200	0.996
750	1.010	1250	1.001	1750	1.008	2250	0.995
800	1.003	1300	1.007	1800	1.006	2300	0.997
850	1.007	1350	1.005	1850	1.000	2350	0.997
900	1.009	1400	1.009	1900	1.001	2400	1.002
950	0.998	1450	1.005	1950	1.000	2450	0.998
1000	0.988	1500	1.010	2000	0.996	2500	1.000

For the flat prior with domain on the whole real line, the convergence of the simulation is so much slower that we have to choose less dispersed starting values for the three chains to reach convergence. Using the new starting points, the three chains start to mix well after about 27,000 iterations. This can be observed from the history plot of $tau[1,1]$ in Figure 5.4 as an example.

The values of the refined potential scale reduction factors are consistent with

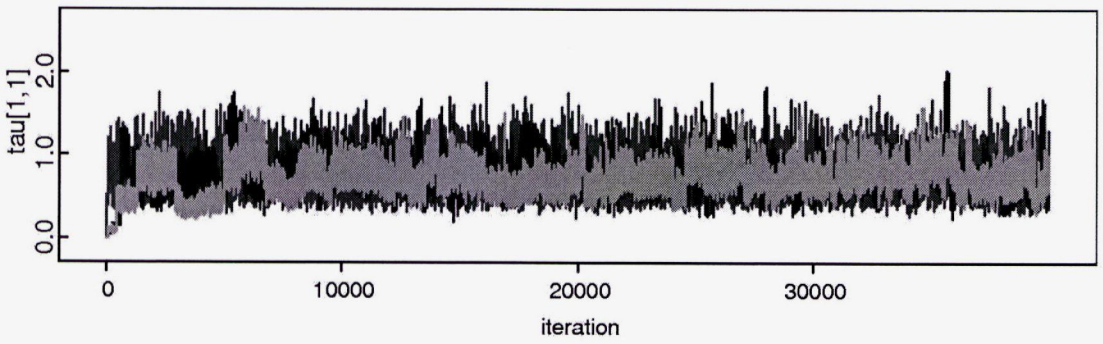

Figure 5.4: History Plot of $tau[1,1]$ with Flat Prior

what we observe in the history plot. The refined potential scale reductions are close to 1 after about 30,000 iterations. The values of the refined potential scale reduction factors for iterations from 550 to 31,500 are listed in Table 5.7. All of the other parameters also seem to converge after about 30,000 iterations.

Based on the convergence results described above, we will use the simulated values for iterations 20,001 to 30,000 from the three chains for the estimation of parameters and reserves in the case of the proportional prior. Iterations 2001 to 12,000 will be used for the posterior inference of the model using the uniform prior. For the model with the flat prior, iterations 30,001 to 40,000 will be used for estimating the parameters as well as the reserves.

Estimation for Parameters

The posterior summaries for the main parameters from the models with the three different diffuse prior density specifications are listed in Table 5.8. From this table,

Table 5.7: Refined Potential Scale Reduction of *tau*[1, 1] with Flat prior

Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$
550	2.176	10050	1.205	30050	1.137	30800	1.014
600	1.522	10100	1.141	30100	1.098	30850	1.013
650	1.519	10150	1.222	30150	1.031	30900	1.009
700	1.450	10200	1.122	30200	1.021	30950	1.015
750	1.333	10250	1.136	30250	1.049	31000	1.019
800	1.332	10300	1.144	30300	1.045	31050	1.016
850	1.325	10350	1.144	30350	1.031	31100	1.017
900	1.328	10400	1.150	30400	1.029	31150	1.017
950	1.327	10450	1.149	30450	1.036	31200	1.009
1000	1.305	10500	1.146	30500	1.026	31250	1.004
1050	1.304	10550	1.165	30550	1.015	31300	1.003
1100	1.300	10600	1.136	30600	1.031	31350	1.003
1150	1.291	10650	1.132	30650	1.031	31400	1.003
1200	1.309	10700	1.111	30700	1.027	31450	0.997
⋮	⋮	⋮	⋮	30750	1.018	31500	1.000

we can see that the results are very close to each other. The type of the diffuse prior selected for the *sige* parameter has little influence on our estimation.

Estimation of Reserves

The reserve estimates from the models with the three different diffuse priors are listed in Tables 5.9, 5.10, and 5.11. As was the case with the posterior parameter summaries, these results are all very close to one another. We can conclude from our results that the type of diffuse prior on *sige* has little influence on our model estimation.

Table 5.8: Estimates of Parameters for the Magnitude Model with Different Priors

	Proportional		Uniform		Flat		Kunkler 2006	
Parameter	mean	std	mean	std	mean	std	mean	std
α_1^+	7.804	0.227	7.798	0.227	7.798	0.235	7.800	0.226
α_2^+	7.207	0.374	7.192	0.372	7.194	0.385	7.195	0.368
γ_1^+	0.554	0.224	0.548	0.236	0.556	0.237	0.552	0.229
γ_2^+	0.023	0.232	0.033	0.242	0.022	0.242	0.022	0.236
γ_3^+	-0.606	0.242	-0.614	0.253	-0.609	0.260	-0.614	0.25
γ_4^+	-0.225	0.141	-0.230	0.144	-0.229	0.146	-0.225	0.142
γ_5^+	-0.676	0.105	-0.677	0.109	-0.676	0.108	-0.677	0.106
α_1^-	5.255	0.374	5.245	0.384	5.247	0.391	5.251	0.376
γ_1^-	0.819	0.237	0.817	0.248	0.819	0.252	0.816	0.242
γ_2^-	-0.613	0.106	-0.613	0.110	-0.614	0.113	-0.614	0.109
ι	0.066	0.047	0.068	0.048	0.067	0.049	0.068	0.047

5.5 Specification of Logit Model

Kunkler's Specification

For the binomial mixture model, recall that in Subsection 5.1.1 we let

$$\begin{aligned} z_{ij} &\sim Bernoulli(\lambda_j), \qquad i = 1, 2, \ldots, n_a;\ j = 1, 2, \ldots, n_a \\ \text{logit}(\lambda_j) &= \delta_0 + (j-5)\,\delta_1 I_{(j>5)}, \qquad j = 1, 2, \ldots, n_a \\ \delta_0 &\sim N(0,\ 100) \\ \delta_1 &\sim N(0,\ 100)\ . \end{aligned} \tag{5.5}$$

This in not exactly the same specification used in Kunkler (2006). In particular, Kunkler (2006) is vague on the prior he used for δ_0 and δ_1, except to say that it was noninformative. Also note, whereas we coded this part of the model specification directly in BUGS, Kunkler used the *probit_g*() function (LeSage, 1999, Chapter 7)

Table 5.9: Reserve Estimates with Proportional Prior

Year	mean	sd	MC error	2.50%	5.00%	median	95.00%	97.50%
1	0	0	0	0	0	0	0	0
2	152	167	1.10	-64	-48	145	442	530
3	481	373	2.73	-115	-81	458	1114	1316
4	1184	731	5.71	-141	47	1143	2451	2797
5	2671	1388	14.38	205	557	2578	5041	5739
6	3106	1570	11.80	360	754	2981	5830	6606
7	5261	2334	16.08	1135	1796	5097	9307	10440
8	8069	3279	26.69	2128	3108	7862	13680	15170
9	13030	5307	50.15	3044	4794	12760	21990	24290
10	18370	7126	91.56	5791	7783	17820	30710	33970
Total	52320	13070	173.6	29830	33040	51260	75440	80900

in MatLab to implement the Gibbs sampling for this part of his model. We discuss this latter point in further detail below.

Albert and Chib's Specification

The *probit_g*() function described in LeSage (1999) makes use of the Albert and Chib (1993) approach to the estimation of logit and probit models. Essentially, Albert and Chib (1993) proposed augmenting the binary 0 and 1 type of data appearing in logit and probit models with variables drawn from an underlying continuous distribution. Albert and Chib (1993) show that the underlying truncated normal distributions are the theoretically appropriate ones to use. A big advantage of this approach is that the resulting model is sometimes easier to code in the context of a MCMC simulation.

The approach proposed by Albert and Chib (1993) is implemented in the *probit_g*() function developed by LeSage (1999) in the following manner. Suppose $\mathbf{z}$ is a vector of binary observations. Let $\mathbf{y}$ denote the corresponding vector of augmented obser-

Table 5.10: Reserve Estimates with Uniform Prior

Year	mean	sd	MC error	2.50%	5.00%	median	95.00%	97.50%
1	0	0	0	0	0	0	0	0
2	153	171	1.20	-65	-49	144	448	548
3	487	386	2.68	-120	-84	463	1154	1349
4	1196	760	6.85	-147	30	1149	2492	2861
5	2693	1458	14.80	161	516	2581	5234	5944
6	3112	1613	11.54	310	725	2978	5933	6764
7	5275	2400	15.19	1063	1725	5072	9479	10690
8	8111	3399	25.05	2024	3051	7888	13920	15720
9	13230	5604	46.95	2864	4724	12920	22700	25420
10	18810	7444	84.02	6055	7948	18210	31750	35340
Total	53070	13630	161.10	29920	33140	51830	77160	83040

vations, such that $y_i < 0$ if $z_i = 0$ and $y_i >= 0$ if $z_i = 1$. Let $\mathbf{X}$ denote the matrix of covariate values and b the vector of regression parameters. The *probit_g*() function assumes this model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}$$

$$\mathbf{e} \sim N(0, \mathbf{V})$$

$$\mathbf{V} = diag(v_1, v_2, \ldots, v_n)$$

$$\frac{r}{v_i} \sim i.i.d. \frac{\chi^2(r)}{r}$$

$$r \sim gamma(m, k)$$

$$\mathbf{b} \sim N(\mathbf{c}, \mathbf{T}) .$$

Albert and Chib (1993) show that this modelling approach encompasses both the traditional logit and probit model structures. LeSage (1999) notes that the resulting posterior estimates for **b** should be close to those resulting from a traditional probit model when r is large (say, $r = 100$) and diffuse prior is adopted for **b**. LeSage (1999) also notes that setting r around 7 and adopting a diffuse prior for **b** should

Table 5.11: Reserve Estimates with Flat Prior

Year	mean	sd	MC error	2.50%	5.00%	median	95.00%	97.50%
1	0	0	0	0	0	0	0	0
2	156	173	1.12	-66	-49	147	453	553
3	488	386	2.88	-119	-84	461	1148	1352
4	1200	764	6.83	-142	40	1148	2524	2916
5	2687	1448	15.14	146	534	2576	5173	5882
6	3117	1644	12.59	331	750	2963	5987	6858
7	5293	2401	18.01	1099	1758	5100	9451	10550
8	8115	3459	28.15	1958	2962	7866	14040	15650
9	13200	5583	55.65	3084	4775	12850	22690	25280
10	18680	7466	96.51	5770	7813	18030	31700	35350
Total	52930	13730	190.40	29500	32860	51620	77330	83150

produce posterior estimates for **b** close to those from a traditional logit model. We can easily use this same approach towards probit/logit modelling in BUGS.

Implementing in BUGS

The BUGS code for the logit model using Albert and Chib's (1993) approach is as follows.

```
# Unobserved z[i,j]
  for (i in 2:10) {
    for (j in (12-i):10) {
      z[i,j] <- step( z.y[i,j] )
      z.y[i,j] ~ dnorm(p[j], tau0[i,j])
     }
       }

# Observed z[i,j]
```

```
for(i in 1:10) {
  for(j in 1:(11-i)) {
    z.y[i,j] ~ dnorm(p[j], tau0[i,j]) I( z.ylow[i,j], z.yupp[i,j] )
    z.ylow[i,j] <- -10000000 * ( 1 - z[i,j] )
    z.yupp[i,j] <- 10000000 * z[i,j]
   }
     }

for (j in 1:10) {
  p[j] <- delta[1]+ (j-5)*step(j-5.5)* delta[2]
     }

for(i in 1:10) {
  for(j in 1:10) {
    tau0[i,j] <- 1 / ( v0[i,j] )
    v0[i,j] <- r0*r0/c0[i,j]
    c0[i,j] ~ dchisqr(r0)
   }
     }
r0 <- 7
```

We use the same magnitude model as in Section 5.2 and the uniform prior for *sige* as in Subsection 5.4.1 to ease the convergence. We ran the model in BUGS

with $r = 7$ and got the estimates of parameters and reserves very close to those from the previous sections. The simulation converges very quickly. We monitored the convergence of all the parameters and used the simulated values from iterations 11,001 to 21,000 for our posterior analysis. The estimates of the reserves are listed in Tables 5.12.

Table 5.12: Reserve Estimates Using Albert and Chib's Approach

Year	mean	sd	MC error	2.50%	5.00%	median	95.00%	97.50%
1	0	0	0	0	0	0	0	0
2	155	172	1.14	-64	-48	146	447	547
3	488	382	3.03	-116	-80	462	1156	1346
4	1190	749	6.24	-144	35	1142	2481	2852
5	2668	1450	14.26	152	532	2560	5152	5886
6	3116	1622	11.48	324	742	2964	5949	6852
7	5285	2406	16.71	1131	1773	5089	9409	10660
8	8049	3374	24.51	2003	3026	7821	13860	15440
9	13140	5553	53.67	2988	4673	12780	22660	25300
10	18650	7404	86.38	5865	7940	17990	31690	35280
Total	52750	13490	162.70	29640	33050	51630	76480	82680

As we use the same magnitude model as before, the estimates of parameters for the magnitude should not change when a different model specification is used for the mixture data. The estimates of parameters for the binomial mixture model are listed in Table 5.13. From the table we observe that there are notable differences in the estimates of parameters from the different specifications of the logit model. However from the reserve estimates in Table 5.12 we can conclude that the effect on the reserve estimates is very small.

The Influence of r

LeSage (1999) notes that with different values of r (i.e., $r = 2, 25, 50, 100$), the

Table 5.13: Estimates for Mixture Model Using Albert and Chib's Approach

Specification	Parameter	mean	sd	MC error	2.50%	median	97.50%
Albert & Chib	δ_0	3.692	0.850	0.015	2.168	3.640	5.533
	δ_1	-0.307	0.533	0.008	-1.272	-0.333	0.841
BUGS Logit	δ_0	2.189	0.507	0.006	1.299	2.155	3.262
	δ_1	-0.197	0.310	0.004	-0.763	-0.213	0.435

resulting posterior estimates for **b** could be used to approximate those from the traditional logit and probit models. Here we run the model in BUGS with different values of r to study the influence of r on the binomial logit model in Equation (5.5). The same magnitude model as in Section 5.2 and the proportional prior for *sige* as in Subsection 5.4.1 are used.

The converged sequences from 3 different chains are used for our posterior inferences. The estimates of all the parameters converge after iteration 11,001, for all values of r considered. We use the simulated values from iterations 11,001 to 21,000 for our posterior analysis of parameters and reserves.

From Table 5.14, we observe that the posterior estimates of the parameters δ_0 and δ_1 in Equation (5.5) vary for different values of r. The larger the difference in r the larger is the difference in the parameter estimates.

The reserve estimates and standard deviations using the values $r = 2$, 25, 50, 100 are listed in Tables 5.15, 5.16, 5.17 and 5.18 for comparison. We observe that the estimates of reserve tend to decrease as the value of r increases. The other percentiles of the reserves also have the same trend.

Table 5.14: Estimates for Mixture Model with Different Values of r

Value of r	Parameter	mean	sd	MC error	2.50%	median	97.50%
2	δ_0	2.855	0.916	0.034	1.467	2.721	4.894
	δ_1	-0.137	0.598	0.021	-1.072	-0.224	1.297
7	δ_0	3.692	0.850	0.015	2.168	3.640	5.533
	δ_1	-0.307	0.533	0.008	-1.272	-0.333	0.841
25	δ_0	6.387	1.320	0.019	3.890	6.350	9.078
	δ_1	-0.586	0.853	0.012	-2.189	-0.613	1.176
50	δ_0	8.691	1.816	0.025	5.290	8.649	12.370
	δ_1	-0.747	1.184	0.014	-2.983	-0.780	1.694
100	δ_0	11.890	2.479	0.034	7.247	11.810	16.980
	δ_1	-0.952	1.623	0.020	-4.019	-0.998	2.352

Table 5.15: Reserve Estimates Using Albert and Chib's Approach ($r = 2$)

Year	mean	sd	MC error	2.50%	5.00%	median	95.00%	97.50%
1	0	0	0	0	0	0	0	0
2	163	174	1.75	-63	-46	152	456	559
3	511	388	4.11	-113	-77	484	1176	1371
4	1259	743	7.74	-111	118	1212	2546	2936
5	2822	1437	17.97	350	702	2693	5315	6053
6	3230	1613	14.11	487	877	3078	6053	6900
7	5450	2412	18.73	1298	1931	5241	9679	10880
8	8312	3378	28.75	2322	3255	8063	14170	15800
9	13440	5439	55.16	3596	5228	13100	22650	25240
10	19040	7342	103.50	6374	8305	18380	31810	35360
Total	54220	13520	199.90	31070	34370	53020	77970	84210

In this chapter, we used BUGS to obtain results very close to those in Kunkler (2006). The results in Kunkler (2006) were originally obtained using MatLab, along with the Econometrics Toolbox of econometric functions (LeSage, 1999) for use in MatLab. Although MatLab is a very good programming environment, arguably BUGS is better represented in the statistics and actuarial literature, and is probably more accessible to the average actuarial practitioner.

Table 5.16: Reserve Estimates Using Albert and Chib's Approach ($r = 25$)

Year	mean	sd	MC error	2.50%	5.00%	median	95.00%	97.50%
1	0	0	0	0	0	0	0	0
2	153	171	1.12	-68	-51	144	448	551
3	484	388	2.96	-123	-84	459	1150	1354
4	1183	753	6.41	-156	17	1140	2490	2837
5	2651	1436	13.95	140	485	2545	5138	5841
6	3049	1621	11.44	268	672	2914	5858	6713
7	5219	2390	15.35	1040	1701	5016	9357	10530
8	8005	3414	24.62	1866	2905	7763	13910	15470
9	13010	5562	50.00	2875	4527	12690	22390	24930
10	18510	7325	86.18	5772	7817	17850	31390	34950
Total	52260	13440	161.60	29310	32710	51050	75890	82160

In the next chapter, we will propose a multinomial model for a more general situation than that is considered in Kunkler (2006). Specifically, we will develop a multinomial model for the situation when there are both zeros and negatives in the loss triangle. In a subsequent chapter, we will implement this model in BUGS.

Table 5.17: Reserve Estimates Using Albert and Chib's Approach ($r = 50$)

Year	mean	sd	MC error	2.50%	5.00%	median	95.00%	97.50%
1	0	0	0	0	0	0	0	0
2	153	172	1.10	-65	-50	145	451	548
3	485	385	2.77	-118	-81	458	1156	1355
4	1183	754	6.36	-146	22	1142	2498	2842
5	2655	1458	14.18	106	470	2549	5136	5877
6	3072	1629	12.12	317	695	2931	5876	6685
7	5213	2426	15.09	1027	1660	5009	9400	10650
8	7990	3442	24.60	1882	2859	7748	13900	15510
9	12950	5607	46.34	2644	4349	12650	22480	25000
10	18360	7450	90.81	5502	7540	17740	31360	34880
Total	52060	13590	162.10	28880	32080	50830	75940	82170

Table 5.18: Reserve Estimates Using Albert and Chib's Approach ($r = 100$)

Year	mean	sd	MC error	2.50%	5.00%	median	95.00%	97.50%
1	0	0	0	0	0	0	0	0
2	157	172	1.13	-65	-49	147	452	552
3	485	384	2.82	-118	-82	459	1151	1354
4	1177	749	6.31	-141	30	1125	2491	2838
5	2644	1443	14.51	121	508	2534	5143	5873
6	3042	1634	11.67	224	630	2906	5881	6696
7	5190	2436	17.20	992	1621	4989	9369	10640
8	7911	3437	26.48	1751	2742	7715	13770	15400
9	12850	5607	48.68	2544	4240	12530	22350	24950
10	18130	7425	90.03	5182	7273	17540	31020	34520
Total	51580	13550	164.90	28380	31690	50500	75270	81640

Chapter 6

A Bayesian Mixture Model as a Solution

6.1 The Model

The models introduced in Chapter Four 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 together.

Based on the two mixture models introduced by Kunkler (2004, 2006), a Bayesian mixture model will be proposed in this chapter 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. The multinomial distribution will be used to model the mixture data which indicate the sign of losses, while distributions such as the lognormal (Kremer, 1982), over-dispersed Poisson (Renshaw and Verrall, 1998), and negative binomial (Verrall, 2000) can be assumed for the magnitude data for both the positives and negatives. GLM structures can be incorporated in the model as well.

6.2 Model for the Mixture Data

6.2.1 Modelling Mixture Data

As in Chapters Four and Five, 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

$$S^{(-)} = \{y_{ij} : y_{ij} < 0\}$$
$$S^{(0)} = \{y_{ij} : y_{ij} = 0\}$$
$$S^{(+)} = \{y_{ij} : y_{ij} > 0\} .$$

A mixture data triangle $\mathbf{z} = \{z_{ij} : i = 1, 2, \ldots, n_a;\ j = 1, 2, \ldots, n_a - i + 1\}$ can be defined for modelling the sign of the incremental loss triangle where

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

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). Denoting $P(y_{ij} < 0) = \lambda_{1j}$, $P(y_{ij} = 0) = \lambda_{2j}$, then $P(y_{ij} > 0) = 1 - \lambda_{1j} - \lambda_{2j}$. It is easy to verify that z_{ij} has a multinomial distribution, which is

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

where

$$\lambda_j = \begin{pmatrix} \lambda_{1j} \\ \lambda_{2j} \\ 1-\lambda_{1j}-\lambda_{2j} \end{pmatrix} .$$

6.2.2 Distribution of Mixture Data

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

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

Assuming 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

$$p(\mathrm{z}_j|\lambda_j) = \begin{pmatrix} n_{a_j} \\ z_{1j}\ z_{2j}\ z_{3j} \end{pmatrix} \lambda_{1j}^{z_{1j}} \lambda_{2j}^{z_{2j}} (1-\lambda_{1j}-\lambda_{2j})^{n_{a_j}-z_{1j}-z_{2j}} ,$$

where

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

6.2.3 Generalized Linear Models

The probabilities of negatives and zeros often seem to depend on the development year. 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). For demonstration purposes, only the logistic link is used for our analysis.

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. To solve the problem of over parameterization, a simple linear regression on $j-1$ can be used as a special case of the above model. That is,

$$\log\left(\frac{\lambda_{lj}}{\lambda_{3j}}\right) = \delta_{l0} + (j-1)\delta_{l1}, \quad l = 1, 2 \, . \tag{6.1}$$

Denoting $l_l(\lambda_j) = \log\left(\frac{\lambda_{lj}}{\lambda_{3j}}\right)$ (where $l = 1, 2$), the model can be expressed in matrix form by

$$l(\Lambda) = \mathbf{X}_{\Delta}^{\cdot} \Delta \, ,$$

where

$$\Lambda = \begin{pmatrix} \lambda_1 & \dots & \lambda_{n_d} \end{pmatrix}$$

$$l(\Lambda) = \begin{pmatrix} l_2(\lambda_1) & l_3(\lambda_1) \\ l_2(\lambda_2) & l_3(\lambda_2) \\ \vdots & \vdots \\ l_2(\lambda_{n_d}) & l_3(\lambda_{n_d}) \end{pmatrix}$$

$$\mathbf{X}_\Delta = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & n_d - 1 \end{pmatrix}$$

$$\Delta = \begin{pmatrix} \delta_{10} & \delta_{20} \\ \delta_{11} & \delta_{21} \end{pmatrix} .$$

6.3 Model for the Magnitude Data

6.3.1 Modelling Magnitude Data

To ease the analyses for this section, simplified notations for the distributions of the magnitudes of the positive and negative data are introduced as

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

$$p(\mathbf{y}^+ | \, \theta_3) = p(y_{ij} | z_{ij} = 1, \theta_3) \, .$$

As discussed in Chapter Two, many distributions such as the lognormal (Kremer, 1982; de Alba, 2002a, 2006; Kunkler, 2004, 2006), over-dispersed Poisson (Renshaw and Verrall, 1998), and negative binomial (Verrall, 2000) 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,

$$\log(\mathbf{y}^-)|\theta_1 \sim N(\mathbf{X}_{\beta_1}\beta_1,\ \sigma_1^2\mathbf{I}_1)$$
$$\log(\mathbf{y}^+)|\theta_3 \sim N(\mathbf{X}_{\beta_2}\beta_2,\ \sigma_2^2\mathbf{I}_2)\ ,$$

where

$\theta_1 = (\beta_1, \sigma_1^2)$

$\theta_3 = (\beta_2, \sigma_2^2)$

$\mathbf{I}_1$ is an identity matrix of $n_{\mathrm{y}-} \times n_{\mathrm{y}-}$

$\mathbf{I}_2$ is an identity matrix of $n_{\mathrm{y}+} \times n_{\mathrm{y}+}$

$$n_{\mathrm{y}-} = \sum_{i=1}^{n_a} \sum_{j=1}^{n_a-i+1} \mathbf{I}_{(y_{ij}<0)}$$

$$n_{\mathrm{y}+} = \sum_{i=1}^{n_a} \sum_{j=1}^{n_a-i+1} \mathbf{I}_{(y_{ij}>0)}$$

β_1 is a parameter vector of $k_{\beta_1} \times 1$

β_2 is a parameter vector of $k_{\beta_2} \times 1$

$\mathbf{X}_{\beta_1}$ is a design matrix of $n_{\mathrm{y}-} \times k_{\beta_1}$

$\mathbf{X}_{\beta_2}$ is a design matrix of $n_{\mathrm{y}+} \times k_{\beta_2}$.

6.3.2 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. Observe that the observed data in the loss triangle provide no information concerning the η_{lt} parameters for $t \geq n$. 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

$$(\mathbf{X}_{\beta_l}\beta_l)_{ij} = \alpha_{li} + \sum_{d=1}^{j-1}\gamma_{ld}, \quad l = 1, 2\,, \tag{6.2}$$

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 to overcome the problem of over parameterization, 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 (Zehnwirth, 1985; Renshaw 1994a; Wright, 1990) can be introduced to avoid the problem of over parameterization. One of the choices may be the Hoerl curve (Zehnwirth, 1985) given by

$$(\mathbf{X}_{\beta_l}\beta_l)_{ij} = c_l + a_{li} + b_{li}\log(j) + r_{li}\, j \ , \tag{6.3}$$

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

A special case of the model in Equation (6.3) 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

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

6.4 Bayesian Inference

6.4.1 Model Basics

In the framework of Bayesian inference, the claim reserve will be estimated based on the posterior predictive distributions of the magnitude and mixture data for future incremental losses. In the analysis of this section, the future incremental data triangle

and future mixture data triangle will be denoted as

$$\tilde{\mathbf{y}} = \{\tilde{y}_{ij} : i = 2, \ldots, n_a;\ j = n_a - i + 1, \ldots, n_d\}$$
$$\tilde{\mathbf{z}} = \{\tilde{z}_{ij} : i = 2, \ldots, n_a;\ j = n_a - i + 1, \ldots, n_d\} .$$

Bayesian inference based on the style of analysis given in Chapter Three can be performed here. Assuming that the parameters are also random variables, the joint probability distribution of $\mathbf{y}$, $\mathbf{z}$, θ and Λ can be written as the product of the joint prior distribution $\pi(\theta, \Lambda)$ and the joint sampling distribution, i.e.

$$p(\mathbf{y}, \mathbf{z}, \theta, \Lambda) = \pi(\theta, \Lambda)\, p(\mathbf{y}, \mathbf{z}|\theta, \Lambda) .$$

Applying Bayes' Theorem, the joint posterior distribution for θ and Λ can be written as

$$\pi(\theta, \Lambda|\mathbf{y}, \mathbf{z}) = \frac{\pi(\theta, \Lambda) p(\mathbf{y}, \mathbf{z}|\theta, \Lambda)}{p(\mathbf{y}, \mathbf{z})} ,$$

where $p(\mathbf{y}, \mathbf{z}) = \sum_{\theta,\Lambda} \pi(\theta, \Lambda) p(\mathbf{y}, \mathbf{z}|\theta, \Lambda)$ or $p(\mathbf{y}, \mathbf{z}) = \iint \pi(\theta, \Lambda) p(\mathbf{y}, \mathbf{z}|\theta, \Lambda)\, \mathrm{d}\theta\, \mathrm{d}\Lambda$. From the above formula, we can obtain the unnormalized joint posterior distribution by

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

The joint sampling distribution for $\mathbf{y}$ and $\mathbf{z}$ in the above formulas can be obtained with the multinomial sampling mixture distribution of $p(z_{ij}|\lambda_j)$ and the conditional sampling distribution $p(y_{ij}|z_{ij}, \theta)$ with

$$p(y_{ij}, z_{ij}|\theta, \lambda_j) = p(z_{ij}|\lambda_j)\, p(y_{ij}|z_{ij}, \theta) .$$

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

$$\begin{aligned}
p(y_{ij}|\theta,\lambda_j) =& P(z_{ij}=-1|\lambda_j)\,p(y_{ij}|z_{ij}=-1,\theta_1) + P(z_{ij}=0|\lambda_j)\,p(y_{ij}|z_{ij}=0,\theta_2) \\
&+ P(z_{ij}=1|\lambda_j)\,p(y_{ij}|z_{ij}=1,\theta_3) \\
=& \lambda_{1j}\,p(y_{ij}|z_{ij}=-1,\theta_1) + \lambda_{2j}\,p(y_{ij}|z_{ij}=0,\theta_2) \\
&+ (1-\lambda_{1j}-\lambda_{2j})\,p(y_{ij}|z_{ij}=1,\theta_3) \\
=& \lambda_{1j}\,p(y_{ij}|z_{ij}=-1,\theta_1) + \lambda_{2j}\,I_{(y_{ij}=0)} + (1-\lambda_{1j}-\lambda_{2j})\,p(y_{ij}|z_{ij}=1,\theta_3)\,.
\end{aligned}$$

In the above formula, we can verify that

$$\begin{aligned}
p(y_{ij}|z_{ij}=0,\theta_2) &= \begin{cases} 0 & \text{if } y_{ij}>0 \text{ or } y_{ij}<0 \\ 1 & \text{if } y_{ij}=0 \end{cases} \\
&= I_{(y_{ij}=0)}\,.
\end{aligned}$$

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 by (similar to Kunkler, 2004, 2006)

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

where $\lambda_{1j}=p(\tilde{z}_{ij}=-1)$, $\lambda_{2j}=p(\tilde{z}_{ij}=0)$ and $1-\lambda_{1j}-\lambda_{2j}=p(\tilde{z}_{ij}=1)$.

6.4.2 Posterior Analysis for Mixture Parameters

Given the prior mixture distribution $\pi(\lambda_j)$ and the multinomial sampling mixture distribution $p(\mathbf{z}_j|\lambda_j)$, the posterior mixture distribution for Λ can be written as

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

Under the GLM structure given in Subsection 6.2.3, the posterior mixture distribution can also be written in terms of the GLM parameters $\Delta = \begin{pmatrix} \delta_{10} & \delta_{20} \\ \delta_{11} & \delta_{21} \end{pmatrix}$. We have

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

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

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

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

$$\lambda_1(\Delta) = \frac{\exp[\delta_{10} + (j-1)\delta_{11}]}{\exp[\delta_{10} + (j-1)\delta_{11}] + \exp[\delta_{20} + (j-1)\delta_{21}] + 1}$$

$$\lambda_2(\Delta) = \frac{\exp[\delta_{20} + (j-1)\delta_{21}]}{\exp[\delta_{10} + (j-1)\delta_{11}] + \exp[\delta_{20} + (j-1)\delta_{21}] + 1} \; .$$

6.4.3 Conjugate Analysis for Mixture Parameters

A conjugate prior distribution for the multinomial is Dirichlet. So the analysis becomes easier if λ_j is assumed to follow a Dirichlet distribution with its probability distribution function given by

$$\pi(\lambda_j) = \frac{\Gamma(\alpha_{1j} + \alpha_{2j} + \alpha_{3j})}{\Gamma(\alpha_{1j}) + \Gamma(\alpha_{2j}) + \Gamma(\alpha_{3j})} \lambda_{1j}^{\alpha_{1j}-1} \lambda_{2j}^{\alpha_{2j}-1} \lambda_{3j}^{\alpha_{3j}-1} \; .$$

The posterior distribution for Λ can be obtained by

$$\begin{aligned}
\pi(\Lambda|\mathbf{z}) &\propto \prod_{j=1}^{n_d} \pi(\lambda_j)\, p(\mathbf{z}_j|\lambda_j) \\
&= \prod_{j=1}^{n_d} \lambda_{1j}^{\alpha_{1j}-1} \lambda_{2j}^{\alpha_{2j}-1} \lambda_{3j}^{\alpha_{3j}-1} \lambda_{1j}^{z_{1j}} \lambda_{2j}^{z_{2j}} \lambda_{3j}^{n_{a_j}-z_{1j}-z_{2j}} \\
&= \prod_{j=1}^{n_d} \lambda_{1j}^{\alpha_{1j}+z_{1j}-1} \lambda_{2j}^{\alpha_{2j}+z_{2j}-1} \lambda_{3j}^{\alpha_{3j}+n_{a_j}-z_{1j}-z_{2j}-1} \\
&= \prod_{j=1}^{n_d} \lambda_{1j}^{\alpha_{1j}^*-1} \lambda_{2j}^{\alpha_{2j}^*-1} \lambda_{3j}^{\alpha_{3j}^*-1} ,
\end{aligned}$$

which is a new Dirichlet distribution with parameters $\alpha_{1j}^* = \alpha_{1j} + z_{1j}$, $\alpha_{2j}^* = \alpha_{2j} + z_{2j}$ and $\alpha_{3j}^* = \alpha_{3j} + n_{a_j} - z_{1j} - z_{2j}$. The details of choosing an appropriate prior can be found in Gelman et al. (2005) and its references.

6.4.4 Posterior Analysis for Sampling Distribution

Under the lognormal GLM model formulated in Subsection 6.3.2, the magnitude data for both the positives and negatives follow lognormal distributions:

$$\begin{aligned}
\log(\mathbf{y}^-)|\theta_1 &\sim N(\mathbf{X}_{\beta_1}\beta_1,\ \sigma_1^2\mathbf{I}_1) \\
\log(\mathbf{y}^+)|\theta_3 &\sim N(\mathbf{X}_{\beta_2}\beta_2,\ \sigma_2^2\mathbf{I}_2) .
\end{aligned}$$

A noninformative uniform prior on $(\beta_l,\, \log\sigma_l)$ can be assumed in order to make the normal regression model easier to analyze. As in Gelman et al. (2005, Chapter 14, Pages 356-357), this leads to

$$\pi(\beta_l, \sigma_l^2) \propto \sigma_l^{-2} , \qquad l = 1,\, 2 .$$

The joint posterior distribution for β_l and σ_l $(l = 1, 2)$ can factored into

$$\pi(\beta_1, \sigma_1^2|\mathrm{y}^-) = \pi(\beta_1|\sigma_1^2, \mathrm{y}^-)\,\pi(\sigma_1^2|\mathrm{y}^-)$$
$$\pi(\beta_2, \sigma_2^2|\mathrm{y}^+) = \pi(\beta_2|\sigma_2^2, \mathrm{y}^+)\,\pi(\sigma_2^2|\mathrm{y}^+)\,,$$

with which the conditional posterior distributions for β_1 and β_2 can be obtained (Gelman et al., 2005, Page 356) as

$$\beta_1|\sigma_1^2, \mathrm{y}^- \sim N\left(\hat{\beta}_1, \left(\mathbf{X}_{\beta_1}^T\mathbf{X}_{\beta_1}\right)^{-1}\sigma_1^2\right)$$
$$\beta_2|\sigma_2^2, \mathrm{y}^+ \sim N\left(\hat{\beta}_2, \left(\mathbf{X}_{\beta_2}^T\mathbf{X}_{\beta_2}\right)^{-1}\sigma_2^2\right)\,.$$

For σ_1 and σ_2, the marginal posterior distributions can be obtained by

$$\pi(\sigma_1^2|\mathrm{y}^-) = \frac{\pi(\beta_1,\, \sigma_1^2|\,\mathrm{y}^-)}{\pi(\beta_1|\,\sigma_1^2,\, \mathrm{y}^-)}$$
$$\pi(\sigma_2^2|\mathrm{y}^+) = \frac{\pi(\beta_2,\, \sigma_2^2|\,\mathrm{y}^+)}{\pi(\beta_2|\,\sigma_2^2,\, \mathrm{y}^+)}\,.$$

These can be shown to follow scaled inverse-χ^2 distributions in the forms of

$$\sigma_1^2|\mathrm{y}^- \sim \text{Inv-}\chi^2(n_{\mathrm{y}^-} - k_{\beta_1},\, \hat{\sigma}_1^2)$$
$$\sigma_2^2|\mathrm{y}^+ \sim \text{Inv-}\chi^2(n_{\mathrm{y}^+} - k_{\beta_2},\, \hat{\sigma}_2^2)\,,$$

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

$$\hat{\beta}_1 = \left(\mathbf{X}_{\beta_1}^T\mathbf{X}_{\beta_1}\right)^{-1}\mathbf{X}_{\beta_1}^T\log(\mathrm{y}^-)$$
$$\hat{\beta}_2 = \left(\mathbf{X}_{\beta_2}^T\mathbf{X}_{\beta_2}\right)^{-1}\mathbf{X}_{\beta_2}^T\log(\mathrm{y}^+)$$
$$\hat{\sigma}_1^2 = \frac{1}{n_{\mathrm{y}^-} - k_{\beta_1}}\left(\log(\mathrm{y}^-) - \mathbf{X}_{\beta_1}\hat{\beta}_1\right)^T\left(\log(\mathrm{y}^-) - \mathbf{X}_{\beta_1}\hat{\beta}_1\right)$$
$$\hat{\sigma}_2^2 = \frac{1}{n_{\mathrm{y}^+} - k_{\beta_2}}\left(\log(\mathrm{y}^+) - \mathbf{X}_{\beta_2}\hat{\beta}_2\right)^T\left(\log(\mathrm{y}^+) - \mathbf{X}_{\beta_2}\hat{\beta}_2\right)\,.$$

With the above analytical results for the posterior distributions and posterior predictive distributions, we will be able to implement our multinomial mixture model in computer languages such as C and FORTRAN. MCMC simulation methods such as the Gibbs sampler discussed in Chapter Three can be used for our posterior simulation. Detailed steps of the posterior sampling algorithm for our model will be similar to those given in Kunkler (2004, pages 29-30).

Chapter 7

Model Implementation

The use of the specialized Bayesian software BUGS makes it easier for implementing our complicated multinomial mixture model. As illustrated in Chapter Three, Bayesian models including Bayesian GLMs can be implemented in BUGS simply by specifying the sampling distributions, prior distributions and the regression functions. Hence, the posterior analysis given in Chapter Six for our multinomial mixture model will not be needed for our model fitting in BUGS.

In this chapter, 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).

7.1 The Data

7.1.1 Loss Triangle

For illustrative purposes, the original loss triangle from the 'Historical Loss Development Study' (1991) by the Reinsurance Association of America listed in Table 2.1 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 7.1: Adjusted Incremental Loss Triangle with Zeros and Negatives

	Development year									
Accident 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									

7.1.2 Mixture Data

Based on the above data of the adjusted incremental loss triangle, we can get the loss triangles of the mixture data $\mathbf{z}$ as listed in Table 7.2.

From the mixture data triangle, it is not difficult to get the values of $\mathbf{z}_{1j}$, $\mathbf{z}_{2j}$ and $\mathbf{z}_{3j}$ for our simulation purpose of the mixture data. These values are listed in Table 7.3 below.

Table 7.2: Mixture Data Triangle for Zeros and Negatives

	Development year									
Accident year	1	2	3	4	5	6	7	8	9	10
1	1	1	1	-1	1	1	1	1	-1	1
2	-1	1	-1	1	1	1	-1	0	1	
3	1	1	1	1	1	1	0	1		
4	1	1	1	1	1	1	1			
5	1	1	1	1	1	-1				
6	1	1	1	1	1					
7	-1	1	1	1						
8	1	1	1							
9	1	1								
10	1									

Table 7.3: Multinomial Data for Modelling Zeros and Negatives

Development year (j)	1	2	3	4	5	6	7	8	9	10
$\mathbf{z}_{1j}$	2	0	1	1	0	1	1	0	1	0
$\mathbf{z}_{2j}$	0	0	0	0	0	0	1	1	0	0
$\mathbf{z}_{3j}$	8	9	7	6	6	4	2	2	1	1

7.1.3 Magnitude Data

The observations of the positive and negative magnitudes $\mathbf{y}^+$ and $\mathbf{y}^-$ are listed in Table 7.4 and Table 7.5.

7.2 Model Construction

7.2.1 Modelling Mixture Data

The Model

In Chapter Six, we proposed using the multinomial distribution to model the mixture data, i.e.

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

Table 7.4: Loss Magnitude Triangle for Positive Losses

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

Table 7.5: Negative Losses with Accident and Development Years

y^-	898	54	106	1111	103	225	557
i	1	1	2	2	2	5	7
j	4	9	1	3	7	6	1

where

$$\lambda_j = \begin{pmatrix} \lambda_{1j} \\ \lambda_{2j} \\ 1 - \lambda_{1j} - \lambda_{2j} \end{pmatrix}.$$

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 Equation (5.5). The model for the negative probability is given by

$$\text{logit}(\lambda_{1j}) = \delta_{10} + (j-5)\,\delta_{11} I_{(j>5)}, \qquad j = 1, 2, \ldots, n_a\,. \tag{7.1}$$

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:

$$\text{logit}(\lambda_{2j}) = \delta_{20} + (j-6)\,\delta_{21} I_{(j>6)}, \qquad j = 1, 2, \ldots, n_a\,. \tag{7.2}$$

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

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

The BUGS code for this part of our model is given below.

```
# Model for z


delta[3, 1] <- 0

delta[3, 2] <- 0


for (j in 1:10) {
   for (i in 1:10) {
     z[j, i, 1:3] ~ dmulti(p[j, 1:3], 1)
        }
          }


for (j in 1:10)  {
   for (i in 1:3) {
     p[j, i] <- phi[j, i] / sum(phi[j, 1:3])
    }
   log(phi[j, 1]) <- delta[1, 1]+(j-5)*step(j-6)* delta[1,2]
   log(phi[j, 2]) <- delta[2, 1]+(j-6)*step(j-7)* delta[2,2]
```

```
    log(phi[j, 3]) <- delta[3, 1]+(j-5)*step(j-6)* delta[3,2]
          }

for (i in 1:2)  {
   for (j in 1:2) {
     delta[i, j] ~ dnorm(0, 0.01)
    }
      }
```

7.2.2 Modelling Magnitude Data

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

$$|y_{ij}| \sim LN\left(\mu_{ij}^{-}, \frac{\sigma_{ij}^2}{\omega^{-}}\right) \quad \text{if } y_{ij} < 0 \tag{7.3}$$

$$|y_{ij}| \sim LN\left(\mu_{ij}^{+}, \frac{\sigma_{ij}^2}{\omega^{+}}\right) \quad \text{if } y_{ij} > 0\,. \tag{7.4}$$

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 (6.2) with the same calendar trend factor ι from Kunkler (2006). The model structure is given by

$$\mu_{ij}^{+} = \alpha_i^{+} + \sum_{d=1}^{j-1} \gamma_d^{+} + (i+j-2)\iota\,. \tag{7.5}$$

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 (ii_1 and jj_1 for the positive magnitudes in the BUGS code) are used to store the values of i and j for the magnitude data. The model is coded in BUGS using the following lines of code.

```
# Model for y+

  for (i in 1:n1[1]) {
    y1[i] ~ dlnorm(mu1[i], tao1[ii1[i], jj1[i]])
    mu1[i] <- alphap[ii1[i]]+(ii1[i]+jj1[i]-2)*iota
    }

  for (i in n1[1]+1: n1[2]) {
    y1[i] ~ dlnorm(mu1[i], tao1[ii1[i], jj1[i]])
    mu1[i] <- alphap[ii1[i]]+sum(gammap[1:jj1[i]-1])
              +(ii1[i]+jj1[i]-2)*iota
    }

  for (i in 1: max1-1) {
    gammap[i] ~ dnorm(0, 1.0E-3)
    alphap[i] ~ dnorm(0, 1.0E-3)
    }
  alphap[max1] ~ dnorm(0, 1.0E-3)
```

Negative Magnitude

For the magnitude data of the negatives, a simplified model structure needs to be used to solve the problem of over parameterization. Since we assume the same negative losses as in Kunkler (2006), the model in Kunkler (2006) given in Section 5.2 can be used to model the mean of negative magnitude. The model structure for the negative magnitude is given by

$$\mu_{ij}^{-} = \alpha_1^{-} + \left[I_{(j\leq 3)}(j-1) + I_{(j>3)}2\right]\gamma_1^{-} + I_{(j>3)}(j-3)\gamma_2^{-} + (i+j-2)\iota\,. \qquad (7.6)$$

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. The BUGS code for the negative magnitude model is given by the following lines.

```
# Model for y-

  for (i in 1:n2) {
    y2[i] ~ dlnorm(mu2[ii2[i], jj2[i]], tao2[ii2[i], jj2[i]])
    mu2[ii2[i], jj2[i]]<-alphan+step(3-jj2[i])*(jj2[i]-1)*gamman[1]
                          +step(jj2[i]-4)*(2*gamman[1]+(jj2[i]-3)*
                           gamman[2])+(ii2[i]+jj2[i]-2)*iota
    }

  alphan ~ dnorm(0, 1.0E-3)
  for (i in 1: 2) {
```

```
    gamman[i] ~ dnorm(0, 1.0E-3)
    }
  iota ~ dnorm(0, 1.0E-3)
```

Modelling Variance

For the variance parameters σ_{ij}^2 in Equations (7.3) and (7.4), we will use a model specification similar to that in the *ols_g*() function in LeSage (1999, page 176). Also, see Chapter Five of this thesis. In our simulation, the prior density specification of σ_{ij}^2 is defined in this way:

$$\begin{aligned} \sigma_{ij}^2 &= sige \times v_{ij} \\ \frac{r}{v_{ij}} &\sim i.i.d. \ \frac{\chi^2(r)}{r} \\ sige &\sim U(0,\, 100) \\ r &= 100\,. \end{aligned} \tag{7.7}$$

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 $U(0,\, L)$ as discussed in Chapter Five of this thesis, with $L = 100$.

We estimate the parameters ω^+ and ω^- in Equations (7.3) and (7.4) 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

(7.7). The estimates are given by

$$\omega^{+} = \frac{1}{\text{var}(\mathbf{r}^{+})}$$
$$\omega^{-} = \frac{1}{\text{var}(\mathbf{r}^{-})},$$

where $\mathbf{r}^{+}$ and $\mathbf{r}^{-}$ are the residual vectors for the positive and negative log magnitudes. That is,

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

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 (7.5) and (7.6) using the posterior estimates of the α, γ and ι parameters.

The residual values $\mathbf{r}^{+}$ and $\mathbf{r}^{-}$ are calculated and listed in Tables 7.6 and 7.7.

Table 7.6: Posterior Estimates of Positive Residuals $\mathbf{r}^{+}$

Accident	Development year									
year	1	2	3	4	5	6	7	8	9	10
1	-0.69	0.33	0.62		0.33	-0.13	-0.41	-0.05		0.00
2		0.22		-0.49	-0.13	0.37			-0.01	
3	-0.20	-0.09	0.11	0.33	0.03	-0.30		0.05		
4	-0.63	-0.07	0.34	-0.48	0.29	0.05	0.39			
5	1.08	-0.37	0.01	-0.55	-0.20					
6	0.29	-0.30	-0.29	0.61	-0.41					
7		0.03	-0.59	0.48						
8	0.53	-0.30	-0.32							
9	-0.49	0.43								
10	-0.04									

Table 7.7: Posterior Estimates of Negative Residuals $\mathbf{r}^-$

$\mathbf{y}^-$	0.468	0.387	-0.410	-0.025	-0.219	-0.464	0.450
i	1	1	2	2	2	5	7
j	4	9	1	3	7	6	1

For the variance calculation, we use the n normalized sample variance. That is, for the data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ 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 in Tables 7.6 and 7.7, we obtained the estimated values of ω^+ and ω^- as

$$\hat{\omega}^+ = 6.3880$$

$$\hat{\omega}^- = 5.1547 .$$

We will use these estimated values for our model implementation in BUGS. The BUGS code for the positive and negative variances is as follows.

```
# Model for inverse variance
  for (i in 1:10) {
    for (j in 1:10) {
      tao1[i,j]<-6.3880*tau[i,j]
      tao2[i,j]<-5.1547*tau[i,j]
      tau[i,j] <- 1/(sige*v[i,j])
```

```
    v[i,j] <- r*r/c[i,j]
    c[i,j] ~ dchisqr(r)
   }
    }
r<- 100
L<-100
sige ~ dunif(0, L)
```

7.3 Estimation and Prediction

7.3.1 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 history plots of the parameter $tau[1, 1]$ are given in Figure 7.1 as an example. The history plots of all the other parameters or quantities of interest behave similarly.

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

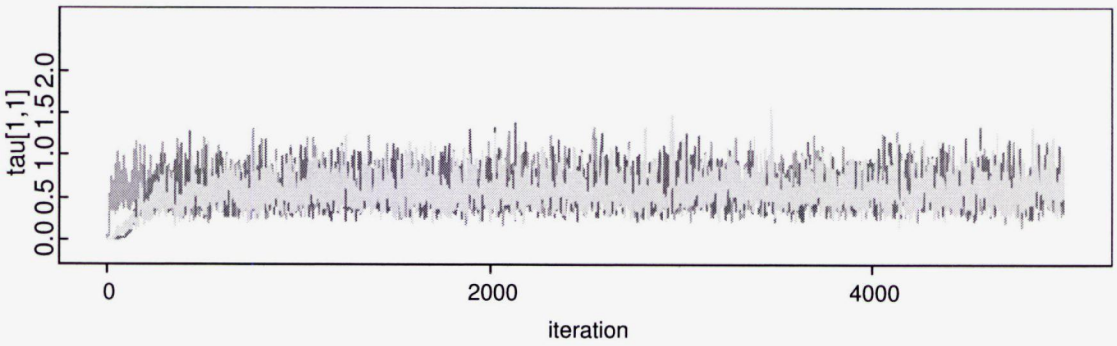

Figure 7.1: History Plot of the Precision Parameter $tau[1,1]$, Multinomial Model

after 2,000 iterations. The values of the refined potential scale reduction factors for the precision parameter $tau[1,1]$ are listed in Table 7.8 as an example. The refined potential scale reduction factors for all the other quantities of interest are very similar.

Table 7.8: Refined Potential Scale Reduction of $tau[1,1]$, Multinomial Model

Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$	Iteration	$\hat{R}_c$
550	1.016	1050	1.007	1550	1.008	2050	1.003
600	1.033	1100	1.002	1600	1.004	2100	1.003
650	1.001	1150	0.995	1650	1.001	2150	1.003
700	0.986	1200	0.996	1700	1.009	2200	1.001
750	0.995	1250	0.998	1750	1.006	2250	1.003
800	0.985	1300	1.001	1800	1.003	2300	1.003
850	1.000	1350	1.002	1850	1.004	2350	1.004
900	0.997	1400	1.006	1900	1.003	2400	1.002
950	1.006	1450	0.999	1950	1.001	2450	1.002
1000	1.010	1500	1.006	2000	1.003	2500	1.001

We will use the simulated values from iterations 2,001 to 12,000 from all three chains for our posterior analysis in the following subsection. A total of 30000 posterior samples will be used for our posterior analysis of parameters and reserves.

7.3.2 Mixture Model

For the multinomial mixture model in Subsection 7.2.1, the parameters for the logit models in Equations (7.1) and (7.2) are estimated by posterior simulation. The estimates of the parameters, their standard deviations and percentiles are listed in Table 7.9.

Table 7.9: Estimates for Multinomial Mixture Model

Model	Parameter	mean	sd	MC error	2.50%	median	97.50%
Multinomial	δ_{10}	-2.181	0.496	0.012	-3.200	-2.157	-1.280
	δ_{11}	0.268	0.323	0.008	-0.397	0.281	0.865
	δ_{20}	-4.126	1.106	0.029	-6.658	-3.990	-2.380
	δ_{21}	0.832	0.602	0.015	-0.389	0.846	1.996
Binomial	δ_0	2.189	0.507	0.006	1.299	2.155	3.262
	δ_1	-0.197	0.310	0.004	-0.763	-0.213	0.435

From the above table 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 7.10, 7.11 and 7.12 respectively for negatives, zeros and positives. 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 for the posterior mean of the positive probability, 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 7.10: Posterior Mean for Probability of Negatives, Multinomial Model

Accident	Development year									
year	1	2	3	4	5	6	7	8	9	10
1										
2										0.250
3									0.237	0.244
4								0.208	0.237	0.248
5							0.167	0.208	0.237	0.245
6						0.133	0.167	0.208	0.236	0.248
7					0.105	0.132	0.171	0.209	0.236	0.243
8				0.104	0.107	0.133	0.166	0.210	0.239	0.248
9			0.105	0.109	0.110	0.134	0.170	0.214	0.244	0.251
10		0.109	0.109	0.110	0.110	0.133	0.169	0.211	0.238	0.249

Table 7.11: Posterior Mean for Probability of Zeros, Multinomial Model

Accident	Development year									
year	1	2	3	4	5	6	7	8	9	10
1										
2										0.303
3									0.178	0.308
4								0.085	0.180	0.305
5							0.040	0.084	0.180	0.310
6						0.022	0.039	0.084	0.181	0.308
7					0.022	0.022	0.037	0.086	0.179	0.304
8				0.021	0.020	0.022	0.038	0.083	0.178	0.307
9			0.023	0.022	0.023	0.021	0.038	0.087	0.173	0.306
10		0.022	0.023	0.022	0.022	0.021	0.040	0.084	0.177	0.304

Table 7.12: Posterior Mean for Probability of Positives, Multinomial Model

Accident	Development year									
year	1	2	3	4	5	6	7	8	9	10
1										
2										0.447
3									0.585	0.448
4								0.707	0.583	0.447
5							0.792	0.708	0.584	0.445
6						0.844	0.794	0.707	0.583	0.444
7					0.873	0.846	0.792	0.705	0.584	0.452
8				0.875	0.873	0.845	0.796	0.706	0.584	0.445
9			0.872	0.869	0.868	0.845	0.793	0.699	0.583	0.443
10		0.869	0.868	0.869	0.868	0.847	0.791	0.705	0.585	0.447

7.3.3 Magnitude Model

Positive Magnitude

For the lognormal positive magnitude model specified in Equations (7.4) and (7.5), 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 7.13.

Negative Magnitude

Similar to the previous section, the posterior predictive means for the α^- and γ_i^- parameters in Equation (7.2) and (7.6) can be estimated using the simulated samples from BUGS. The estimates of the parameters, their standard deviations and percentiles are listed in Table 7.14. For comparison purposes, the estimates of the same parameters for the negative magnitude model of Kunkler (2006) are also listed. From Table 7.14 we observe that the estimates for the parameters are very close, since we use the same model and same negative data for our simulation.

Table 7.13: Parameter Estimation of Positive Magnitude, Multinomial Model

Parameter	mean	sd	MC error	2.50%	median	97.50%
α_1^+	7.834	0.298	0.005	7.242	7.837	8.427
α_2^+	7.816	0.394	0.009	7.035	7.816	8.601
α_3^+	7.647	0.396	0.014	6.875	7.643	8.449
α_4^+	7.574	0.494	0.021	6.613	7.568	8.568
α_5^+	7.488	0.616	0.027	6.301	7.484	8.728
α_6^+	6.868	0.732	0.034	5.452	6.858	8.370
α_7^+	6.694	0.888	0.040	4.986	6.676	8.522
α_8^+	6.705	0.989	0.047	4.792	6.694	8.742
α_9^+	6.371	1.130	0.053	4.192	6.351	8.683
α_{10}^+	6.256	1.310	0.060	3.719	6.246	8.928
γ_1^+	0.437	0.311	0.009	-0.172	0.435	1.052
γ_2^+	-0.079	0.303	0.008	-0.675	-0.081	0.521
γ_3^+	-0.692	0.336	0.009	-1.357	-0.687	-0.038
γ_4^+	-0.325	0.347	0.008	-1.004	-0.327	0.366
γ_5^+	-0.196	0.382	0.008	-0.945	-0.198	0.566
γ_6^+	-0.792	0.507	0.010	-1.789	-0.798	0.211
γ_7^+	-0.919	0.586	0.010	-2.062	-0.920	0.246
γ_8^+	-0.338	0.759	0.012	-1.835	-0.336	1.141
γ_9^+	-1.159	0.864	0.012	-2.861	-1.157	0.548

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 variance. The posterior mean, standard deviation and percentiles of the calendar parameters are listed in Table 7.15. The posterior means of the precision parameters are listed in Table 7.16. The posterior means of the precision parameters for different accident years and development years are very close to each other, which is consistent with the independently identical distribution assumption.

Table 7.14: Parameter Estimation of Negative Magnitude, Multinomial Model

Model	Parameter	mean	sd	MC error	2.50%	median	97.50%
Multinomial	α^-	4.958	0.622	0.024	3.758	4.957	6.205
	γ_1^-	0.819	0.309	0.005	0.220	0.816	1.432
	γ_2^-	-0.694	0.180	0.007	-1.051	-0.694	-0.336
Kunkler2006	α^-	5.255	0.387	0.014	4.491	5.258	6.021
	γ_1^-	0.805	0.241	0.007	0.338	0.807	1.293
	γ_2^-	-0.609	0.108	0.003	-0.818	-0.609	-0.395

Table 7.15: Parameter Estimation of Calendar Trend Factor, Multinomial Model

Parameter	mean	sd	MC error	2.50%	median	97.50%
ι	0.153	0.132	0.007	-0.128	0.153	0.408

7.3.4 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 7.17.

The posterior means, standard deviations and percentiles for the total reserve and reserves by accident year are listed in Table 7.18. From the table, we observe that the reserves for our multinomial mixture model are smaller than those for the binomial model in Table 5.10 using the same prior distributions for the parameters. 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.

Table 7.16: Posterior Mean for Precision Parameters, Multinomial Model

Acc	Development year									
year	1	2	3	4	5	6	7	8	9	10
1	0.575	0.582	0.580	0.584	0.582	5.802	0.579	0.581	0.581	0.581
2	0.581	0.581	0.580	0.581	0.581	0.582	0.580	0.581	0.581	0.580
3	0.579	0.583	0.583	0.579	0.583	0.580	0.584	0.581	0.580	0.581
4	0.584	0.584	0.586	0.582	0.585	0.582	0.582	0.583	0.581	0.580
5	0.577	0.585	0.583	0.580	0.582	0.585	0.580	0.580	0.581	0.581
6	0.558	0.581	0.585	0.579	0.585	0.579	0.580	0.580	0.581	0.580
7	0.583	0.582	0.582	0.575	0.580	0.580	0.580	0.581	0.580	0.581
8	0.578	0.584	0.576	0.579	0.581	0.581	0.581	0.581	0.580	0.580
9	0.576	0.582	0.582	0.581	0.582	0.580	0.580	0.581	0.581	0.581
10	0.581	0.581	0.580	0.581	0.581	0.582	0.580	0.581	0.581	0.581

Table 7.17: Mean Reserve by Accident & Development Years, Multinomial Model

Accident	Development year									
year	1	2	3	4	5	6	7	8	9	10
1										
2										116
3									424	111
4								604	460	118
5							1603	647	504	127
6						1895	978	384	296	69
7					1982	1911	957	374	287	65
8				2688	2322	2225	1130	432	341	72
9			3676	2136	1903	1866	918	331	266	48
10		4226	3922	2351	2131	2087	1018	361	299	46

Table 7.18: Mean, STD and Percentiles of Reserve Estimates, Multinomial Model

Year	mean	sd	MC error	2.50%	5.00%	median	95.00%	97.50%
1	0	0	0	0	0	0	0	0
2	116	248	2.19	-78	-53	532	735	2001
3	534	805	7.39	-162	-109	1877	2477	2001
4	1182	1234	10.46	-265	-157	3324	4139	2001
5	2880	2456	20.73	-304	-9	7311	8940	2001
6	3622	2746	22.74	-343	244	8286	9871	2001
7	5577	4051	31.52	-272	711	12760	15240	2001
8	9210	6501	59.28	-636	1258	20050	23730	2001
9	11140	10160	111.80	-5555	-852	26900	32510	2001
10	16440	17820	214.30	-8590	-1744	44630	57080	2001
Total	50710	25030	378.80	9571	19440	90850	104600	2001

Chapter 8

Conclusions

In the previous chapters, we have investigated numerous stochastic models in loss reserving, particularly those dealing with zeros and negatives in the loss triangle. Papers such as de Alba (2002a, 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. After a review of the literature and methodologies, we implemented the model of Kunkler (2006) in BUGS with slightly different specifications, in order to test the model and reproduce Kunkler's results. 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 model was implemented in BUGS with prior distributions and data similar to those in Kunkler (2006).

8.1 The Model

The Bayesian multinomial mixture model we proposed in Chapter Six and Seven seems to work very well in dealing with zeros and negatives in stochastic loss reserving. From the simulation results in Chapter Seven, 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.

8.2 The Software

Application of the Bayesian software package BUGS facilitates the model implementation for our model. The programming language and grammar are easy to use and flexible for model coding. For Bayesian generalized linear modelling, any type of link function can be specified in the model equation in addition to the well known types. By specifying noninformative priors for all the parameters, we are able to implement the classical models, such as the classical generalized linear models in BUGS. As an open source program, a large number of researchers are contributing to the development of BUGS, which keeps BUGS up to date with the latest developments in Bayesian statistics.

8.3 Future Work

The Bayesian mixture model implemented in Chapter Seven 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 as reviewed in Chapter Two 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.

Bibliography

[1] ALBERT, J. H. AND CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679.

[2] ANTONIO, K., BEIRLANT, J., HOEDEMAKERS, T., AND VERLAAK, R. (2006). Lognormal mixed models for reported claim reserves. *North American Actuarial Journal* **10**, 1, 30–48.

[3] BAYES, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 330–418. Reprinted, with biographical note by G. A. Barnard, in *Biometrika* **45**, 293–315 (1958).

[4] BORNHUETTER, R. L. AND FERGUSON, R. E. (1972). The actuary and IBNR. *Proceedings of the Casualty Actuarial Society* **59**, 181–195.

[5] BROOKS, S. AND GELMAN, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**, 4, 434–455.

[6] DE ALBA, E. (2002a). Claims reserving when there are negative values in the runoff triangle. *37th Actuarial Research Conference, ARCH 2003 Proceedings.*

[7] DE ALBA, E. (2002b). Bayesian estimation of outstanding claim reserves. *North American Actuarial Journal* **6**, 4, 1–20.

[8] 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.

[9] DOBSON, A. J. (2002). *An Introduction to Generalized Linear Models*, Second ed. Chapman & Hall/CRC, Boca Raton, FL.

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

[11] ENGLAND, P. D. AND VERRALL, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal* **8**, 3, 519–544.

[12] FISHER, R. A. (1935). *The Design of Experiments.* Edinburgh: Oliver & Boyd.

[13] GELFAND, A. E. (2000). Gibbs sampling. *Journal of the American Statistical Association* **95**, 452, 1300–1304.

[14] GELFAND, A. E. AND SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

[15] GELMAN, A., CARLIN, J. B., STERN, H. S., AND RUBIN, D. B. (2004). *Bayesian Data Analysis*, Second ed. Chapman & Hall/CRC, Boca Raton, FL.

[16] GELMAN, A. AND RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–472.

[17] GEMAN, S. AND GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

[18] HAASTRUP, S. AND ARIAS, E. (1996). Claims reserving in continuous time: a nonparametric Bayesian approach. *ASTIN Bulletin* **26**, 2, 139–164.

[19] HARNEK, R. F. (1966). Formula loss reserves. *Insurance Accounting and Statistical Proceedings*.

[20] HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

[21] HESS, K. T. AND SCHMIDT, K. D. (2002). A comparison of models for the chain-ladder method. *Insurance: Mathematics and Economics* **31**, 3, 351–364.

[22] KREMER, E. (1982). IBNR claims and the two way model of ANOVA. *Scandinavian Actuarial Journal*, 47–55.

[23] KUNKLER, M. (2004). Modelling zeros in stochastic reserving models. *Insurance: Mathematics and Economics* **34**, 1, 23–35.

[24] KUNKLER, M. (2006). Modelling negatives in stochastic reserving models. *Insurance: Mathematics and Economics* **38**, 3, 540–555.

[25] LEONARD, T. AND HSU, J. S. J. (1999). *Bayesian Methods.* Cambridge University Press, Cambridge. An Analysis for Statisticians and Interdisciplinary Researchers.

[26] LeSage, J. P. (1999). Applied econometrics using MATLAB. *Available at http://www.spatial-econometrics.com/*.

[27] Mack, T. (1991). A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bulletin* **22**, 1, 93–109.

[28] Mack, T. (1993). Distribution-free calculation of the standard error of chain-ladder reserve estimates. *ASTIN Bulletin* **23**, 2, 213–225.

[29] Mack, T. (1994). Which stochastic model is underlying the chain ladder method? *Insurance: Mathematics and Economics* **15**, 2-3, 133–138.

[30] Mack, T. and Venter, G. (2000). A comparison of stochastic models that reproduce chain ladder reserve estimates. *Insurance: Mathematics and Economics* **26**, 1, 101–107.

[31] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Second ed. Chapman & Hall/CRC, Boca Raton, FL.

[32] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.

[33] Metropolis, N. and Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association* **44**, 335–341.

[34] Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A* **135**, 3, 370–384.

[35] NTZOUFRAS, I. AND DELLAPORTAS, P. (2002). Bayesian modelling of outstanding liabilities incorporating claim count uncertainty. *North American Actuarial Journal* **6**, 113–128.

[36] NTZOUFRAS, I., KATSIS, A., AND KARLIS, D. (2005). Bayesian assessment of the distribution of insurance claim counts using reversible jump MCMC. *North American Actuarial Journal* **9**, 90–108.

[37] RENSHAW, A. E. (1994a). Claims reserving by joint modelling. Actuarial Research Paper No. 72, Department of Actuarial Science and Statistics, City University, London.

[38] RENSHAW, A. E. AND VERRALL, R. J. (1998). A stochastic model underlying the chain-ladder techniques. *British Actuarial Journal* **4**, 4, 903–923.

[39] SCOLLNIK, D. P. M. (1993). A Bayesian analysis of a simultaneous equations model for insurance rate-making. *Insurance: Mathematics and Economics* **12**, 3, 265–286.

[40] SCOLLNIK, D. P. M. (1995). Bayesian analysis of generalized Poisson models for claim frequency data utilising Markov chain Monte Carlo methods. *Actuarial Research Clearing House* **1995**, 1, 339–356.

[41] SCOLLNIK, D. P. M. (1996). An introduction to Markov chain Monte Carlo methods and their actuarial applications. *Proceedings of the Casualty Actuarial Society* **83**, 114–165.

[42] SCOLLNIK, D. P. M. (2001). Actuarial modeling with MCMC and BUGS. *North American Actuarial Journal* **5**, 2, 96–125.

[43] SCOLLNIK, D. P. M. (2002a). 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.

[44] SCOLLNIK, D. P. M. (2002b). Regression models for bivariate loss data. *North American Actuarial Journal* **6**, 4, 67–80.

[45] SCOLLNIK, D. P. M. (2002c). Modeling size of loss distributions for exact data in WinBUGS. *Journal of Actuarial Practice* **10**, 1-2, 193–218.

[46] SCOLLNIK, D. P. M. (2004). Bayesian reserving models inspired by chain ladder methods and implemented using WinBUGS. *Actuarial Research Clearing House* **2004.2**, 1, 1–43.

[47] SCOLLNIK, D. P. M. (2005). Comments on: "A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving," by R. J. Verrall. *North American Actuarial Journal* **9**, 3, 146–149.

[48] 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.

[49] SPIEGELHALTER, D. J., THOMAS, A., BEST, N. G., AND GILKS, W. R. (1996). BUGS: Bayesian inference Using Gibbs Sampling, version 0.5.

[50] SPIEGELHALTER, D. J., THOMAS, A., BEST, N. G., AND LUNN, D. (2004). Winbugs user manual. *Available at http://www.mrc-bsu.cam.ac.uk/bugs.*

[51] TIERNEY, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics* **22**, 4, 1701–1762.

[52] VERRALL, R. J. (1989). A state space representation of the chain ladder linear model. *Journal of the Institute of Actuaries* **116**, 589–610.

[53] VERRALL, R. J. (1996). Claims reserving and generalised additive models. *Insurance: Mathematics and Economics* **19**, 31–43.

[54] VERRALL, R. J. (2000). An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: Mathematics and Economics* **26**, 1, 91–99.

[55] 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.

[56] VERRALL, R. J. (2005). Comments on: "A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving," by R. J. Verrall. *North American Actuarial Journal* **9**, 3, 149.

[57] VERRALL, R. J. AND ENGLAND, P. D. (2000). Comments on: "A comparison of stochastic models that reproduce chain ladder reserve estimates", by T. Mack and G. Venter. *Insurance: Mathematics and Economics* **26**, 1, 109–111.

[58] VERRALL, R. J. AND ENGLAND, P. D. (2005). Incorporating expert opinion into a stochastic model for the chain-ladder technique. *Insurance: Mathematics and Economics* **37**, 2, 355–370.

[59] WRIGHT, T. S. (1990). A stochastic method for claims reserving in general insurance. *Journal of the Institute of Actuaries* **117**, 677–731.

[60] ZEHNWIRTH, B. (1985). *ICRFS Version 4 Manual and Users Guide.* Benhar Nominees Pty Ltd, Turramurra, NSW, Australia.

[61] 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.