A.A.S. **1**, II, 221-270 (2006)

PREDICTIVE DISTRIBUTIONS OF OUTSTANDING LIABILITIES IN GENERAL INSURANCE

By P. D. England and R. J. Verrall

ABSTRACT

This paper extends the methods introduced in England & Verrall (2002), and shows how predictive distributions of outstanding liabilities in general insurance can be obtained using bootstrap or Bayesian techniques for clearly defined statistical models. A general procedure for bootstrapping is described, by extending the methods introduced in England & Verrall (1999), England (2002) and Pinheiro *et al.* (2003). The analogous Bayesian estimation procedure is implemented using Markov-chain Monte Carlo methods, where the models are constructed as Bayesian generalised linear models using the approach described by Dellaportas & Smith (1993). In particular, this paper describes a way of obtaining a predictive distribution from recursive claims reserving models, including the well known model introduced by Mack (1993). Mack's model is useful, since it can be used with data sets which exhibit negative incremental amounts. The techniques are illustrated with examples, and the resulting predictive distributions from both the bootstrap and Bayesian methods are compared.

KEYWORDS

Bayesian; Bootstrap; Chain Ladder; Dynamic Financial Analysis; Generalised Linear Model; Markov Chain Monte Carlo; Reserving Risk; Stochastic Reserving

CONTACT ADDRESS

Dr P. D. England, EMB Consultancy, Saddlers Court, 64-74 East Street, Epsom KT17 1HB, U.K.
E-mail: peter.england@emb.co.uk

1. Introduction

The 'holy grail' of stochastic reserving techniques is to obtain a predictive distribution of outstanding liabilities, incorporating estimation error from uncertainty in the underlying model parameters and process error due to the underlying claims generating process. With many of the stochastic reserving models which have been proposed to date, it is not possible to obtain that distribution analytically, since the distribution of the sum of random variables is required, taking account of estimation error. Where an analytic solution is not possible, progress can still be made by adopting simulation methods.

Two methods have been proposed which produce a simulated predictive

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

distribution: bootstrapping; and Bayesian methods implemented using Markov chain Monte Carlo techniques. We are unaware of any papers in the academic literature comparing the two approaches until now, and as such, this paper aims to fill that gap, and to highlight the similarities and the differences between the approaches. Bootstrapping has been considered by Ashe (1986), Taylor (1988), Brickman *et al.* (1993), Lowe (1994), England & Verrall (1999), England (2002), England & Verrall (2002) and Pinheiro *et al.* (2003), amongst others. Bayesian methods for claims reserving have been considered by Haastrup & Arjas (1996), de Alba (2002), England & Verrall (2002), Ntzoufras & Dellaportas (2002), Verrall (2004) and Verrall & England (2005).

England & Verrall (2002) laid out some of the basic modelling issues, and in this paper we explore further the methods which provide predictive distributions. A general framework for bootstrapping is set out, and illustrated by applying the procedure to recursive models, including Mack's model (Mack, 1993). With Bayesian methods, we set out the theory and show that, with non-informative prior distributions, predictive distributions can be obtained which are very similar to those obtained using bootstrapping methods. Thus, Bayesian methods can be seen as an alternative to bootstrapping in practical applications. We limit ourselves to using non-informative prior distributions to highlight the similarities to bootstrapping, in the hope that a good understanding of the principles and application of Bayesian methods in the context of claims reserving will help the methods to be more widely applied, and make it easier to move on to applications where the real advantages of Bayesian modelling become apparent. By focusing on non-informative prior distributions, we acknowledge that we are presenting a very limited view of the possibilities and power of Bayesian inference.

We believe that Bayesian methods offer considerable advantages in practical terms, and deserve greater attention than they have received so far in practice. Hence, a further aim of this paper is to show that the Bayesian approach with no prior information is only a short step away from the popular bootstrapping methods. Once that step has been made, the Bayesian framework can be used to explore alternative modelling strategies (such as modelling claim numbers and amounts together), and incorporating prior opinion (for example, in the form of manual intervention, or a stochastic Bornhuetter-Ferguson method). Some of these ideas have been explored in the Bayesian papers cited above, and we believe that there is scope for actuaries to progress from the basic stochastic reserving methods, which have now become better-understood, to more sophisticated approaches.

Bootstrapping has proved to be a popular method for a number of reasons, including:

— the ease with which it can be applied;
— the fact that bootstrap estimates can often be obtained in a spreadsheet; and

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

— the possibility of obtaining predictive distributions when combined with simulation for the process error.

However, it is not without its difficulties, for example:

— a small number of sets of 'pseudo' data may be incompatible with the underlying model, and may require modification;
— models which require statistical software to fit them, and do not have an equivalent traditional method, are more difficult to implement;
— there is a limited number of combinations of residuals which can be used when generating pseudo data, which is a potential issue with smaller data sets; and
— the method is open to manipulation, and may not always be implemented appropriately.

The final item in the list above could also be seen as a benefit, and partly explains the popularity of the method, since actuaries can extend the methodology, while broadly obeying its spirit, but losing any clear link between the bootstrapping procedure and a well specified statistical model.

When using bootstrapping to help to obtain a predictive distribution of outstanding claims, it is a common misunderstanding that the approach is 'distribution free'. Furthermore, since the publication of England & Verrall (1999), some readers have incorrectly associated 'bootstrapping', in this context, exclusively with the model presented in that paper (the chain ladder model represented as the over-dispersed Poisson model described in Renshaw & Verrall (1998)). One of the aims of this paper is to correct those misconceptions, and to describe bootstrapping as a general procedure, which, if applied consistently, can be used to obtain the estimation error (standard error) of well specified models. In addition, England (2002) showed that when forecasting into the future, bootstrapping can be supplemented by a simulation approach to incorporate process error, giving a full predictive distribution. The procedure for using bootstrap methods to obtain a predictive distribution for outstanding claims is summarised in Figure 1.

The procedure outlined in this paper for obtaining predictive distributions using Bayesian techniques has many similarities to bootstrapping, and is summarised in Figure 2. The starting point is also a well-specified statistical model. However, instead of using bootstrapping to incorporate estimation error, Markov chain Monte Carlo (MCMC) techniques can be used to provide distributions of the underlying parameters instead. The final forecasting stage is identical in both paradigms.

Comparison with Figure 1 shows that the principal difference between the two approaches is at the second stage, and that, as long as the underlying statistical model can be adequately defined, either methodology could be used. In this paper, we stress the importance of starting with a well-defined statistical model, and show that where the procedures in

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

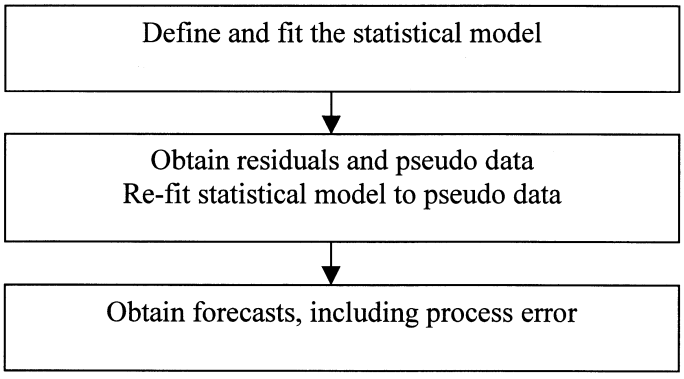


Figure 1. Conceptual framework for obtaining predictive distributions of outstanding liabilities using bootstrap methods

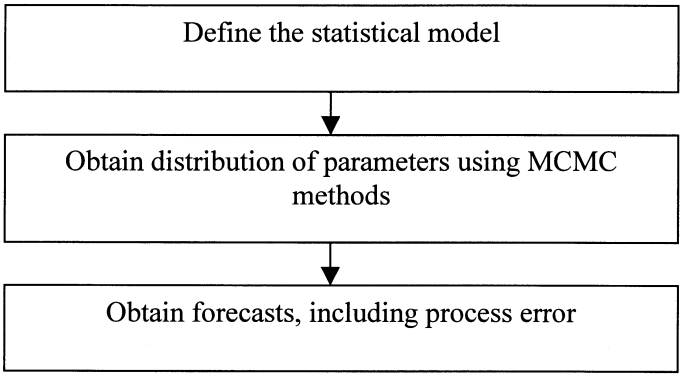


Figure 2. Conceptual framework for obtaining predictive distributions of outstanding liabilities using Bayesian methods

Figure 1 and Figure 2 are followed, it is possible to apply bootstrapping and Bayesian techniques to models which hitherto have not been tried, such as Mack's model (Mack, 1993).

Several stochastic models used for claims reserving can be embedded within the framework of generalised linear models (GLMs). This includes models for the chain ladder technique, that is, the over-dispersed Poisson and negative binomial models, and the method suggested by Mack (1993). It also applies to some models including parametric curves, such as the Hoerl

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

curve, and models based on the lognormal distribution (see Section 8). In all cases, a similar procedure can be followed in order to apply bootstrap and Bayesian methods to obtain the estimation error of the reserve estimates. If the process error is included in a way which is consistent with the underlying model, the results will be analogous to results obtained analytically from the same underlying model. A further aim of this paper is to illustrate this by example, comparing results obtained analytically with results obtained using bootstrap and Bayesian approaches.

This paper is set out as follows. Section 2 contains some basic definitions. Section 3 briefly outlines the stochastic reserving methods which are considered in this paper, and Section 4 summarises how predictions and prediction errors can be calculated analytically. Section 5 considers a general procedure for bootstrapping generalised linear models, and describes how the procedure can be implemented for the models introduced in Section 3. Section 6 considers Bayesian modelling and Gibbs sampling generally, before introducing the application to Bayesian generalised linear models. Section 6 also describes how the Bayesian procedure can be implemented for the models introduced in Section 3. Examples are provided in Section 7, where the results of the bootstrap and Bayesian approaches are compared. A discussion appears in Section 8, and concluding remarks in Section 9.

For readers only interested in bootstrapping, Section 6 can be ignored, and for readers only interested in Bayesian methods, Section 5 can be ignored.

2. The Chain Ladder Technique

For ease of exposition, we assume that the data consist of a triangle of observations. The stochastic methods described in this paper can also be applied to other shapes of data, and the assumption of a triangle does not imply any loss of generality. Thus, we assume that the data consist of a triangle of incremental claims:

$$\begin{aligned} &C_{1,1}, C_{1,2}, \ldots, C_{1,n} \\ &C_{2,1}, \ldots, C_{2,n-1} \\ &\vdots \\ &C_{n,1}. \end{aligned}$$

This can be also written as $\{C_{ij} : i = 1, \ldots, n; j = 1, \ldots, n - i + 1\}$, where n is the number of origin years. C_{ij} is used to denote incremental claims, and D_{ij} is used to denote the cumulative claims, defined by:

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

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

The aim of the exercise is to populate the missing lower portion of the triangle, and to extrapolate beyond the maximum development period where necessary. One traditional actuarial technique which has been developed to do this is the chain ladder technique, which forecasts the cumulative claims recursively, using:

$$\hat{D}_{i,n-i+2} = D_{i,n-i+1}\hat{\lambda}_{n-i+2}$$

and

$$\hat{D}_{i,j} = \hat{D}_{i,j-1}\hat{\lambda}_j \qquad j = n-i+3, n-i+4, \ldots, n$$

where the fitted development factors, denoted by $\{\hat{\lambda}_j : j = 2, \ldots, n\}$, are given by:

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

The fitted development factors may also be written in terms of a weighted average of observed development factors, which are defined as $f_{ij} = \frac{D_{ij}}{D_{i,j-1}}$, giving:

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} D_{i,j-1} f_{ij}}{\sum_{i=1}^{n-j+1} D_{i,j-1}}. \tag{2.1}$$

3. Claims Reserving Models as Stochastic Models

England & Verrall (2002) provide a review of stochastic reserving models for claims reserving based (for the most part) on generalised linear models. This includes models which are related to the chain ladder technique, methods which fit curves to enable extrapolation, and models based on observed development factors. Kaas *et al.* (2002) also present claims reserving models in the framework of generalised linear models.

In this section we provide a brief overview of three stochastic models which can be expressed within the framework of generalised linear models, and which give exactly the same forecasts as the chain ladder technique when parameterised appropriately. This is useful, since it provides a link to traditional actuarial techniques, which can later be generalised.

The distributional assumptions of generalised linear models are usually

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

expressed in terms of the first two moments only, such that, for each 'unit' u of a random variable X:

$$\mathrm{E}[X_u] = m_u \quad \text{and} \quad \mathrm{Var}[X_u] = \frac{\phi V(m_u)}{w_u} \tag{3.1}$$

where ϕ denotes a scale parameter, $V(m_u)$ is the so-called variance function (a function of the mean) and w_u are weights (often set to one for all observations). The choice of distribution dictates the values of ϕ and $V(m_u)$ (see McCullagh & Nelder, 1989).

3.1 *The Over-Dispersed Poisson Model*

The over-dispersed Poisson model is formulated as a 'non-recursive' model, since the forecast claims are fully specified by the model, without requiring knowledge of the cumulative claims at the previous time period. The over-dispersed Poisson model assumes that the incremental claims C_{ij} are distributed as independent over-dispersed Poisson random variables, with mean and variance:

$$\mathrm{E}\big[C_{ij}\big] = m_{ij} \quad \text{and} \quad \mathrm{Var}\big[C_{ij}\big] = \phi m_{ij}. \tag{3.2}$$

The specification is completed by providing a parametric structure for the mean m_{ij}. For example, forecast values consistent with the chain ladder technique (under suitable conditions) can be obtained using:

$$\log\big(m_{ij}\big) = c + \alpha_i + \beta_j. \tag{3.3}$$

In the terminology of generalised linear models, we use a log link function with a predictor structure which has a parameter for each row i, and a parameter for each column j. As a generalised linear model, it is easy to obtain maximum likelihood parameter estimates using standard software packages. Note that constraints have to be applied to the sets of parameters, which could take a number of different forms. For example, the corner constraints put $\alpha_1 = \beta_1 = 0$.

Over-dispersion is introduced through the scale parameter ϕ, which is unknown and estimated from the data (see the Appendix), although usually then treated as a 'plug-in' estimate and not counted as a parameter. Allowing for over-dispersion does not affect estimation of the parameters, but has the effect of increasing their standard errors. Full details of this model can be found in Renshaw & Verrall (1998).

The restriction that the scale parameter is constant for all observations can be relaxed. It is common to allow the scale parameters to depend on development period j, in which case, in a maximum likelihood setting, the

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

scale parameters ϕ_j can be estimated as part of an extended fitting procedure known as 'joint' modelling (see the Appendix).

Although the model in this section is based on the Poisson distribution, this does not imply that it is only suitable for data consisting exclusively of positive integers. That constraint can be overcome using a 'quasi-likelihood' approach (see McCullagh & Nelder, 1989), which can be applied to non-integer data. With quasi-likelihood, in this context, the likelihood is the same as a Poisson likelihood up to a constant of proportionality. For data consisting entirely of positive integers, and using a constant scale parameter, identical parameter estimates are obtained using the full or quasi-likelihood. In modelling terms, the crucial assumption is that the variance is proportional to the mean, and the data are not restricted to being positive integers. The derivation of the quasi-log-likelihood for this model is considered in Section 6.1.

3.2 *The Over-Dispersed Negative Binomial Model*

The over-dispersed negative binomial (ONB) model is formulated as a 'recursive' model, since the forecast claims are a multiple of the cumulative claims at the previous time period. Building on the over-dispersed Poisson chain ladder model, Verrall (2000) developed the over-dispersed negative binomial chain ladder model and showed that the same predictive distribution can be obtained. The model developed by Verrall (2000) uses a recursive approach, where the incremental claims, C_{ij}, have mean and variance:

$$\mathrm{E}\left[C_{ij} \mid D_{i,j-1}\right] = \left(\lambda_j - 1\right)D_{i,j-1} \text{ and } \mathrm{Var}\left[C_{ij} \mid D_{i,j-1}\right] = \phi\lambda_j\left(\lambda_j - 1\right)D_{i,j-1} \text{ for } j \geq 2.$$

It should be noted that the incremental claims are conditionally independent of the cumulative claims at the previous time period, and assumed to be independent of the incremental claims in other origin periods. By adding the previous cumulative claims, the equivalent model for cumulative claims, D_{ij}, has mean and variance:

$$\mathrm{E}\left[D_{ij} \mid D_{i,j-1}\right] = \lambda_j D_{i,j-1} \quad \text{and} \quad \mathrm{Var}\left[D_{ij} \mid D_{i,j-1}\right] = \phi\lambda_j\left(\lambda_j - 1\right)D_{i,j-1} \quad \text{for} \quad j \geq 2.$$

Because the incremental claims at time j are conditionally independent of the cumulative claims at the previous time period, the cumulative claims at time j are also conditionally independent of the cumulative claims at time $(j-1)$.

It is convenient to write this model in terms of the observed development factors, f_{ij}, where:

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

$$f_{ij} = \frac{D_{ij}}{D_{i,j-1}}$$

such that the development factors f_{ij} are conditionally independent and have mean and variance:

$$\mathrm{E}\left[f_{ij} \mid D_{i,j-1}\right] = \lambda_j \quad \text{and} \quad \mathrm{Var}\left[f_{ij} \mid D_{i,j-1}\right] = \frac{\phi\lambda_j(\lambda_j - 1)}{D_{i,j-1}} \quad \text{for} \quad j \geq 2. \quad (3.4)$$

The specification is completed by providing a parametric structure for the expected development factors λ_j. For example, forecast values consistent with the chain ladder technique (under suitable conditions) can be obtained using:

$$\log(\log(\lambda_j)) = \gamma_j. \quad (3.5)$$

Use of the log-log link function ensures that the fitted development factors are greater than one, otherwise the variance is undefined.

Again, over-dispersion is introduced through the scale parameter ϕ, which is estimated from the data (see the Appendix), and usually then treated as a 'plug-in' estimate. Again, the assumption that the scale parameter is constant for all observations can be relaxed, and it is common to allow the scale parameters to depend on development period j, in which case, in a maximum likelihood setting, the scale parameters can be estimated using 'joint' modelling (see the Appendix).

Like the over-dispersed Poisson model, a quasi-likelihood approach is adopted that can be applied to non-integer data. The derivation of the quasi-log-likelihood for this model is considered in Section 6.2.

3.3 *Mack's Model*

The model introduced by Mack (1993) is also a recursive model. Mack focused on the cumulative claims D_{ij} as the response with mean and variance:

$$\mathrm{E}\left[D_{ij} \mid D_{i,j-1}\right] = \lambda_j D_{i,j-1} \quad \text{and} \quad \mathrm{Var}\left[D_{ij} \mid D_{i,j-1}\right] = \sigma_j^2 D_{i,j-1} \quad \text{for} \quad j \geq 2.$$

Like the negative binomial model, Mack's model assumes that the cumulative claims at time j are conditionally independent of the cumulative claims at time $(j-1)$. Mack considered the model to be 'distribution free', since only the first two moments of the cumulative claims are specified, not the full distribution. Mack also derived expressions for the estimators of λ_j and σ_j^2. England & Verrall (2002) showed that the same estimators are obtained assuming the cumulative claims D_{ij} are normally distributed.

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

England & Verrall (2002) also showed that an equivalent formulation can be obtained using the observed development factors f_{ij}, with mean and variance:

$$\mathrm{E}\left[f_{ij} \mid D_{i,j-1}\right] = \lambda_j \quad \text{and} \quad \mathrm{Var}\left[f_{ij} \mid D_{i,j-1}\right] = \frac{\sigma_j^2}{D_{i,j-1}} \quad \text{for} \quad j \geq 2. \tag{3.6}$$

The specification is completed by providing a parametric structure for the expected development factors λ_j. For example, forecast values consistent with the chain ladder technique can be obtained using:

$$\log(\lambda_j) = \gamma_j. \tag{3.7}$$

Use of the log link function ensures that the fitted development factors are greater than zero, otherwise the model does not make sense in the context of claims reserving. This formulation, along with the assumption of normality, allows modelling with negative incremental claims without difficulty, making the methods suitable for use with incurred data, which often exhibit negative incrementals in later development periods due to earlier over-estimation of case reserves. In England & Verrall (2002), the model was fitted as a weighted normal regression model, with weights $D_{i,j-1}$ (assumed to be fixed and known). The derivation of the log-likelihood for this model is considered in Section 6.3.

4. Predictions, Prediction Errors and Predictive Distributions

Claims reserving is a predictive process; given the data, we try to predict future claims. In Section 3 different models have been outlined from which future claims can be predicted. In this context, we use the expected value as the prediction. In classical statistics, the expected value is usually evaluated using maximum likelihood parameter estimates. When using Bayesian statistics, or when bootstrapping, the expected value of the predictive distribution is used. Obtaining the predictive distribution requires an additional simulation step when forecasting, to include the process error (see the final step in Figure 1). The way that this additional step is incorporated differs for recursive and non-recursive models, and is covered in Sections 5 and 6.

When considering variability, in classical statistics the root mean square error of prediction (RMSEP) can be obtained, also known as the prediction error. When using Bayesian statistics, or when bootstrapping, the analogous measure is the standard deviation of the predictive distribution. It should be possible to compare the results from the different approaches, and explain any observed differences.

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

In classical statistics, the RMSEP may not be straightforward to obtain. For a single value in the future C_{ij}, say (where $j > n - i + 1$), the mean squared error of prediction (MSEP) is the expected squared difference between the actual outcome and the predicted value:

$$\begin{aligned}\mathrm{E}\left[\left(C_{ij} - \hat{C}_{ij}\right)^2\right] &= \mathrm{E}\left[\left(\left(C_{ij} - \mathrm{E}[C_{ij}]\right) - \left(\hat{C}_{ij} - \mathrm{E}[C_{ij}]\right)\right)^2\right] \\ &\simeq \mathrm{E}\left[\left(C_{ij} - \mathrm{E}[C_{ij}]\right)^2\right] + \mathrm{E}\left[\left(\hat{C}_{ij} - \mathrm{E}[\hat{C}_{ij}]\right)^2\right].\end{aligned}$$

That is, the prediction variance = process variance + estimation variance, and the problem reduces to estimating the two components. It should be noted that the independence assumptions in all the models proposed in Section 3 imply that the future observations are (conditionally) independent of past data, and as such, the relations shown above hold.

Whilst it is possible to calculate these quantities for a single forecast C_{ij}, the prediction variance for sums of observations is useful in the reserving process. For example, the row sum of predicted values and the overall reserve (up to development year n) are:

$$\sum_{j=n-i+2}^{n} C_{ij} \quad \text{and} \quad \sum_{i=2}^{n} \sum_{j=n-i+2}^{n} C_{ij}, \text{ respectively.} \tag{4.1}$$

The prediction variances for these quantities may not be straightforward to calculate directly, and can be a deterrent to the practical application of stochastic reserving. England & Verrall (2002) show how the quantities can be calculated for the models given in Section 3. In a bootstrap or Bayesian context, the quantities are straightforward to evaluate; they are simply the variances of the respective simulated predictive distributions. It is preferable to have a full predictive distribution, rather than just the first two moments, since any measure on the predictive distribution can be evaluated, and the predictive distribution can be used, for example, in capital modelling. Bootstrap and Bayesian methods have the advantage that a predictive distribution is generated automatically.

5. Bootstrapping Generalised Linear Models

When bootstrapping generalised linear models, the first stage is defining and fitting the statistical model (see Figure 1). This is straightforward for any of the models described in Section 3. In the case of models which give the same estimates as the chain ladder technique, this is particularly easy, because the chain ladder method itself can be used to obtain fitted values. In that

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

special case it is possible to avoid using any specialist software; the calculations can be carried out in a spreadsheet.

The second stage is when bootstrapping is applied, which involves creating new sets of 'pseudo' data, using the data in the original triangle. A key requirement of bootstrapping is that the 'observations' used for bootstrapping must be independent and identically distributed (see Efron & Tibshirani, 1993). With regression-type problems, the data are usually assumed to be independent, but are not identically distributed, since the means (and possibly the variances) depend on covariates. Therefore, with regression-type models, it is common to bootstrap the residuals rather than the data themselves, since the residuals are approximately independent and identically distributed, or can be made so. The residual definition must be consistent with the model being fitted, and it is usual to use Pearson residuals in this context. A random sample of the residuals is taken (using sampling with replacement), together with the fitted values, and new 'pseudo' data values are obtained by inverting the definition of the residuals. This is repeated many times, and the model refitted to each set of pseudo data, giving a distribution of parameter estimates.

The final forecasting stage extends bootstrapping to provide forecast values (based on the distribution of parameter estimates), incorporating process error. The exact details of this process differ slightly depending on the type of model that has been used, and further details are given in Sections 5.1, 5.2 and 5.3.

For linear regression models with homoscedastic normal errors, the residuals are simply the observed values less the fitted values, but for GLMs an extended definition of residuals is required which have (approximately) the usual properties of normal theory residuals. Several different types of residuals have been suggested for use with GLMs, for example Deviance, Pearson and Anscombe residuals, where the precise form of the residual definitions is dictated by the distributional assumptions. In this paper, we have used the scaled (or 'modified') Pearson residuals when bootstrapping, defined as:

$$r_u = r_{PS}\left(X_u, \hat{m}_u, w_u, \hat{\phi}\right) = \frac{X_u - \hat{m}_u}{\sqrt{\dfrac{\hat{\phi} V(\hat{m}_u)}{w_u}}}. \tag{5.1}$$

When performing diagnostic checks, the scaled Pearson residuals have the usual interpretation that approximately 95% of scaled residuals are expected to lie in the interval $(-2, +2)$ for a reasonable model.

The bootstrapping process involves sampling, with replacement, from the set of actual residuals $\{r_u : u = 1, \ldots, N\}$, to produce a bootstrap sample of residuals $\{r_u^B : u = 1, \ldots, N\}$, where $N = \frac{1}{2}n(n+1)$ for the over-dispersed Poisson model and $N = \frac{1}{2}n(n-1)$ for the recursive models.

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

This provides a sample of residuals for a single bootstrap iteration. A set of pseudo data is then obtained, using the bootstrap sample together with the fitted values, by backing out the residual definition. The Pearson residuals are useful in this context since they can usually be inverted analytically, such that $X_u^B = r_{PS}^{-1}\left(r_u^B, \hat{m}_u, w_u, \hat{\phi}\right)$, giving:

$$X_u^B = r_u^B \sqrt{\frac{\hat{\phi} V(\hat{m}_u)}{w_u}} + \hat{m}_u. \tag{5.2}$$

The same model is then fitted to the pseudo data (using exactly the same model definition used to obtain the residuals), to obtain bootstrap parameter estimates for the first iteration. The process is then repeated many times to give a bootstrap distribution of parameter estimates.

The bootstrap distribution of parameter estimates can be used in a number of ways, for example to obtain: a bootstrap estimate of the standard error of the parameters (by taking the standard deviation of the distribution of each parameter in turn); a bootstrap estimate of the covariance matrix of the parameters; or a bootstrap estimate of the standard error of the fitted values. When used to forecast into the future, a bootstrap estimate of the prediction error can be obtained when combined with an additional step to incorporate the process error.

When bootstrapping models with a constant scale parameter it is not necessary to scale the residuals (by the square root of the scale parameter), since the scaling is unwound when inverting the definition of the residual when constructing the pseudo data. However, if non-constant scale parameters are used (such that ϕ is replaced by ϕ_u), it is essential to scale the residuals first. Further details of the estimation of the scale parameters appear in the Appendix. Adjustments to the residuals to take account of the trade-off between goodness-of-fit and degrees of freedom are also described in the Appendix, which enable a comparison on a consistent basis between bootstrap results and results obtained analytically.

England & Verrall (1999) and England (2002) used the approach described above to provide bootstrap estimates of the prediction error of outstanding liabilities associated with the over-dispersed Poisson model described in Section 3.1. England & Verrall (1999) also stated that the approach could be used for other models, such as log-normal and Gamma models. Pinheiro *et al.* (2003) provided a more general description of bootstrapping for claims reserving models within the generalised linear model framework, and provided illustrative examples for the over-dispersed Poisson model and Gamma models (with constant scale parameters). In this paper the methods are extended further to consider models with non-constant scale parameters and recursive models, such as the model introduced by Mack (1993). Mack's model is popular amongst practitioners,

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

although it only provides prediction errors calculated analytically. This paper describes how to bootstrap Mack's model, providing a simulated predictive distribution, which hitherto has not been available, in addition to (approximate) prediction errors.

Precise details for the models described in Sections 3.1, 3.2 and 3.3 are contained in the following sections.

5.1 *The Over-Dispersed Poisson Model*

Since bootstrapping the over-dispersed Poisson model has been described previously (in England & Verrall, 1999; England, 2002), a brief description only is included here. For the over-dispersed Poisson model, the response variable is the incremental claims C_{ij}, and from equation (3.2):

$$\mathrm{E}[C_{ij}] = m_{ij} \quad \text{and} \quad \mathrm{Var}[C_{ij}] = \phi \mathrm{E}[C_{ij}].$$

Therefore, in terms of equation (3.1), $X_u = C_{ij}$ and $V(m_u) = m_{ij}$. Then, from equation (5.1), the scaled Pearson residuals are defined as:

$$r_{ij} = r_{PS}(C_{ij}, \hat{m}_{ij}, \hat{\phi}) = \frac{C_{ij} - \hat{m}_{ij}}{\sqrt{\hat{\phi}\hat{m}_{ij}}}.$$

The pseudo data are then defined as:

$$C_{ij}^{B} = r_{ij}^{B}\sqrt{\hat{\phi}\hat{m}_{ij}} + \hat{m}_{ij}$$

and the model used to obtain the residuals can be fitted to each triangle of pseudo data.

When the model has been fitted using the linear predictor defined in equation (3.3), giving the same forecasts as the chain ladder model, a number of short cuts can be made, and the process can be implemented in a spreadsheet, as described in England (2002); that is, the fitted values can be obtained by backwards recursion using the traditional chain ladder development factors, and the chain ladder model can be used to fit the model and obtain forecasts at each bootstrap iteration. If alternative predictor structures have been used, such as predictors including calendar year terms or parametric curves, the model must be fitted using suitable software capable of fitting GLMs.

Since this is a non-recursive model, bootstrap forecasts $\tilde{C}_{ij}$, excluding process error, can be obtained for the complete lower triangle of future values, that is:

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

$$\tilde{C}_{ij} = \hat{m}_{ij}^{B} \quad \text{for} \quad i = 2, \ldots, n \quad \text{and} \quad j = n - i + 2, n - i + 3, \ldots, n.$$

Extrapolation beyond the final development period can be used where curves have been fitted in order to estimate tail factors.

To add the process error, a forecast value C_{ij}^{*} can then be simulated from an over-dispersed Poisson distribution with mean $\tilde{C}_{ij}$ and variance $\hat{\phi}\tilde{C}_{ij}$. There is a number of ways in which this can be achieved, and England (2002) makes several suggestions. In this paper, we simply use a gamma distribution with the target mean and variance as a reasonable approximation.

The forecasts can then be aggregated using equation (4.1) to provide predictive distributions of the outstanding liabilities.

When non-constant scale parameters are used, the procedure is identical, except that the constant scale parameter ϕ is replaced by ϕ_j.

5.2 *The Over-Dispersed Negative Binomial Model*

From Section 3.2, using the development ratios f_{ij} as the response variable, this gives:

$$\mathrm{E}\left[f_{ij} \mid D_{i,j-1}\right] = \lambda_j \quad \text{and} \quad \mathrm{Var}\left[f_{ij} \mid D_{i,j-1}\right] = \frac{\phi\lambda_j(\lambda_j - 1)}{D_{i,j-1}} \quad \text{for} \quad j \geq 2.$$

Therefore, in terms of equation (3.1), $X_u = f_{ij}$, $m_u = \lambda_j$, $w_u = D_{i,j-1}$ and $V(m_u) = \lambda_j(\lambda_j - 1)$. Then from equation (5.1), the scaled Pearson residuals are defined as:

$$r_{ij} = r_{PS}\left(f_{ij}, \hat{\lambda}_j, w_{ij}, \hat{\phi}\right) = \frac{\sqrt{w_{ij}}\left(f_{ij} - \hat{\lambda}_j\right)}{\sqrt{\hat{\phi}\hat{\lambda}_j(\hat{\lambda}_j - 1)}}.$$

Notice that this is now a model of the ratios f_{ij}, so the pseudo data are defined as:

$$f_{ij}^{B} = r_{ij}^{B}\frac{\sqrt{\hat{\phi}\hat{\lambda}_j(\hat{\lambda}_j - 1)}}{\sqrt{w_{ij}}} + \hat{\lambda}_j.$$

The model used to obtain the residuals can be fitted to each triangle of pseudo data to obtain new fitted development factors $\tilde{\lambda}_j$.

When the model has been fitted using the linear predictor defined in equation (3.5), giving the same forecasts as the chain ladder model, a number of short cuts can be made, and the process can be implemented in a spreadsheet; that is, the fitted values $\hat{\lambda}_j$ are the traditional chain ladder development factors and can be obtained using equation (2.1). Furthermore,

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

at each bootstrap iteration, the bootstrap development factors can be obtained as a weighted average of the bootstrap ratios using:

$$\tilde{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} w_{ik} f_{ik}^B}{\sum_{i=1}^{n-j+1} w_{ik}}.$$

The reason for re-naming $D_{i,j-1}$ as w_{ij} is to emphasise that it is treated as a known weight, and not re-sampled in the bootstrapping process. This is a crucial point to note when bootstrapping recursive models: the (weighted) residuals are calculated from an underlying generalised linear model using weights that are assumed to be fixed and known; and the same model must also be fitted in each bootstrap iteration. Alternatively, it might be tempting to use as weights the pseudo values created in each bootstrap iteration, but this does not give results which are consistent with prediction variances obtained analytically from the same model.

If alternative predictor structures have been used, such as predictors including parametric curves, the bootstrap development factors $\tilde{\lambda}_j$ must be fitted using suitable software capable of fitting GLMs.

If we simply required the standard error of the development factors, we could stop at this point and calculate the standard deviation of the bootstrap sample of development factors $\tilde{\lambda}_j$. However, the aim is to obtain a predictive distribution of the outstanding liabilities, using the final forecasting step of Figure 1, including process error. The way in which this is implemented for the negative binomial model is different from the method used for the Poisson model, since the negative binomial model is a recursive model.

With recursive models, forecasting proceeds one step at a time. Starting from the latest cumulative claims, the one-step-ahead forecasts can be obtained for each bootstrap iteration by drawing a sample from the underlying process distribution; that is, for $i = 2, 3, \ldots, n$:

$$D^*_{i,n-i+2} \mid D_{i,n-i+1} \sim ONB\big(\tilde{\lambda}_j D_{i,n-i+1}, \hat{\phi}\tilde{\lambda}_j\big(\tilde{\lambda}_j - 1\big)D_{i,n-i+1}\big).$$

Again, there is a number of ways in which this can be achieved, and in this paper we simply use a gamma distribution with the target mean and variance as a reasonable approximation.

It should be noted that the original data $D_{i,n-i+1}$ are used for the one-step-ahead forecast rather than pseudo values. This is a direct result of using a recursive model, and is required to give prediction errors which are consistent with those obtained analytically from the same model.

The two-steps-ahead forecasts, and beyond, are obtained in a similar way, except that the previous simulated forecast cumulative claims are used, including the process error added at the previous step; that is, for $i = 3, 4, \ldots, n$ and $j = n - i + 3, n - i + 4, \ldots, n$, D^*_{ij} is simulated using:

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

$$D^*_{i,j} \mid D^*_{i,j-1} \sim ONB\big(\tilde{\lambda}_j D^*_{i,j-1}, \hat{\phi}\tilde{\lambda}_j(\tilde{\lambda}_j - 1)D^*_{i,j-1}\big).$$

Note that this procedure includes both the estimation error, through bootstrapping, and the process error, because a forecast value is simulated at each step. In contrast, if the aim was solely to calculate the estimation error (standard error), it would be sufficient just to project forward from the latest cumulative to ultimate claims using:

$$\tilde{D}_{i,n} = D_{i,n-i+1}\tilde{\lambda}_{n-i+2}\tilde{\lambda}_{n-i+3}\ldots\tilde{\lambda}_n.$$

It can be seen that the difference is that, in order to obtain prediction errors which are consistent with prediction errors obtained analytically from the same model, the process error is included at each step before proceeding.

The forecast incremental claims can be obtained by differencing in the usual way, and can then be aggregated using equation (4.1) to provide predictive distributions of the outstanding liabilities.

Like the over-dispersed Poisson model, when non-constant scale parameters are used, the procedure is identical, except that the constant scale parameter ϕ is replaced by ϕ_j.

5.3 *Mack's Model*

The procedure for bootstrapping Mack's model is almost identical to the procedure for the negative binomial model, since it is also a recursive model. The differences are in the underlying distributional assumptions, which dictate the definition used for the residuals, and hence the calculation of scale parameters. This highlights that, in this context, bootstrapping cannot strictly be considered 'distribution free', since distributional assumptions must be made when defining the statistical models (see Figure 1) and obtaining estimators of key parameters.

From equation (3.6), using the development ratios f_{ij} as the response variable, gives:

$$\mathrm{E}\big[f_{ij} \mid D_{i,j-1}\big] = \lambda_j \quad \text{and} \quad \mathrm{Var}\big[f_{ij} \mid D_{i,j-1}\big] = \frac{\sigma_j^2}{D_{i,j-1}} \quad \text{for} \quad j \geq 2.$$

Therefore, in terms of equation (3.1), $X_u = f_{ij}$, $m_u = \lambda_j$, $w_u = D_{i,j-1}$ and $V(m_u) = 1$, and the model is defined using non-constant scale parameters $\phi_j = \sigma_j^2$. Then from equation (5.1), the scaled Pearson residuals are defined as:

$$r_{ij} = r_{PS}\big(f_{ij}, \hat{\lambda}_j, w_{ij}, \hat{\sigma}_j\big) = \frac{\sqrt{w_{ij}}\big(f_{ij} - \hat{\lambda}_j\big)}{\hat{\sigma}_j},$$

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

giving pseudo data:

$$f_{ij}^B = r_{ij}^B \frac{\hat{\sigma}_j}{\sqrt{w_{ij}}} + \hat{\lambda}_j.$$

The model used to obtain the residuals can be fitted to each triangle of pseudo data, to obtain new fitted development factors $\tilde{\lambda}_j$.

When the model has been fitted using the linear predictor defined in equation (3.7), giving the same forecasts as the chain ladder model, a number of short cuts can be made, and the process can be implemented in a spreadsheet, as described in Section 5.2. Again, the reason for re-naming $D_{i,j-1}$ as w_{ij} is to emphasise that it is treated as a weight which is fixed and known.

If alternative predictor structures have been used, such as predictors including parametric curves, the bootstrap development factors $\tilde{\lambda}_j$ must be fitted using suitable software capable of fitting weighted normal regression models.

Like the negative binomial model, forecasting proceeds one step at a time. Starting from the latest cumulative claims, the one-step-ahead forecasts can be obtained for each bootstrap iteration by drawing a sample from the underlying process distribution. That is, for $i = 2, 3, \ldots, n$:

$$D^*_{i,n-i+2} \mid D_{i,n-i+1} \sim \text{Normal}\big(\tilde{\lambda}_j D_{i,n-i+1}, \hat{\sigma}_j^2 D_{i,n-i+1}\big).$$

The two-step ahead forecasts, and beyond, are obtained using:

$$D^*_{i,j} \mid D^*_{i,j-1} \sim \text{Normal}\big(\tilde{\lambda}_j D^*_{i,j-1}, \hat{\sigma}_j^2 D^*_{i,j-1}\big) \quad \text{for} \quad i = 3, 4, \ldots, n$$
$$\text{and} \quad j = n - i + 3, n - i + 4, \ldots, n.$$

Notice that use of a Normal distribution implicitly allows the simulation of negative cumulative claims (for large σ_j^2), which is an undesirable property. Where this is likely to occur, a practical compromise is to use a Gamma distribution instead, say, with the same mean and variance. Use of a Gamma distribution would still allow negative incremental claims, since the cumulative claims could reduce while still being positive.

Again, the forecast incremental claims can be obtained by differencing in the usual way, and can then be aggregated using equation (4.1) to provide predictive distributions of the outstanding liabilities.

6. Bayesian Generalised Linear Models

When implementing Bayesian generalised linear models, the first stage is also defining the statistical model (see Figure 2), and again this is straightforward for any of the models described in Section 3.

The second stage involves obtaining a distribution of parameters. This

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

has been simplified enormously in recent years due to the advent of numerical methods based on Markov chain Monte Carlo (MCMC) techniques. An excellent overview of MCMC methods with applications in actuarial science is provided by Scollnik (2001), although Klugman (1992), Makov *et al.* (1996) and Makov (2001) also discuss Bayesian methods in actuarial science. Dey *et al.* (2000) provide a theoretical overview of generalised linear models from a Bayesian perspective.

The final forecasting stage extends the methodology to provide forecast values (based on the distribution of parameters), incorporating the process error. This stage is exactly the same for the bootstrap and Bayesian approaches.

Since the use of Bayesian methods is still uncommon in actuarial applications, a brief overview is included here. In general terms, given a random variable X with corresponding density $f(x_u \mid \theta)$, with parameter vector θ, the likelihood function for the parameters, given the data, is given by $L(\theta \mid X) = \prod_u f(x_u \mid \theta)$. In Bayesian modelling, the likelihood function is combined (using Bayes' Theorem) with prior information on the parameters in the form of prior density $\pi(\theta)$, to obtain a posterior joint distribution of the parameters: $f(\theta \mid X) \propto L(\theta \mid X)\pi(\theta)$. MCMC techniques obtain samples from the posterior distribution of the parameters by simulating in a particular way. In this paper, we consider MCMC techniques implemented using Gibbs sampling.

Gibbs sampling is straightforward to apply, and involves simply populating a grid with values, where the rows of the grid relate to iterations of the Gibbs sampler, and the columns relate to parameters. For example, if t iterations of the Gibbs sampler are required, and there are k parameters, then it is necessary to populate a t by k grid. Given parameter vector $\theta = (\theta_1, \ldots, \theta_k)$, and arbitrary starting values $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_k^{(0)})$, the first iteration of Gibbs sampling proceeds one parameter at a time, by making random draws from the full conditional distribution of each parameter, as follows:

$$
\begin{aligned}
&\theta_1^{(1)} \sim f\left(\theta_1 \mid \theta_2^{(0)}, \ldots, \theta_k^{(0)}\right)\\
&\theta_2^{(1)} \sim f\left(\theta_2 \mid \theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_k^{(0)}\right)\\
&\vdots\\
&\theta_j^{(1)} \sim f\left(\theta_j \mid \theta_1^{(1)}, \ldots, \theta_{j-1,}^{(1)}\theta_{j+1}^{(0)}, \ldots, \theta_k^{(0)}\right)\\
&\vdots\\
&\theta_k^{(1)} \sim f\left(\theta_k \mid \theta_1^{(1)}, \ldots, \theta_{k-1}^{(1)}\right).
\end{aligned}
$$

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

This completes a single iteration of the Gibbs sampler, populates the first row of the grid, and defines the transition from $\theta^{(0)}$ to $\theta^{(1)}$. The process starts again for the transition from $\theta^{(1)}$ to $\theta^{(2)}$. Note that for each parameter, the most recent information to date for the other parameters is always used (hence it is a Markov chain), and random draws are made for each parameter in turn, breaking down a multiple parameter problem into a sequence of one parameter problems. After a sufficiently large number of iterations, $\theta^{(t+1)}$ is considered a random sample from the underlying joint distribution. In theory, the whole process should be repeated, starting from new arbitrary starting values, and the new $\theta^{(t+1)}$ retained as another sample from the underlying joint distribution. In practice, it is more common to continue beyond t for another m iterations (once the Markov chain has 'converged'), and retain $\theta^{(t+1)}, \cdots, \theta^{(t+m)}$ as a simulated sample from the underlying joint posterior distribution, rejecting $\theta^{(1)}, \cdots, \theta^{(t)}$ as a 'burn-in' sample of size t.

Although Gibbs sampling itself is a straightforward process to apply, the difficulty arises in making random draws from the full conditional distribution of each parameter. Even factorising the full joint posterior distribution into the conditional distributions may be troublesome (or impossible), and it is often not possible to recognise the conditional distributions as standard distributions. However, since the conditional distributions are proportional to the joint posterior distribution, it is often easier to simply treat the joint posterior distribution sequentially as a function of each parameter (the other parameters being fixed), combined with a generic sampling algorithm for obtaining the random samples.

Several generic samplers exist for efficiently generating random samples from a given density function, for example Adaptive Rejection Sampling (Gilks & Wild, 1992) and Adaptive Rejection Metropolis Sampling (Gilks *et al.*, 1995). Gibbs sampling is usually used in combination with a generic sampler to make the random draws from the conditional distributions $f(\theta_j \mid \theta_{-j})$.

Dellaportas & Smith (1993) showed that Gibbs sampling, combined with Adaptive Rejection Sampling (Gilks & Wild, 1992), provides a straightforward computational procedure for Bayesian inference with generalised linear models. Dellaportas & Smith illustrated their approach with an example based on a GLM with a binomial error structure and a quadratic predictor. Generalising their example, the posterior log-likelihood can be written as:

$$\text{Log L}(\theta \mid \text{X}) = \text{Log}\,(\pi(\theta)) + \sum_u \text{Log}\,(f(x_u \mid \theta)) + \text{constant}$$

where the first component in the sum relates to the prior distribution of the parameters and the final component is the standard log-likelihood of the GLM. Dellaportas & Smith used a multivariate normal prior, giving:

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

$$\text{Log L}(\theta \mid \text{X}) = -\tfrac{1}{2}(\theta - \theta_0)' D_0^{-1}(\theta - \theta_0) + \sum_u \text{Log}\,(f(x_u \mid \theta)) + \text{constant} \quad (6.1)$$

where θ_0 is a prior mean vector and D_0 is a prior covariance matrix. The first expression in the sum simply represents the kernel of a multivariate normal distribution. With independent normal priors, the expression simplifies further.

Using independent non-informative uniform priors, the posterior log-likelihood is simply:

$$\text{Log L}(\theta \mid \text{X}) = \sum_u \text{Log}\,(f(x_u \mid \theta)) + \text{constant}. \quad (6.2)$$

Dellaportas & Smith sampled from the full conditional distribution of each parameter, up to proportionality, by taking the form of the joint posterior likelihood and regarding it successively as functions of each parameter in turn, treating the other parameters as fixed.

Using a similar approach, we have successfully used both multivariate normal and uniform priors, although the results reported in this paper use uniform priors only. In this paper, we use Adaptive Rejection Metropolis Sampling (ARMS) within Gibbs sampling, using the joint posterior distribution, $f(\theta \mid X)$, treated sequentially as a function of each parameter. The methodology was implemented using Igloo Professional with ExtrEMB (2005), although early prototypes were implemented using Excel as a front-end to the ARMS program (written in C) described in Gilks *et al.* (1995) and freely available on the internet. We have also implemented some of the models using WinBUGS (Spiegelhalter *et al.*, 1996), again freely available on the internet.

When maximum likelihood estimates of the underlying GLM would be obtained using a quasi-likelihood approach, we use the quasi-likelihood to construct the posterior log-likelihood. Also, when dispersion parameters are required, these are treated as fixed and known 'plug-in' estimates; that is, a prior distribution for the dispersion parameters is not supplied and they are not sampled within the Gibbs sampling procedure. A discussion of both of these points appears in Section 8.

A derivation of the log-likelihood (or quasi-log-likelihood), $\text{Log}\,(f(x_u \mid \theta))$, for the models specified in Section 3 follows in the next sections.

6.1 *The Over-Dispersed Poisson Model*

For a random variable X, with $\text{E}[X] = m$ and $\text{Var}[X] = \sigma^2 V(m)$, McCullagh & Nelder (1989) define the quasi-log-likelihood $Q(x; m)$ for a single component $x \in X$ as:

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

$$Q(x; m, \sigma) = \int_x^m \frac{x - t}{\sigma^2 V(t)} dt. \tag{6.3}$$

Following on from Section 3.1, and writing $\sigma^2 = \phi$, the quasi-log-likelihood is given by:

$$\begin{aligned} Q(C_{ij}; m_{ij}, \phi) &= \int_{C_{ij}}^{m_{ij}} \frac{C_{ij} - t}{\phi t} dt \\ &= \phi^{-1}\big(C_{ij} \log(m_{ij}) - m_{ij} - C_{ij} \log(C_{ij}) + C_{ij}\big). \end{aligned}$$

Collecting together terms which involve the parameters only gives:

$$\text{Log L}_{ODP} = \sum_{i=1}^{n} \sum_{j=1}^{n-j+1} \phi^{-1}\big(C_{ij} \log(m_{ij}) - m_{ij}\big) + \text{constant}. \tag{6.4}$$

The (quasi-)log-likelihood has been written in general form to allow for any model structure, including structures which incorporate parametric curves, smoothers, and terms relating to calendar periods.

Equation (6.4) can then be used with equation (6.1) or (6.2), and Gibbs sampling used to provide a distribution of parameter estimates, which can then be used in the forecasting procedure.

In a Bayesian context, forecasting proceeds in exactly the same way as described in Section 5.1 for bootstrapping. That is, given the simulated posterior distribution of parameters from Gibbs sampling, the parameters can be combined for each iteration to give an estimate of the future claims $\tilde{C}_{ij}$. To add the process error, a forecast value, C^*_{ij}, can then be simulated from an over-dispersed Poisson distribution with mean $\tilde{C}_{ij}$ and variance $\hat{\phi}\tilde{C}_{ij}$. The forecasts can then be aggregated using equation (4.1) to provide predictive distributions of the outstanding liabilities.

When non-constant scale parameters are used, the procedure is identical, except that the constant scale parameter ϕ is replaced by ϕ_j in the construction of the quasi-log-likelihood, and when forecasting.

6.2 *The Over-Dispersed Negative Binomial Model*

From Section 3.2, using the development ratios f_{ij} as the response variable, and writing $\sigma^2 = \frac{\phi}{D_{i,j-1}}$ in equation (6.3), the quasi-log-likelihood is given by:

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

$$Q(f_{ij};\lambda_j,\phi,D_{i,j-1}) = \int_{f_{ij}}^{\lambda_j} \frac{D_{i,j-1}(f_{ij}-t)}{\phi t(t-1)}\,dt$$
$$= \frac{D_{i,j-1}}{\phi}\big((f_{ij}-1)\log(\lambda_j-1) - f_{ij}\log(\lambda_j) - (f_{ij}-1)\log(f_{ij}-1)$$
$$+ f_{ij}\log(f_{ij})\big).$$

Collecting together terms that involve the parameters only gives:

$$\text{Log}\,L_{ONB} = \sum_{i=1}^{n-1}\sum_{j=2}^{n-i+1} \frac{D_{i,j-1}}{\phi}\big((f_{ij}-1)\log(\lambda_j-1) - f_{ij}\log(\lambda_j)\big) + \text{constant}. \quad (6.5)$$

Again the (quasi-)log-likelihood has been written in general form to allow for any model structure, including structures that incorporate parametric curves and smoothers.

Equation (6.5) can then be used with equation (6.1) or (6.2), and Gibbs sampling used to provide a distribution of parameter estimates, which can be combined to provide a distribution of development factors $\tilde{\lambda}_j$ used in the forecasting procedure.

In a Bayesian context, forecasting proceeds in exactly the same way as described in Section 5.2 for bootstrapping. That is, for $i = 2, 3, \ldots, n$:

$$D^*_{i,n-i+2} \sim ONB\big(\tilde{\lambda}_j D_{i,n-i+1}, \hat{\phi}\tilde{\lambda}_j(\tilde{\lambda}_j-1)D_{i,n-i+1}\big)$$

and for $i = 3, 4, \ldots, n$ and $j = n-i+3, n-i+4, \ldots, n$

$$D^*_{i,j} \sim ONB\big(\tilde{\lambda}_j D^*_{i,j-1}, \hat{\phi}\tilde{\lambda}_j(\tilde{\lambda}_j-1)D^*_{i,j-1}\big).$$

Again, the forecast incremental claims can be obtained by differencing in the usual way, and can then be aggregated using equation (4.1) to provide predictive distributions of the outstanding liabilities.

Like the over-dispersed Poisson model, when non-constant scale parameters are used, the procedure is identical, except the constant scale parameter ϕ is replaced by ϕ_j in the construction of the quasi-log-likelihood, and when forecasting.

6.3 *Mack's Model*

Following on from Section 3.3, and considering Mack's model as a weighted normal regression model, then:

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

$$f_{ij};\, \sigma_j^2,\, D_{i,j-1} \sim \text{Normal}\left(\lambda_j, \frac{\sigma_j^2}{D_{i,j-1}}\right)$$

and it is straightforward to show that:

$$\text{Log}\, L_N = \sum_{i=1}^{n-1} \sum_{j=2}^{n-i+1} 0.5 \times \left(\log\left(\frac{D_{i,j-1}}{\sigma_j^2}\right) - \frac{D_{i,j-1}}{\sigma_j^2}\left(f_{ij} - \lambda_j\right)^2\right) + \text{constant}. \quad (6.6)$$

Notice that, in this case, it is not necessary to use quasi-likelihood in the derivation of the log-likelihood, and the model is defined using non-constant scale parameters. Again, the log-likelihood has been written in general form to allow for any model structure, including structures which incorporate parametric curves and smoothers.

Equation (6.6) can then be used with equation (6.1) or (6.2), and Gibbs sampling used to provide a distribution of parameters, which can be combined to provide a distribution of development factors $\tilde{\lambda}_j$ used in the forecasting procedure.

In a Bayesian context, forecasting proceeds in exactly the same way as described in Section 5.3 for bootstrapping. That is, for $i = 2, 3, \ldots, n$:

$$D^*_{i,n-i+2} \sim \text{Normal}\left(\tilde{\lambda}_j D_{i,n-i+1},\, \hat{\sigma}_j^2 D_{i,n-i+1}\right)$$

and for $i = 3, 4, \ldots, n$ and $j = n-i+3, n-i+4, \ldots, n$

$$D^*_{i,j} \sim \text{Normal}\left(\tilde{\lambda}_j D^*_{i,j-1},\, \hat{\sigma}_j^2 D^*_{i,j-1}\right).$$

Again, the forecast incremental claims can be obtained by differencing in the usual way, and can then be aggregated using equation (4.1) to provide predictive distributions of the outstanding liabilities.

7. Illustrations

To illustrate the methodology, consider the claims amounts in Table 1, shown in incremental form. These are the data from Taylor & Ashe (1983), also used in England & Verrall (1999) and England (2002). Also shown are the standard chain ladder development factors and reserve estimates. The models described in Section 3 were fitted to these data using maximum likelihood methods, Bayesian and bootstrap methods, and the results are compared below.

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

Table 1. Claims data from Taylor & Ashe (1983)

Incremental claim amounts											Reserves
357,848	766,940	610,542	482,940	527,326	574,398	146,342	139,950	227,229	67,948		0
352,118	884,021	933,894	1,183,289	445,745	320,996	527,804	266,172	425,046			94,634
290,507	1,001,799	926,219	1,016,654	750,816	146,923	495,992	280,405				469,511
310,608	1,108,250	776,189	1,562,400	272,482	352,053	206,286					709,638
443,160	693,190	991,983	769,488	504,851	470,639						984,889
396,132	937,085	847,498	805,037	705,960							1,419,459
440,832	847,631	1,131,398	1,063,269								2,177,641
359,480	1,061,648	1,443,370									3,920,301
376,686	986,608										4,278,972
344,014											4,625,811
										Total	18,680,856
Development factors											
	3.4906	1.7473	1.4574	1.1739	1.1038	1.0863	1.0539	1.0766	1.0177		

7.1 *The Over-Dispersed Poisson Model*

Initially, consider using an over-dispersed Poisson generalised linear model, with a logarithmic link function, constant scale parameter and linear predictor given by equation (3.3). The maximum likelihood parameter estimates and their standard errors obtained by fitting this model are shown in Table 2, using a constant Pearson scale parameter evaluated using the methods shown in the Appendix.

The forecast expected values obtained from this model for the outstanding liabilities in each origin period and in total are shown in Table 3, and are identical to the chain ladder reserve estimates. Also shown are the prediction errors calculated analytically (using the methods described in England & Verrall, 2002), and the prediction error shown as a percentage of the mean.

The same model was fitted as a Bayesian model using non-informative uniform priors. As such, the posterior log-likelihood represented by equation (6.2) is simply equation (6.4). The scale parameter given by the maximum likelihood analysis (and used in the bootstrap analysis) was used as the plug-in scale parameter in the Bayesian analysis, and the maximum likelihood parameter estimates were used as the initial parameter values in the Gibbs sampling. The expected value of the parameters and their standard errors

Table 2. ODP parameter estimates and standard errors

	Maximum likelihood		Bayesian	
	Parameter estimate	Standard error	Expected value	Standard error
c	12.506	0.173	12.490	0.171
$\alpha(2)$	0.331	0.154	0.336	0.154
$\alpha(3)$	0.321	0.158	0.325	0.157
$\alpha(4)$	0.306	0.161	0.309	0.159
$\alpha(5)$	0.219	0.168	0.222	0.166
$\alpha(6)$	0.270	0.171	0.271	0.168
$\alpha(7)$	0.372	0.174	0.375	0.173
$\alpha(8)$	0.553	0.187	0.554	0.185
$\alpha(9)$	0.369	0.239	0.360	0.239
$\alpha(10)$	0.242	0.428	0.178	0.440
$\beta(2)$	0.913	0.149	0.918	0.150
$\beta(3)$	0.959	0.153	0.963	0.154
$\beta(4)$	1.026	0.157	1.031	0.156
$\beta(5)$	0.435	0.184	0.435	0.185
$\beta(6)$	0.080	0.215	0.074	0.214
$\beta(7)$	−0.006	0.238	−0.015	0.239
$\beta(8)$	−0.394	0.310	−0.425	0.314
$\beta(9)$	0.009	0.320	−0.021	0.326
$\beta(10)$	−1.380	0.897	−1.582	0.811
Scale parameter	52,601		52,601	

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

Table 3. ODP with constant scale parameter — maximum likelihood method

	Expected reserves	Prediction error	Prediction error %
Year 1	0	0	–
Year 2	94,634	110,099	116%
Year 3	469,511	216,042	46%
Year 4	709,638	260,871	37%
Year 5	984,889	303,549	31%
Year 6	1,419,459	375,012	26%
Year 7	2,177,641	495,376	23%
Year 8	3,920,301	789,957	20%
Year 9	4,278,972	1,046,508	24%
Year 10	4,625,811	1,980,091	43%
Total	18,680,856	2,945,646	16%

using 25,000 Gibbs iterations are also shown in Table 2, calculated as the mean and standard deviation of the marginal distributions. The results can be compared to the maximum likelihood (ML) values, although we do not expect perfect agreement, since the ML estimates are derived when the likelihood is maximised, and the standard errors are derived using approximate asymptotic methods. Nevertheless the comparison is useful to perform as a consistency check.

Table 4 shows the expected values for the outstanding liabilities in each origin period and in total, together with the prediction errors, given by the Bayesian analysis. Also shown are the equivalent values of a bootstrap analysis using 25,000 bootstrap iterations for the same model, using the

Table 4. ODP with constant scale parameter — Bayesian and bootstrap methods

	Bayesian			Bootstrap		
	Expected reserves	Prediction error	Prediction error %	Expected reserves	Prediction error	Prediction error %
Year 1	0	0	–	0	0	–
Year 2	103,966	113,099	109%	95,828	114,115	119%
Year 3	480,365	218,870	46%	472,499	219,785	47%
Year 4	720,581	262,230	36%	714,692	265,620	37%
Year 5	998,793	305,889	31%	991,940	307,732	31%
Year 6	1,431,804	376,636	26%	1,428,229	376,606	26%
Year 7	2,201,136	498,725	23%	2,191,141	496,376	23%
Year 8	3,949,499	792,227	20%	3,943,107	803,720	20%
Year 9	4,312,636	1,057,277	25%	4,302,341	1,048,451	24%
Year 10	4,705,145	2,026,746	43%	4,706,454	2,034,830	43%
Total	18,903,924	2,974,783	16%	18,846,231	3,010,585	16%

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

Table 5. ODP with constant scale parameter — percentiles of total outstanding liabilities

	Bayesian	Bootstrap
1st percentile	13,065,635	12,662,941
5th percentile	14,529,350	14,284,445
10th percentile	15,358,672	15,179,072
25th percentile	16,800,019	16,749,569
50th percentile	18,622,505	18,634,618
75th percentile	20,718,854	20,673,617
90th percentile	22,859,777	22,790,582
95th percentile	24,179,717	24,188,186
99th percentile	27,038,068	26,892,337

methods described in Section 5.1. Tables 3 and 4 can be compared, and show a good degree of similarity between the maximum likelihood, Bayesian and bootstrap approaches. Percentiles of the predictive distribution of total outstanding liabilities for the Bayesian and bootstrap analyses are shown in Table 5, and Figure 3 shows a graphical comparison of the equivalent densities. Again, the comparisons show a good degree of similarity between the bootstrap and Bayesian methods.

Using non-constant scale parameters instead, calculated using the

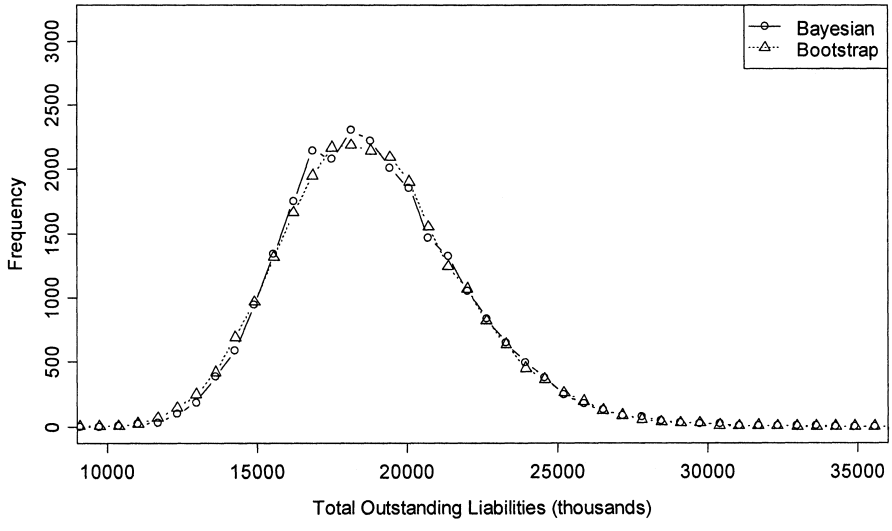


Figure 3. ODP with constant scale parameter — density chart of total outstanding liabilities

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

Table 6. ODP with non-constant scale parameters — Bayesian and bootstrap methods

	Bayesian			Bootstrap		
	Expected reserves	Prediction error	Prediction error %	Expected reserves	Prediction error	Prediction error %
Year 1	0	0	–	0	0	–
Year 2	99,653	43,172	43%	95,996	43,286	45%
Year 3	492,551	101,586	21%	473,654	109,636	23%
Year 4	715,933	125,293	18%	715,675	141,615	20%
Year 5	1,020,251	241,930	24%	989,792	257,410	26%
Year 6	1,461,234	378,691	26%	1,428,257	388,702	27%
Year 7	2,203,159	488,763	22%	2,189,413	522,002	24%
Year 8	3,894,073	728,236	19%	3,940,037	732,093	19%
Year 9	4,298,306	814,895	19%	4,297,357	813,324	19%
Year 10	4,648,130	1,292,539	28%	4,659,678	1,296,265	28%
Total	18,833,290	2,179,794	12%	18,789,859	2,203,228	12%

methods described in the Appendix, gives the expected outstanding liabilities and associated standard errors shown in Table 6, with the simulated density shown in Figure 4. Again, the results of the bootstrap and Bayesian methods are reassuringly close. Comparison of Tables 4 and 6 shows that the prediction errors are lower when using non-constant scale

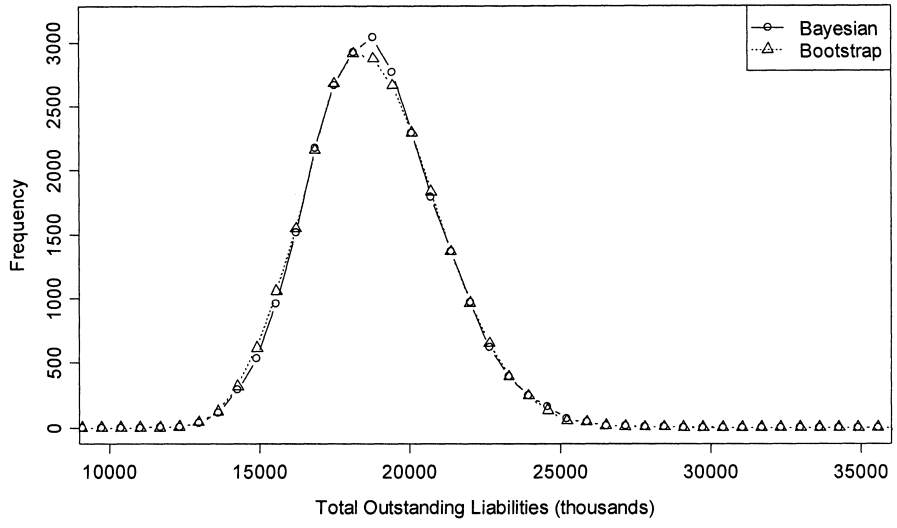


Figure 4. ODP with non-constant scale parameters — density chart of total outstanding liabilities

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

parameters, reflecting the tendency for lower scale parameters in the later development periods.

7.2 *The Over-Dispersed Negative Binomial Model*

Again, we initially consider obtaining maximum likelihood estimates by fitting the model as a generalised linear model. A log-log link function has been used, with a constant scale parameter and linear predictor given by equation (3.5). The maximum likelihood parameter estimates and their standard errors obtained by fitting this model are shown in Table 7, using a constant Pearson scale parameter evaluated using the methods shown in the Appendix.

The forecast expected values obtained from this model for the outstanding liabilities in each origin period and in total are shown in Table 8, and are identical to the chain ladder reserve estimates. Also shown are the prediction errors calculated analytically (using the methods described in England & Verrall, 2002), and the prediction error shown as a percentage of the mean. Comparison with Table 3 shows that the prediction errors calculated analytically for the over-dispersed negative binomial model are very close to the prediction errors calculated analytically using the over-dispersed Poisson model, the remaining differences being due to the differences in the scale parameters only. This is because the negative binomial model is the recursive equivalent of the Poisson model (see Verrall, 2000).

The same model was fitted as a Bayesian model using non-informative uniform priors. As such, the posterior log-likelihood represented by equation (6.2) is simply equation (6.5). The scale parameter given by the maximum likelihood analysis (and used in the bootstrap analysis) was used as the plug-in scale parameter in the Bayesian analysis, and the maximum likelihood parameter estimates were used as the initial parameter values in the Gibbs

Table 7. ONB parameter estimates and standard errors

	Maximum likelihood		Bayesian	
	Parameter estimate	Standard error	Expected value	Standard error
$\gamma(2)$	0.223	0.084	0.221	0.085
$\gamma(3)$	−0.583	0.083	−0.587	0.084
$\gamma(4)$	−0.976	0.087	−0.980	0.087
$\gamma(5)$	−1.831	0.127	−1.838	0.129
$\gamma(6)$	−2.315	0.167	−2.327	0.169
$\gamma(7)$	−2.492	0.194	−2.511	0.196
$\gamma(8)$	−2.947	0.275	−2.987	0.281
$\gamma(9)$	−2.607	0.282	−2.646	0.287
$\gamma(10)$	−4.042	0.874	−4.260	0.800
Scale parameter	51,957		51,957	

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

Table 8. ONB with constant scale parameter — maximum likelihood method

	Expected reserves	Prediction error	Prediction error %
Year 1	0	0	–
Year 2	94,634	109,422	116%
Year 3	469,511	214,721	46%
Year 4	709,638	259,280	37%
Year 5	984,889	301,700	31%
Year 6	1,419,459	372,733	26%
Year 7	2,177,641	492,374	23%
Year 8	3,920,301	785,197	20%
Year 9	4,278,972	1,040,221	24%
Year 10	4,625,811	1,968,331	43%
Total	18,680,856	2,928,201	16%

sampling. The expected value of the parameters and their standard errors using 25,000 Gibbs iterations are also shown in Table 7, calculated as the mean and standard deviation of the marginal distributions. The results can be compared to the maximum likelihood values, although again we do not expect perfect agreement, for the same reasons as those given in Section 7.1.

Table 9 shows the expected values for the outstanding liabilities in each origin period and in total, together with the prediction errors, given by the Bayesian analysis. Also shown are the equivalent values of a bootstrap analysis using 25,000 bootstrap iterations for the same model, using the methods described in Section 5.2. Tables 8 and 9 can be compared, and show a good degree of similarity between the maximum likelihood, Bayesian and

Table 9. ONB with constant scale parameter — Bayesian and bootstrap methods

	Bayesian			Bootstrap		
	Expected reserves	Prediction error	Prediction error %	Expected reserves	Prediction error	Prediction error %
Year 1	0	0	–	0	0	–
Year 2	103,744	112,623	109%	98,840	105,488	107%
Year 3	476,681	217,105	46%	475,338	215,236	45%
Year 4	718,556	260,850	36%	711,994	257,229	36%
Year 5	993,864	301,246	30%	989,970	301,597	30%
Year 6	1,429,092	374,113	26%	1,423,089	373,673	26%
Year 7	2,196,312	494,051	22%	2,183,293	495,844	23%
Year 8	3,943,310	790,830	20%	3,928,999	789,301	20%
Year 9	4,302,016	1,052,099	24%	4,291,783	1,038,105	24%
Year 10	4,687,977	1,995,053	43%	4,636,926	1,982,566	43%
Total	18,851,551	2,962,210	16%	18,740,232	2,942,040	16%

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

Table 10. ONB with constant scale parameter — percentiles of total outstanding liabilities

	Bayesian	Bootstrap
1st percentile	12,970,475	12,721,825
5th percentile	14,417,696	14,270,887
10th percentile	15,281,165	15,157,706
25th percentile	16,775,171	16,688,754
50th percentile	18,626,984	18,540,998
75th percentile	20,681,172	20,584,904
90th percentile	22,689,320	22,560,234
95th percentile	24,073,540	23,887,939
99th percentile	26,945,287	26,527,418

bootstrap approaches. Percentiles of the predictive distribution of total outstanding liabilities for the Bayesian and bootstrap analyses are shown in Table 10, and Figure 5 shows a graphical comparison of the equivalent densities. Again, the comparisons show a good degree of similarity between the bootstrap and Bayesian methods. Furthermore, comparison of Tables 9 and 10 with Tables 4 and 5 shows a very good degree of similarity between the negative binomial and Poisson models.

Using non-constant scale parameters instead, calculated using the methods

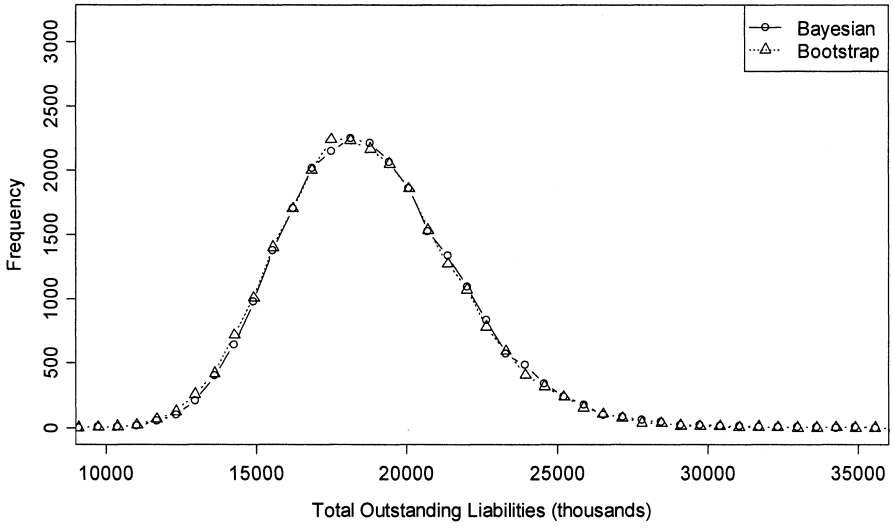


Figure 5. ONB with constant scale parameter — density chart of total outstanding liabilities

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

Table 11. ONB with non-constant scale parameters — Bayesian and bootstrap methods

	Bayesian			Bootstrap		
	Expected reserves	Prediction error	Prediction error %	Expected reserves	Prediction error	Prediction error %
Year 1	0	0	–	0	0	–
Year 2	94,762	38,988	41%	94,531	38,894	41%
Year 3	470,018	84,501	18%	469,608	84,363	18%
Year 4	709,860	98,459	14%	709,321	99,408	14%
Year 5	988,093	243,000	25%	984,733	240,896	24%
Year 6	1,424,811	403,025	28%	1,420,551	396,914	28%
Year 7	2,181,605	558,522	26%	2,178,469	553,241	25%
Year 8	3,922,132	890,250	23%	3,917,507	882,356	23%
Year 9	4,295,119	994,132	23%	4,272,722	986,619	23%
Year 10	4,652,618	1,433,435	31%	4,626,889	1,429,897	31%
Total	18,739,019	2,447,121	13%	18,674,331	2,431,532	13%

described in the Appendix, gives the expected outstanding liabilities and associated standard errors shown in Table 11, with the simulated density shown in Figure 6. Again, the results of the bootstrap and Bayesian methods are reassuringly close. Comparison of Tables 9 and 11 shows that the prediction errors are lower when using non-constant scale parameters, reflecting the tendency for lower scale parameters in the later development

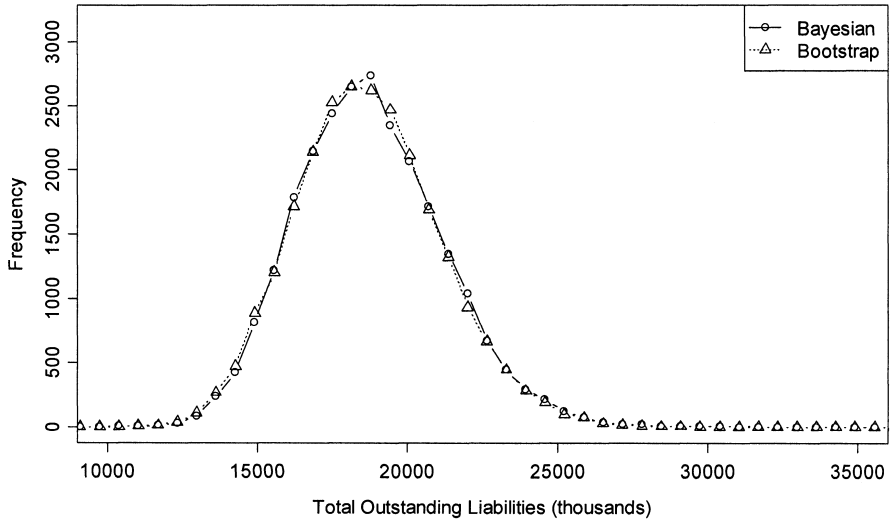


Figure 6. ONB with non-constant scale parameters — density chart of total outstanding liabilities

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

periods. Furthermore, comparison of Tables 6 and 11 shows a good degree of similarity between the negative binomial and Poisson models.

7.3 *Mack's Model*

Again, we initially consider obtaining maximum likelihood estimates by fitting the model as a generalised linear model, using the weighted normal regression method described in England & Verrall (2002). A log link function has been used with non-constant scale parameters and linear predictor given by equation (3.7). The maximum likelihood parameter estimates and their standard errors obtained by fitting this model are shown in Table 12, using non-constant scale parameters evaluated using the methods shown in the Appendix.

The forecast expected values obtained from this model for the outstanding liabilities in each origin period and in total are shown in Table 13, and are identical to the chain ladder reserve estimates. Also shown are the prediction errors calculated analytically (using the methods described in England & Verrall, 2002), and the prediction error shown as a percentage of the mean. Comparison with Tables 6 and 11 shows that the prediction errors calculated analytically for Mack's model are very similar to the over-dispersed Poisson and negative binomial models when non-constant scale parameters are used, the biggest differences being in the oldest origin periods. This is consistent with England & Verrall's observation that Mack's model could be seen as a normal approximation to the negative binomial model.

The same model was fitted as a Bayesian model using non-informative uniform priors. As such, the posterior log-likelihood represented by equation (6.2) is simply equation (6.6). The scale parameters given by the maximum likelihood analysis (and used in the bootstrap analysis) were used as the plug-in scale parameters in the Bayesian analysis, and the maximum likelihood

Table 12. Mack's model: parameter estimates and standard errors

	Maximum likelihood		Bayesian	
	Parameter estimate	Standard error	Expected value	Standard error
$\gamma(2)$	1.250	0.063	1.244	0.063
$\gamma(3)$	0.558	0.035	0.556	0.035
$\gamma(4)$	0.377	0.036	0.375	0.036
$\gamma(5)$	0.160	0.025	0.160	0.025
$\gamma(6)$	0.099	0.025	0.098	0.025
$\gamma(7)$	0.083	0.021	0.082	0.021
$\gamma(8)$	0.052	0.006	0.052	0.006
$\gamma(9)$	0.074	0.011	0.073	0.011
$\gamma(10)$	0.018	0.011	0.017	0.011

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

Table 13. Mack's model — maximum likelihood method

	Expected reserves	Prediction error	Prediction error %
Year 1	0	0	–
Year 2	94,634	75,535	80%
Year 3	469,511	121,699	26%
Year 4	709,638	133,549	19%
Year 5	984,889	261,406	27%
Year 6	1,419,459	411,010	29%
Year 7	2,177,641	558,317	26%
Year 8	3,920,301	875,328	22%
Year 9	4,278,972	971,258	23%
Year 10	4,625,811	1,363,155	29%
Total	18,680,856	2,447,095	13%

parameter estimates were used as the initial parameter values in the Gibbs sampling. The expected value of the parameters and their standard errors using 25,000 Gibbs iterations are also shown in Table 12, calculated as the mean and standard deviation of the marginal distributions. The results can be compared to the maximum likelihood values, although again we do not expect perfect agreement for the same reasons as those given in Section 7.1.

The forecast incremental claims can be obtained by differencing in the usual way, and can then be aggregated using equation (4.1) to provide predictive distributions of the outstanding liabilities. Table 14 shows the expected values for the outstanding liabilities in each origin period and in total, together with the prediction errors, given by the Bayesian analysis. Also shown are the equivalent values of a bootstrap analysis using 25,000 bootstrap iterations for the same model, using the methods described in

Table 14. Mack's model — Bayesian and bootstrap methods

	Bayesian			Bootstrap		
	Expected reserves	Prediction error	Prediction error %	Expected reserves	Prediction error	Prediction error %
Year 1	0	0	–	0	0	–
Year 2	93,569	75,933	81%	94,408	75,112	80%
Year 3	469,115	121,639	26%	468,390	121,159	26%
Year 4	707,275	133,119	19%	709,029	134,110	19%
Year 5	979,076	261,210	27%	984,371	260,546	26%
Year 6	1,412,494	410,378	29%	1,416,331	409,134	29%
Year 7	2,172,518	561,310	26%	2,174,282	556,773	26%
Year 8	3,909,446	876,336	22%	3,919,355	880,047	22%
Year 9	4,259,164	966,688	23%	4,273,149	965,707	23%
Year 10	4,589,769	1,359,048	30%	4,614,590	1,350,376	29%
Total	18,592,427	2,444,583	13%	18,653,905	2,426,621	13%

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

Table 15. Mack's model — percentiles of total outstanding liabilities

	Bayesian	Bootstrap
1st percentile	13,326,448	13,411,527
5th percentile	14,730,817	14,827,236
10th percentile	15,520,550	15,618,751
25th percentile	16,877,221	16,965,524
50th percentile	18,510,764	18,553,804
75th percentile	20,196,677	20,230,157
90th percentile	21,796,350	21,830,650
95th percentile	22,749,126	22,791,594
99th percentile	24,610,620	24,700,021

Section 5.3. Tables 13 and 14 can be compared, and show a good degree of similarity between the maximum likelihood, Bayesian and bootstrap approaches. Percentiles of the predictive distribution of total outstanding liabilities for the Bayesian and bootstrap analyses are shown in Table 15, and Figure 7 shows a graphical comparison of the equivalent densities. Again, the comparisons show a good degree of similarity between the bootstrap and Bayesian methods. Furthermore, comparison of Table 14 with Tables 6 and 11 shows a very good degree of similarity between Mack's model and the negative binomial and Poisson models when non-constant scale parameters are used.

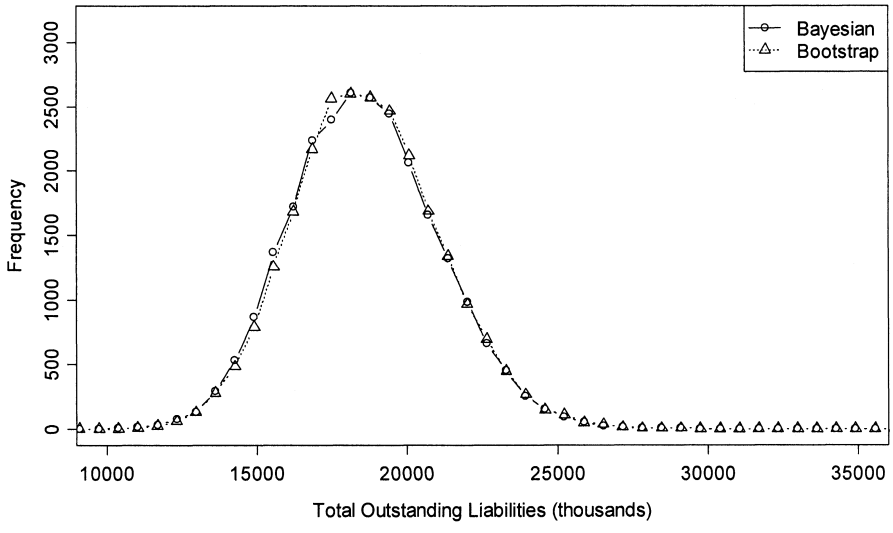


Figure 7. Mack's model — density chart of total outstanding liabilities

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

8. Discussion

In the examples in Section 7, we have made a comparison between paradigms (maximum-likelihood, bootstrap and Bayesian) and models (over-dispersed Poisson, negative binomial and Mack), where an appropriate comparison can be made. Figures 3 to 7 show a comparison of the densities using the Bayesian and bootstrap approaches for each model, showing remarkable similarity, given the differences in the procedures used. Figure 8 shows a comparison between models, using the densities of the total outstanding liabilities given by the Bayesian method for the Poisson model, the negative binomial model and Mack's model. Again, the graph shows a high degree of similarity between the Poisson and negative binomial models when constant scale parameters are used, and between Mack's model and the negative binomial model with non-constant scale parameters. The graph also shows a higher peak for Mack's model, and when non-constant scale parameters are used for the Poisson and negative binomial models, reflecting the lower prediction errors in those cases.

One advantage of Mack's model, which is based on the normal distribution, is that it can be used with incurred data, which often include negative incremental values at later development periods, resulting in chain ladder development factors which are less than one. Where this occurs, Mack's model should be used. Where incurred data are used, predictive

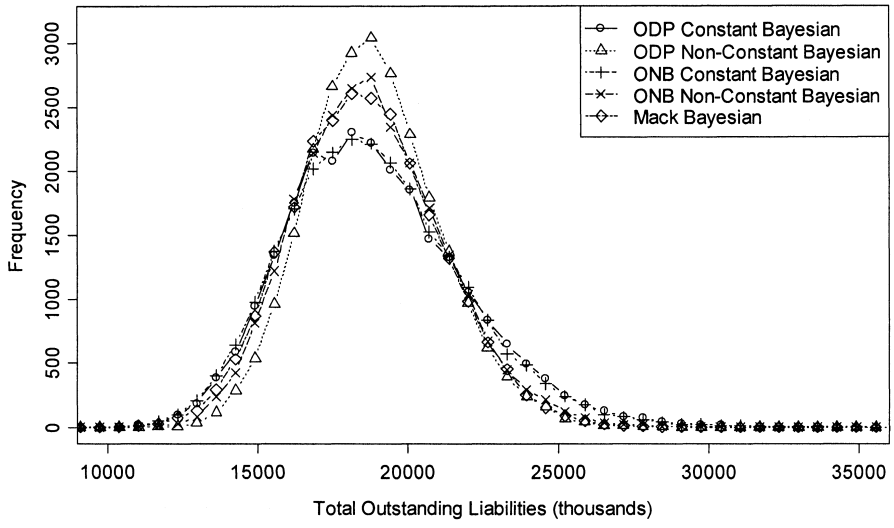


Figure 8. All models — density charts of total outstanding liabilities

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

distributions of the ultimate cost of claims are obtained. The observed paid-to-date for each origin period can be subtracted to give predictive distributions of the outstanding liabilities. However, where cash flows are required (for example, within DFA models), the distribution of outstanding liabilities must be combined with a payment pattern for the proportion of outstanding which emerges in each development period. That payment pattern should be simulated, taking account of estimation error, otherwise the resultant variability of the cash flows will be underestimated. A complete model based primarily on incurred data, taking account of the pattern to provide cash flows, could be developed within a Bayesian framework.

It is straightforward to extend models based on incremental claims to include calendar year components, to model the effect of claims inflation. However, in a DFA model it is usually undesirable to project claims inflation into the future using the modelled inflation rates, since a DFA model will usually include an economic scenario generator (ESG), and it is important that any dependence between reserving risk and inflation from the ESG is incorporated. One solution is to inflation-adjust the data to remove the effects of historic price inflation, and to use calendar year components to model super-imposed claims inflation. Forecasts can then be made on an inflation-adjusted basis (but including superimposed inflation), before adding the effect of price inflation linked to the ESG. Clearly, there are several variations on this theme.

When forecasting, it is important to simulate in a way that is broadly consistent with the underlying process distribution. In the examples shown in Section 7, we have simulated from a standard distribution with approximately similar characteristics, and such that the first two moments, at least, are maintained. Where this could result in simulated values that are inconsistent with the problem of claims reserving (for example, using a normal distribution with Mack's model resulting in negative cumulative claims), we adopt the use of other distributions which behave better, as a practical compromise, while still maintaining the spirit of the model as far as possible. In a bootstrapping context, instead of simulating directly from the assumed process distribution when forecasting, Pinheiro *et al.* (2003) re-sample again from the residuals, thereby mirroring the approach used to obtain the pseudo data at the previous stage. However, using that approach, the extremes will be limited according to the most extreme residuals observed.

In the examples, for simplicity, and to provide a connection to traditional actuarial techniques, we have used predictor structures which give expected values which are the same as the chain ladder technique. This does not imply that we think that the chain ladder technique should always be used. The model specifications in Sections 3, 5 and 6 are sufficiently general that any predictor structure could be used. For example, for the over-dispersed Poisson model outlined in Section 5.1, we could use:

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

$$\log(m_{ij}) = c + \alpha_i + \beta_j \qquad \text{for } j \leq \tau$$

$$\log(m_{ij}) = c + \alpha_i + b_1(j - \tau) + b_2 \log(j - \tau) \qquad \text{for } j > \tau.$$

That is, standard chain ladder factors could be used for the early development periods, and a parametric curve (Hoerl curve) then used in the later development periods, which can be extrapolated beyond the latest observed development period if required. This is still a linear model, since it is linear in the parameters, and therefore falls within the framework of generalised linear models. As such, it is straightforward to implement using the methods described in this paper (together with software that fits GLMs when bootstrapping).

As a further example, for Mack's model outlined in Section 3.3, we could use:

$$\lambda_j = \exp(\gamma_j) \qquad \text{for } j \leq \tau$$

$$\lambda_j = 1 + a\exp(b(j - \tau)) \qquad \text{for } j > \tau.$$

That is, standard chain ladder factors could be used for the early development periods, and a parametric curve then used in the later development periods, which can be extrapolated beyond the latest observed development period if required. This is now non-linear in the parameters. In a Bayesian context, it is still straightforward to implement this model, since Gibbs sampling is indifferent to linear and non-linear structures. However, software that can fit non-linear models is required when bootstrapping.

There are many alternatives which could be tried, including models which are piecewise linear, and models involving non-parametric smoothers.

The models presented in this paper have been applied to aggregate claims triangles, but they could equally be applied to other types of data. For example, the methods could be applied to triangles of numbers of reported claims to identify the number of claims incurred, but not reported (IBNR). Ntzoufras & Dellaportas (2002) and de Alba (2002) consider Bayesian estimation of outstanding claims by separating aggregate triangles into the number of claims and the average cost of claims, and modelling each component separately, before combining. This is a useful alternative approach for practitioners who prefer methods based on the average cost of claims, and is closer in spirit to a pricing analysis (of a motor portfolio, say), where it is routine to separate the cost of claims into frequency and average severity components.

In this paper we do not consider the issues surrounding modelling gross and net claims, although the models which we have presented could be used with gross or net data. However, in a DFA model simulations of gross and

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

net claims are required where each simulated estimate of gross and net outstanding liabilities should be matched (such that dependencies are taken into consideration). In this respect, a good approach would be to simulate outstanding liabilities on a gross basis, and net down each simulation using the appropriate reinsurance programme within each origin period. Where quota-share reinsurance applies, this is straightforward, but excess-of-loss reinsurance requires simulated large losses. For this, a reasonable way forward is to separate an aggregate triangle into attritional claims (high frequency, low severity) and large claims (low frequency, high severity). The gross attritional claims could then be modelled in aggregate using the methods presented in this paper, but the large claims could be modelled individually using a frequency/individual severity approach, and the projected individual large claims could then be netted down by passing them through the appropriate excess-of-loss reinsurance contract. The large claim analysis could be performed by generalising the Bayesian methods proposed by Ntzoufras & Dellaportas (2002) and de Alba (2002) into a frequency/individual severity model instead of frequency/average severity.

The Bayesian models have been implemented using the approach of Dellaportas & Smith (1995), who combined the GLM log-likelihood with prior information on the parameters to form the posterior joint log-likelihood. That posterior likelihood was then used with Gibbs sampling, treating it sequentially as a function of each parameter in turn. Where, a quasi-likelihood would naturally be used in a GLM context to model over-dispersion, we have formed the posterior log-likelihood using the quasi-likelihood. Where dispersion parameters are required, we treat them as plug-in estimates, and use the same values which would be derived from a GLM analysis.

We have shown that, with non-informative prior distributions, the results from the Bayesian analysis are analogous to those from a maximum likelihood approach. From a theoretical perspective, this is well known to statisticians familiar with Bayesian statistics. For example, Gelfand & Ghosh (2000) state:

> "With little or no prior information, an alternative is to use non-informative priors. This implies that the posterior distribution is essentially the likelihood [and] that the Bayesian analysis will be close to a likelihood analysis, possibly attractive to frequentists."

Some readers might object to the use of quasi-likelihoods and plug-in scale parameters, since the models are not 'fully' Bayesian. We have sympathy with those objections, and, in this paper at least, like Dellaportas & Smith, we are "leaving aside philosophical issues — and debates about whether GLMs should be interpreted as genuine models, or simply as exploratory data analysis devices ...". We believe that our approach is useful, and casts light on the assumptions made when adopting alternative

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

stochastic reserving methods. Albert & Pepple (1989) also suggested using the quasi-likelihood in a Bayesian analysis of over-dispersion models, and Congdon (2003) shows that quasi-likelihoods could be used with Gibbs sampling, and that complete likelihoods are not necessarily required (see Chapter 3, p83). Dey & Ravishanker (2000) also consider Bayesian approaches to over-dispersion in generalised linear models.

The desirability of making a comparison, and the insights gained from doing so, was the over-riding reason for treating the dispersion parameters as plug-in estimates in the Bayesian analysis, since we could ensure that the same values were used in all models, thereby removing one area where a difference might obscure the similarities. The methods can be extended to consider modelling the dispersion parameters within the Bayesian analysis (together with their inherent uncertainty), although this would be expected to increase the variability of the predictive distributions of outstanding liabilities.

We recommend using non-constant scale parameters (or at least checking the assumption that using a constant scale parameter is appropriate), and a way of allowing the scale parameters to vary by development period is shown in the Appendix. However, the amount of data on which the scale parameters are estimated reduces as development time increases, so the scale parameters can be volatile. Where relatively large scale parameters occur at later development periods, we recommend that the cause is investigated. This can usually be tracked to a single data point; if there is a problem with the data, then the problem should be rectified, and the analysis repeated. However, if the data are valid, and the volatility is 'real', the practitioner must decide on an appropriate course of action. There are various options, including smoothing the scale parameters, either by fitting a parametric curve or non-parametric smoother, or by making manual adjustments.

When bootstrapping, we also recommend making a bias correction to the residuals (described in the Appendix) to allow for the trade-off between goodness-of-fit and number of parameters, especially where the results will be compared with an analytic approach. Moulton & Zeger (1991) and Pinheiro *et al.* (2003) recommend a further adjustment (known as 'standardisation') to improve the independence of residuals, which may perform slightly better, although that adjustment is not straightforward to accommodate, and has consequently been overlooked in this paper, since any out-performance is outweighed by difficulty of implementation.

As we mentioned in the introduction, there could be sets of pseudo data which are inconsistent with the underlying statistical model, and a number of modifications can be made to overcome the practical difficulties. For example, with Mack's model some values in the pseudo triangle of development ratios could be negative. In practice, we would tolerate them, although we would force the fitted development factors themselves (being a

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

weighted average of the individual development ratios) to be positive for each bootstrap iteration.

We do not view the bootstrap or Bayesian methods as a panacea, or the only approaches which could be used to obtain predictive distributions. Furthermore, we do not consider the model structures introduced in Section 3 as the only ones which could be used. In England & Verrall (2002), other stochastic reserving models were also considered which fall within the framework of generalised linear models, for example models based on the gamma error structure and models based on the log-normal distribution. It is straightforward to apply the procedures outlined in this paper to those models as well. For example, models based on the gamma error structure can be used, but replacing $\mathrm{Var}[C_{ij}] = \phi m_{ij}$ by $\mathrm{Var}[C_{ij}] = \phi m_{ij}^2$ in Sections 5.1 and 6.1. That is, with the gamma error structure, the variance is assumed to be proportional to the mean squared. Mack (1991) suggested using the gamma model in the context of claims reserving, although not within the framework of generalised linear models, and bootstrapping the gamma model was considered by Pinheiro *et al.* (2003).

The log-normal distribution also assumes that the variance of the incremental claims is proportional to the mean squared, but focuses on the log of the incremental claims, $Y_{ij} = \log(C_{ij})$, where $Y_{ij} \sim \mathrm{Normal}(m_{ij}, \sigma^2)$. In this case, when bootstrapping:

$$r_{ij} = r_{PS}(Y_{ij}, \hat{m}_{ij}, \hat{\sigma}) = \frac{Y_{ij} - \hat{m}_{ij}}{\hat{\sigma}}.$$

The pseudo data, still on a log scale, are then defined as:

$$Y_{ij}^B = r_{ij}^B \hat{\sigma} + \hat{m}_{ij}$$

and the model used to obtain the residuals can be fitted to each triangle of pseudo data.

Since this is a non-recursive model, bootstrap forecasts, excluding process error, can be obtained for the complete lower triangle of future values, that is:

$$\tilde{Y}_{ij} = \hat{m}_{ij}^B \quad \text{for} \quad i = 2, 3, \ldots, n \quad \text{and} \quad j = n - i + 2, \ldots, n.$$

Extrapolation beyond the final development period can be used where curves have been fitted to help estimate tail factors.

To add the process error, a forecast value (on a log scale) can then be simulated from a normal distribution with mean $\tilde{Y}_{ij}$ and variance $\hat{\sigma}^2$. Forecasts on the untransformed scale can then be obtained simply by exponentiating. That is:

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

$$C_{ij}^{*} \sim \text{Exp}\big(\text{Normal}\big(\tilde{Y}_{ij}, \hat{\sigma}^2\big)\big).$$

The forecasts can then be aggregated to provide predictive distributions of the outstanding liabilities by origin year and overall. Again, non-constant scale parameters can be used, replacing σ by σ_j. The implementation of the analogous Bayesian model is straightforward, and was considered by Ntzoufras & Dellaportas (2002).

9. Conclusions

In the context of claims reserving in general insurance, this paper has presented bootstrapping and Bayesian methods as procedures, which, if followed carefully, can be used for obtaining predictive distributions of outstanding liabilities of well-specified models which are analogous to results that would be obtained analytically. This was illustrated using representations of some stochastic reserving models, based on the framework of generalised linear models, which have been described previously in other papers, including the well known model of Mack (1993). For some specific examples, the models are analogous to traditional actuarial methods, although the advantage of bootstrap and Bayesian approaches is that full predictive distributions are obtained automatically, essential when building stochastic models for use in capital setting. In this paper we have also emphasised the distinction between recursive and non-recursive models, and shown that the procedures can be applied to both types.

It is reasonable to ask the question: "Which approach is best?" Bootstrapping and Bayesian methods both give a predictive distribution, and, under the same assumptions, give similar results (as we have shown in Section 7). Bootstrapping has the advantage of apparent simplicity, although once set up, Bayesian MCMC methods are very easy to apply and generalise. Even though we recommend starting with a clearly defined statistical model, bootstrapping methods are more amenable to manipulation, although this could be seen as a disadvantage, since that leaves the methods open to abuse. Bayesian methods demand a well-specified model when setting up the posterior log-likelihood, which could be seen as an advantage, since the underlying assumptions are clear. Bayesian methods do not require the construction of many sets of pseudo-data, which some may find intuitively appealing. When constructing predictive distributions of outstanding liabilities using aggregate triangles, for use in capital modelling and dynamic financial analysis, either method could be used. However, when considering models based on individual data, and further generalisations, a Bayesian approach is likely to be the most productive.

We believe that predictive distributions should be required of all stochastic reserving methods. Bootstrapping is a useful approach for

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

obtaining predictive distributions, but a Bayesian framework offers the best way forward. By describing how some previously published stochastic reserving models can be implemented in a Bayesian framework, and highlighting the similarities between the approaches, we hope that others will be encouraged to adopt a Bayesian approach, and develop the models further.

Acknowledgements

We would like to thank the referees for their useful comments and for the additional references which they provided.

References

Albert, J.H. & Pepple, P.A. (1989). A Bayesian approach to some overdispersion models. *Canadian Journal of Statistics*, **17**(3), 333-344.

Ashe, F.R. (1986). An essay at measuring the variance of estimates of outstanding claim payments. *ASTIN Bulletin*, **16S**, 99-113.

Brickman, S., Barlow, C., Boulter, A., English, A., Furber, L., Ibeson, D., Lowe, L., Pater, R. & Tomlinson, D. (1993). Variance in claims reserving. *Proceedings of the 1993 General Insurance Convention*, Institute of Actuaries and Faculty of Actuaries.

Congdon, P. (2003). *Applied Bayesian modelling*. Wiley, Chichester.

De Alba, E. (2002). Bayesian estimation of outstanding claim reserves. *North American Actuarial Journal*, **6**(4), 1-20.

Dellaportas, P. & Smith, A.F.M. (1993). Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *Applied Statistics*, **42**(3), 443-459.

Dey, D.K., Ghosh, S.K. & Mallick, B.K. (2000). *Generalized linear models: a bayesian perspective*. Marcel Dekker.

Dey, D.K. & Ravishanker, N. (2000). Bayesian approaches for overdispersion in generalised linear models. In *generalized linear models: a Bayesian perspective*. Edited by D.K. Dey, S.K. Ghosh & B.K. Mallick. Marcel Dekker.

Efron, B. & Tibshirani, R.J. (1993). *An Introduction to the bootstrap*. Chapman and Hall, London.

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

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

England, P.D. & Verrall, R.J. (2002). Stochastic claims reserving in general insurance (with discussion). *British Actuarial Journal*, **8**, 443-544.

Gelfand, A.E. & Ghosh, M. (2000). Generalised linear models: a Bayesian view. In *Generalized linear models: a Bayesian perspective*. Edited by D.K. Dey, S.K. Ghosh & B.K. Mallick. Marcel Dekker.

Gilks, W.R. & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, **41**(2), 337-348.

Gilks, W.R., Best, N.G. & Tan, K.K.C. (1995). Adaptive rejection metropolis sampling within Gibbs sampling. *Applied Statistics*, **44**(4), 455-472.

Haastrup, S. & Arjas, E. (1996). Claims reserving in continuous time; a non-parametric Bayesian approach. *ASTIN Bulletin*, **26**(2), 139-164.

Igloo Professional with ExtrEMB (2005). Igloo Professional with ExtrEMB v2.3.1. EMB Software Ltd, Epsom, U.K.

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

KAAS, R., GOOVAERTS, M., DHAENE, M. & DENUIT, M. (2002). *Modern actuarial risk theory*. Kluwer.

KLUGMAN, S.A. (1992). *Bayesian statistics in actuarial science*. Kluwer (Boston).

LOWE, J. (1994). A practical guide to measuring reserve variability using: bootstrapping, operational time and a distribution free approach. *Proceedings of the 1994 General Insurance Convention*, Institute of Actuaries and Faculty of Actuaries.

MACK, T. (1991). A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bulletin*, **22**(1), 93-109.

MACK, T. (1993). Distribution-free calculation of the standard error of chain-ladder reserve estimates. *ASTIN Bulletin*, **23**, 213-225.

MAKOV, U.E. (2001). Principal applications of Bayesian methods in actuarial science: a perspective. *North American Actuarial Journal*, **5**(4), 53-73.

MAKOV, U.E., SMITH, A.F.M. & LIU, Y.-H. (1996). Bayesian methods in actuarial science. *The Statistician*, **45**(4), 503-515.

MCCULLAGH, P. & NELDER, J. (1989). *Generalized linear models (2nd edition)*. Chapman and Hall, London.

MOULTON, L.H. & ZEGER, S.L. (1991). Bootstrapping generalized linear models. *Computational Statistics and Data Analysis*, **11**, 53-63.

NTZOUFRAS, I. & DELLAPORTAS, P. (2002). Bayesian modelling of outstanding liabilities incorporating claim count uncertainty. *North American Actuarial Journal*, **6**(1), 113-128.

PINHEIRO, P.J.R., ANDRADE, E., SILVA, J.M. & CENTENO, M.L.C. (2003). Bootstrap methodology in claim reserving. *Journal of Risk and Insurance*, **70**(4), 701-714.

RENSHAW, A.E. (1994). Claims reserving by joint modelling. Actuarial Research Paper No. 72, Department of Actuarial Science and Statistics, City University, London, EC1V 0HB.

RENSHAW, A.E. & VERRALL, R.J. (1998). A stochastic model underlying the chain ladder technique. *British Actuarial Journal*, **4**, 903-923.

SCOLLNIK, D.P.M. (2001). Actuarial modelling with MCMC and BUGS. *North American Actuarial Journal*, **5**(2), 96-125.

SPIEGELHALTER, D.J., THOMAS, A., BEST, N.G. & GILKS, W.R. (1996). *BUGS 0.5: Bayesian inference using Gibbs sampling*. MRC Biostatistics Unit, Cambridge, UK.

TAYLOR, G.C. (1988). Regression models in claims analysis: theory. *Proceedings of the Casualty Actuarial Society*, **74**, 354-383.

TAYLOR, G.C. (2000). *Loss reserving: an actuarial perspective*. Kluwer.

TAYLOR, G.C. & ASHE, F.R. (1983). Second moments of estimates of outstanding claims. *Journal of Econometrics*, **23**(1), 37-61.

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

VERRALL, R.J. (2004). A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving. *North American Actuarial Journal*, **8**, 67-89.

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

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

APPENDIX

ESTIMATION OF SCALE PARAMETERS

A.1 *Overview*

The distributional assumptions of generalised linear models are usually expressed in terms of the first two moments only, such that

$$\mathrm{E}[X_u] = m_u \quad \text{and} \quad \mathrm{Var}[X_u] = \frac{\phi V(m_u)}{w_u} \tag{A1.1}$$

where ϕ denotes a scale parameter, $V(m_u)$ is the so-called variance function (a function of the mean) and w_u are weights (often set to one for all observations). The choice of distribution dictates the values of ϕ and $V(m_u)$ (see McCullagh & Nelder, 1989).

Where a scale parameter is required, it is usually estimated as either the model deviance divided by the degrees of freedom, or the Pearson chi-squared statistic divided by the degrees of freedom, the choice often making little difference. The deviance and Pearson chi-squared statistics are obtained as the sum of the squares of their corresponding (unscaled) residuals.

Dropping the suffix denoting the 'unit', u, the deviance scale parameter is given by:

$$\hat{\phi}_D = \frac{\sum r_D(X, \hat{m}, w)^2}{N - p}$$

and the Pearson scale parameter is given by:

$$\hat{\phi}_P = \frac{\sum r_P(X, \hat{m}, w)^2}{N - p}$$

where N is the total number of observations, p is the number of parameters, and r_P and r_D denote the Pearson and deviance residuals respectively. The summation is over the number (N) of residuals. Note that, traditionally, the degrees of freedom are calculated using the number of parameters in the linear predictor only, and the scale parameter itself is not counted as a parameter.

It should be noted that, for linear regression models with normal errors, the residuals are simply the observed values less the fitted values (weighted where necessary), but for generalised linear models an extended definition of residuals is required which have (approximately) the usual properties of normal theory residuals. The precise form of the residual definitions is dictated by the distributional assumptions. In this paper, when estimating

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

scale parameters, we have used the unscaled weighted Pearson residuals, defined as:

$$r_P(X, \hat{m}, w) = \frac{X - \hat{m}}{\sqrt{\frac{V(\hat{m})}{w}}}. \tag{A1.2}$$

The precise definitions for the models considered in this paper are shown in Sections A1.1, A1.2 and A1.3.

Dropping the suffix denoting the type, we can write the calculation of the scale parameter as:

$$\hat{\phi} = \frac{\sum r(X, \hat{m}, w)^2}{N - p} = \frac{N}{N - p} \times \frac{\sum r(X, \hat{m}, w)^2}{N} = \frac{\sum \left(\left(\frac{N}{N-p} \right)^{\frac{1}{2}} r(X, \hat{m}, w) \right)^2}{N}.$$

That is, the scale parameter can be thought of as the average of the squared residuals multiplied by a bias correction factor $N/(N-p)$, or the average of squared bias adjusted residuals, where each residual is adjusted by a factor $(N/(N-p))^{1/2}$. The latter representation is useful, since it can be used when estimating non-constant scale parameters. The concept of bias adjusted residuals is also useful when bootstrapping, since they can be used to give results from a bootstrap analysis which are analogous to results obtained analytically using maximum likelihood methods (see England, 2002).

The Pearson residuals are more commonly used when bootstrapping, since they can easily be inverted to produce the pseudo-data required in the bootstrap process, although in a Bayesian setting, using plug-in estimates, it is just as easy to use scale parameters based on Pearson or deviance residuals.

If there is evidence that the scale parameter is not constant, simultaneous or 'joint' modelling of the mean and variance can be performed. Using maximum likelihood methods, this is an iterative process, whereby initial parameter estimates are obtained using arbitrary initial values for the scale parameters. The scale parameters are then updated, and revised parameter estimates obtained. The process iterates until convergence, although after the first iteration, the changes are usually small. In the models proposed in this paper, where a scale parameter by development period only is required, a good first approximation can be obtained by calculating the average of the squared bias adjusted (weighted) residuals at each development period, without the need for an iterative procedure. This allows neighbouring development periods to be combined, where appropriate, and in the limit, when all development periods are combined, the original formula for the

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

constant scale parameter emerges. This is a practical expedient which was developed when bootstrapping, since it is easy to apply in a spreadsheet. We adopt the same approach here when calculating 'plug-in' non-constant scale parameters, which enables a comparison between the bootstrap and the Bayesian approaches on a like-for-like basis. Therefore, the non-constant scale parameters can be calculated using:

$$\hat{\phi}_j = \frac{\sum_{i=1}^{n_j} \left(\left(\frac{N}{N-p} \right)^{\frac{1}{2}} r\left(X_{ij}, \hat{m}_{ij}, w_{ij}\right) \right)^2}{n_j} = \frac{N}{N-p} \times \frac{\sum_{i=1}^{n_j} r\left(X_{ij}, \hat{m}_{ij}, w_{ij}\right)^2}{n_j} \qquad \text{(A1.3)}$$

where n_j is the number of residuals in development period j. It should be noted that as j increases there are fewer data points available, which can lead to scale parameters which fluctuate erratically. It is possible to propose more complicated models for the scale parameters which apply parametric curves or other smoothing mechanisms to help ameliorate this, although manual smoothing would often be applied in practice. Further details on joint modelling in the context of stochastic reserving can be found in England & Verrall (2002) and Renshaw (1994). Note that Mack's model is formulated from the outset as a joint model of the mean and variance, using the average of the squared bias adjusted (weighted) residuals at each development period as the variance estimators (see A1.3).

A1.1 *Over-dispersed Poisson model*

For the over-dispersed Poisson model, the response variable is the incremental claims, C_{ij}, and equation (3.2) gives:

$$\mathrm{E}\left[C_{ij}\right] = m_{ij} \quad \text{and} \quad \mathrm{Var}\left[C_{ij}\right] = \phi m_{ij}.$$

Therefore, in terms of equation (A1.1), $X_u = C_{ij}$ and $V(m_u) = m_{ij}$. Then from equation (A1.2), the Pearson residuals are defined as:

$$r_P\left(C_{ij}, \hat{m}_{ij}\right) = \frac{C_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}}}.$$

A1.2 *Over-dispersed negative binomial model*

The over-dispersed negative binomial model can be constructed as a model for the incremental claims C_{ij}, the cumulative claims D_{ij}, or the development ratios f_{ij}. Using the development ratios f_{ij} as the response variable, equation (3.4) gives:

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

$$\mathrm{E}[f_{ij}] = \lambda_j \quad \text{and} \quad \mathrm{Var}[f_{ij}] = \frac{\phi\lambda_j(\lambda_j - 1)}{D_{i,j-1}} \quad \text{for} \quad j \geq 2.$$

Therefore, in terms of equation (A1.1), $X_u = f_{ij}$, $m_u = \lambda_j$, $w_u = D_{i,j-1}$ and $V(m_u) = \lambda_j(\lambda_j - 1)$. Then from equation (A1.2), the Pearson residuals are defined as:

$$r_P(f_{ij}, \hat{\lambda}_j, D_{i,j-1}) = \frac{\sqrt{D_{i,j-1}}(f_{ij} - \hat{\lambda}_j)}{\sqrt{\hat{\lambda}_j(\hat{\lambda}_j - 1)}}.$$

A1.3 *Mack's model*

Mack's model can also be constructed as a model for the incremental claims C_{ij}, the cumulative claims D_{ij}, or the development ratios f_{ij}. Using the development ratios f_{ij} as the response variable, equation (3.6) gives:

$$\mathrm{E}[f_{ij}] = \lambda_j \quad \text{and} \quad \mathrm{Var}[f_{ij}] = \frac{\sigma_j^2}{D_{i,j-1}} \quad \text{for} \quad j \geq 2.$$

Therefore, in terms of equation (A1.1), $X_u = f_{ij}$, $m_u = \lambda_j$, $w_u = D_{i,j-1}$ and $V(m_u) = 1$, using non-constant scale parameters $\phi_j = \sigma_j^2$. Then from equation (A1.2), the Pearson residuals are defined as:

$$r_P(f_{ij}, \hat{\lambda}_j, D_{i,j-1}) = \sqrt{D_{i,j-1}}(f_{ij} - \hat{\lambda}_j).$$

Note that using equation (A1.3) gives:

$$\hat{\phi}_j = \frac{N}{N-p} \times \frac{1}{n_j} \sum_{i=1}^{n_j} D_{i,j-1}(f_{ij} - \hat{\lambda}_j)^2.$$

In Mack (1993) a different bias correction was used, such that:

$$\hat{\phi}_j = \frac{n_j}{n_j - 1} \times \frac{1}{n_j} \sum_{i=1}^{n_j} D_{i,j-1}(f_{ij} - \hat{\lambda}_j)^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} D_{i,j-1}(f_{ij} - \hat{\lambda}_j)^2.$$

When parametric curves are used, it is not clear how the number of parameters would be taken into account using the formulation in Mack (1993), but in the formulation in this paper the number of parameters is taken into account automatically. However, to enable a comparison with results obtained using Mack's model, we have adopted the bias correction

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press

used in Mack (1993) when implementing the bootstrap and the Bayesian versions of Mack's model.

Note that the definition of the residual used in estimating the scale parameters in Mack's model is consistent with the assumption of a weighted normal regression model. Note also that there is insufficient information to estimate the scale parameter in the final development period (since there is only one observation at that point). In that case we have taken the minimum of the scale parameters in the previous two development periods, as suggested by Mack.

https://doi.org/10.1017/S1748499500000142 Published online by Cambridge University Press