Bayesian Claims Reserving When There Are Negative Values in the Runoff Triangle.

Enrique de Alba
ITAM, Mexico
and
University of Waterloo, Canada

and
Marco A. Ramírez Corzo
ITAM, México

Abstract

In this paper we are concerned with the situation that occurs in claims reserving when there are negative values in the development triangle of incremental claim amounts. Typically, these negative values will be the 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. Some of the traditional methods of claims reserving, such as the chain-ladder technique, may produce estimates of the reserves even when there are negative values. However, many methods can break down in the presence of enough (in number and/or size) negative incremental claims if certain constraints are not met. Historically the chain-ladder method has been used as a gold standard (benchmark), due to its generalized use and ease of application. A method that improves on the gold standard is one that can handle situations where there are many negative incremental claims and/or some of these are large. We present a Bayesian model to consider negative incremental values, based on a three parameter log-normal distribution. The model presented here allows the actuary to provide point estimates and measures of dispersion, as well as the complete distribution for outstanding claims from which the reserves can be derived. We apply MCMC using the package WinBUGS.

Bayesian Claims Reserving When There Are Negative Values in the Runoff Triangle.

Enrique de Alba
ITAM, Mexico
and
University of Waterloo, Canada

And
Marco A. Ramírez Corzo
ITAM, Mexico

40 th. Actuarial Research Conference
ITAM, Mexico

August 11-13, 2005

Bayesian Claims Reserving When There Are Negative Values in the Runoff Triangle.

1. INTRODUCTION

The estimation of adequate reserves for outstanding claims is one of the main activities of actuaries in property/casualty insurance. The need to estimate future claims has led to the development of many loss reserving techniques. Probably the oldest and most widely used of these techniques is the well known chain-ladder. It is frequently used as a benchmark, due to its generalized use and ease of application, Hess and Schmidt (2002). In its original form the chain-ladder is a non-stochastic algorithm for producing estimates of outstanding claims. There are many variations of the method and we now provide a description of one them which will be useful in what follows.

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, de Alba (2002b), Hess and Schmidt (2002). The results are applicable to any frequency of claim payments (years, quarters, etc.) and length of pay-off' period.

We use the term claims reserving in its most general sense. Essentially the data would correspond to a typical run-off triangle used in loss reserving. In this paper we use following notation, let Z_{it} = incremental amount of 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: Typical development triangle used in general insurance claims reserving.

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}	-					-

Although the models described in this paper can be applied to more general shapes of claims data, other than a triangle, we assume that we have a conventional triangle of data. Thus, for ease of comparability with other methods, and without loss of generality, assume that the data consists of a triangle of incremental claims: $\{Z_{it} : t = 1,\ldots,k - i + 1; i = 1,\ldots k\}$.

The cumulative claims are defined by $W_{ij} = \sum_{t=1}^{j} Z_{it}$, $i=1,2,...,k$ and the development factors of the chain-ladder technique are denoted by $\{\lambda_j : j = 2,\ldots,k\}$. The estimates of the development factors from the standard chain-ladder technique are

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{k-j+1} W_{ij}}{\sum_{i=1}^{k-j+1} W_{i,j-1}}, \tag{1}$$

see Verrall (1989). These estimates are then applied to the latest cumulative claims available in each row ($W_{i,k-i+1}$) to produce forecasts of future values of cumulative claims:

$$\hat{W}_{i,k-i+2} = W_{i,k-i+1}\hat{\lambda}_{k-i+2} \quad i = 2,\ldots,k, \tag{2}$$

and

$$\hat{W}_{i,j} = \hat{W}_{i,j-1}\hat{\lambda}_j, \; i = 3,\ldots,k, \quad j = k-i+3,\ldots,k. \tag{3}$$

The required reserve for the i^{th} accident year will then be $\hat{R}_i = \hat{W}_i - W_{i,k-i+1}$, and the total reserve is $\hat{R} = \sum_{i=2}^{k} \hat{R}_i$.

Note that the estimators of the development-factors can be written as

$$\hat{\lambda}_j = 1 + \frac{\sum_{i=1}^{k-j+1} Z_{i,j}}{\sum_{i=1}^{k-j+1} W_{i,j-1}} = 1 + r_j \, , \qquad (4)$$

where

$$r_j = \frac{\sum_{i=1}^{k-j+1} Z_{i,j}}{\sum_{i=1}^{k-j+1} W_{i,j-1}} \qquad (5)$$

is the rate of change of the cumulative claims between the development years *(j-1)* and j . Hence we can see that if the numerator is negative and the denominator positive we will have $r_j < 0$ and so $\hat{\lambda}_j < 1$. Furthermore, in the literature on the chain-ladder method, in general it is implicitly assumed that $\hat{\lambda}_j > 0$. Clearly, formulas (1) through (5) can be applied no matter what values we get for r_j and $\hat{\lambda}_j$. However in some cases the resulting estimated reserves may be meaningless. Now, the numerator in (5) will be negative if there are enough (in size and/or number) $Z_{i,j} < 0$ in the j^{th} column.

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. England and Verrall (2002), argue that it is probably better to use, paid claims rather than incurred claims (paid losses and aggregate case reserve estimates combined) since negative values are less likely to appear in the former. That is because case reserve estimates, the amount set aside by the claims handlers (see Chamberlain (1989), or Brown and Gottlieb (2001)) are set individually and often tend to be conservative, resulting in over-estimation in the aggregate. Adjusting for this overestimation in the later stages of development may lead to negative incremental amounts. Whatever their cause, the presence of these negative incremental values in the data may cause problems when applying some claims reserving methods. Thus ideally, before applying claims reserving methods, the actuary will revise and correct the data in order to eliminate negative incremental values. In this respect de Alba and Bonilla (2002) provide a list of potential adjustments frequently used in practice. However, even after correcting the data it is not always possible to eliminate all the negative values. Hence it is convenient to have available claims reserving methods that will allow him/her to compute the necessary reserves even in the presence of the negative values that may remain in the data.

The assumption $W_{i,j} > 0, \quad i = 1,\ldots,k; \;\; j = 1,\ldots,k-i+1$ is implicit in most discussions of the chain-ladder and is sometimes made explicit, Hess and Schmidt (2002), Mack (2004). Under this assumption, the estimates of the development factors will be acceptable even if the sum of the known incremental claims of the j^{th} column is negative, i.e. $\sum_{i=1}^{k-j+1} Z_{i,j} < 0$, for some j's. Thus this method can be applied in the presence of negative incremental claim values. The consequence of which will be to have $\hat{\lambda}_j < 1$ for one or more j's. Other methods are not so 'resistant' and may even break down. In this paper we propose a stochastic model that will yield optimal Bayesian estimates of the reserves for outstanding claims. In particular we are concerned with the situation when there are many negative incremental claims in the development triangle and/or some of these are large.

The remaining of the paper is structured as follows. Section 2 presents stochastic claims reserving models, Bayesian and non-Bayesian, that can be applied when there are negative claim values Section 3 describes a Bayesian model for claim amounts in the presence of negative values. Section 4 describes the prior distributions used in the model. In Section 5 we describe how to use our model to estimate reserves. Its implementation for Markov chain Monte Carlo, as well as an example, is given in Section 6. All models are presented only in discrete time.

2. STOCHASTIC MODELS

Stochastic models to claims reserving improve on the classical approach by allowing the actuary to obtain measures of uncertainty and sometimes the complete distribution of outstanding claims. For a comprehensive, although not exhaustive, review of existing stochastic methods for claims reserving see England and Verrall (2002), or Hess and Schmidt (2002). Most of the methods presented in these references use the point of view of frequentist or classical statistics. Hess and Schmidt (2002) concentrate on stochastic models for the chain-ladder.

Mack (1993) presents one of the earliest attempts at formalizing a stochastic model for claims reserving. He proposes a nonparametric model that reproduces the chain-ladder and obtains distribution-free expressions for the standard of reserve estimates. The use of the model is not limited by the existence of negative incremental claims. What may be considered a limitation is that it is directed at reproducing the chain-ladder reserves. England and Verrall (1999) have proposed the use of bootstraping to compute the prediction errors.

England and Verrall (2002) emphasize the framework of generalized linear models (GLM), Anderson et al. (2004). 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 the models considered by England and Verrall (2002) there are several that can handle negative values: an (over-dispersed) Poisson, a negative binomial and a Normal approximation to the negative binomial. They also mention the log-Normal model which was introduced by Kremer (1982) and analyzed in detail in Verrall (1991) when there are some negative incremental claims.

In the context of generalized linear models (GLM) the first stochastic version of the chain-ladder method that can be applied in the presence of negative incremental claim values is defined as a generalized linear model with an over-dispersed Poisson distribution, Renshaw and Verrall (1998). In the over-dispersed Poisson model the mean and variance are not the same. Using our previous notation $E(Z_{it}) = m_{it}$, with variance function $V(Z_{it}) = \phi m_{it}$ and scale parameter $\phi > 0$, combined with the function $\log(m_{it}) = \mu + \alpha_i + \beta_t$, where α_i and β_t represent row and column effects in the triangle, respectively. Over-dispersion is achieved if $\phi > 1$. The model reproduces the estimates of the classical chain-ladder method. Estimates of the parameters, $\hat{\mu}, \hat{\alpha}_i, \hat{\beta}_t$, are obtained by using a 'quasi-likelihood' approach, Anderson et al. (2004). Renshaw and Verrall (1998) point out that their procedure "is not applicable to all sets of data, and can break down in the presence of a sufficient number of negative incremental claims." Sufficient means that there are enough incremental claims and their values are such that they make $\sum_{i=1}^{k-j+1} Z_{ij} < 0$, for some $j = 1,\ldots, k$. Note that this is the condition for $\hat{\lambda}_j < 1$ in the chain-ladder, see equation (4). They then point out that for the method not to break down it is necessary to have some 'positivity' constraints, essentially that $\sum_{i=1}^{k-j+1} Z_{ij} > 0$ for all $j = 1,\ldots,k$, Verrall (2000). They discuss the relationship between this model and the chain-ladder technique, and show that, under the 'positivity' constraints, the same reserve estimates are produced by both methods.

The negative binomial model is closely related to the Poisson model, 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 $W_{ij} = \sum_{t=1}^{j} Z_{it}$. The parameters $\{\lambda_j : j = 2,\ldots,n\}$ are the typical chain-ladder development factors defined in Section 1. As in the Poisson model, ϕ is an over-dispersion parameter. This model yields essentially the same estimates as the (over-dispersed) Poisson. Again, with enough 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 and the model can not be applied.

A third stochastic model for the chain-ladder, mentioned in England and Verrall (2002) that can be applied in the presence of negative incremental claim values is a Normal approximation to the negative binomial model, under which the chain-ladder results can still be reproduced. The approximation assumes the distribution of incremental claims is Normal with mean $(\lambda_j - 1)W_{i,j-1}$, as in the negative binomial model, but the variance is assumed to be $\phi_j W_{i,j-1}$. This model is seen to be equivalent to one proposed by Mack (1993). In the Normal distribution there is no problem with negative incremental values. However, it is not recommended to use the Normal approximation in all situations, mainly because real claims data are known to be skewed.

The estimate of the variance for the reserves under the Normal approximation can not be obtained from the standard output from statistical packages and must be 'constructed' from part of the computer output and formulas, England and Verrall (2002), which is not immediate. The computation process would require summing terms that involve all the covariances between the parameters and these are not readily available from the statistical packages.

We do not intend to give here an extensive review of the application of Bayesian methods in actuarial science. For general discussion on Bayesian theory and methods see Berger (1985), Bernardo and Smith (1994) or Zellner (1971a).

Bayesian analysis of IBNR reserves has been considered by Jewell (1989,1990), Verrall (1990) and Haastrup and Arjas (1996). See de Alba (2004) for a review. For a discussion of Bayesian methods in actuarial science see Klugman (1992), Makov (1996, 2001), Scollnik (2001, 2002), Ntzoufras and Dellaportas (2002), de Alba (2002b, 2004) and Verrall (2004). Kunkler (2004) provides a Bayesian claims reserving method for the situation when there are zeros in the development triangle, but not negative incremental claim values.

We now refer to one existing Bayesian model, that can be applied to situations where $Z_{it} < 0$ for some $i,t = 1,\ldots,k$. Verrall (2004) presents a Bayesian formulation of the Bornhuetter-Ferguson (B-F) technique. The Bornhuetter-Ferguson is a variant of the chain-ladder that uses external information to obtain an initial estimate for the amount of expected ultimate claims, W_{ik} for each $i = 2,\ldots,k$. This is then combined with the development factors of the chain-ladder technique to estimate outstanding claims, Brown and Gottlieb (2001), Chamberlain (1989). This is clearly well suited for the application of Bayesian methods if the initial information about ultimate claims is given in terms of a prior distribution, Verrall (2004). However, the author points out that the method may break down in the presence of enough negative values (in number and/or size), certainly if any column sum of the incremental claims in the development triangle is negative, i.e. again if $\sum_{i=1}^{k-j+1} Z_{it} < 0$.

We have seen that there are a number of stochastic claims reserving methods that may work in the presence on negative incremental claims, but that they will break down if enough negative values are present. In the next section we propose a Bayesian model that can be applied under very general conditions.

3. 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. 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 claims amounts in the t^{th} development year of accident year i, $i,t=1,\ldots,k$. The Z_{it} are known for $i+t \leq k+1$ and we let

$$Y_{it} = \log(Z_{it} + \delta), \qquad (6)$$

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

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

In our claims reserving problem the threshold parameter $\delta > 0$ adjusts the negative incremental claim values so as to assure $(z_{it} + \delta) > 0$, for $i,t=1,...,k$, with $i+t \le k+1$. We assume in addition that

$$Y_{it} = \log(Z_{it} + \delta) = \mu + \alpha_i + \beta_t + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0, \sigma^2) \qquad (7)$$

$i=1,...,k$, $t=1,...,k$ and $i+t \le k+1$ 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

$$f(z_{it} \mid \mu, \alpha_i, \beta_t, \sigma^2, \delta) = \frac{1}{\sigma(z_{it}+\delta)\sqrt{2\pi}} \exp[-\frac{1}{2\sigma^2}(\log(z_{it}+\delta) - \mu - \alpha_i - \beta_t)^2]. \quad (8)$$

The α_i and the β_t, $i,t = 1,...,k$, represent the accident year (row) and development year (column) effects, respectively. The model in (7) corresponds to an unbalanced two-way analysis of variance (ANOVA) model. 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 $\alpha_1 = \beta_1 = 0$, Verrall (1990).

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

In de Alba (2002a) the threshold parameter δ is first estimated by Maximum Likelihood, "plugged in" to define $y_{it} = \log(z_{it} + \hat{\delta})$, and then the 'profile' likelihood, $L(\boldsymbol{\theta}, \sigma^2 \mid \mathbf{z}, \hat{\delta}) = f(\mathbf{z} \mid \boldsymbol{\theta}, \sigma^2, \hat{\delta})$, is obtained. This is the likelihood function with δ replaced by its ML estimator, say $\hat{\delta}$, Crow and Shimizu (1988, page 123). So the profile likelihood is used instead of the likelihood $L(\boldsymbol{\theta}, \sigma^2, \delta \mid \mathbf{z}) = f(\mathbf{z} \mid \boldsymbol{\theta}, \sigma^2, \delta)$ to carry out the Bayesian analysis, which then is done using results for a two parameter lognormal distribution, Zellner (1971b). The approach followed in de Alba (2002a) has the disadvantage that the variability due to estimating δ is not taken into account in the inference process. This can be potentially troublesome since it is well known that the estimates of δ can be very unstable, see Cohen and Whitten (1980), Johnson et al. (1994, chapter 14) and Hill (1963). Here we use the full likelihood function $L(\boldsymbol{\theta},\sigma^2,\delta \mid \boldsymbol{z})$.

To carry out the Bayesian analysis we must now specify prior distributions for the parameters $\boldsymbol{\theta},\sigma^2$ and δ. We will assume prior independence so that

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

This assumption can be justified by considering that, a priori, we have no reason to believe that μ,σ^2 and δ are influenced by the effects of the development years (the β_t) or by those of the accident years (the α_i), and vice versa, i.e. they are statistically independent. Furthermore we can assume, a priori, that the effects of the development years (the β_t) are not influenced by those of the accident years (the α_i), and vice versa. So the α_i can be assumed independent of the β_t. A standard assumption in claims reserving models is that accident years are independent, so this justifies the assumption of prior mutual independence of the α_i. Even if there is some dependence among the parameters, as may be the case with the development year effects (the β_t), and it is not modeled explicitly a priori so initially we assume they are independent, it will be reflected in the posterior distribution since it will be introduced by the sample data through the likelihood function. Of course it is possible to consider other prior models that include explicitly potential dependencies, if there is enough prior information to specify them. We will assume independence to simplify the exposition.

The posterior distribution will be obtained from

$$f(\boldsymbol{\theta},\sigma^2,\delta \mid \boldsymbol{z}) \propto f(\boldsymbol{z} \mid \boldsymbol{\theta},\sigma^2,\delta)\times f(\mu)\times[\prod_{i=2}^{k} f(\alpha_i)][\prod_{t=2}^{k} f(\beta_t)]\times f(\sigma^2)\times f(\delta). \quad (9)$$

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 $f(\delta)$, $i,t = 2,\dots,k$. 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) or OpenBUGS (at http://mathstat.helsinki.fi/openbugs/). In the next section we describe the specification of the prior distributions.

4. PRIOR DISTRIBUTIONS

Specification of a prior distribution can be made by using any distribution that is reasonable according to the characteristics of the parameter. When there is no agreement on the prior information, or even when there is a total lack of it, we can use what are known as non-informative or reference priors; i.e. the prior distribution $\pi(\boldsymbol{\theta})$ will be chosen to reflect our state of ignorance. Inference under these circumstances is known as objective Bayesian inference. Here, for most of the parameters, we specify what are known as natural conjugate prior distributions. They have the property that the prior and posterior distributions belong to the same family of distributions, Bernardo and Smith (1994). They are easy to use because they lead to well known posterior distributions. Conjugate priors can also be specified so as to reflect lack of information, by assigning specific values to the parameters. In addition, most of the more common conjugate distributions are included in WinBUGS.

Given $\alpha_i, \beta_t, \sigma^2$ and δ, 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 $\alpha_i \sim N(\mu_{\alpha_i}, \sigma^2_{\alpha_i})$ and $\beta_t \sim N(\mu_{\beta_t}, \sigma^2_{\beta_t})$, 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,

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

Bernardo and Smith (1994), denoted by $\sigma^2 \sim GI(\nu, \lambda)$. 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, given the rest of the parameters, μ, α_i, β_t and σ^2. We do not have a natural conjugate prior in this case. Here we specify a Normal prior distribution $\delta \sim N(\mu_\delta, {\sigma_\delta}^2)$.

5. ESTIMATING THE RESERVES

We want to estimate (or obtain the distribution of) aggregate claims for the i^{th} accident year, $i = 2,\ldots,k,$ given the information available in the development triangle. Recall that the cumulative claims are given by $W_{ij} = \sum_{t=1}^{j} Z_{it}$ for $1 \le j \le k$. Hence, in the run-off triangle setup, we are really interested in estimating W_{ik} $i=2,\ldots,k$, given Z_{it}, $i=1,\ldots,k$

$t=1,...,k$, with $i+t \le k+1$. Now let $R_i = W_{ik} - W_{i,k-i+1}$, for $i=2,...,k$ where $W_{i,k-i+1}$ is the accumulation of Z_{it} up to the latest development period and R_i = the total aggregate outstanding claims corresponding to business year i, for $i=2,...,k$. From the distribution of R_i we can obtain the required reserves corresponding to this business year by choosing the measure of central tendency or quantile we want in the distribution of outstanding claims. Finally, from the distribution of total aggregate outstanding claims $R = \sum_{i=2}^{k} R_i$ we can compute the required total reserves.

In the Bayesian approach, when interest is on prediction, as in loss reserving, the past (known) data in the upper triangle, $\boldsymbol{z}$, are used to predict the observations in the lower triangle by means of the posterior predictive distribution for outstanding claims in each cell:

$$f(z_{it} \mid \boldsymbol{z}) = \int f(z_{it} \mid \boldsymbol{\theta}, \sigma^2, \delta) f(\boldsymbol{\theta}, \sigma^2, \delta \mid \boldsymbol{z}) d\boldsymbol{\theta} d\sigma^2 d\delta \,,$$
$$i = 1,...,k, \quad t = 1,...,k, \; with \quad i + t > k + 1 \,.$$

The reserves for the outstanding aggregate claims are estimated as the mean of the predictive distribution. Hence for each cell we must obtain $E(Z_{it} \mid \boldsymbol{z})$. Then the Bayesian estimate of the total outstanding claims for year of business i is $\sum_{t>k-i+1} E(Z_{it} \mid \boldsymbol{z})$. The Bayesian 'estimator' of the variance of outstanding claims (the predictive variance) for that same year is too cumbersome to derive analytically. 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 (9) and then in turn use the resulting values of the parameters in $f(z_{it} \mid \boldsymbol{\theta}, \sigma^2, \delta)$ to generate random values of Z_{it}. This yields random observations for aggregate claims in each cell of the (unobserved) lower right triangle $z_{it}^{(j)}$, $i = 2,.., k$, $t > k - i + 1$, 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 outstanding claims $R^{(j)} = \sum_{i,t} Z_{it}^{(j)}$. The mean and variance can be computed as

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

The standard deviation σ_R thus obtained is an 'estimate' for the prediction error of the 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. However, direct simulation may be very cumbersome to do, if not impossible, when the posterior distribution is not of a known type. It is for this kind of situation where MCMC proves very useful. The MCMC methodology has the added advantage that it is possible to obtain the mean, variance and any quantiles for the predictive distribution of the outstanding claims directly. We illustrate this in the next section with an application.

6. 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.

Table 2: General insurance paid claims from an anonymous insurance company

	1	2	3	4	5	6	7	8	9
1	33250.717	2097.059	78.897	21.117	-18.65	-0.121	-5.072	-1.292	-0.78
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	12.87			
5	35980.821	2892.427	52.478	10.982	-3.496				
6	37518.185	2901.650	-23.61	-39.496					
7	40213.152	3006.438	-14.59						
8	39105.807	3080.126							
9	41184.755								

We compare the results of applying the chain-ladder method; the Normal approximation; the Bayesian method using a 'profile' likelihood with δ replaced with its ML estimate, mentioned in Section 3; and our Bayesian model presented above with MCMC simulation.

When applying MCMC the parameters in the third stage were chosen to reflect lack of information. Those for σ^2_{μ}, $\sigma^2_{\alpha_i}$ $\sigma^2_{\beta_t}$ and σ^2_{δ} 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, $\hat{\delta} = 300.2$, as in de Alba (2002a). Finally, the distribution for δ was set so that a priori δ has a large variance and hence it allows the random values generated in the MCMC simulation to cover a broad range. Its prior expected value is chosen to be between zero and its MLE $\hat{\delta}$.

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, and then 'thinned out' to one out of ten observations to reduce autocorrelations and obtained the results shown below. Various checks were made of the convergence of the Markov chain using BOA, Smith (2005), as well as visual inspection of the sampled values.

Table 3. MCMC estimates of parameters for the log-normal distribution

Parameter	Mean	Std. Dev.	Percentiles		
			2.50%	Median	97.50%
δ	106.40	17.15	81.56	101.20	155.40
μ	10.53	0.0719	10.39	10.53	10.67
$\tau=1/\sigma^2$	26.41	10.57	11.80	24.43	54.62

We obtain some characteristics of the posterior distribution of the parameters δ, μ and the precision $\tau = 1/\sigma^2$. The results are shown in Table 3. The posterior mean of δ is 106.4 while its MLE is $\hat{\delta} = 300.2$. Notice that the MLE is out on the right tail of the posterior distribution, and beyond the 97.5% quantile. Thus if we had this information before trying to apply the Bayesian method with the profile likelihood, this would be a very un likely value to use to 'plug in'. Other options for the prior distribution of δ and this was always the case.

Table 4. Estimation of reserves by accident year using different methods: the chain-ladder; the GLM with a normal approximation and quasi-likelihood estimation; the Bayesian method with profile likelihood; and the Bayesian model with MCMC.

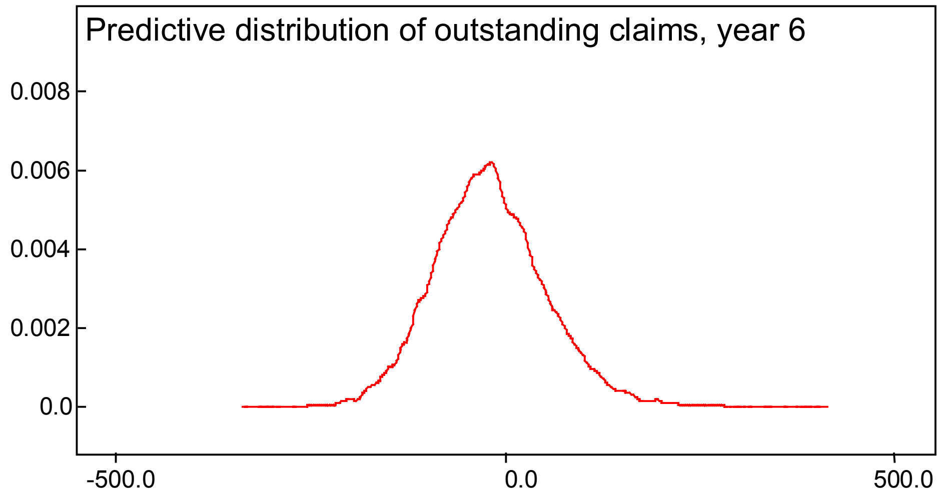
Accident Year	Chain-Ladder	Normal Approximation	Bayesian method with profile likelihood		Bayesian method using MCMC	
	Reserves	Reserves	Reserves	Std. Dev.	Reserves	Std. Dev.
2	-0.86	1.00	1.38	32.25	3.27	34.04
3	-0.91	2.00	21.35	47.19	12.16	45.84
4	-6.60	3.00	5.81	57.88	7.00	52.57
5	-6.02	7.00	64.14	73.00	22.78	63.70
6	-8.72	-12.00	-54.89	81.64	-16.53	68.51
7	-8.82	22.00	77.69	107.55	7.02	75.79
8	9.51	48.00	244.08	148.15	37.71	88.45
9	3041.18	3085.00	3363.20	514.50	2859.00	718.30
TOTAL	3018.77	3182.00	3722.70	620.50	2933.00	768.40

In this example the reserves in both Bayesian methods are not very close. They are given in Table 4. The chain-ladder development factors can be computed without any problem and applied to obtain the reserves, since all the cumulative claims values are positive. Notice that the reserves are negative for all but the last two accident years. This is a result of the fact that three of the development factors turn out to be less than one. In this case the estimates of the reserves obtained with the Bayesian method presented here are lower than those of the non-Bayesian ones and than the others. The variance of the MCMC estimates is larger than those from the direct simulation with the profile likelihood for accident years 2 and 9, and for the total. In the other accident years they are smaller. Apparently allowing δ to vary does not have a systematic effect on the variances. But in any case they are more realistic than when fixing δ as equal to its MLE. The non-informative priors used in the third stage of the hierarchical model seem to allow a better fit to the data. These prediction errors are comparable to the standard deviation of the predictive distribution in the Bayesian models.

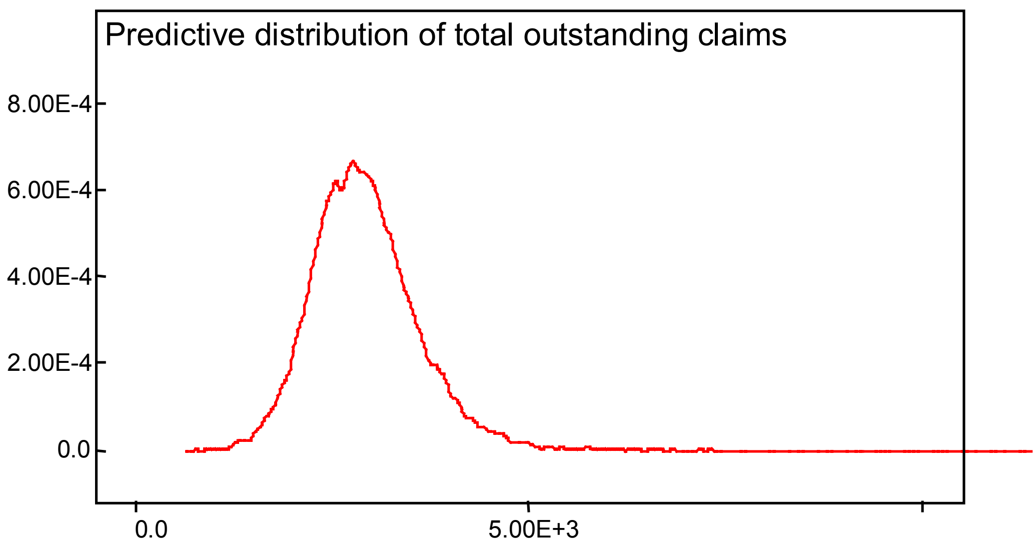
Notice that the reserves for accident year 6 are negative under all methods. So we will plot the complete distribution to see the behavior of the outstanding claims for this year. Figure 1 shows the predictive distribution for accident year 6 (panel a)) and for the total (panel b)). The distribution for reserves corresponding to accident year 6 has a long right tail and yet shows a large probability of negative values. In fact the predictive mean is close to zero. This explains why the reserves are negative. The results for accident years 2 and 4 (not shown) also yield medians very close to zero, but even so the predictive mean is positive. The predictive distribution of total reserves is symmetric but is clearly not Normal, as can be verified by comparing the percentiles with those of the Normal. The 95% predictive interval obtained from the MCMC results is given by (1653,4670) whereas the Normal interval would be (1595,4255). The MCMC predictive distribution has heavier tails, in particular the right one. Hence if one uses a Normal approximation, because the distribution looks symmetric, it is possible that for some accident years there may be a high probability that claims result much larger that the reserves. There is also a probability larger than .05 that claims are larger than the reserves if the chain-ladder method is used, although the total reserves are fairly close: 3018.77 for the chain-ladder and 2933.0 for the Bayesian model.

7. CONCLUDING REMARKS

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 results that are comparable to those of other methods, but does not have their limitations. 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 enough negative values (in number and/or size), even if these make the total sum of one or more of the columns in the development triangle of incremental claims be negative, and so most other methods break down. One other model that does not break down is the one proposed by Mack (1993), but its purpose is to provide a stochastic model that reproduces chain-ladder reserve estimates. The main advantages, for practitioners, of the method presented in this paper are that it yields the complete distribution of outstanding claims and it works when most other methods break down. Methods that do not break down assume Normality, which may not be adequate, since incremental claims are known to have skewed distributions. Or else they reproduce chain-ladder estimates, which in some cases may not be desirable.



a) claims in year 6



b) total claims

Figure 1. Predictive distribution for outstanding claims in accident year 6 and the total, obtained using MCMC with 10,000 simulations.

REFERENCES

(1) Anderson, D., Feldblum S., Modlin, C., Schirmacher, D., Schirmacher, E. and Thandi, N. (2004), *A Practitioner's Guide to Generalized Linear Models*, 2004 Discussion Paper Program, Casualty Actuarial Society.

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

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

(4) Brown, R.L. and Gottlieb, L.R. (2001), *Introduction to Ratemaking and Loss Reserving for Property and Casualty Insurance*, 2nd Ed., ACTEX Publications.

(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) Chamberlain, G.F. (ed.) (1989), *Claims Reserving Manual*, Vol. 1, Institute of Actuaries

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

(8) 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

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

(10) de Alba, E. (2004), Bayesian Claims Reserving, *Encyclopedia of Actuarial Science,* John Wiley and Sons, Ltd., London,

(11) de Alba, E. and R. Bonilla (2002), Un Modelo Para el Tratamiento de Valores Negativos en el Triángulo de Desarrollo Utilizado en la Estimación de Reservas para SONR, *Transactions of the 27th International Congress of Actuaries*, Cancún, México.

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

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

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

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

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

(17) 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

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

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

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

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

(22) Kremer, (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, 33-35.

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

(25) Mack, T. (2004), Chain-ladder Method, *Encyclopedia of Actuarial Science,* John Wiley and Sons, Ltd., London.

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

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

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

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

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

(31) Scollnik, D.P.M. (2002), Implementation of Four Models for Outstanding Liabilities in WinBUGS: A Discussion of a Paper by Ntzoufras and Dellaportas, *North American Actuarial Journal* 6(1), 128-136.

(32) Smith, B. (2005), Bayesian Output Analysis Program (BOA), Version 1.1.5 for R, http://www.public-health.uiowa.edu/boa/

(33) 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/

(34) 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*.

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

(36) Thomas, A. (2005), OpenBUGS, http://mathstat.helsinki.fi/openbugs/

(37) Verrall, R. J. (ed.) (1989), *Claims Reserving Manual*, Vol. 2, Institute of Actuaries.

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

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

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

(41) Verrall, R.J. (2004), A Bayesian Generalised Linear Model for the Bornhuetter-Ferguson Method of Claims Reserving, *North American Actuarial Journal* 8(3), 67-89.

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

(43) 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