A BAYESIAN GENERALIZED LINEAR MODEL FOR THE BORNHUETTER-FERGUSON METHOD OF CLAIMS RESERVING

R. J. Verrall*

ABSTRACT

This paper shows how Bayesian models within the framework of generalized linear models can be applied to claims reserving. The author demonstrates that this approach is closely related to the Bornhuetter-Ferguson technique. Benktander (1976) and Mack (2000) previously studied the Bornhuetter-Ferguson technique and advocated using credibility models. The present paper uses a Bayesian parametric model within the framework of generalized linear models.

1. INTRODUCTION

The Bornhuetter-Ferguson method (Bornhuetter and Ferguson 1972) has proved useful for certain classes of general insurance business. In particular, when the data are very unstable, a method such as the chain-ladder technique can produce unsatisfactory results. To stabilize the results, the Bornhuetter-Ferguson method uses an external initial estimate of ultimate claims. This is then used with the development factors of the chain-ladder technique, or something similar, to estimate outstanding claims. This method has been investigated by a number of authors, and the recent paper by Mack (2000) provides an excellent summary of this work. Mack gives details of a similar approach to that advocated in this paper, using a credibility theory approach first suggested by Benktander (1976) and, hence, called Benktander's method.

Some say the Bornhuetter-Ferguson method has some similarities to a Bayesian procedure, in particular because of the initial estimate of ultimate claims, which is supplied as prior information. However, it was not defined as a Bayesian method, and makes no probabilistic statements about the data or any prior distribution. The purpose of this paper is to attempt to connect the Bornhuetter-Ferguson method with generalized linear models by applying Bayesian estimation. The present paper is based very much on generalized linear models, and the theory in this paper is not applicable to all sets of data (in particular, it may break down for negative incremental claims). A review of stochastic reserving in general insurance is given in England and Verrall (2002) and Taylor (2000), and other papers using a Bayesian approach include Ntzoufras and Dellaportas (2002) and de Alba (2002).

This paper shows how the Bornhuetter-Ferguson method is related to the *generalized linear models* approach to claims reserving, using a Bayesian approach. The advantages of this are as follows:

- It provides further help for the actuary to understand what assumptions are made when the Bornhuetter-Ferguson method is used.
- It clarifies the connection between the Bornhuetter-Ferguson method and the chain-ladder technique, particularly for the case when generalized linear models are used.
- It allows mean square prediction errors to be calculated and indicates how the Bornhuetter-Ferguson method could be incorporated into a Dynamic Financial Analysis (DFA) exercise.
- It provides a range of Bayesian generalized linear models, which could be used in a claims-reserving

---

* R. J. Verrall, Ph.D., is Associate Dean and Head of Faculty at the Faculty of Actuarial Science and Statistics, Cass Business School, City University, 106 Bunhill Row, London EC1Y 8T2, email: r.j.verrall@city.ac.uk.

10exercise, of which the Bornhuetter-Ferguson method and the chain-ladder technique are special cases.

The approach taken in this paper to the Bornhuetter-Ferguson method is based on the approach of Verrall (2000), in which it was shown that the chain-ladder technique could be expressed in a number of different ways as stochastic models. This paper uses the same framework and examines the Bornhuetter-Ferguson method, giving a number of insights into this method. The stochastic model developed in Section 5 encompasses the Bornhuetter-Ferguson method, as it is usually implemented by actuaries, as a special case.

The paper is set out as follows. Section 2 describes briefly the Bornhuetter-Ferguson method, as currently used by actuaries. Section 3 defines a Bayesian framework for the over-dispersed Poisson model. In Section 4, the relationship between Bayesian models, the Bornhuetter-Ferguson method, and the chain-ladder technique is examined. Section 5 considers the estimation of the parameters and defines a Bayesian model that has the Bornhuetter-Ferguson model as a special case. In Section 6, I show how the parameters can be estimated using WinBUGS. There are two appendices. Appendix A contains the proofs of the results stated in the paper, and Appendix B contains the code for implementing the Bayesian models in WinBUGS.

2. THE BORNHUETTER-FERGUSON METHOD

Without loss of generality, we assume that the data consist of a triangle of incremental claims:

$$\begin{array}{llll} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & \cdots & C_{2,n-1} & \\ \vdots & & & \\ C_{n1} & & & \end{array}$$

This can be also written as $\{C_{ij} : j = 1, \ldots, n - i + 1; i = 1, \ldots, n\}$, where n is the number of accident years. It should be emphasized that the assumption that the data consist of a triangle is made so the notation will not get too complicated, and the methods can also be applied to other shapes of data.

Corresponding with the triangle of incremental data is a triangle of cumulative claims, defined by:

$$D_{ij} = \sum_{k=1}^{j} C_{ik}.$$

The development factors of the chain-ladder technique are denoted by $\{\lambda_j : j = 2, \ldots, n\}$, and the estimates of the development factor from the standard chain-ladder technique are

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} D_{ij}}{\sum_{i=1}^{n-j+1} D_{i,j-1}}.$$

No tail factors are applied, and claims are only forecast up to the latest development year (n) so far observed. It would be possible to extend this to allow a tail factor, using the same methods, but no specific modeling is carried out of the shape of the runoff beyond the latest development year. Thus, “ultimate claims” refers to

$$D_{in} = \sum_{k=1}^{n} C_{ik}.$$

The Bornhuetter-Ferguson method can be summarized as follows.

1. Obtain an initial estimate of ultimate claims, D_{in}, for each accident year, i.
2. Estimate the proportion of ultimate claims that are outstanding for each accident year, using, for example, the chain-ladder technique.
3. Apply the proportion from No. 2 to the initial estimate of ultimate claims from No. 1 to obtain the estimate of outstanding claims.

The usual way of expressing this is as follows. Let the initial estimate of ultimate claims for accident year i be M_i, for example taken from the premium calculation. The estimate of outstanding claims for accident year i is

$$M_i\left(1 - \frac{1}{\lambda_{n-i+2}\lambda_{n-i+3}\cdots\lambda_n}\right) = M_i \frac{1}{\lambda_{n-i+2}\lambda_{n-i+3}\cdots\lambda_n}(\lambda_{n-i+2}\lambda_{n-i+3}\cdots\lambda_n - 1).$$

Thus,

$$M_i \frac{1}{\lambda_{n-i+2}\lambda_{n-i+3}\cdots\lambda_n}$$

replaces the latest cumulative claims for accident year i, to which the usual chain-ladder parameters are applied to obtain the estimate of outstanding claims. For the chain-ladder technique, the estimate of outstanding claims is $D_{i,n-i+1}(\lambda_{n-i+2}\lambda_{n-i+3}\cdots\lambda_n - 1)$.

Thus, it can be seen that the difference between the Bornhuetter-Ferguson method and the chain-ladder technique is the factor that is used to multiply the development factors. For the chain-ladder technique, this is $D_{i,n-i+1}$, and for the Bornhuetter-Ferguson method, this is

$$M_i \frac{1}{\lambda_{n-i+2}\lambda_{n-i+3}\cdots\lambda_n}.$$

3. A BAYESIAN FRAMEWORK FOR THE OVER-DISPERSED POISSON MODEL

This section defines a Bayesian generalized linear model using the over-dispersed Poisson model, defined by Renshaw and Verrall (1998). Renshaw and Verrall discussed the relationship between this model and the chain-ladder technique, and showed that, under certain positivity constraints, the same reserve estimates are produced by each.

The term "over-dispersed" requires some explanation. It is used here in connection with the Poisson distribution, and it means that if $X \sim$ Poisson (μ), then $Y = \varphi X$ follows the over-dispersed Poisson distribution, with $E(Y) = \varphi\mu$ and $V(Y) = \varphi E(X) = \varphi^2\mu$. φ is usually greater than 1—hence, the term "over-dispersed"—but this is not a necessity. It can also be used for other distributions, and we make use of it for the negative binomial distribution. As with the Poisson distribution, the over-dispersed negative binomial distribution is defined such that if $X \sim$ negative binomial then $Y = \varphi X$ follows the over-dispersed negative binomial distribution. Furthermore, a quasi-likelihood approach is taken so that the claims data are not restricted to the positive integers.

The over-dispersed Poisson model for the chain-ladder technique can be written as follows:

$$C_{ij}|x, y, \varphi \sim \text{independent over-dispersed Poisson, with mean } x_i y_j \text{, and } \sum_{k=1}^{n} y_k = 1. \tag{3.1}$$

Here $x = \{x_1, x_2, \ldots, x_n\}$ and $y = \{y_1, y_2, \ldots, y_n\}$ are parameter vectors relating to the rows (accident years) and columns (development years), respectively, of the runoff triangle. The parameter $x_i = E[D_{in}]$, and so represents expected ultimate cumulative claims (up to the latest development year so far observed, n) for the i-th accident year. The column parameters, y_j, can be interpreted as the proportions of ultimate claims that emerge in each development year. This model applies to all the data, both observed and future observations. The estimation is based on the observed data, and we require predictive distributions for the future observation.

In the chain-ladder model, no prior assumptions are made about the row parameters, $\{x_i : i = 1, \ldots, n\}$. The key assumption of the Bornhuetter-Ferguson method is that there is prior knowledge about these parameters. For this reason, we use a Bayesian approach and allow the inclusion of prior information about the row parameters, which can be summarized in the following prior distributions:

$$x_i|\alpha_i,\ \beta_i \sim \text{independent } \Gamma(\alpha_i,\ \beta_i).$$

The column parameters, y_j, and the over-dispersion parameter, φ, also have to be estimated, and there are a number of possibilities. We use a full Bayesian model for the column parameters, giving them improper prior distributions. It is also possible to use "plug-in" estimates, which could be derived from elsewhere. This is closest to what happens when the Bornhuetter-Ferguson technique is applied and so we consider this in Section 4 for comparison purposes. The use of plug-in estimates in the Bornhuetter-Ferguson technique can be seen in Section 2, where the development factors, λ_j, are simply those from the chain-ladder technique. Section 5 gives more detail of the estimation of the column parameters. It is shown there that, in order to follow the Bornhuetter-Ferguson technique, the column parameters must be estimated first, before prior distributions are applied to the row parameters.

Finally, we consider the nuisance parameter φ. For this we use a plug-in estimate, in line with the approach taken in classical methods (in England and Verrall 2002, for example). The value used is that obtained from the straightforward application of the over-dispersed Poisson model, estimating the row and column parameters using maximum likelihood estimation (we use S-Plus). A fully Bayesian approach would assign a prior distribution for φ as well, and integrate it out.

The posterior distribution for the row and column parameters is obtained via Bayes theorem:

$$f(x, y|C_{ij},\ i = 1, \ldots, n, j = 1, \ldots, n - i + 1,\ \varphi) \propto \prod_{i=1}^{n} \prod_{j=1}^{n-i+1} f(C_{ij}|x, y, \varphi) \prod_{i=1}^{n} f(x_i)f(y_i). \quad (3.2)$$

For the full Bayesian model, we should also include a prior distribution for φ. For the future observations, we require the predictive distributions:

$$f(C_{kl}|C_{ij},\ i = 1, \ldots, n, j = 1, \ldots, n - i + 1,\ \varphi) = \iint f(C_{kl}|x, y, \varphi)\, f(x, y|C_{ij},\ i = 1, \ldots, n, j = 1, \ldots, n - i + 1,\ \varphi)\, dx\, dy. \quad (3.3)$$

The Bayesian models can be implemented (and, indeed, are probably best implemented) using a Markov chain Monte Carlo (MCMC) approach, through the software WinBUGS (Spiegelhalter et al. 1996). The advantages of using the simulation-based MCMC approach are discussed in de Alba (2002). Although it is possible to make progress analytically, we have found that the simulation approach is usually easier to implement, particularly when considering the predictive distribution of quantities such as the overall reserve. Illustrations of the implementation of the Bayesian models are given in Section 6. However, before looking at the estimation of the parameters, it is instructive to examine the relationship between the Bornhuetter-Ferguson technique, the chain-ladder technique, and the Bayesian model, and this is done in Section 4. In Section 5 we consider the estimation of the parameters and define a Bayesian model for the Bornhuetter-Ferguson method.

4. THE RELATIONSHIP BETWEEN THE BORNHUETTER-FERGUSON METHOD AND THE CHAIN-LADDER TECHNIQUE

This section examines the relationship between the Bayesian model defined in Section 3, the Bornhuetter-Ferguson technique, and the chain-ladder technique. To do this, consider again equation (3.3);

$$
\begin{aligned}
&f(C_{kl}|C_{ij},\, i = 1, \ldots, n, j = 1, \ldots, n - i + 1,\, \varphi)\\
&\qquad = \iint f(C_{kl}|x, y, \varphi) f(x, y|C_{ij},\, i = 1, \ldots, n, j = 1, \ldots, n - i + 1,\, \varphi)\, dx\, dy\\
&\qquad = \int \left(\int f(C_{kl}|x, y, \varphi) f(x|C_{ij},\, i = 1, \ldots, n, j = 1, \ldots, n - i + 1, y, \varphi)\, dx \right) f(y|C_{ij},\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1, \ldots, n, j = 1, \ldots, n - i + 1,\, \varphi)\, dy.
\end{aligned}
$$

The key to comparing the methods is to look at the form of

$$\int f(C_{kl}|x, y, \varphi) f(x|C_{ij},\, i = 1, \ldots, n, j = 1, \ldots, n - i + 1, y, \varphi)\, dx.$$

Note that this is the predictive distribution, assuming that the row parameters have been estimated, but that the column parameters and φ have not yet been estimated:

$$
\begin{aligned}
&\int f(C_{kl}|x, y, \varphi) f(x|C_{ij},\, i = 1, \ldots, n, j = 1, \ldots, n - i + 1, y, \varphi)\, dx\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = f(C_{ij}|C_{i1}, C_{i2}, \ldots, C_{i,j-1}, y, \varphi).
\end{aligned}
$$

To compare the Bayesian model with the Bornhuetter-Ferguson technique and the chain-ladder technique, and to gain understanding of the Bornhuetter-Ferguson method, we derive the distribution of $C_{ij}|C_{i1}, C_{i2}, \ldots, C_{i,j-1}, y, \varphi$.

We only need to use the data in row i when considering C_{ij}, since we are only estimating x_i. In Appendix A, it is shown that the distribution of $C_{ij}|C_{i1}, C_{i2}, \ldots, C_{i,j-1}, y, \varphi$ is an over-dispersed negative binomial distribution with parameters $k = \alpha_i + (D_{i,j-1}/\varphi)$ and $p = (\beta_i\varphi + S_{j-1})/(\beta_i\varphi + S_j)$, where $S_j = \sum_{k=1}^{j} y_k$.

Verrall (2000) showed that the distribution of $C_{ij}|C_{i1}, C_{i2}, \ldots, C_{i,j-1}, y, \varphi$ under the chain ladder model is also a negative binomial distribution, and we can make comparisons by looking at the means and variances. In particular, the means show clearly the differences in the assumptions made by the two approaches. For the chain-ladder technique, the mean of the distribution of $C_{ij}|C_{i1}, C_{i2}, \ldots, C_{i,j-1}, y, \varphi$ is $D_{i,j-1}y_j/S_{j-1}$. Note that, since $\lambda_j = S_j/S_{j-1}$, this can also be written as $(\lambda_j - 1)D_{i,j-1}$. Also, since $D_{i,j} = D_{i,j-1} + C_{i,j}$, the mean of the predictive distribution for cumulative claims is $\lambda_j D_{i,j-1}$.

The mean of the distribution of $C_{ij}|C_{i1}, C_{i2}, \ldots, C_{i,j-1}, y, \varphi$ for the Bayesian method is

$$
\begin{aligned}
\varphi\, \frac{\left(\alpha_i + \dfrac{D_{i,j-1}}{\varphi}\right) \dfrac{y_j}{\beta_i\varphi + S_j}}{\dfrac{\beta_i\varphi + S_{j-1}}{\beta_i\varphi + S_j}} &= \varphi\, \frac{\left(\alpha_i + \dfrac{D_{i,j-1}}{\varphi}\right) y_j}{\beta_i\varphi + S_{j-1}} = \left(\frac{S_{j-1}}{\beta_i\varphi + S_{j-1}} \frac{D_{i,j-1}}{S_{j-1}} + \frac{\beta_i\varphi}{\beta_i\varphi + S_{j-1}} \frac{\alpha_i}{\beta_i} \right) y_j\\
&= \left(Z_{ij} \frac{D_{i,j-1}}{S_{j-1}} + (1 - Z_{ij}) \frac{\alpha_i}{\beta_i} \right) y_j \quad \text{where} \quad Z_{ij} = \frac{S_{j-1}}{\beta_i\varphi + S_{j-1}}.
\end{aligned}
$$

It can be seen that this is in the form of what actuaries call "a credibility formula." In modern statistical terms, it is a natural trade-off between two competing estimates for the row parameter. Note that y_j is the proportion of ultimate claims that emerge in development year j. This is then multiplied by the prior mean of the ultimate claims, α_i/β_i, for the Bornhuetter-Ferguson method, or an estimate of ultimate claims from the data, $D_{i,j-1}/S_{j-1}$, for the chain-ladder technique. We have here a combination of these two, with the chain-ladder as one extreme (which assumes there is no prior information about the row parameters), and the Bornhuetter-Ferguson method as the other extreme (exact prior information about the row parameters, so that the variance of the prior distribution is 0). It is interesting to note that the Bornhuetter-Ferguson method does not use the data at all for the estimation of the row parameters.

The mean of the distribution of $C_{ij}|C_{i1}, C_{i2}, \ldots, C_{i,j-1}, y, \varphi$ for the Bayesian method can also be written as

$$\left(Z_{ij}D_{i,j-1} + (1 - Z_{ij})\frac{\alpha_i}{\beta_i}S_{j-1}\right)\frac{y_j}{S_{j-1}}.$$

Since $\lambda_j = S_j/S_{j-1}$ and $S_n = 1$, $\lambda_j\lambda_{j+1}\cdots\lambda_n = (S_j/S_{j-1})(S_{j+1}/S_j)\cdots(S_n/S_{n-1}) = 1/S_{j-1}$. Hence, we may also write this mean as

$$\left(Z_{ij}D_{i,j-1} + (1 - Z_{ij})\frac{\alpha_i}{\beta_i}\frac{1}{\lambda_j\lambda_{j+1}\cdots\lambda_n}\right)\frac{y_j}{S_{j-1}} = \left(Z_{ij}D_{i,j-1} + (1 - Z_{ij})\frac{\alpha_i}{\beta_i}\frac{1}{\lambda_j\lambda_{j+1}\cdots\lambda_n}\right)(\lambda_j - 1).$$

It can then be seen that the two values used for the row parameter are the equivalent of those in Section 2:

$$D_{i,j-1} \text{ and } \frac{\alpha_i}{\beta_i}\frac{1}{\lambda_j\lambda_{j+1}\cdots\lambda_n} = M_i\frac{1}{\lambda_j\lambda_{j+1}\cdots\lambda_n}.$$

The credibility factor, $Z_{ij} = S_{j-1}/(\beta_i\varphi + S_{j-1})$, governs the trade-off between the prior mean and the data. Notice that the farther through the development we are, the larger S_{j-1} is, and the more weight is given to the chain-ladder estimate. The choice of β_i is governed by the prior precision of the initial estimate for ultimate claims, and this should be chosen with due regard given to the over-dispersion parameter (an initial estimate of which could be obtained from the over-dispersed Poisson model of Renshaw and Verrall 1998).

This section has looked at the distribution of $C_{ij}|C_{i1}, C_{i2}, \ldots, C_{i,j-1}, y, \varphi$ in order to make comparisons between the Bayesian method and the Bornhuetter-Ferguson and chain-ladder techniques. It has been shown that different specifications of the prior distributions can be used for the chain-ladder technique, the Bornhuetter-Ferguson method, or for a complete spectrum of methods between these two extremes. We have yet to consider the estimation of the column parameters, other than to point out that the Bornhuetter-Ferguson method, being deterministic, simply plugs in the chain-ladder parameter estimates. Section 5 considers this issue in more detail and defines a Bayesian approach to the Bornhuetter-Ferguson method.

5. ESTIMATION OF THE COLUMN PARAMETERS

In Section 3, the Bayesian framework for the over-dispersed Poisson model was defined as follows:

$$C_{ij}|x, y, \varphi \sim \text{independent over-dispersed Poisson, with mean } x_i y_j, \text{ and } \sum_{k=1}^{n} y_k = 1, \tag{5.1}$$

and

$$x_i|\alpha_i, \beta_i \sim \text{independent } \Gamma(\alpha_i, \beta_i). \tag{5.2}$$

Section 4 discussed the specification of the prior distributions for the row parameters, with particular reference to the chain-ladder technique and the Bornhuetter-Ferguson method. This section considers the estimation of the column parameters.

One option is simply to use plug-in estimates, obtained, for example, from the straightforward chain-ladder technique. This is the approach used in the deterministic application of the Bornhuetter-Ferguson method, but it is not suitable here since we would prefer a stochastic approach.

A second option is to define improper prior distributions for the column parameters and estimate all the parameters together. This option is illustrated in Section 6 in order to show the properties of the estimators when this approach is taken. If all the parameters are estimated together, the prior distributions used will affect the estimates for *all* parameters. In other words, if informative prior distributions are used for the row parameters, the estimates of the column parameters will be affected, and will no longer be those implied by the chain-ladder technique.

It may be that this is appropriate, and the effect is illustrated in Section 6, but it is not what happens in the Bornhuetter-Ferguson method. There, the chain-ladder parameter estimates are used and, for this reason, we show how this can be done using a Bayesian model. To do this, we estimate the column parameters *first,* before applying prior distributions for the row parameters and estimating these. It is not required to include any information about the column parameters; hence, informative priors are not used. We use improper gamma distributions for the column parameters and derive the posterior distributions of these using a standard prior-posterior analysis.

Note that we have two different ways to obtain the posterior distributions, which will (in most cases) give different results. The first is to apply prior distributions for the row and column parameters and obtain the posterior distributions simultaneously. The second is to first apply prior distributions to the column parameters, obtain estimators of these, and then apply the prior distributions for the row parameters. The advantage of the second approach for the purposes of this paper is that the column parameter estimators are fixed to be the same as those implied by the chain-ladder technique (since we always use improper prior distributions for the column parameters).

I now show how to apply improper prior distributions to the column parameters in the over-dispersed Poisson model. The easiest way to do this is to replace the constraint that the column parameters sum to 1 by the alternative constraint that the row parameters sum to 1. It should be noted that the constraint in equation (3.1), $\sum_{k=1}^{n} y_k = 1$, is used only so that the parameters are identifiable, and it is just one of many forms that could be used without changing the model. The particular constraint chosen is influenced by the interpretation of the parameters. Thus, we can alter this constraint without changing the predictive distributions, although the parameter estimates and their interpretations will change. In other words, an equivalent stochastic model for the chain-ladder technique, which could be used in place of (3.1) is

$$C_{ij}|x', y', \varphi \sim \text{independent over-dispersed Poisson, with mean } x_i'y_j', \text{ and } \sum_{k=1}^{n} x_k' = 1. \tag{5.3}$$

I have added dashes to the parameters, just to emphasize that they will have different values and different interpretations from the parameterization in (3.1). However, *it is simply a reparameterization of the model;* it will still give the chain-ladder results if improper prior distributions are used for the parameters.

Applying a prior-posterior analysis to the column parameters gives

$$f(y'|C_{ij}, i = 1, \ldots, n, j = 1, \ldots, n - i + 1, x', \varphi) \propto \prod_{i=1}^{n} \prod_{j=1}^{n-i+1} f(C_{ij}|x', y', \varphi) \prod_{i=1}^{n} f(y_i') \tag{5.4}$$

We then integrate out the column parameters from the sampling distribution of the data:

$$f(C_{kl}|C_{ij}, i = 1, \ldots, n, j = 1, \ldots, n - i + 1, x', \varphi)$$
$$= \int f(C_{kl}|x', y', \varphi) f(y'|C_{ij}, i = 1, \ldots, n, j = 1, \ldots, n - i + 1, x', \varphi)\, dy'. \tag{5.5}$$

Once this has been obtained, the prior distributions discussed in Section 4 can be used for the row parameters, and the appropriate posterior distributions can be obtained. There is obviously now the added complication of converting the prior distributions for the row parameters, x, into prior distributions for the new parameterization, and we show how to do this below.

It is now required to carry out the analysis implied by (5.4) and (5.5). However, this is straightforward because of the complete symmetry of the problem. This analysis is identical to that described in Section 4 and Appendix A, and in Verrall (2000), but with the row and column parameters interchanged. Hence, the result will be the same, and the distribution $f(C_{kl}|C_{ij}, i = 1, \ldots, n, j = 1, \ldots, n - i + 1, x', \varphi)$ is also a negative binomial distribution whose form looks very similar to that of the chain-ladder model. Recall that the distribution of $C_{ij}|C_{i1}, C_{i2}, \ldots, C_{i,j-1}, y, \varphi$ under the chain-ladder model is a negative binomial distribution and is written in recursive form. This distribution has parameters $1/\lambda_j$ and $D_{i,j-1}$, and its mean is $(\lambda_j - 1)D_{i,j-1}$. It is obtained by applying improper prior distributions to the row parameters, x, in equation (3.1) and integrating these out.

The chain-ladder technique is completely symmetric in the rows and columns, and this has enabled us to express the stochastic model as in (5.3). We can now appeal to the symmetry of the model to show that the distribution defined in (5.5) also has a recursive form, but it is recursive in i instead of j. I call this Bayesian model the "negative binomial model" and give illustrations of the results in Sections 6.3 and 6.4.

To summarize, the negative binomial Bayesian model for the data is as given in (5.6) below, and this takes into account the estimation of the column parameters using the chain-ladder assumptions. The Bornhuetter-Ferguson technique can be reproduced by using strong prior information for the row parameters, and the chain-ladder technique can be reproduced by using improper priors for the row parameters.

$$C_{ij}|C_{1,j}, C_{2,j}, \ldots, C_{i-1,j}, x', \varphi \sim \text{over-dispersed negative binomial, with parameters } \frac{1}{\gamma_i} \text{ and } \sum_{m=1}^{i-1} C_{mj}. \tag{5.6}$$

The mean of this distribution is $(\gamma_i - 1) \sum_{m=1}^{i-1} C_{mj}$. Comparing this to the mean of the chain-ladder model, $(\lambda_j - 1)D_{i,j-1} = (\lambda_j - 1) \sum_{m=1}^{j-1} C_{i,m}$, it can be seen that they are identical in form, with the recursion either being across the rows, or down the columns.

In the context of the Bornhuetter-Ferguson method, as discussed in Section 4, prior distributions will be defined for the parameters x_i. However, as the model has been reparameterized, it is necessary to define the relationship between the new parameters, γ_i, and the original parameters, x_i. This is given in equation (5.7), which can be used to find values of γ_i from the values of x_i given in the prior distributions.

$$\gamma_i = \frac{x_i\left(1 - \dfrac{1}{\prod_{k=n-i+2}^{n} \lambda_k}\right)}{\sum_{k=n-i+2}^{n} \left[\left(\prod_{l=1}^{i+k-n-1} \gamma_l\right) \sum_{m=1}^{n-k+1} C_{m,k}\right]} + 1, \tag{5.7}$$

with $\gamma_1 = 1$.

This equation is derived by considering the estimate of outstanding claims for the Bornhuetter-Ferguson method and equating this to the estimate of outstanding claims for the negative binomial model:

$$x_i\left(1 - \frac{1}{\prod_{k=n-i+2}^{n} \lambda_k}\right) = \sum_{k=n-i+2}^{n} \left[\left(\prod_{l=1}^{i+k-n-1} \gamma_l\right) (\gamma_i - 1) \sum_{m=1}^{n-k+1} C_{m,k}\right]$$

(where $\gamma_1 = 1$).

Note that the left-hand side of this equation makes use of the chain-ladder development factors to convert the prior estimate of ultimate claims to an estimate of outstanding claims. To avoid using these parameter estimates, which are estimated separately from the Bayesian model, it would be more satisfactory to use a prior distribution for the outstanding claims, rather than the ultimate claims. While this would be a slight departure from the Bornhuetter-Ferguson technique as it is usually described in the literature, I believe that it may be closer to what actually happens in practice.

This has now defined a stochastic version of the Bornhuetter-Ferguson technique. Since the column parameters (the development factors) are dealt with first, using improper prior distributions, their estimates will be those implied by the chain-ladder technique. These are then integrated out to give the recursive negative binomial model, (5.6). Prior information can be defined in terms of distributions for the parameters x_i, which can then be converted into values for the parameters γ_i in an MCMC estimation process, using (5.7). This is implemented in Section 6.

As a side issue, it is interesting to note that the chain-ladder estimates of outstanding claims can be obtained by either applying the deterministic chain-ladder technique in the conventional way, recursively from column to column, or alternatively by applying the same deterministic chain-ladder technique, recursively from row to row. Another way of looking at this is that the incremental claims triangle can be transformed so that the rows become the columns and vice versa:

$$C^*_{ij} = C_{ji}.$$

This gives a new triangle of incremental claims data to which the deterministic chain-ladder technique can be applied. The development factors will be estimates of the parameters γ_i, and estimates of outstanding claims can be found by finding the predicted incremental claims and summing down each column (instead of across each row):

$$\sum_{j=n-i+2}^{n} C_{ij} = \sum_{j=n-i+2}^{n} C^*_{ji}.$$

This section has defined the negative binomial model, (5.6), which can be used for the Bornhuetter-Ferguson method. This is illustrated in Section 6, which also considers the over-dispersed Poisson model.

6. IMPLEMENTATION

The Bayesian models described in Sections 3 and 5 can be implemented using WinBugs (Spiegelhalter et al. 1996), a freely available piece of software (http://www.mrc-bsu.cam.ac.uk/bugs) that uses the MCMC method. An excellent overview of MCMC methods, with applications in actuarial science is provided by Scollnick (2001), and further examples are available from http://www.math.ucalgary.ca/~scollnik. Appendix B contains code for the WinBUGS implementation illustrated in this section.

For illustration purposes, consider the data given in Table 1. This data was also used to illustrate stochastic reserving methods in England and Verrall (2002). Table 1 also shows the standard chain-ladder results.

We consider two models. The first was described in Section 3, and is the over-dispersed Poisson model. In the first model, the column parameters, y_j, will be given improper prior distributions and will be estimated at the same time as the row parameters, x_i. This model is:

$$C_{ij}|x, y, \varphi \sim \text{independent over-dispersed Poisson, with mean } x_i y_j, \text{ and } \sum_{k=1}^{n} y_k = 1$$

$$x_i|\alpha_i, \beta_i \sim \text{independent } \Gamma(\alpha_i, \beta_i) \text{ and improper prior distributions are used for } y_j.$$

Table 1

Historical Loss Development Study (1991) Automatic Facultative General Liability Data (Excluding Asbestos and Environmental)

Claim Payments										Reserves
$5,012	$3,257	$2,638	$ 898	$1,734	$2,642	$1,828	$599	$ 54	$172	$ 0
106	4,179	1,111	5,270	3,116	1,817	−103	673	535		154
3,410	5,582	4,881	2,268	2,594	3,479	649	603			617
5,655	5,900	4,211	5,500	2,159	2,658	984				1,636
1,092	8,473	6,271	6,333	3,786	225					2,747
1,513	4,932	5,257	1,233	2,917						3,649
557	3,463	6,926	1,368							5,435
1,351	5,596	6,165								10,907
3,133	2,262									10,650
2,063										16,339
									Overall	52,135
Development Factors	2.999	1.624	1.271	1.172	1.113	1.042	1.033	1.017	1.009	

The WinBUGS code is given in Appendix B(i). We consider a number of specifications for the prior distributions of the row parameters, x_i, in Sections 6.1 and 6.2.

The second model was described in Section 5, and the column parameters have already been estimated (using improper prior distributions), and integrated out of the data. This model is the recursive negative binomial model:

$$C_{ij}|C_{1,j}, C_{2,j}, \ldots, C_{i-1,j}, x', \varphi \sim \text{over-dispersed negative binomial, with parameters } \frac{1}{\gamma_i} \text{ and } \sum_{m=1}^{i-1} C_{mj}.$$

The mean of this distribution is $(\gamma_i - 1) \sum_{m=1}^{i-1} C_{mj}$. Again, we consider a number of different specifications for the prior distributions of the row parameters, which are defined in terms of x_i and then related to γ_i using (5.7). The WinBUGS code is given in Appendix B(ii). We consider a number of specifications for the prior distributions of the row parameters, x_i, in Sections 6.3 and 6.4. For all the illustrations in this section, I used an initial burn-in of 10,000 iterations (the results of which were discarded) to remove any effect attributable to the initial conditions and allow the simulations to converge. I also examined the results using a number of different initial conditions to ensure that these had no undue effect on the results. I then ran another 10,000 simulations 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.

6.1 Over-Dispersed Poisson Model with Improper Prior Distributions (Chain-Ladder)

We begin with improper prior distributions and obtain results that are very close to those from the chain-ladder technique. These results are for reference purposes; Sections 6.2 and 6.3 illustrate the use of informative prior distributions. This section uses the first set of code in Appendix B, using large values for the variances of the row parameters. Note that it is also possible to obtain results that are very close to the chain-ladder technique using the negative binomial model with large values for the variances of the row parameters. As was pointed out by de Alba (2002), the Bayesian implementation makes the production of the predictive distribution, and prediction errors, enormously much easier than the analytic approach (MCMC is a simulation-based method).

Table 2 shows the results of using improper prior distributions for all the parameters, together with the analytic results from the equivalent over-dispersed Poisson model obtained using S-Plus (see

Table 2

Over-Dispersed Poisson Model: Bayesian Chain-Ladder Model with Improper Priors: Mean and Prediction Error of Reserves

	Bayesian Mean Reserve	Bayesian Prediction Error	Bayesian Prediction Error %	Analytic Mean Reserve	Analytic Prediction Error	Analytic Prediction Error %
Year 2	152	579	381%	154	556	361%
Year 3	633	1,186	187	617	1,120	181
Year 4	1,665	1,842	111	1,636	1,775	109
Year 5	2,778	2,320	84	2,747	2,231	81
Year 6	3,636	2,520	69	3,649	2,440	67
Year 7	5,493	3,214	59	5,435	3,124	57
Year 8	11,020	5,261	48	10,907	5,032	46
Year 9	10,760	6,268	58	10,650	6,075	57
Year 10	17,340	14,090	81	16,339	12,987	79
Overall	53,470	19,200	36	52,135	18,193	35

England and Verrall 2002, for more details). The Bayesian Mean Reserve and the Bayesian Prediction Error are the mean and standard deviation of the posterior distribution sampled by WinBUGS. To use an improper prior distribution, the value of β_i must be small. For example, if the mean and standard deviation of x_i are 10 and 1,000, respectively, then $\beta_i = 0.00001$, and $\alpha_i = 0.0001$. Note that the values of α_i and β_i can be obtained from the mean and standard deviation of x_i as follows:

$$\beta_i = \frac{(\text{standard deviation of } x_i)^2}{\text{mean of } x_i} \text{ and } \alpha_i = \beta_i \times \text{mean of } x_i.$$

It can be seen from Table 2 that the results of the Bayesian model are similar to the analytic results, but it would be desirable if the results were closer still. The differences are due to a number of factors, including simulation error, choice of prior for the parameters, and choice of the form of the model linking the parameters to the mean. However, for ease of implementation the Bayesian model has great advantages. MCMC is a simulation-based procedure and it is, therefore, only necessary to specify the quantities to be monitored in order to obtain the results in Table 2. By contrast, the analytic method requires careful programming to obtain the prediction errors. Since the implementation uses a simulation-based technique, the results will vary slightly each time the program is run.

The value used for φ is that obtained from fitting the over-dispersed Poisson model, using, for example, S-Plus. It will be noticed that, if the first WinBUGS program given in Appendix B (i) is run and densities of the reserve estimates for the rows are examined, the simulated outcomes for each incremental claims are always multiples of the dispersion parameter. The properties of the distributions are correct (means, variances, etc.), but, particularly for the early rows, they can look a little strange. England and Verrall (2002) noticed the same effect in the context of bootstrapping, and suggested replacing the over-dispersed Poisson by a gamma distribution with the same mean and variance if this observed effect was considered undesirable. This approach is used in the second program in Appendix B (see Sections 6.4 and 6.5) and the same could also be done for this model, or a smoother (such as a kernel smoother) could be applied to the densities before they are presented.

This has shown that results similar to the over-dispersed Poisson model can be obtained using improper prior distributions. However, it is assumed that the real purpose here is to incorporate prior information about the row parameters. Section 6.2 illustrates the application of strong prior information, in a similar way to the deterministic Bornhuetter-Ferguson technique.

6.2 Over-Dispersed Poisson Model with Strong Prior Distributions

In this section the column parameters are again given improper distributions, and so are estimated from the data. As was shown in Section 4, the Bornhuetter-Ferguson technique does not allow the data to

Table 3

Over-Dispersed Poisson Model: Bayesian Model with Precise Prior for One Row: Mean and Prediction Error of Reserves

	Bayesian Mean Reserve	Bayesian Prediction Error	Bayesian Prediction Error %	Bornhuetter-Ferguson Reserve
Year 2	150	581	387%	154
Year 3	626	1,174	187	617
Year 4	1,647	1,828	111	1,636
Year 5	2,772	2,351	85	2,747
Year 6	3,680	2,515	68	3,649
Year 7	5,451	3,215	59	5,435
Year 8	11,080	5,321	48	10,907
Year 9	10,850	6,338	58	10,650
Year 10	14,220	3,915	28	14,206
Overall	50,470	12,900	26	50,002

influence the estimates of the row parameters at all, and so we begin with an example with a very low standard deviation for the prior distribution. We begin by considering applying prior information for a single accident year. Suppose we believe the predicted ultimate for year 10 is too high, due to the payment in the first development period being unusually high. We can put an informative prior on x_{10} and use a distribution that has mean 16,000 and standard deviation 1.

The reserve results are shown in Table 3, together with the standard chain-ladder estimates and the Bornhuetter-Ferguson estimates. Notice that the reserve estimate for year 10 is similar to the result of the Bornhuetter-Ferguson technique, as might be expected with a precise prior. Notice also that the prediction errors have reduced substantially, reflecting the degree of precision of the prior; it could be questioned whether use of such a strong prior is appropriate.

We next consider applying strong prior information again using the over-dispersed Poisson model, but this time to the other accident years as well (rows 2 to 10). In this example, we use prior distributions for the row parameters that all have standard deviation 1 and whose means are:

$$\begin{array}{ccccccccc} x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} \\ 17{,}500 & 25{,}000 & 30{,}000 & 30{,}000 & 25{,}000 & 25{,}000 & 25{,}000 & 25{,}000 & 25{,}000 \end{array}$$

The Bornhuetter-Ferguson estimates of outstanding claims, and the results from the Bayesian model are shown in Table 4.

It can be seen that there are differences between the Bayesian results and the results from the

Table 4

Over-Dispersed Poisson Model: Bayesian Model with Precise Priors for All Rows: Mean and Prediction Error of Reserves

	Bayesian Mean Reserve	Bayesian Prediction Error	Bayesian Prediction Error %	Bornhuetter-Ferguson Reserve
Year 2	251	710	283%	160
Year 3	917	1,395	152	641
Year 4	2,169	1,984	91	1,710
Year 5	3,457	2,342	68	2,849
Year 6	5,343	2,687	50	4,678
Year 7	8,422	3,231	38	7,656
Year 8	12,060	3,774	31	11,353
Year 9	17,170	4,432	26	16,594
Year 10	22,400	4,938	22	22,197
Overall	72,190	11,330	16	67,837

Bornhuetter-Ferguson technique. The reason for this is that the prior distributions for the row parameters have affected the posterior distributions for the column parameters, which has had a consequential effect on the reserve estimates. In contrast, the Bornhuetter-Ferguson technique uses the chain-ladder estimate of the development factors, and these are not changed when prior information about the rows is used. To do the same in the Bayesian context, it is necessary to use the negative binomial model from Section 5, and this is illustrated in Section 6.3.

6.3 Negative Binomial Model with Strong Prior Distributions (Bornhuetter-Ferguson)

In this section, we use the same prior distribution for the row parameters as was used in Section 6.2 and assume there is strong prior information. This means that we use prior distributions for the row parameters that all have standard deviation 1 and whose means are:

x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
17,500	25,000	30,000	30,000	25,000	25,000	25,000	25,000	25,000

We now use the negative binomial model from Section 5. The WinBUGS code for this is given in the second part of Appendix B. The Bornhuetter-Ferguson estimates of outstanding claims, and the results from the Bayesian model are shown in Table 5.

In this case, it can be seen that the results are very close to those of the Bornhuetter-Ferguson technique. Thus, if it is desirable to use the Bornhuetter-Ferguson method within this stochastic framework, this is the approach that should be used. The added information available is the prediction errors. Furthermore, it is possible to generate predictive distributions rather than just the mean and prediction error.

The Bornhuetter-Ferguson technique assumes that there is strong prior information about the row parameters, so that the standard deviations of the prior distributions used in this example are small. As was shown in Section 6.1, the other end of the spectrum is constituted by the chain-ladder technique, when large standard deviations are used for the prior distributions (it is possible to do the same for the negative binomial model, and get the chain-ladder results in that way as well). Between these two extremes is a whole range of possible models, which can be specified by using different standard deviations.

We illustrate this in Section 6.4, using the negative binomial model. As was noted in this section and Section 6.2, the difference between the negative binomial model and the Poisson model is whether the prior distributions on the row parameters should be allowed to affect the estimation of the column parameters. Since the negative binomial model does not allow this to happen, and therefore takes a

Table 5

Negative Binomial Model: Bayesian Model with Precise Priors for All Rows: Mean and Prediction Error of Reserves

	Bayesian Mean Reserve	Bayesian Prediction Error	Bayesian Prediction Error %	Bornhuetter-Ferguson Reserve
Year 2	163	569	349%	160
Year 3	637	1,156	182	641
Year 4	1,697	1,810	107	1,710
Year 5	2,855	2,232	78	2,849
Year 6	4,694	2,701	58	4,678
Year 7	7,683	3,311	43	7,656
Year 8	11,280	3,943	35	11,353
Year 9	16,600	4,665	28	16,594
Year 10	22,190	5,351	24	22,197
Overall	67,790	13,920	21	67,837

similar approach to the Bornhuetter-Ferguson technique, we illustrate this here. It is possible to use similar prior distributions for the Poisson model.

6.4 Negative Binomial Model with Prior Distributions between Chain-Ladder and Bornhuetter-Ferguson

I now illustrate the results when less strongly informative prior distributions are used for the row parameters. We use the same prior means as in Section 6.3, but this time use a standard deviation of 5000. We are incorporating prior belief about the ultimate claims for each year, but allowing for uncertainty. The associated reserve results are shown in Table 6. Notice that the reserves are between the chain-ladder and Bornhuetter-Ferguson results (except for year 2 because of simulation error). Notice also that the precision of the prior has influenced the prediction errors, but to a lesser extent.

Table 6

Negative Binomial Model: Bayesian Model with Informative Priors: Mean and Prediction Error of Reserves

	Bayesian Mean Reserve	Bayesian Prediction Error	Bayesian Prediction Error %	Bornhuetter-Ferguson Reserve	Chain-Ladder Reserve
Year 2	163	632	387%	160	154
Year 3	634	1,183	187	641	617
Year 4	1,680	1,886	112	1,710	1,636
Year 5	2,845	2,377	84	2,849	2,747
Year 6	3,784	2,533	67	4,678	3,649
Year 7	5,873	3,266	56	7,656	5,435
Year 8	11,290	4,886	43	11,353	10,907
Year 9	14,450	5,662	39	16,594	10,650
Year 10	21,670	7,215	33	22,197	16,339
Overall	62,390	14,820	24	67,837	52,135

7. CONCLUSIONS

This paper has shown how the Bornhuetter-Ferguson method can be written in terms of Bayesian models within the framework of generalized linear models. It has also shown that the method as currently implemented by actuaries can be regarded, within the framework of generalized linear models, as an extreme case of a Bayesian model, which assumes that the prior information about the row parameters is the only information that should be used for estimating these. It would perhaps be more sensible to use a slightly less exact prior distribution in practice, and thus apply a model somewhere between the Bornhuetter-Ferguson method and the chain-ladder technique.

The theory derived in this paper shows how the approach to Bornhuetter-Ferguson method described in Mack (2000) can be applied when a generalized linear model is used. The Bayesian models derived in this paper may break down if there are negative incremental claims values and is, therefore, probably only suitable for paid data. This is certain to happen if any column sum of incremental claims is negative, although the model can cope with some negative values (as in the data set used in Section 6).

Two Bayesian models were used. The difference between these is in the estimation of the column parameters. In the over-dispersed Poisson model all the parameters are estimated at the same time, and the prior distributions applied to the row parameters will affect the estimation of the column parameters. In contrast, for the negative binomial model the column parameters are estimated first, before prior distributions are applied to the row parameters. Thus, since improper prior distributions are used for the column parameters, they are not altered from the values that would be used by the chain-ladder method. It is this latter approach which reproduces the Bornhuetter-Ferguson method, since that also uses the chain-ladder parameter estimates. Since both methods have been implemented, they can now be examined in more detail in practical situations to decide which is the more appropriate approach to use.

APPENDIX A

This appendix shows how the results in this paper are derived. In particular, it shows how the predictive distribution of $C_{ij}|C_{i1}, C_{i2}, \ldots, C_{i,j-1}, y, \varphi$ is derived. To derive these results, a reformulation of the over-dispersed Poisson model is required. Recall that it is assumed, in (3.1), that

$$C_{ij}|x, y, \varphi \sim \text{independent over-dispersed Poisson, with mean } x_i y_j, \text{ and } \sum_{k=1}^{n} y_k = 1.$$

This model can be reparameterized as follows:

$$C_{ij}|z_{ij}, y, \varphi \sim \text{independent over-dispersed Poisson, with mean } \frac{z_{ij} y_j}{\sum\limits_{k=1}^{j} y_k} = \frac{z_{ij} y_j}{S_j}.$$

In this case, $z_{i,j} = E[D_{ij}]$, which is the expected value of cumulative claims up to the latest development year observed in accident year i. Note that $z_{in} = E[D_{in}] = x_i$.

The following theorem gives the posterior distribution of $z_{ij}|C_{i1}, C_{i2}, \ldots, C_{ij}, y, \varphi$, and is needed mainly for the corollary, which gives the predictive distribution of $C_{ij}|C_{i1}, C_{i2}, \ldots, C_{i,j-1}, y, \varphi$.

Theorem

Suppose $C_{ij}|z_{ij}, y, \varphi \sim$ independent over-dispersed Poisson, with mean $z_{ij} y_j / S_j$ and $x_i|\alpha_i, \beta_j \sim$ independent $\Gamma(\alpha_i, \beta_i)$. Then

$$z_{ij}|C_{i1}, C_{i2}, \ldots, C_{ij}, y, \varphi \sim \Gamma\left(\alpha_i + \frac{D_{i,n-i+1}}{\varphi}, \frac{\beta_i \varphi + S_{n-i+1}}{\varphi S_{n-i+1}}\right)$$

PROOF

Now,

$$z_{ij} = E[D_{ij}] = E[D_{i,j-1}] + E[C_{ij}] = z_{i,j-1} + \frac{z_{ij} y_j}{S_j}$$

and, hence,

$$z_{ij} = z_{i,j-1}\left(1 - \frac{y_j}{S_j}\right)^{-1} = z_{i,j-1}\left(\frac{S_j - y_j}{S_j}\right)^{-1} = z_{i,j-1}\left(\frac{S_{j-1}}{S_j}\right)^{-1}.$$

Thus, $z_{ij} = z_{i,j-1}(S_j/S_{j-1})$ and the conditional distribution of C_{ij} given $z_{i,j-1}, y, \varphi$ is an over-dispersed Poisson distribution with mean $z_{i,j-1} y_j / S_{j-1}$.

Also, it can be seen that

$$x_i = z_{in} = z_{i,n-1}\frac{S_n}{S_{n-1}} = z_{i,n-2}\frac{S_{n-1}}{S_{n-2}}\frac{S_n}{S_{n-1}} = \ldots = z_{i,1}\frac{S_n}{S_1} = z_{i,1}\frac{1}{y_1}$$

since $S_n = \sum_{k=1}^{n} y_k = 1$. The prior distribution of x_i is $x_i|\alpha_i, \beta_i \sim \Gamma(\alpha_i, \beta_i)$ and, hence, the prior distribution of $z_{i,1}$ is $z_{i,1}|\alpha_i, \beta_i \sim \Gamma(\alpha_i, \beta_i/y_1)$. A standard Bayesian prior-posterior analysis gives the distribution of $z_{i,1}$, given $C_{i,1}$:

$$f(z_{i,1}|C_{i,1}, y, \varphi) \propto \left(\frac{z_{i,1}}{\varphi}\right)^{C_{i,1}/\varphi} e^{-(z_{i,1}/\varphi)} z_{i,1}^{\alpha_i - 1} e^{-(\beta_i/y_1) z_{i,1}},$$

from which it can be seen that

$$z_{i,1}|C_{i,1},\, y,\, \varphi \sim \Gamma\left(\alpha_i + \frac{C_{i,1}}{\varphi}, \frac{\beta_i}{y_1} + \frac{1}{\varphi}\right);$$

that is,

$$z_{i,1}|C_{i,1},\, y,\, \varphi \sim \Gamma\left(\alpha_i + \frac{D_{i,1}}{\varphi}, \frac{\beta_i\varphi + S_1}{\varphi S_1}\right).$$

We can now prove the theorem by induction. Suppose that

$$z_{i,j-1}|C_{i,1},\, C_{i,2},\, \ldots,\, C_{i,j-1},\, y,\, \varphi \sim \Gamma\left(\alpha_i + \frac{D_{i,j-1}}{\varphi}, \frac{\beta_i\varphi + S_{j-1}}{\varphi S_{j-1}}\right).$$

We have shown above that this is true for $z_{i,1}$, and we now prove that if it holds for $z_{i,j-1}$ it also holds for z_{ij}. Hence, we will have proved the result of the theorem.

Since the conditional distribution of C_{ij} given $z_{i,j-1}, y, \varphi$ is an over-dispersed Poisson distribution with mean $z_{i,j-1}y_j/S_{j-1}$, a standard Bayesian analysis gives the posterior distribution of $z_{i,j-1}$ as

$$z_{i,j-1}|C_{i1},\, C_{i2},\, \ldots,\, C_{i,j-1},\, C_{i,j},\, y,\, \varphi \sim \Gamma\left(\alpha_i + \frac{D_{i,j-1}}{\varphi} + \frac{C_{i,j}}{\varphi}, \frac{\beta_i\varphi + S_{j-1}}{\varphi S_{j-1}} + \frac{y_j}{\varphi S_{j-1}}\right);$$

that is,

$$z_{i,j-1}|C_{i1},\, C_{i2},\, \ldots,\, C_{ij},\, y,\, \varphi \sim \Gamma\left(\alpha_i + \frac{D_{ij}}{\varphi}, \frac{\beta_i\varphi + S_j}{\varphi S_{j-1}}\right).$$

Since we have a relationship between $z_{i,j}$ and $z_{i,j-1}$, we can obtain the distribution of $z_{i,j}$, conditional on the information received up to development year j by a straightforward transformation:

$$\text{If } z_{i,j-1} \sim \Gamma(a, b), \text{ then } z_{ij} \sim \Gamma\left(a, \frac{bS_{j-1}}{S_j}\right).$$

Hence,

$$z_{i,j}|C_{i1},\, C_{i2},\, \ldots,\, C_{ij},\, y,\, \varphi \sim \Gamma\left(\alpha_i + \frac{D_{ij}}{\varphi}, \frac{\beta_i\varphi + S_j}{\varphi S_{j-1}}\frac{S_{j-1}}{S_j}\right);$$

that is,

$$z_{ij}|C_{i1},\, C_{i2},\, \ldots,\, C_{ij},\, y,\, \varphi \sim \Gamma\left(\alpha_i + \frac{D_{ij}}{\varphi}, \frac{\beta_i\varphi + S_j}{\varphi S_j}\right),$$

which completes the recursive proof required.

Corollary

The predictive distribution of $C_{ij}|C_{i1},\, C_{i2},\, \ldots,\, C_{i,j-1},\, y,\, \varphi$ is an over-dispersed negative binomial distribution, with parameters

$$k = \alpha_i + \frac{D_{i,j-1}}{\varphi} \quad \text{and} \quad p = \frac{\beta_i\varphi + S_{j-1}}{\beta_i\varphi + S_j}.$$

PROOF

$$f\left(\frac{C_{ij}}{\varphi}\middle| C_{i1}, C_{i2}, \ldots, C_{i,j-1}, y, \varphi\right) = \int f\left(\frac{C_{ij}}{\varphi}\middle| z_{i,j-1}, y, \varphi\right) f(z_{i,j-1}|C_{i1}, C_{i2}, \ldots, C_{i,j-1}, y, \varphi)\, dz_{i,j-1}$$

$$= \int \frac{\left(\dfrac{z_{i,j-1}y_j}{\varphi S_{j-1}}\right)^{C_{ij}/\varphi} e^{-(z_{i,j-1}y_j/\varphi S_{j-1})}}{\dfrac{C_{ij}}{\varphi}!} \frac{\left(\dfrac{\beta_i\varphi + S_{j-1}}{\varphi S_{j-1}}\right)^{\alpha_i+(D_{i,j-1}/\varphi)}}{\Gamma\left(\alpha_i + \dfrac{D_{i,j-1}}{\varphi}\right)} z_{i,j-1}^{\alpha_i+(D_{i,j-1}/\varphi)-1} e^{-(\beta_i\varphi+S_{j-1}/\varphi S_{j-1})\, z_{i,j-1}}\, dz_{i,j-1}$$

$$= \frac{\left(\dfrac{y_j}{\varphi S_{j-1}}\right)^{C_{ij}/\varphi}\left(\dfrac{\beta_i + S_{j-1}}{S_{j-1}}\right)^{\alpha_i+D_{i,j-1}}}{\dfrac{C_{ij}}{\varphi}!\,\Gamma\left(\alpha_i + \dfrac{D_{i,j-1}}{\varphi}\right)} \int z_{i,j-1}^{\alpha+(D_{j-1}/\varphi)+(C_j/\varphi)-1} e^{-[(\beta_i\varphi+S_{j-1}/\varphi S_{j-1})+(y_j/\varphi S_{j-1})]\, z_{i,j-1}}\, dz_{i,j-1}$$

$$= \frac{\left(\dfrac{y_j}{\varphi S_{j-1}}\right)^{C_{ij}/\varphi}\left(\dfrac{\beta_i\varphi + S_{j-1}}{\varphi S_{j-1}}\right)^{\alpha_i+(D_{i,j-1}/\varphi)}}{\dfrac{C_{ij}}{\varphi}!\,\Gamma\left(\alpha_i + \dfrac{D_{i,j-1}}{\varphi}\right)} \frac{\Gamma\left(\alpha_i + \dfrac{D_{ij}}{\varphi}\right)}{\left(\dfrac{\beta_i\varphi + S_j}{\varphi S_{j-1}}\right)^{\alpha_i+(D_{ij}/\varphi)}}$$

$$= \frac{\Gamma\left(\alpha_i + \dfrac{D_{ij}}{\varphi}\right)}{\dfrac{C_{ij}}{\varphi}!\,\Gamma\left(\alpha_i + \dfrac{D_{i,j-1}}{\varphi}\right)} \left(\frac{y_j}{\beta_i\varphi + S_j}\right)^{C_{ij}/\varphi}\left(\frac{\beta_i\varphi + S_{j-1}}{\beta_i\varphi + S_j}\right)^{\alpha_i+(D_{i,j-1}/\varphi)},$$

which is a negative binomial distribution with parameters

$$k = \alpha_i + \frac{D_{i,j-1}}{\varphi} \quad \text{and} \quad p = \frac{\beta_i\varphi + S_{j-1}}{\beta_i\varphi + S_j}.$$

APPENDIX B

The code for WinBugs is shown below for the models used in Section 6. This can be cut and pasted directly into WinBugs.

(i) Over-Dispersed Poisson Model

```
model
{
# Model for data:
 for( i in 1 : 55 ) {
  Z[i] <- Y[i]/1000
  log(mu[i]) <- alpha[row[i]] + beta[col[i]]
# Zeros trick:
 zeros[i]<- 0
 zeros[i] ~ dpois(phi[i])
 phi[i] <- (mu[i]-Z[i]*log(mu[i])+loggam(Z[i]+1))/scale # MINUS log likelihood
  }
# Model for future observations:
 for( i in 56 : 100 ) {
  mu2[i] <- mu[i]/scale
```

```
  Y[i] ~ dpois(mu2[i]);
  log(mu[i]) <- alpha[row[i]] + beta[col[i]]
  Z[i] <- scale*Y[i]
  }
 for( i in 1 : 100 ) {
  fit[i] <- Z[i]*1000
  }
 scale <- 1.08676
 a[1] <- 18.834
 alpha[1] <- log(a[1])
# Prior distributions for row parameters:
 for (k in 2:10) {
  a[k] ~ dgamma(ulta[k],ultb[k])
  alpha[k] <- log(a[k])
  ulta[k] <- pow(ultmean[k-1], 2)/pow(ultsd[k-1],2)
  ultb[k] <- ultmean[k-1]/pow(ultsd[k-1], 2)
   }
# The prior distribution can be changed by changing the data input values for the
vectors ultmean and ultsd
# Prior distributions for column parameters:
 for (k in 1:10) {
   p1[k] ~ dgamma(0.0001,0.001)
    }
   s<-sum(p1[1:10])
   for (k in 1:10) {
   p[k]<-p1[k]/s
  beta[k] <- log(p[k])
  }
 pc[1]<-p[1]
for (k in 2:10) {pc[k]<-pc[k-1]+p[k]}
 for (k in 1:9) {lambda[k]<-pc[k+1]/pc[k]}
#Row Totals and Overall Reserve:
 R[1] <- 0
 R[2] <- fit[56]
 R[3] <- sum(fit[57:58])
 R[4] <- sum(fit[59:61])
 R[5] <- sum(fit[62:65])
 R[6] <- sum(fit[66:70])
 R[7] <- sum(fit[71:76])
 R[8] <- sum(fit[77:83])
 R[9] <- sum(fit[84:91])
 R[10] <- sum(fit[92:100])
 Total <- sum(R[2:10])
 }
# DATA
list(row=c(1,1,1,1,1,1,1,1,1,1,
2,2,2,2,2,2,2,2,2,3,
3,3,3,3,3,3,3,4,4,4,
4,4,4,4,5,5,5,5,5,5,
```

```
6,6,6,6,6,7,7,7,7,8,
8,8,9,9,10,2,3,3,4,4,
4,5,5,5,5,6,6,6,6,6,
7,7,7,7,7,7,8,8,8,8,
8,8,8,9,9,9,9,9,9,9,
9,10,10,10,10,10,10,10,10,10),
col=c(1,2,3,4,5,6,7,8,9,10,
1,2,3,4,5,6,7,8,9,1,
2,3,4,5,6,7,8,1,2,3,
4,5,6,7,1,2,3,4,5,6,
1,2,3,4,5,1,2,3,4,1,
2,3,1,2,1,10,9,10,8,9,
10,7,8,9,10,6,7,8,9,10,
5,6,7,8,9,10,4,5,6,7,
8,9,10,3,4,5,6,7,8,9,
10,2,3,4,5,6,7,8,9,10),
Y=c(5012,3257,2638,898,1734,2642,1828,599,54,
172,106,4179,1111,5270,3116,1817,-103,673,535,3410,
5582,4881,2268,2594,3479,649,603,5655,5900,4211,
5500,2159,2658,984,1092,8473,6271,6333,3786,225,
1513,4932,5257,1233,2917,557,3463,6926,1368,1351,
5596,6165,3133,2262,2063,
NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA,NA),
ultmean=c(17.5,25,30,30,25,25,25,25,25),
ultsd=c(10000,10000,10000,10000,10000,10000,10000,10000,10000))
# INITIAL VALUES
list(a = c(NA,20,20,20,20,20,20,20,20,20),
p1=c(0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1 ),
Y=c(NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0))
```

(ii) Negative Binomial Model

```
model
 {
# Model for Data
 for( i in 1 : 45 ) {
 Z[i] <- Y[i]/1000
 pC[i]<-D[i]/1000
```

```
# Zeros trick
 zeros[i]<- 0
 zeros[i] ~ dpois(phi[i])
 phi[i]<-(-pC[i]*log(1/(1+g[row[i]]))-Z[i]*log(g[row[i]]/(1+g[row[i]])))/
scale
  }
# Cumulate down the columns:
 DD[3]<-DD[1]+Y[46]
 for( i in 1 : 2 ) {DD[4+i]<-DD[4+i-3]+Y[49+i-3]}
 for( i in 1 : 3 ) {DD[7+i]<-DD[7+i-4]+Y[52+i-4]}
 for( i in 1 : 4 ) {DD[11+i]<-DD[11+i-5]+Y[56+i-5]}
 for( i in 1 : 5 ) {DD[16+i]<-DD[16+i-6]+Y[61+i-6]}
 for( i in 1 : 6 ) {DD[22+i]<-DD[22+i-7]+Y[67+i-7]}
 for( i in 1 : 7 ) {DD[29+i]<-DD[29+i-8]+Y[74+i-8]}
 for( i in 1 : 8 ) {DD[37+i]<-DD[37+i-9]+Y[82+i-9]}
# Needed for the denominator in (5.7):
 E[3]<-E[1]*gamma[1]
 for( i in 1 : 2 ) {E[4+i]<-E[4+i-3]*gamma[2]}
 for( i in 1 : 3 ) {E[7+i]<-E[7+i-4]*gamma[3]}
 for( i in 1 : 4 ) {E[11+i]<-E[11+i-5]*gamma[4]}
 for( i in 1 : 5 ) {E[16+i]<-E[16+i-6]*gamma[5]}
 for( i in 1 : 6 ) {E[22+i]<-E[22+i-7]*gamma[6]}
 for( i in 1 : 7 ) {E[29+i]<-E[29+i-8]*gamma[7]}
 for( i in 1 : 8 ) {E[37+i]<-E[37+i-9]*gamma[8]}
 EC[1]<-E[1]/1000
 EC[2]<-sum(E[2:3])/1000
 EC[3]<-sum(E[4:6])/1000
 EC[4]<-sum(E[7:10])/1000
 EC[5]<-sum(E[11:15])/1000
 EC[6]<-sum(E[16:21])/1000
 EC[7]<-sum(E[22:28])/1000
 EC[8]<-sum(E[29:36])/1000
 EC[9]<-sum(E[37:45])/1000
# Model for future observations
  for( i in 46 : 90 ) {
   a1[i]<- a[row[i]]*DD[i-45]/(1000*scale)
   b1[i]<- 1/(gamma[row[i]]*1000*scale)
   Z[i]~dgamma(a1[i],b1[i])
   Y[i]<-Z[i]
   fit[i]<-Y[i]
      }
 scale <- 1.08676
#Convert row parameters to gamma using (5.7)
  for (k in 1:9) {
   gamma[k]<-1+g[k]
   g[k]<-u[k]/EC[k]
    a[k]<-g[k]/gamma[k]
       }
```

```
# Prior distributions for row parameters.
  for (k in 1:9) {
    u[k]~dgamma(au[k],bu[k])
    au[k]<-bu[k]*(ultm[k+1]*(1-1/f[k]))
    bu[k]<-(ultm[k+1]*(1-1/f[k]))/ultv[k+1]
      }
# The prior distribution can be changed by changing the data input values for
the
# vectors ultm and ultv
# Row totals and overall reserve
  R[1] <- 0
  R[2] <- fit[46]
  R[3] <- sum(fit[47:48])
  R[4] <- sum(fit[49:51])
  R[5] <- sum(fit[52:55])
  R[6] <- sum(fit[56:60])
  R[7] <- sum(fit[61:66])
  R[8] <- sum(fit[67:73])
  R[9] <- sum(fit[74:81])
  R[10] <- sum(fit[82:90])
  Total <- sum(R[2:10])
  }
# DATA
list(
row=c(1,1,1,1,1,1,1,1,1,1,
2,2,2,2,2,2,2,2,
3,3,3,3,3,3,3,
4,4,4,4,4,4,
5,5,5,5,5,
6,6,6,6,
7,7,7,
8,8,
9,
1,
2,2,
3,3,3,
4,4,4,4,
5,5,5,5,5,
6,6,6,6,6,6,
7,7,7,7,7,7,7,
8,8,8,8,8,8,8,8,
9,9,9,9,9,9,9,9,9),
Y=c(106,4179,1111,5270,3116,1817,-103,673,535,
3410,5582,4881,2268,2594,3479,649,603,
5655,5900,4211,5500,2159,2658,984,
1092,8473,6271,6333,3786,225,
1513,4932,5257,1233,2917,
557,3463,6926,1368,
1351,5596,6165,
3133,2262,
2063,
```

```
NA,
NA,NA,
NA,NA,NA,
NA,NA,NA,NA,
NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA),
D=c(5012,3257,2638,898,1734,2642,1828,599,54,
5118,7436,3749,6168,4850,4459,1725,1272,
8528,13018,8630,8436,7444,7938,2374,
14183,18918,12841,13936,9603,10596,
15275,27391,19112,20269,13389,
16788,32323,24369,21502,
17345,35786,31295,
18696,41382,
21829,
NA,
NA,NA,
NA,NA,NA,
NA,NA,NA,NA,
NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA),
DD=c(172,
589,NA,
1875,NA,NA,
3358,NA,NA,NA,
10821,NA,NA,NA,NA,
16306,NA,NA,NA,NA,NA,
22870,NA,NA,NA,NA,NA,NA,
37460,NA,NA,NA,NA,NA,NA,NA,
43644,NA,NA,NA,NA,NA,NA,NA,NA),
E=c(172,
589,NA,
1875,NA,NA,
3358,NA,NA,NA,
10821,NA,NA,NA,NA,
16306,NA,NA,NA,NA,NA,
22870,NA,NA,NA,NA,NA,NA,
37460,NA,NA,NA,NA,NA,NA,NA,
43644,NA,NA,NA,NA,NA,NA,NA,NA),
f=c(1.009217, 1.026309, 1.060448, 1.104917, 1.230198, 1.441392, 1.831848,
2.974047, 8.920234),
ultm=c(NA,17.5,25,30,30,25,25,25,25,25),
ultv=c(NA,25,25,25,25,25,25,25,25,25))
```

```
#INITIAL VALUES
list(u = c(17.5,25,30,30,25,25,25,25,25),
Z=c(NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,
NA,NA,NA,NA,
NA,NA,NA,
NA,NA,
NA,
0,
0,0,
0,0,0,
0,0,0,0,
0,0,0,0,0,
0,0,0,0,0,0,
0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0))
```

REFERENCES

BENKTANDER, GUNNAR. 1976. "An Approach to Credibility in Calculating IBNR for Casualty Excess Reinsurance," *The Actuarial Review* (April): 7.

BORNHUETTER, RONALD L., AND RONALD E. FERGUSON. 1972. "The Actuary and IBNR." In *Proceedings of the Casualty Actuarial Society* 59: 181–95.

DE ALBA, ENRIQUE. 2002. "Bayesian Estimation of Outstanding Claims Reserves," *North American Actuarial Journal* 6(4): 1–20.

ENGLAND, PETER D., AND R. J. VERRALL. 2002. "Stochastic Claims Reserving in General Insurance" (with discussion), *British Actuarial Journal* (8): 443–544.

NTZOUFRAS, IOANNIS, AND PETROS DELLAPORTAS. 2002. "Bayesian Modelling of Outstanding Liabilities Incorporating Claim Count Uncertainty," *North American Actuarial Journal* 6(1): 113–28.

MACK, THOMAS. 2000. "Credible Claims Reserves: The Benktander Method," *ASTIN Bulletin* 30: 333–47.

RENSHAW, ARTHUR E., AND R. J. VERRALL. 1998. "A Stochastic Model Underlying the Chain Ladder Technique," *British Actuarial Journal* 4(4): 903–23.

SCOLLNIK, DAVID P. M. 2001. "Actuarial Modeling with MCMC and BUGS," *North American Actuarial Journal* 5(2): 96–124.

SPIEGELHALTER, DAVID J., ANDREW THOMAS, NICKY G. BEST, AND WALLY R. GILKS. 1996. *BUGS 0.5: Bayesian inference using gibbs sampling*. Cambridge, U.K.: MRC Biostatistics Unit.

TAYLOR, GREG C. 2000. *Loss Reserving: An Actuarial Perspective*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

VERRALL, RICHARD J. 2000. "An Investigation into Stochastic Claims Reserving Models and the Chain-Ladder Technique," *Insurance: Mathematics & Economics* 26: 91–9.

Discussions on this paper can be submitted until January 1, 2005. The author reserves the right to reply to any discussion. Please see the Submission Guidelines for Authors on the inside back cover for instructions on the submission of discussions.