Claims Reserving When There Are Negative Values in the Runoff Triangle: Bayesian analysis using the three-parameter log-normal distribution

Enrique de Alba

ITAM, Mexico
and
University of Waterloo, Canada

and

Gilberto Atondo
ITAM, Mexico

39 th. Actuarial Research Conference
The University of Iowa,

August 5-7, 2004

1. INTRODUCTION

The many uncertainties involved in the payment of losses makes the estimation of the required reserves difficult. Yet, some of the existing methods, such as the popular chain-ladder, are simple to apply. However, it has become evident that there is a need for better ways not only to estimate the reserves, but also to obtain some measures of their variability. This has led to the development of stochastic reserving models, Taylor (2000), Kaas et. al. (2001), England and Verrall (2002), England (2002), de Alba (2002b).

The chain-ladder is used as a benchmark in several of the references mentioned above, due to its generalized use and ease of application, see also Hess and Schmidt (2002). This facilitates comparison between methods. However, in this paper our aim is not to develop Bayesian methods that provide results close to those of the chain-ladder method. Rather, we aim at developing 'objective' Bayesian methods using some common assumptions and to use the resulting predictive distributions to estimate loss reserves, allowing for the possibility of negative values in the data.

In this paper we present an application of Bayesian forecasting methods to the estimation of reserves for outstanding claims. We assume that the time (number of periods) it takes for the claims to be completely paid is fixed and known, that payments are made annually and that the development of partial payments follows a stable pay-off pattern. This is in agreement with many existing models for claims reserving in non-life (general) insurance that assume, explicitly or implicitly, that the proportion of claim payments, payable in the *j-th* development period, is the same for all periods of origin, Hess and Schmidt (2002). The results are applicable to any frequency of claim payments (years, quarters, etc.) and length of pay-off' period. We present a Bayesian approach to forecasting total aggregate claims given data on some development years for several occurrence years. Essentially the data would correspond to a typical run-off triangle used in loss reserving. We use the term claims reserving in its most general sense. In particular we are concerned with the situation when there are negative values in the development triangle of the incremental claim amounts.

We use standard notation, so that Z_{it} = incremental number (or amount) of events (claims) in the *t-th* development year corresponding to year of origin (or accident year) *i*. Thus $\{Z_{it}; i=1,...,k,\ t=1,...,s\}$ where s = maximum number of years (sub periods) it takes to completely pay out the total number (or amount) of claims corresponding to a given exposure year. In this paper we do not assume $Z_{it} > 0$ for all $i = 1,...,k$ and $t = 1,...,s$. Most claims reserving methods usually assume that $s=k$ and that we know the values Z_{it} for $i+t \le k+1$. The known values are presented in the form of a run-off triangle, Table 1.

Table 1

Year of origin	Development Year 1	2	......	t	...	k-1	k
1	Z_{11}	Z_{12}	...	Z_{1t}		$Z_{1,k-1}$	Z_{1k}
2	Z_{21}	Z_{22}	...	Z_{2t}		$Z_{2,k-1}$	-
3	Z_{31}	Z_{32}	...	Z_{3t}		-	
:						-	-
k-1	$Z_{k-1,1}$	$Z_{k-1,2}$				-	-
k	Z_{k1}	-					-

Negative incremental values can arise in the run-off triangle as a result of salvage recoveries, payments from third parties, total or partial cancellation of outstanding claims, due to initial over-estimation of the loss or to possible favorable jury decision in favor of the insurer, rejection by the insurer, or just plain errors. It could be argued that the problem is more with the data than with the methods. Typically these negative values will be the Hence the ideal situation would be to correct the data before applying claims reserving methods so as to eliminate any negative values. In this respect de Alba and Bonilla (2002) provide a list of potential adjustments frequently used in practice. Although the estimation procedures can be applied both to incurred (paid losses and aggregate case estimates combined) or paid claims, it is probably better to use the latter, since negative values are less likely to appear, England and Verrall (2002). That is because case estimates are set individually and often tend to be conservative, resulting in over-estimation in the aggregate. This leads to negative incremental amounts in the later stages of development. Also, sometimes the data should be adjusted before applying these methods to satisfy regulatory requirements. However, corrections are not always possible and it is convenient to have options available to work with the negative values.

We present a full Bayesian model. It is extended from de Alba (2002b) to consider negative incremental values. The model presented here allows the actuary to provide point estimates and measures of dispersion, as well as the complete distribution for the reserves.

The paper is structured as follows. Section 2 gives a brief description of previous results relevant to our approach. Section 3 introduces some Bayesian concepts and their applications in actuarial science. Section 4 describes a Bayesian model for claim amounts in the presence of negative values. Section 5 describes the prior distributions used in the model. In Section 6 we describe how to use our model to estimate reserves and its implementation for Markov chain Monte Carlo is given in Section 7. An example is given in Section 8. All models are presented only in discrete time.

2. BACKGROUND

For a comprehensive, although not exhaustive, review of existing stochastic methods that can handle the existence of negative incremental values see England and Verrall (2002). Although they provide some Bayesian results, most of the methods presented there approach the problem from the point of view of frequentist or classical statistics and in the framework of generalized linear models (GLM). They provide predictions and prediction errors for the different methods discussed and show how the predictive distributions may be obtained by bootstrapping and Monte Carlo methods. Among those that can handle negative values, from the classical viewpoint, they mainly consider three models: an (over-dispersed) Poisson, a negative binomial and a Normal approximation to the latter. They also mention the standard log-Normal model which was introduced by Kremer (1982) and analyzed in detail in Verrall (1991).

England and Verrall (2002) emphasize "that some of the methods presented … are better suited for modeling paid amounts or number of claims, since incurred data, which may include over-estimation of case estimates, leading to negative incremental values, may cause problems." After describing the stochastic basis for the chain-ladder method, they indicate that "the Normal model has the advantage that it can produce estimates for a wide range of data sets, and is less affected by the presence of negatives"

There are several stochastic formulations of the chain-ladder method, Hess and Schmidt (2002). The stochastic version of the chain-ladder method that can handle some negative values is defined as a generalized linear model (GLM) with an over-dispersed Poisson distribution, Renshaw and Verrall (1998). In the over-dispersed Poisson model the mean and variance are not the same. In our previous notation $m_{ij} = E(Z_{ij})$, with a variance function $V(Z_{ij}) = \phi m_{ij}$ and scale parameter $\phi > 0$, combined with the log 'link' function $log(m_{ij}) = \mu + \alpha_i + \beta_j$. Over-dispersion is achieved through ϕ. This model reproduces the estimates of the classical chain-ladder method.

Estimates of the parameters, $\hat{\mu}, \hat{\alpha}_i, \hat{\beta}_j$, are obtained by using a 'quasi-likelihood' approach. Renshaw and Verrall (1998) suggest the use of Pearson residuals in the GLM when there are negative values. They point out that it "is not applicable to all sets of data, and can break down in the presence of a sufficient number of negative incremental claims." In addition, the Poisson assumption seems inadequate for continuous variables, like claims amounts. They discuss the relationship between this model and the chain-ladder technique, and show that, under certain positivity constraints, the same reserve estimates are produced by each.

The negative binomial model is closely related to the previous one, Verrall (2000). The distribution in the GLM is now assumed to be a negative binomial with mean $(\lambda_j - 1)W_{i,j-1}$ and variance $\phi\lambda_j(\lambda_j - 1)W_{i,j-1}$, where $W_{ij} = \sum_{k=1}^{j} Z_{ik}$. The parameters $\{\lambda_j : j = 2,...,n\}$ are the typical chain-ladder development factors. As before, ϕ is an over-dispersion parameter. This model yields essentially the same estimates as the (over-dispersed) Poisson. With a sufficient number of negative incremental claims, it is possible that some of the λ's (one would be enough) become less than one and so that clearly the variance would not exist. It is then necessary and possible to use a Normal approximation, and the chain-ladder results can still be reproduced. It is not recommended to use the Normal approximation in all situations, mainly because real claims data are skewed, even though its application is likely to be less troublesome in practice. The normal approximation assumes the distribution is normal with the same mean as before and variance $\phi_j W_{i,j-1}$. The link function remains the same in all cases. This is seen to be equivalent to one proposed by Mack (1993). These models have the additional disadvantage that they incorporate *n* new parameters (the ϕ_j) that must also be estimated, but this is the price one must pay to estimate the reserves in the presence of negative values.

In a recent paper Verrall (2004) presents a Bayesian model for the Bornhuetter-Ferguson method. He indicates that the Bayesian models derived in that paper may break down if there are negative incremental claims values, and is therefore probably only suitable for paid data. He further says that this is certain to happen if any column sum of incremental claims is negative and that the model can cope with some (few?) negative values.

In this paper we follow the approach set out in de Alba (2002a) where use is made of the three parameter log-normal distribution to estimate outstanding claims reserves in the presence of negative incremental claims. Let the random variable Z_{it} represent the value of incremental aggregate claims in the *t-th* development year of accident year *i*, $i,t=1,\dots,k$. The Z_{it} are known for $i+t \le k+1$ and we let

$$Y_{it} = \log(Z_{it} + \delta). \tag{1}$$

If $Y_{it} \sim N(\mu,\sigma^2)$ then Z_{it} has a three parameter log-normal distribution and its density is

$$f(z_{it} \mid \mu,\sigma^2) = \begin{cases} \dfrac{1}{\sigma(z_{it}+\delta)\sqrt{2\pi}} \exp\{-\dfrac{1}{2\sigma^2}(\log(z_{it}+\delta)-\mu)^2\} & \text{for} \quad z_{it} > -\delta \\ 0 & \text{for} \quad z_{it} \le -\delta \end{cases} \tag{2}$$

In de Alba (2002a) the threshold parameter is first estimated by Maximum Likelihood and then it is “plugged in” to use the ‘profile’ likelihood with δ replaced by its ML estimator, say $\hat{\delta}$, Crow and Shimizu (1988, page 123), and define $y_{it} = \log(z_{it} + \hat{\delta})$. This way, if some of the incremental claims are negative, the addition of $\hat{\delta}$ will make the resulting value positive and the log-transformation can be used. Then the Bayesian analysis is carried out using results for a two parameter lognormal distribution, Zellner (1971b). The approach followed in that paper has the disadvantage that the variability due to estimating δ is not taken into account. Furthermore, it is well known that the estimates of δ can be very unstable, see Cohen and Whitten (1980) and Johnson et al. (1994, chapter 14). Hill (1963) explains the cause of the instability and how it can be avoided by the use of Bayesian methods. Here, we shall use a full Bayesian structure for the model that solves these problems.

3. BAYESIAN MODELS

We do not intend to give here an extensive review of Bayesian methods. Rather we will describe them very briefly and discuss their applications in actuarial science, specifically in loss reserving. Bayesian analysis of IBNR reserves has been considered before by Jewell (1989,1990), Verrall (1990) and Haastrup and Arjas (1996). For general discussion on Bayesian theory and methods see Berger (1985), Bernardo and Smith (1994) or Zellner (1971a). For a discussion of Bayesian methods in actuarial science see Klugman (1992), Makov (1996, 2001), Scollnik (2001), Ntzoufras and Dellaportas (2002), de Alba (2002b) and Verrall (2004). Here, we refer only to those Bayesian models that can be applied to situations where $Z_{it} < 0$ for some $i,t = 1,\ldots,k$.

Verrall (1990) approaches the subject of predicting outstanding claims using hierarchical Bayesian linear models, considering the fact that the chain-ladder technique is based on a linear model: the two-way analysis of variance model (ANOVA). He essentially carries out a Bayesian analysis of the two-way ANOVA model to obtain Bayes and empirical Bayes estimates. The latter are given a credibility interpretation. Two alternative formulations are considered, one with no prior information and another where he uses a specific prior distribution for the parameters.

More recently, Bayesian results are provided in England and Verrall (2002), notably for the Bornhuetter-Ferguson (B-F) technique. The Bornhuetter-Ferguson technique is useful when there is instability in the proportion of ultimate claims paid in the early development years, so that the chain-ladder technique yields unsatisfactory results. The idea in the B-Fmethod is to use external information to obtain an initial estimate of ultimate claims. In the traditional B-F method use is made explicitly of perfect prior (expert) knowledge of 'row' parameters, ultimate claims. This is then used with the development factors of the chain-ladder technique to estimate outstanding claims. This is clearly well suited for the application of Bayesian methods when knowledge is not perfect. Verrall (2004) further explores this method in a Bayesian framework. In both references it is pointed out that they may break down in the presence of negative values, certainly if any column sum of incremental claims in the development triangle is negative.

Ntzoufras and Dellaportas (2002) consider various competing models using Bayesian theory and Markov chain Monte Carlo methods to simulate. Claim counts are used in order to add a further hierarchical stage in the model with log-normally distributed claim amounts. In a recent paper, de Alba (2002a) presents a model for aggregate claims by separating number of claims and average claims, which are also assumed log-normally distributed.

A standard measure of variability is prediction error, defined as the standard deviation of the distribution of possible reserves. In the Bayesian context the usual measure of variability is the standard deviation of the predictive distribution of the reserves. This is a natural way of doing analysis in the Bayesian approach. In this paper our aim is to obtain not only this standard deviation, but also show the complete predictive distribution. For this purpose we will use Markov chain Monte Carlo simulation (MCMC) that will be implemented with the package WinBUGS 1.4 (Spiegelhalter et al., 2001).

4. A BAYESIAN MODEL FOR AGGREGATE CLAIMS

In this section we present a model for the unobserved aggregate claim amounts and hence the necessary reserves for outstanding claims. Let the random variable Z_{it} represent the value of incremental aggregate claims in the *t-th* development year of accident year *i*, $i,t=1,...,k$. The Z_{it} are known for $i+t \leq k+1$. Let Y_{it} be defined as in equation (1) and recall that if $Y_{it} \sim N(\mu,\sigma^2)$ then Z_{it} has a three parameter log-normal distribution and its density is as given in (2).

The "threshold" parameter $\delta > 0$ corrects the values so as to assure $(z_{it}+\delta) > 0$, for $i,t=1,...,k$, with $i+t \leq k+1$. We assume in addition that

$$y_{it} = \log(z_{it}+\delta) = \mu + \alpha_i + \beta_t + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0,\sigma^2) \tag{3}$$

$i=1,...,k$, $t=1,...,k$ and $i+t \leq k+1$ so that Z_{it} follows a three parameter log-normal distribution, denoted by $Z_{it} \sim LN(\mu_{it},\sigma^2,\delta)$ with $\mu_{it} = \mu + \alpha_i + \beta_t$ and

$$f(z_{it} \mid \mu,\alpha_i,\beta_t,\sigma^2,\delta) = \frac{1}{\sigma(z_{it}+\delta)\sqrt{2\pi}} \exp[-\frac{1}{2\sigma^2}(\log(z_{it}+\delta)-\mu-\alpha_i-\beta_t)^2]. \tag{4}$$

It is well known in ANOVA that certain restrictions must be imposed on the parameters in order to attain estimability in (3). We use the assumption that $\alpha_1 = \beta_1 = 0$. Also, we define $T_U = (k+1)k/2$ = number of cells with known claim information in the upper triangle; and $T_L = (k-1)k/2$ = number of cells in the lower triangle, whose claims are unknown.

As mentioned in Section 2, one approach would be to use the 'profile' likelihood with δ replaced by its ML estimator as given in Crow and Shimizu (1988, page 123) or some other estimator, say $\hat{\delta}$, define $y_{it} = \log(z_{it}+\hat{\delta})$ and then carry out the rest of the analysis with this value replaced in (4). In this paper we will present a full Bayesian analysis and compare results.

Let $\underline{z} = \{z_{it}; i,t=1,...,k, i+t \leq k+1\}$ be a T_U-dimension vector that contains all the observed values of Z_{it}, where $Y_{it} = \log(Z_{it}+\delta)$, and $\underline{\theta} = (\mu,\alpha_2,...,\alpha_k,\beta_2,...,\beta_k)'$ is the $((2k-1)\times 1)$ vector of parameters. The likelihood function will be

$$f(\underline{z} \mid \underline{\theta}, \sigma^2, \delta) = \frac{\sigma^{-T_U}(2\pi)^{-T/2}}{\prod (z_{it} + \delta)} \exp[-\frac{1}{2\sigma^2} \sum_i \sum_t (\log(z_{it} + \delta) - \mu - \alpha_i - \beta_t)^2], \quad (5)$$

where the product and the summations are over the known z_{it} values, $i, t = 1, ..., k$, $i + t \leq k + 1$. To carry out the Bayesian analysis we must specify prior distributions for the parameters $\underline{\theta}, \sigma^2$ and δ . We will assume prior independence so that

$$f(\underline{\theta}, \sigma^2, \delta) = f(\underline{\theta}) f(\sigma^2) f(\delta) = f(\mu) \times [\prod_{i=2}^{k} f(\alpha_i)][\prod_{t=2}^{k} f(\beta_t)] \times f(\sigma^2) \times f(\delta)$$

and the posterior distribution will be

$$f(\underline{\theta}, \sigma^2, \delta \mid \underline{z}) \propto f(\underline{z} \mid \underline{\theta}, \sigma^2, \delta) f(\mu) \times [\prod_{i=2}^{k} f(\alpha_i)][\prod_{t=2}^{k} f(\beta_t)] \times f(\sigma^2) \times f(\delta) .$$

We will use a hierarchical model in the following sections. In hierarchical models, at a first stage the data is specified to come from a given distribution, $f(z_{it} \mid \mu, \alpha_i, \beta_t, \sigma^2, \delta)$ in our case. At the second stage, the parameters are assumed to follow their own (prior) distributions, here $f(\mu)$, $f(\alpha_i)$, $f(\beta_t)$, $f(\sigma^2)$ and $f(\delta)$, $i,t = 2,...,k$. Each one of these distributions may in turn depend on unknown 'hyperparameters' whose values or distribution will need to be specified. In the latter case it will be necessary to still add another stage, Klugman (1992, page 66). We will then use Markov chain Monte Carlo (MCMC) simulation to generate samples from the posterior distributions of the parameters as well as the predictive distribution of the reserves. This can be implemented with the package WinBUGS 1.4 (Spiegelhalter et al., 2001). In the next section we describe the specification of the prior distributions.

5. PRIOR DISTRIBUTIONS

Assuming α_i, β_t, σ^2 and δ are known, then a natural conditionally-conjugate prior for μ in the model (3) will be $\mu \sim N(\mu_0, {\sigma_0}^2)$. Analogously, the natural conjugate priors for α_i and β_t are $\alpha_i \sim N(\mu_\alpha, {\sigma_\alpha}^2)$ and $\beta_t \sim N(\mu_\beta, {\sigma_\beta}^2)$, respectively. The values of the parameters in these distributions must be specified or else they must be assumed to follow a distribution. We will do the latter specifying a distribution that reflects little or no information, Zellner (1971a). This is easily done in WinBUGS, Scollnik (2001).

Now, given μ, α_i, β_t and δ the conjugate prior distribution for σ^2 will be an inverse gamma distribution,

$$f(\sigma^2) = \frac{\lambda^\nu}{\Gamma(\nu)} (\sigma^2)^{-(\nu+1)} \exp\{-\lambda / \sigma^2\}, \qquad \sigma^2 > 0, \qquad (6)$$

Bernardo and Smith (1994), denoted by $\sigma^2 \sim IG(\nu, \lambda)$. Again, we will assume (ν, λ) follow distributions that reflect little knowledge about them.

Finally we must specify the prior distribution for the threshold parameter δ, assuming the rest of the parameters, μ, α_i, β_t and σ^2, are given. The only reference on this is Hill (1963). He specifies a prior distribution for δ that reflects one's subjective beliefs or knowledge about the parameters in order to limit the range of values they can take and thus avoid the strange results he describes. Alternative non-Bayesian methods for estimating δ, based on order statistics use the first order statistic of the sample $z_{(1)} = \min\{z_{it}; i, t = 1,..., k\ , i + t \leq k + 1\}$ in their estimation process to estimate δ, Cohen and Whitten (1980). Here we specify a Pareto distribution as a prior for δ

$$f(\delta) = ac^a \delta^{-(a+1)}, \quad c > 0 \quad a > 0 \quad \delta \geq c, \qquad (7)$$

where we shall use $c = -z_{(1)}$ and a suitably specified distribution for *a*, so we use a further stage in the hierarchical model. We will use the notation $\delta \sim \mathrm{Par}(a, c)$. The reason for using this prior is that since we have negative values in the sample (at least one) we must make sure that $\delta > z_{(1)}$ and $(z_{it} + \delta) > 0$. Also, this distribution does not assign zero probability to any intervals above *c*, so that the estimates that would result from alternative estimation methods are not excluded a-priori. Hence we can analyze alternative estimates of δ in relation to its posterior distribution under this model.

6. ESTIMATING THE RESERVES

We want to estimate (or obtain the distribution of) aggregate claims for accident year *i* given information on at least one year that has fully developed and perhaps on *m* previous completely known accident years. Let $Z^*_{it} = \sum_{j=1}^{t} Z_{ij}$ for $1 \leq t \leq k$. Hence, in the run-off triangle setup, we are really interested in estimating Z^*_{ik} $i=2,...,k$, given Z^*_{1k}, and Z_{it}, $i=1,...,k$ $t=1,...,k$, with $i+t \leq k+1$. Now let $R_i = Z^*_{ik} - Z^*_{ia_i}$, for $i=2,...,k$ with $a_i=k-i+1$, so that $Z^*_{ia_i}$ is the accumulation of Z_{it} up to the latest development period and R_i = the total aggregate outstanding claims for the development years for which it is unknown, both corresponding to business year *i*, i.e. the required reserves corresponding to this business year.

We assume the parameters are independent a-priori and specify conjugate priors for $\underline{\theta}$ and σ and a Pareto prior for δ, as indicated in section 5. The joint posterior distribution is then seen to be

$$f(\underline{\theta},\sigma^2,\delta \mid \underline{z}) \propto f(z \mid \underline{\theta},\sigma^2,\delta) f(\mu) \times [\prod_{i=2}^{k} f(\alpha_i)][\prod_{t=2}^{k} f(\beta_t)] \times f(\sigma^2) \times f(\delta)$$

and using (5)

$$f(\underline{\theta},\sigma^2,\delta \mid \underline{z}) \propto \frac{\sigma^{-T_U}}{\prod (z_{it}+\delta)} \exp[-\frac{1}{2\sigma^2}\sum_i \sum_t (\log(z_{it}+\delta)-\mu-\alpha_i-\beta_t)^2] \times f(\mu) \times [\prod_{i=2}^{k} f(\alpha_i)]$$

$$\times [\prod_{t=2}^{k} f(\beta_t)] \times f(\sigma^2) \times f(\delta) \quad . \qquad (8)$$

In the Bayesian approach, when interest centers on inference about the parameters it can be carried out using $f(\underline{\theta},\sigma^2,\delta \mid \underline{z})$. When interest is on prediction, as in loss reserving, the past (known) data in the upper portion of the triangle, $\underline{z}$, are used to predict the observations in the lower triangle z_{it} by means of the posterior predictive distribution

$$f(z_{it} \mid \underline{z}) = \int f(z_{it} \mid \underline{\theta},\sigma^2,\delta) f(\underline{\theta},\sigma^2,\delta \mid \underline{z}) d\underline{\theta} d\sigma^2 d\delta \,,$$

$$i = 1,\ldots,k, \quad t = 1,\ldots,k, \quad with \quad i+t > k+1 .$$

Hence to compute the reserves for the outstanding aggregate claims we need to estimate the lower portion of the triangle. We can do this by obtaining the mean and variance of the predictive distribution. Hence for each cell we must obtain $E(Z_{it} \mid \underline{z})$. Then the Bayes estimate of outstanding claims for year of business i is $\sum_{t>k-i+1} E(Z_{it} \mid \underline{z})$. The Bayes 'estimator' of the variance (the predictive variance) for that same year is too cumbersome to derive. One alternative would be to use direct simulation from the posterior distributions to generate a set of *N* randomly generated values for the parameters from (8) and then in turn use the resulting values of the parameters in $f(z_{it} \mid \underline{\theta},\sigma^2,\delta)$. This yields random observations for aggregate claims in each cell of the (unobserved) lower right triangle $z_{it}^{(j)}$, $i = 2,\ldots,k$, $t > k-i+1$, for *j=1,…,N,* de Alba (2002b). The resulting values will include both parameter variability and process variability. Thus we can compute a random value of the total required reserves $R^{(j)} = \sum_{i,t} Z_{it}^{(j)}$. The mean and variance can be computed as

$$\sigma_R^2 = \sum_{j=1}^{N} \frac{(R^{(j)} - \overline{R})^2}{N} \quad and \quad \overline{R} = \frac{1}{N}\sum_{j=1}^{N} R^{(j)} \quad .$$

The standard deviation σ_R thus obtained is an 'estimate' for the prediction error of the number of claims to be paid. The simulation process has the added advantage that it is not necessary to obtain explicitly the covariances that may exist between parameters. They are dealt with implicitly. The direct simulation may be very cumbersome, if not impossible, to do if the posterior distribution is not of a known type. It is for this kind of situation where MCMC proves very useful. We turn to this in the next section.

7. IMPLEMENTING MCMC IN BUGS

To implement the simulation by Markov chain Monte Carlo we use the package WinBUGS 1.4, Spiegelhalter et al. (2001). We need to specify each of the stages in the hierarchical model. According to the specification of the model in Section 5, in the first stage the likelihood function is:

$$Y_{it} = log(Z_{it} + \delta) \sim N(\mu_{it}, \sigma^2)$$
$$\mu_{it} = \mu + \alpha_i + \beta_j$$
$$\alpha_1 = \beta_1 = 0$$
$$i = 1,...,k, \quad j = 1,...,k, \quad y \quad i + j \le k + 1$$

k = number of accident years = number of development years

In the second stage the prior distribution is specified for the unknown parameters. In our model:

$$\mu \sim N(0, \sigma^2_\mu)$$
$$\alpha_i \sim N(0, \sigma^2_{\alpha_i})$$
$$\beta_t \sim N(0, \sigma^2_{\beta_t})$$
$$\sigma^2 \sim GI(v, \lambda)$$
$$\delta \sim \text{Par}(a, c)$$

The prior distributions for μ, α_i, and β_t have been chosen as to allow the simulation process to converge to any possible value, positive or negative. This will be further enhanced by the choice of their variances or the distribution of their variances. If these are allowed to be large the prior distributions will be of the non-informative kind. In the prior for σ^2 we choose each of the parameters (v, λ) to follow a distribution such that the prior reflects ignorance about the possible value of σ^2, Hill (1963), Gelman (2004). This is done at the third stage. The prior distribution for δ is specified, at a second stage, as indicated in Section 5, with $c = -z_{(1)}$. A distribution is also specified for the parameter a in the Pareto distribution in a third stage.

In the third stage we specify the distributions for some of the (hyper) parameters used in specifying the distributions in stage two, Klugman (1992). These are usually given values that reflect lack of information, Scollnik (2001) and Zellner (1971a). Thus

$$\sigma_{\mu}^{2} \sim GI(b,c)$$
$$\sigma_{\alpha_i}^{2} \sim GI(d,e)$$
$$\sigma_{\beta_t}^{2} \sim GI(f,g)$$
$$v \sim G(h,i)$$
$$\lambda \sim G(j,k)$$
$$a \sim G(l,m)$$

where b through m are given values, and $G(j,k)$ denotes a Gamma distribution with parameters j and k. As before, $GI(b,c)$ denotes an inverse Gamma distribution with parameters b and c.

8. APPLICATION

In this section we present a set of data that contains many negative values. Table 2 presents the data which was provided by Prof. R. L. Brown and was kindly made available by an (anonymous) American insurance company. It includes a fairly large number of negative values. This set of data is interesting because there are actually four columns in the development triangle whose sum is negative and this causes all the stochastic claims reserving methods to break down. Perhaps the only one that will produce results is the generalized linear model with over-dispersed Poisson distribution and using quasi-likelihood estimation, Renshaw and Verrall (1998). But this method implies the use of a Normal approximation, which may not be adequate with skewed distributions. Also, the standard (deterministic) chain ladder method does provide results, but it is not possible to obtain estimates of the variances. Tables 3 and 4 present the results of applying different methods.

Table 2

	1	2	3	4	5	6	7	8	9
1	33250.717	2097.059	78.897	21.117	-18.654	-0.121	-5.072	-1.292	-0.775
2	36717.578	2583.632	-34.240	19.080	10.120	-3.699	-2.492	1.259	
3	38155.786	2705.212	38.503	-0.247	6.442	-6.669	-9.525		
4	36180.233	2601.743	21.501	-8.662	-6.250	12.865			
5	35980.821	2892.427	52.478	10.982	-3.496				
6	37518.185	2901.650	-23.612	-39.496					
7	40213.152	3006.438	-14.591						
8	39105.807	3080.126							
9	41184.755								

We compare the results of applying the chain-ladder method, the over-dispersed Poisson of Renshaw and Verrall (1998) with quasi-likelihood estimation, the Bayesian method using a 'profile' likelihood with δ replaced with its ML estimate, as in de Alba (2002a), and our fully Bayesian model presented above with MCMC simulation. The three stages of the hierarchical model were specified as follows:

Stage 1:

$$Y_{it} = log(Z_{it} + \delta) \sim N(\mu_{it}, \sigma^2)$$
$$\mu_{it} = \mu + \alpha_i + \beta_j$$
$$\alpha_1 = \beta_1 = 0$$
$$i = 1,...,k, \quad j = 1,...,k, \quad y \quad i + j \leq k+1$$
$$k = 9$$

Stage 2:

$$\mu \sim N(0, \sigma_\mu^2)$$
$$\alpha_i \sim N(0, \sigma_{\alpha_i}^2)$$
$$\beta_t \sim N(0, \sigma_{\beta_t}^2)$$
$$\sigma^2 \sim GI(\nu, \lambda)$$
$$\delta \sim Par(a, 39.5)$$

Stage 3:

$$\sigma_\mu^2 \sim GI(0.1, 0.1)$$
$$\sigma_{\alpha_i}^2 \sim GI(0.001, 0.001)$$
$$\sigma_{\beta_j}^2 \sim GI(0.001, 0.001)$$
$$\nu \sim G(2.5, 0.001)$$
$$\lambda \sim G(2, 0.1)$$
$$a \sim G(0.001, 0.001)$$

The parameters in the third stage were chosen to reflect lack of information. Those for $\sigma_\mu^2, \sigma_{\alpha_i}^2$ and $\sigma_{\beta_t}^2$ are specified as in Spiegelhalter et al. (2003) and Verrall (2004). The priors for ν and λ are specified so that the prior expected value of σ^2 corresponds approximately to the value that results when applying OLS to the data after correcting by adding the MLE of δ, $\hat{\delta} = 300.2$. Finally, the distribution for a was set so that a-priori δ has a large variance and so it allows the random values generated in the MCMC simulation to cover a broad range of values. Its prior expected value is between $z_{(1)}$ and its MLE $\hat{\delta}$. In addition it is also one of the typical non-informative priors used with WinBUGS, Spiegelhalter et al. (2003).

We used an initial burn-in sample of 10000 iterations. Two parallel chains were generated, each one with a burn in sample of 5000. The results of these observations were discarded, to remove any effect from the initial conditions and allow the simulations to converge. We also examined the results using a number of different initial conditions to ensure that these had no effect on the results. We then ran a further 50000 simulations for each of the two chains to obtain the results shown below. Various checks were made of the convergence of the Markov chain, including a visual inspection of the sampled values.

We obtain some characteristics of the posterior distribution of the parameters δ, μ and σ^2. They are shown in Table 3. The posterior mean of δ is 182.0 while its MLE is $\hat{\delta} = 300.2$. In this example the reserves in both Bayesian methods are not very close. They are given in Table 4. The standard deviations from the full Bayesian model smaller. Again, the chain-ladder does not seem to be affected by the negative values. In this case the estimates of the reserves obtained with the fully Bayesian method presented here are lower than those of the non-Bayesian ones and than the others. Also, the variance of the MCMC estimates is smaller that those from the direct simulation.

Table 3

Parameter	Mean	Std. Dev.	Percentiles		
			2.50%	Median	97.50%
δ	182.0	50.0	102.5	176.6	302.3
μ	10.53	.0460	10.44	10.53	10.62
σ^2	.0167	.0091	.0068	.0145	.0397

Figure 1 shows the predictive distribution for accident year 6 (top panel) and for the total (bottom panel). The latter has a long right tail. The distribution for reserves corresponding to accident year 6 shows a large probability of negative values and in fact the predictive mean close to zero. The results for accident years 2 and 4 (not shown) also yield medians very close to zero. This is not the case for total reserves. It appears that the use of MCMC, as opposed to plugging in the MLE of δ decreases the variance of the predictive distribution of the reserves. This probably has to do with some reduction in variance of the other parameters. The non-informative priors used in the third stage seems to allow a better fit to the data. For some accident years there may be a very high probabil be set aside if the chain-la .

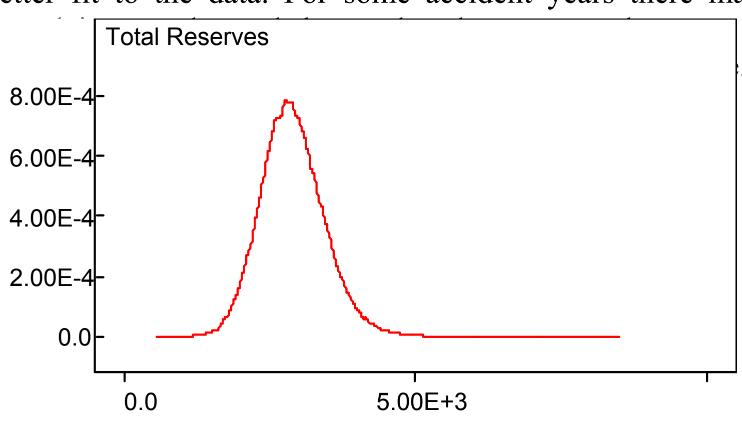

Figure 1

Table 4

Year	Chain-Ladder	OD Poisson	Bayesian		Bayesian MCMC	
	Reserves	Reserves	Reserves	Std. Dev.	Reserves	Std. Dev.
2	-0.860	1.000	1.381	32.246	0.3587	32.22
3	-0.912	2.000	21.348	47.190	9.383	44.54
4	-6.601	3.000	5.810	57.877	1.187	52.91
5	-6.024	7.000	64.140	73.000	22.20	63.51
6	-8.715	-12.000	-54.890	81.640	-27.65	69.47
7	-8.817	22.000	77.690	107.550	7.945	79.73
8	9.513	48.000	244.080	148.150	49.17	94.72
9	3041.181	3085.000	3363.200	514.500	2835.0	468.8
TOTAL	3018.766	3182.000	3722.700	620.500	2897.0	545.3

9. CONCLUDING REMARK

The Bayesian method presented here constitutes an appealing alternative to claims reserving methods in the presence of negative values in incremental claims for some cells of the development triangle. It yields good results. Furthermore, the model is based on fairly standard and widely used assumptions. However, the main advantage is that this method will not break down even in the presence of a considerable number of negative values.

REFERENCES

(1) de Alba, E. (2002a), Claims Reserving When There Are Negative Values in the Runoff Triangle, *37th Actuarial Research Conference*, *The University of Waterloo,* ARCH 2003 Proceedings, The Society of Actuaries

(2) de Alba, E. (2002b), Bayesian Estimation of Outstanding Claim Reserves, *North American Actuarial Journal* 6(4), 1-20.

(3) Berger, J.O. (1985), *Statistical Decision Theory and Bayesian Analysis*, 2nd. Ed., Springer-Verlag, New York.

(4) Bernardo, J.M. and A.F.M. Smith (1994), *Bayesian Theory*, John Wiley & Sons, New York.

(5) Cohen, A.C. and Whitten, B.J. (1980), "Estimation in the three parameter log-normal distribution", *Journal of the American Statistical Association*, 75,370, 399-404

(6) Crow, E.L. and Shimizu, K. (1988), *Lognormal Distributions. Theory and Applications*, Marcel Dekker, New York.

(7) England, P. (2002), Addendum to "Analytic and bootstrap estimates of prediction errors in claims reserving", *Insurance: Mathematics and Economics 31*, 461-466.

(8) England, P. and R. J. Verrall (1999), Analytic and bootstrap estimates of prediction errors in claims reserving, *Insurance: Mathematics and Economics* 25, 281-293.

(9) England, P. and R. J. Verrall (2002), Stochastic Claims Reserving in General Insurance (with discussion), *British Actuarial Journal* 8, 443-544.

(10) Gelman, A., J.B. Carlin, H.S. Stern and D.B. Rubin (2004), *Bayesian Data Analysis*, 2nd Ed., Chapman & Hall/CRC

(11) Haastrup, S. and E. Arjas (1996), Claims Reserving in Continuous Time; A Nonparametric Bayesian Approach, *ASTIN Bulletin* 26(2), 139-164.

(12) Hess, K.T. and Schmidt, K.D. (2002), A comparison of models for the chain-ladder method, *Insurance: Mathematics and Economics 31*, 351-364.

(13) Hill, B. M. (1963), ``The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic", *Journal of the American Statistical Association*, 58 , 72-84

(14) Jewell, W.S. (1989), Predicting IBNYR Events and Delays. I Continuous Time, *ASTIN Bulletin* 19(1), 25-56.

(15) Jewell, W.S. (1990), Predicting IBNYR Events and Delays. II DiscreteTime, *ASTIN Bulletin* 20(1), 93-111.

(16) Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994), *Continuous Univariate Distributions in Statistics Vol. I*, Wiley: New York.

(17) Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2001), *Modern Actuarial Risk Theory*, Kluwer Academic Publishers.

(18) Klugman, S.A. (1992) *Bayesian Statistics in Actuarial Science,* Kluwer: Boston.

(19) Kremer, (1982), IBNR claims and the two-way model of ANOVA*, Scandinavian Actuarial Journal*, 47-55.

(20) Mack, T. (1993), Distribution-Free calculation of the standard error of chain ladder reserve structure, *ASTIN Bulletin* 23, 213-225.

(21) Mack, T. (1994), Which Stochastic Model is Underlying the Chain Ladder Method?, *Insurance: Mathematics and Economics* 15, 133-138.

(22) Makov, U.E., A.F.M. Smith & Y.-H. Liu (1996), "Bayesian Methods in Actuarial Science", *The Statistician*, Vol. 45,4, pp. 503-515.

(23) Makov, U.E. (2001), "Principal applications of Bayesian Methods in Actuarial Science: A Perspective", *North American Actuarial Journal* 5(4)*,* 53-73.

(24) Ntzoufras, I. and Dellaportas, P. (2002), Bayesian Modeling of Outstanding Liabilities Incorporating Claim Count Uncertainty*, North American Actuarial Journal* 6(1), 113-136.

(25) Renshaw, A.E. and R. J. Verrall (1998), A stochastic model underlying the chain-ladder technique, *British Actuarial Journal* 4,(IV) 905-923.

(26) Scollnik, D.P.M. (2001), Actuarial Modeling With MCMC and BUGS, *North American Actuarial Journal* 5(2), 96-124.

(27) Spiegelhalter, D. J., Thomas, A., Best, N. G., Gilks, W. R., and Lunn, D. (1994, 2003). BUGS: Bayesian inference using Gibbs sampling. MRC Biostatistics Unit, Cambridge, England. www.mrc-bsu.cam.ac.uk/bugs/

(28) Spiegelhalter, D.J., Thomas, A., Best, N.G. and Gilks, W.R. (2001) Win*BUGS 1.4: Bayesian Inference using Gibbs Sampling*, Imperial College and MRC Biostatistics Unit, Cambridge, UK. *http://www.mrc-bsu.cam.ac.uk/bugs*.

(29) Taylor, G.C. (2000), *Claim Reserving. An Actuarial Perspective,* Elsevier Science Publishers, New York.

(30) Verrall, R. J. (1990), Bayes and Empirical Bayes Estimation for the Chain Ladder Model, *ASTIN Bulletin* 20(2), 217-243.

(31) Verrall, R. J. (1991), Negative Incremental Claims: Chain ladder and their Models, *Journal of the Institute of Actuaries* 120(I), 171-183.

(32) Verrall, R. J. (2000), An investigation into stochastic claims reserving models and the chain-ladder technique, *Insurance: Mathematics and Economics*, Vol. 26, 91-99.

(33) Verrall, R.J. (2004), A Bayesian Generalised Linear Model for the Bornhuetter-Ferguson Method of Claims Reserving, *North American Actuarial Journal* 8, forthcoming.

(34) Zellner, A. (1971a), *An Introduction to Bayesian Inference in Econometrics*, Wiley: New York.

(35) Zellner, Arnold (1971b), ``Bayesian and non-Bayesian analysis of the log-normal distribution and log-normal regression'', *Journal of the American Statistical Association*, 66 , 327-330