STOCHASTIC CLAIMS RESERVING IN GENERAL INSURANCE

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

[Presented to the Institute of Actuaries, 28 January 2002]

ABSTRACT

This paper considers a wide range of stochastic reserving models for use in general insurance, beginning with stochastic models which reproduce the traditional chain-ladder reserve estimates. The models are extended to consider parametric curves and smoothing models for the shape of the development run-off, which allow extrapolation for the estimation of tail factors. The Bornhuetter-Ferguson technique is also considered, within a Bayesian framework, which allows expert opinion to be used to provide prior estimates of ultimate claims. The primary advantage of stochastic reserving models is the availability of measures of precision of reserve estimates, and in this respect, attention is focused on the root mean squared error of prediction (prediction error). Of greater interest is a full predictive distribution of possible reserve outcomes, and different methods of obtaining that distribution are described. The techniques are illustrated with examples throughout, and the wider issues discussed, in particular, the concept of a 'best estimate'; reporting the variability of claims reserves; and use in dynamic financial analysis models.

KEYWORDS

Stochastic Reserving; General Insurance; Chain-Ladder; Bornhuetter-Ferguson; Generalised Linear Model; Generalised Additive Model; Bayesian; Bootstrap; Simulation; Markov Chain Monte Carlo; Dynamic Financial Analysis

CONTACT ADDRESS

Dr P. D. England, EMB Consultancy, Nightingale House, 46-48 East Street, Epsom, Surrey KT17 1HP, U.K. E-mail: peter.england@emb.co.uk

1. Introduction

1.1 Although the last twenty years have witnessed increasing interest in stochastic claims reserving methods, they are still only used by a limited number of practitioners. A number of reasons for this could be suggested, including: a general lack of understanding of the methods; lack of flexibility in the methods; lack of suitable software; and so on. However, the main reason is probably lack of need for the methods, when traditional methods suffice for the calculation of a best estimate of outstanding claims reserves. More recently, greater interest has been expressed in estimating the downside potential of claims reserves, in addition to a best estimate. For that, it is

necessary to be able to estimate the variability of claims reserves, and ideally, to be able to estimate a full distribution of possible outcomes, from which percentiles (or other measures) of that distribution can be obtained. Stochastic claims reserving methods extend traditional techniques to allow those additional measures to be estimated.

1.2 The aims of this paper are threefold: to review some of the stochastic models which have been suggested, highlighting the connections between them; to show how the methods can be implemented in practice; and to discuss the characteristics of the models, interpretation of the results, and their wider usefulness.

1.3 It is not the aim of this paper to present a panacea for claims reserving. There will be sets of data for which it is difficult to justify any standard model-based approach. There will also be cases where none of the approaches set out will be appropriate. However, the models discussed have wide applicability, and we believe that a thorough and complete understanding of the stochastic basis is essential for the actuarial profession.

1.4 It is important to emphasise at the outset that, for the most part, this paper considers claims reserving techniques *applied mechanically and without judgement* (once a particular technique has been selected). This does not imply that the authors believe that judgement should not be used, nor that judgement is not possible in the context of stochastic methods. Indeed, using Bayesian methods, prior judgement can be incorporated quite naturally, and this concept is explored later. Nevertheless, the practical application is illustrated only to the extent of showing how the models can be fitted, not how expert opinion can be incorporated to fully satisfy the reserving practitioner. Before considering how judgement and expert knowledge can be incorporated, it is first helpful to have a theoretical framework which can then be extended to make the theory meet the practice. The theoretical framework is the subject of this paper, and it is hoped that further research in this area will be stimulated.

1.5 It is also important to emphasise that some of the methods in this paper are better suited to modelling paid amounts or numbers of claims, since incurred data, which may include over-estimation of case estimates, leading to negative incremental values, may cause problems. Negative incremental values can also arise due to timing of reinsurance or salvage recoveries, or premiums being included as negative loss amounts. It could be argued that the problem is more with the data than with the methods, since different processes are being combined and therefore polluting the data. The limitations of the various models are discussed as they are introduced, and further discussion appears in Section 10.

1.6 It is sometimes rather naïvely hoped that stochastic methods will provide solutions to problems when deterministic methods fail. Indeed, sometimes stochastic models are judged on whether they can help when

simple deterministic models fail. This rather misses the point. The usefulness of stochastic models is that they can, in many circumstances, provide more information which may be useful in the reserving process and in the overall management of the company.

1.7 There is a large and growing body of literature concerning loss reserving, some of which is concerned with business and management aspects. It is not the purpose of this paper to provide a detailed exposition of those issues. It is also not the aim of this paper to provide a complete survey of the literature, but we would direct the reader to a number of particular references. The book by Taylor (2000) provides a useful summary of deterministic loss reserving techniques, together with much useful material on stochastic methods and practical issues. The Casualty Actuarial Society web site is a useful resource. In particular, the papers from the annual loss reserving forum are available, and can be downloaded freely. Finally, there is the *Claims Reserving Manual*, published by the Institute of Actuaries.

1.8 The very nature of this paper means that a high technical content is unavoidable. We have generally tried to avoid long derivations and proofs intruding into the flow, since the application of the stochastic models is the main focus, not the proofs of statements made about the models. Where necessary, proofs and long derivations appear in appendices, or reference is made to papers where the mathematics is provided.

1.9 Because of its pre-eminent position in claims reserving, and because it is well-known, widely used, and easy to apply, we begin by concentrating on the chain-ladder technique. Very often, the chain-ladder technique is the first method to be applied, followed by manual smoothing of the resultant development factors, then adjustment of the results in line with expert opinion combined with additional information.

1.10 There has been a large number of papers investigating the statistical basis of the chain-ladder technique, notably, Kremer (1982), Taylor & Ashe (1983), Renshaw (1989), Verrall (1989, 1990, 1991a, 1991b, 1994, 1996, 2000), Mack (1993, 1994a, 1994b), Murphy (1994), Schmidt & Schnaus (1996), Renshaw & Verrall (1998), Barnett & Zehnwirth (1998), Mack & Venter (2000), and England & Verrall (1999, 2001). These have made significant advances in the understanding of the chain-ladder technique, and this paper aims to bring these together in a convenient form, and to show how extensions to the models are possible. Attention is focused primarily on stochastic methods based on the framework of generalised linear models.

1.11 This paper is structured as follows. Section 2 covers models which reproduce reserve estimates given by the chain-ladder technique. Section 3 covers models which usually provide results which are similar to chain-ladder estimates, but, for some data sets, there could be material differences. Extensions to the chain-ladder model using parametric curves, which reduce the number of parameters and allow extrapolation to help estimate tail

factors, are considered in Section 4. Smoothing models are considered in Section 5, which allow a seamless transition between the chain-ladder type models and the models based on parametric curves. Bayesian methods and the Bornhuetter-Ferguson technique are outlined in Section 6, which provide a way in which judgement can be incorporated. In Section 7, the predictions and prediction errors from the various models are considered and illustrated with examples. The examples provide an important part of the paper, since the practical implementation is illustrated, and the characteristics of the models are highlighted. Predictive distributions are considered in Section 8, which change the focus of stochastic claims reserving from the first two moments (the mean and prediction error), to the full distribution of predicted reserves. Hitherto, this has received little attention, but is of vital importance in dynamic financial analysis (DFA), which is the topic of Section 9. A discussion of the wider issues appears in Section 10.

1.12 An initial overview of the paper, ignoring most of the mathematical descriptions, can be obtained by reading Sections 1, 2.1, 2.2, 3.1, 4.1, 5.1, 6.1, 7.1, 8.1, 9 and 10.

2. Chain-Ladder Models

2.1 *Introduction*

2.1.1 The straightforward chain-ladder technique uses cumulative data, and derives a set of 'development factors' or 'link ratios'. We will show that to a large extent, it is irrelevant whether incremental or cumulative data are used when considering claims reserving in a stochastic context, and it is easier for the explanations here to use incremental. In order to keep the exposition as straightforward as possible, and without loss of generality, we assume that the data consist of a triangle of incremental claims. This is the simplest shape of data that can be obtained, and it is often the case that data from early origin years are considered fully run-off or that other parts of the triangle are missing. Using a triangle simply avoids us having to introduce complicated notation to cope with all possible situations. Thus, we assume that we have the following set of incremental claims data:

$$\{C_{ij} : i = 1, \ldots, n;\ j = 1, \ldots, n - i + 1\}.$$

The suffix i refers to the row, and could indicate accident year or underwriting year, for example. The suffix j refers to the column, and indicates the delay, here assumed also to be measured in years. It is straightforward to consider data collected more frequently for all models discussed in this paper.

2.1.2 The cumulative claims are defined by:

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

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

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

2.1.3 These are then applied to the latest cumulative claims in each row ($D_{i,n-i+1}$) to produce forecasts of future values of cumulative claims:

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

$$\hat{D}_{i,k} = \hat{D}_{i,k-1}\hat{\lambda}_k, \quad k = n-i+3, n-i+4, \ldots, n.$$

2.1.4 Thus, the chain-ladder technique, in its simplest form, consists of a way of obtaining forecasts of ultimate claims only. Here 'ultimate' is interpreted as the latest delay year so far observed, and does not include any tail factors. From a statistical viewpoint, given a point estimate, the natural next step is to develop estimates of the likely variability in the outcome so that assessments can be made, for example, of whether extra reserves should be held for prudence, over and above the predicted values. In this respect, the measure of variability commonly used is the *prediction error*, defined as the standard deviation of the distribution of possible reserve outcomes. It is also desirable to take account of other factors, such as the possibility of unforeseen events occurring which might increase the uncertainty, but which are difficult to model.

2.1.5 The first step to obtaining the prediction error is to formulate an underlying statistical model making assumptions about the data. If the aim is to provide a stochastic model which is analogous to the chain-ladder technique, then an obvious first requirement is that the predicted values should be the same as those of the chain-ladder technique. There are two ways in which this has been attempted: specifying distributions for the data; or just specifying the first two moments. We begin by considering models which specify distributions for the data, and consider the (over-dispersed) Poisson distribution, the negative binomial distribution, and the Normal distribution in this context. Before describing the models in detail, a summary is provided, which sets the scene.

2.2 *Summary of Chain-Ladder Models*

2.2.1 The aim of all of the models in this section is to give the same reserve estimates as the chain-ladder technique, and the differences between them are mostly implementational. The negative binomial model of Section 2.4 is derived from the Poisson model of Section 2.3, and, as such, is very closely related, but with a different parameterisation. We regard the Normal model in Section 2.5 (and by implication Mack's model) as an approximation to the negative binomial model, and we examine, through examples in Sections 7.2 to 7.5, the implications for the prediction errors of using these different assumptions (the predictions are the same in each case). The Normal model has the advantage that it can produce estimates for a wide range of data sets, and is less affected by the presence of negatives.

2.2.2 Murphy (1994) also considers the chain-ladder technique within a Normal linear regression context, and various extensions are suggested by Barnett & Zehnwirth (1998). The exposition in this paper differs in that we derive the Normal model as an approximation to the negative binomial, and make explicit reference to non-constant error components estimated by 'joint modelling' (see Section 7.4), which provides a link to Mack's model.

2.2.3 Renshaw & Verrall (1998) were not the first to notice the link between the chain-ladder technique and the Poisson distribution, but were the first to implement the model using standard methodology in statistical modelling, and to provide a link with the analysis of contingency tables. Wright (1990) also describes a similar model, including a term to model claims inflation, but did not consider the model in detail. Mack (1991) also points out that the chain-ladder estimates can be obtained by maximising a Poisson likelihood by appealing to the so called 'method of marginal totals'.

2.2.4 A discussion of the stochastic basis of chain-ladder models can be found in Mack (1994a), Verrall (2000), Mack & Venter (2000) and Verrall & England (2000). At the heart of the discussion is the relationship between the various models, and whether they can justifiably be used to add value to the deterministic chain-ladder technique. Some of the questions raised include the issue of whether incremental or cumulative claims should be modelled, how many parameters are estimated and what happens when there are negative incremental claims. We hope that, by setting out the models and providing a set of examples in later sections, we are able to clarify some of these issues. Some of the questions raised are irrelevant. For example, it does not matter in principle whether incremental or cumulative claims are used; whatever data are available can be used with the corresponding model, and the same results are obtained. Counting the number of parameters being estimated explicitly can be misleading; the negative binomial model has fewer parameters requiring estimation than the Poisson model, but the predictive distributions are the same. With a sufficient number of negative incremental claims, it is necessary to use a Normal approximation, and the chain-ladder results can still be reproduced. We would not recommend using

the Normal approximation in all situations, even though it is likely to be less troublesome in practice, but would use it if features of the data required its use (or diagnostic tests showed it to be more suitable). Our view on this, as on many other issues, is a pragmatic one: reserving is a practical data analysis exercise, and we should try to understand and learn from the data rather than impose the same approach in all situations.

2.3 *Over-Dispersed Poisson Model*

2.3.1 The over-dispersed Poisson distribution differs from the Poisson distribution in that the variance is not *equal* to the mean, but, instead, is *proportional* to the mean. In claims reserving, 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:

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

where:

$$\sum_{k=1}^{n} y_k = 1.$$

Here, x_i is the expected ultimate claims (where ultimate means up to the latest development year observed in the triangle), and y_j is the proportion of ultimate claims to emerge in each development year. Over-dispersion is introduced through the parameter ϕ, which is unknown and estimated from the data. Allowing for over-dispersion does not affect estimation of the parameters, but does have the effect of increasing their standard errors. Full details of this model can be found in Renshaw & Verrall (1998).

2.3.2 It should be noted that, since y_j appears in the variance, the restriction that y_j must be positive is automatically imposed. This implies that the sum of incremental claims in column j must also be positive, which is a limitation of the model. Note that some negative incrementals are allowed, as long as any column sum is not negative.

2.3.3 In this formulation, the mean has a multiplicative structure, that is, it is the product of the row effect and the column effect. Both the row effect and the column effect have specific interpretations (being the expected ultimate claims in each origin year and proportion of ultimate to emerge in each development year, respectively), and it is sometimes useful to preserve the model in this form.

2.3.4 However, for estimation purposes, it is often better to re-parameterise the model so that the mean has a linear form. In the terminology of generalised linear models, we use a log link function so that:

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

This predictor structure is still a chain-ladder type, in the sense that there is a parameter for each row i, and a parameter for each column j. There are some advantages and some disadvantages to this form of the model. As a generalised linear model, it is easy to estimate, and standard software packages can be used; the estimates should be well behaved. However, the parameter values themselves will be harder to interpret, making it necessary to convert them back into more familiar quantities. 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 would put $\alpha_1 = \beta_1 = 0$.

2.3.5 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, positive and negative. 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, identical parameter estimates are obtained using the full or quasi-likelihood. Many statistical packages fit GLMs using quasi-likelihood by default, the user being entirely unaware. 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.

2.4 *Negative Binomial Model*

2.4.1 A model which is closely related to the over-dispersed Poisson model is the negative binomial model (Verrall, 2000). This is more intuitive for the chain-ladder technique, since parameters appear to be more 'like' the chain-ladder development factors. The negative binomial model can be expressed as a model for either incremental or cumulative claims; C_{ij} has an over-dispersed negative binomial distribution, with mean and variance:

$$(\lambda_j - 1)D_{i,j-1} \quad \text{and} \quad \phi\lambda_j(\lambda_j - 1)D_{i,j-1}, \quad \text{respectively,}$$

where λ_j is analogous to the standard chain ladder development factor. Again, there is an unknown dispersion parameter ϕ in the variance, making the distribution 'over-dispersed'. Henceforth, this label is dropped when referring to the negative binomial model, for convenience.

2.4.2 Note that $D_{ij} = D_{i,j-1} + C_{ij}$, and it is assumed in this (recursive) approach that $D_{i,j-1}$ is known. This means that we can also write this model in terms of the cumulative claims; D_{ij} has an over-dispersed negative binomial distribution, with mean and variance:

$$\lambda_j D_{i,j-1} \quad \text{and} \quad \phi\lambda_j(\lambda_j - 1)D_{i,j-1}, \quad \text{respectively.}$$

2.4.5 Note that the variance holds only where $D_{i,j-1}$ is known, and the recursive procedure outlined in Section 7.3 and Appendix 1 is required where cumulative claims are forecast.

2.4.6 The negative binomial model was first derived by Verrall (2000), by integrating out the row parameters from the Poisson model, and full details can be found in that paper. Since the negative binomial model is derived from the Poisson model, the predictive distributions are essentially the same, and give identical predicted values.

2.5 *Normal Approximation to the Negative Binomial Model*

2.5.1 Notice that if $\lambda_j < 1$, the variance of incremental claims in Section 2.4 is negative, and the negative binomial model breaks down. $\lambda_j < 1$ implies that the sum of incremental claims in column j is negative. To make any progress in this case, it is necessary to use a distribution whose support is not restricted to the positive real line, and a suitable candidate is the Normal distribution, for that reason. We can imagine that some refinements of this are likely to be suggested, to allow for the fact that the distribution of the data is unlikely to be symmetrical. However, as a first step, it is possible to replace the negative binomial by a Normal distribution, whose mean is unchanged, but whose variance is altered to accommodate the case when $\lambda_j < 1$. Preserving as much of $\lambda_j(\lambda_j - 1)D_{i,j-1}$ as possible, we would expect the variance to be proportional to $D_{i,j-1}$, with the constant of proportionality depending on j. Using a Normal approximation for the distribution of incremental claims C_{ij} is approximately Normally distributed, with mean and variance:

$$D_{i,j-1}(\lambda_j - 1) \quad \text{and} \quad \phi_j D_{i,j-1}, \quad \text{respectively,}$$

or D_{ij} is approximately Normally distributed, with mean and variance:

$$\lambda_j D_{i,j-1} \quad \text{and} \quad \phi_j D_{i,j-1}, \quad \text{respectively.}$$

Again this holds where $D_{i,j-1}$ is known, and a recursive approach to estimating the variance is required where cumulative claims are forecast.

2.5.2 These models have unknown parameters in the variance, which can be estimated using an iterative estimation procedure known as 'joint modelling', since it involves estimating parameters for the mean and the variance, which deviates from the usual practice of modelling the mean alone. Details can be found in McCullagh & Nelder (1989), and an overview is given in Section 7.4.

2.5.3 Both the negative binomial model, and the Normal approximation are recursive models. This means that the likelihood is written down by taking the data in a certain (obvious) order. Another model which takes a similar recursive approach, and which is closely related to the Normal approximation to the negative binomial, is Mack's model.

2.6 *Mack's Model*

2.6.1 One of the earliest stochastic models which reproduced chain-ladder estimates is the model of Mack (1993), which made limited assumptions as to the distribution of the underlying data, preferring simply to specify the first two moments only. According to Mack (1993), the mean and variance of D_{ij} are:

$$\lambda_j D_{i,j-1} \quad \text{and} \quad \sigma_j^2 D_{i,j-1}, \quad \text{respectively.}$$

2.6.2 Mack produced estimators of the unknown parameters λ_j and σ_j^2 and, making further limited assumptions, provided formulae for the prediction errors of predicted payments and reserve estimates. Readers of Mack's papers should be aware that we have changed notation to be consistent within this paper and other papers by the same authors.

2.6.3 Mack considers the model to be distribution-free, since the full distribution of the underlying data is not specified. While this has the advantage of simplicity, it limits analysis of the distribution of outstanding reserves to the first two moments only. Further assumptions are necessary if the results are used in a dynamic financial analysis exercise where the distribution of outstanding reserves might be simulated.

2.6.4 Comparison of the mean and variance of D_{ij} shown in this section with the mean and variance of D_{ij} in the Normal approximation to the negative binomial in the previous section reveals a striking similarity, with the unknown scale parameters ϕ_j of the Normal approximation being replaced by σ_j^2 in Mack's model. In fact, parameter estimates, predicted values and prediction errors provided by the two models are essentially identical, as shown in Sections 7.4 and 7.5.

3. Other Modelling Distributions

3.1 *Introduction*

Stochastic reserving models are not restricted to the modelling distributions used in Section 2. The log-Normal distribution, which was used in early work on stochastic claims reserving, has received considerable attention. In some ways this is a historical accident, since the methods were first used when generalised linear modelling software was in its infancy, but fitting Normal errors models by minimising the residual sum of squares was

straightforward. This does not necessarily mean that the log-Normal distribution should never be used. Indeed it remains useful, since a predictive distribution of reserves can be obtained by simulating from the parameters, as described in Section 8. It might also be preferable in situations where the downside potential is very large, since it tends to result in a predictive distribution which is heavily skewed (see Section 8). A further distribution, which has received little attention in stochastic claims reserving, is the Gamma distribution. Why the Gamma distribution has been largely overlooked is not clear. The log-Normal and Gamma distributions are considered in Sections 3.2 and 3.3 respectively.

3.2 *Log-Normal Model*

3.2.1 Early work on stochastic models for the chain-ladder technique focused on the logarithm of the incremental claims amounts $Y_{ij} = \log(C_{ij})$ and the log-Normal class of models $Y_{ij} = m_{ij} + \varepsilon_{ij}$, with ε_{ij} as independent Normal random errors. That is:

$$\varepsilon_{ij} \sim \text{IN}(0, \sigma^2) \quad \text{or} \quad Y_{ij} \sim \text{IN}(m_{ij}, \sigma^2).$$

3.2.2 The use of the logarithmic transform immediately imposes a limitation on this class of models, in that incremental claim amounts must be positive. The Normal responses Y_{ij} are assumed to decompose (additively) into a deterministic non-random component with mean $m_{ij} = \eta_{ij}$ and Normally distributed random error components about a zero mean. Using the chain-ladder type structure for the mean, introduced in Section 2.3, gives:

$$\eta_{ij} = c + \alpha_i + \beta_j.$$

3.2.3 This log-Normal model was introduced by Kremer (1982) and used by Renshaw (1989), Verrall (1989), Zehnwirth (1989) and Christofides (1990), amongst others. Use of this model usually produces predicted values close to those from the simple chain-ladder technique, but it is not guaranteed, and there can be material differences.

3.2.4 The log-Normal distribution has the advantage that it can be implemented without the need for specialist software. For example, Christofides (1990) showed how spreadsheets could be used to analyse the data using log-Normal models.

3.2.5 Another advantage is that other statistical techniques can also be used to allow different assumptions to be incorporated concerning the run-off pattern and the connections between origin years. Since Y_{ij} is Normally distributed, it is possible, for example, to apply the Kalman filter (Verrall, 1989). This is an example of the possibility of moving beyond the simple

chain-ladder technique. The Kalman filter allows, for example, a way of smoothing the row parameters, instead of treating the rows as completely separate. In general, the Normal distribution is easier to use in this context, and there are many possibilities for imaginative use to be made of the existing theory. Care must be taken with the predictions from this model, as discussed by Verrall (1991a) and Doray (1996).

3.3 *Gamma Model*

Mack (1991) suggested using the same linear predictor as Kremer (and therefore the same as Renshaw & Verrall), but proposed using a Gamma distribution for claim amounts. As Renshaw & Verrall (1998) note, the same model can be fitted using the GLM described in Section 2.3, but replacing $\text{Var}[C_{ij}] = \phi m_{ij}$ by $\text{Var}[C_{ij}] = \phi m_{ij}^2$. Standard statistical software packages can then be used to obtain maximum likelihood parameter estimates. Again, the predicted values provided by this model are usually close to the chain-ladder estimates, but it is not guaranteed.

3.4 *Other Modelling Distributions*

In general, we could consider models where the variance is proportional to the kth power of the mean, and assess different values for k. The choice of the value of k could be made *a priori,* for example $k = 0$ implies a Normal distribution, $k = 1$ implies a Poisson and $k = 2$ a Gamma, or it could be chosen after looking at the data. In principle, we are not restricted to integer values for k.

4. Other Parametric Models

4.1 *Introduction*

One of the criticisms sometimes made of the chain-ladder technique is that it is over-parameterised, in the sense that a separate parameter is used for each development period. There have been various suggestions as to how this could be remedied, usually by fitting a parametric curve to the run-off pattern. Of the curves suggested, the most popular appears to be the gamma curve, also known as the Hoerl curve (Wright, 1990, Renshaw, 1994a, Zehnwirth, 1985). The Hoerl curve is useful, since the general shape of the curve is similar to the usual shape of the run-off of incremental claims, that is, increasing rapidly to a peak, then dying off exponentially.

4.2 *The Hoerl Curve*

4.2.1 For the log-Normal model of Section 3.2, or models with a log link function (such as the over-dispersed Poisson model of Section 2.3, and the Gamma model of Section 3.3), the Hoerl curve is provided by replacing

the chain-ladder type linear predictor which has a parameter for each development year with:

$$\eta_{ij} = c + \alpha_i + \beta_i \log(j) + \gamma_i j \qquad (j > 0). \tag{4.1}$$

The parametric form of the model on the untransformed scale can be seen by exponentiating equation 4.1, giving:

$$\exp(\eta_{ij}) = A_i j^{\beta_i} e^{\gamma_i j} \quad \text{where:} \quad A_i = \exp(c + \alpha_i).$$

4.2.2 Here, development year j is considered as a continuous covariate, and the run-off pattern follows a fixed parametric form, being linear in development time and log development time on a log scale. The advantage of working on a log scale is that parameters can be readily estimated. An advantage of treating development time as a continuous covariate is that extrapolation is possible beyond the range of development time observed. This is beneficial, since it helps with estimating tail factors.

4.2.3 A special case is created by setting $\beta_i = \beta$ and $\gamma_i = \gamma$ for all i, where the decay pattern is the same for all origin years and is represented by only two parameters. De Jong & Zehnwirth (1983) adopted the Kalman filter, which provides a way of passing information between origin years and provides smoothed estimates of the parameters β_i and γ_i.

4.2.4 It is unlikely that the Hoerl curve will fit well over the entire range of development time. The fit can often be improved dramatically by allowing the first few development years to have their own level, and imposing a parametric form from that point on.

4.3 *Wright's Model*

4.3.1 Significant advances were made in stochastic claims reserving with the publication of a paper by Wright (1990), in which the systematic and random components of the underlying model were based on a risk theoretic model of the claims generating process.

4.3.2 Wright considered the incremental paid claims C_{ij} to be the sum of N_{ij} (independent) claims of amount X_{ij}. The claim numbers N_{ij} were assumed to be Poisson random variables, where:

$$\mathrm{E}[N_{ij}] = e_i a_j \kappa_i j^{A_i} \exp(-b_i j)$$

and

$$\mathrm{Var}[N_{ij}] = \mathrm{E}[N_{ij}]$$

where κ, A and b are unknown constants to be estimated, e is a measure of

exposure, and a is a *known* adjustment term needed on technical grounds. The values a are specified in Appendix 1 of Wright for each value of j. (Note: Wright also recommended a technical adjustment to development time j, which has been ignored here for simplicity.) The variance of the number of claims equals the mean as a result of using the Poisson distribution.

4.3.3 Claim amounts X_{ij} were considered to be Gamma type random variables where:

$$\mathrm{E}\left[X_{ij}\right] = \exp(\delta t) k j^{\lambda}$$

and

$$\mathrm{Var}\left[X_{ij}\right] = \upsilon\left\{\mathrm{E}\left[X_{ij}\right]\right\}^2$$

where k and λ are unknown constants. The optional term $\exp(\delta t)$ is included to allow for possible claims inflation, where $t = i + j$ represents calendar time and δ is the estimated constant force of claims inflation. Wright chose not to assume that the claim amounts are actually Gamma distributed, only that the variance exists and is proportional to the mean squared. This is a subtle technicality which makes no practical difference when claim amounts are all positive.

4.3.4 This formulation is interesting, because it uses the same model specification in the claims reserving context as in pricing, that is, claim numbers are modelled as Poisson random variables and claim severities are modelled as Gamma random variables.

4.3.5 Combining the models for claims frequency and severity, using standard results for risk models, gives:

$$\mathrm{E}\left[C_{ij}\right] = m_{ij} = e_i a_j \kappa_i j^{A_i} \exp(-b_i j) \exp(\delta t) k j^{\lambda}$$

and

$$\mathrm{Var}\left[C_{ij}\right] = (1 + \upsilon) k j^{\lambda} \exp(\delta t) \mathrm{E}\left[C_{ij}\right].$$

4.3.6 Wright showed that with a suitable reparameterization, this represents a generalised linear model, and went on to use the Kalman filter to pass information between origin years to produce smoothed parameter estimates, thus avoiding problems associated with the excessive parameterisation.

4.3.7 The formulation of the problem as a GLM and the fitting method adopted by Wright are not easy to follow, but a simpler derivation was given by Renshaw (1994a), who showed that the model for incremental claims can be written as follows:

$$\mathrm{E}[C_{ij}] = \exp(u_{ij} + c + \alpha_i + \beta_i \log(j) + \gamma_i j + \delta t)$$

and

$$\mathrm{Var}[C_{ij}] = \phi_{ij}\mathrm{E}[C_{ij}] = \phi_{ij} m_{ij}.$$

This is a generalised linear model in which the response C_{ij} is modelled with a logarithmic link function and the variance is proportional to the mean. The ϕ_{ij} are unknown scale parameters to be estimated by the model, via the process of 'joint modelling' (see Section 7.4).

4.3.8 Note that we can write:

$$\log(\mathrm{E}[C_{ij}]) = \eta_{ij}, \quad \text{where:} \quad \eta_{ij} = u_{ij} + c + \alpha_i + \beta_i \log(j) + \gamma_i j + \delta t.$$

Ignoring the u_{ij} terms, for the moment, and the optional term for claims inflation (δt), this represents the familiar Hoerl curve which appeared in Section 4.2. Wright effectively uses the same linear predictor as Zehnwirth (1985), with the inclusion of an optional term to model possible claims inflation. The u_{ij} terms are known, and represent small technical adjustments. They are declared as *offsets* when fitting the model using standard statistical software packages. The important differences between the model used by Zehnwirth (1985) and the model proposed by Wright are that Zehnwirth uses the logarithm of the incremental claims as the response, and links the linear predictor to the expected value of the response through the identity link function, whereas Wright treats the incremental claims themselves as the responses, and links (essentially) the same predictor to the expected value of the response through the logarithmic link function. This has implications for the necessity of bias correction factors. A useful critique of the differing assumptions concerning the distributional assumptions can be found in Appendix 4 of Wright (1990).

5. Non-Parametric Smoothing Models

5.1 *Introduction*

5.1.1 It is often the case that parametric curves are too rigid (in some ways the opposite problem to the chain-ladder technique, which assumes no prior shape on the run-off), and England & Verrall (2001) proposed using non-parametric smoothing methods as an alternative. England & Verrall showed that it is possible to use a wide range of models with a non-parametric approach, with the chain-ladder technique at one end of the range, and the Hoerl curve at the other. The non-parametric smoothing models move seamlessly between these two extremes, and allow the practitioner to choose a model somewhere between the two. It is straightforward to examine the effect on the run-off pattern. Another example of the use of non-parametric

smoothing was given in Verrall (1996). In that paper, the stochastic chain-ladder model of Renshaw & Verrall (1998) was extended to incorporate smoothing of parameter estimates over origin years, while leaving the model describing the run-off pattern alone.

5.1.2 Reserving specialists are probably more familiar with the practice of first fitting a chain-ladder model (or variation thereof), then smoothing the resultant development factors using a model with a fixed parametric form (see, for example, Sherman, 1984). Using that approach, the development factors themselves become the focus, and a model is fitted to development factors with equal weight (usually) being given to each development factor. The non-parametric smoothers described in this section differ, in that the incremental claims themselves are the focus, and smoothing is performed at the same time as model fitting. No prior form is assumed for the shape of the run-off, the smoothing procedure allows the model to follow the trends in the data. The implied development factors can be calculated from the fitted values of the model, which will automatically be smoothed.

5.2 *Smoothing Models*

5.2.1 The smoothing models are implemented using generalised *additive* models (GAMs) (Hastie & Tibshirani, 1990). GAMs differ from GLMs in the way in which the relationship between the response variable and the covariates is modelled. In GLMs the relationship is parametric; in GAMs the response is assumed to vary smoothly with the covariates through the introduction of a smoothing procedure.

5.2.2 GAMs can be regarded as extensions of GLMs, with the linear predictor being replaced by a non-parametric smoother:

$$\eta_{ij} = \sum_{v=1}^{p} s_v(x_{ij})$$

where $s(x)$ represents a non-parametric smoother on x, which may be chosen from several different types of smoother, such as locally weighted regression smoothers (loess), cubic smoothing splines and kernel smoothers. Note that we are not restricted to using a smoother for all covariates; the predictor may comprise a mixture of parametric and non-parametric components.

5.2.3 We concentrate here on the cubic smoothing spline, which is well-known in the actuarial literature for graduating mortality rates (Whittaker, 1923). When data are Normally distributed, the (univariate) cubic smoothing spline $\hat{y} = s(x)$ is found by minimising the penalised residual sum of squares:

$$\sum_i (y_i - s(x_i))^2 + \theta \int (s''(t))^2 dt.$$

The second part of this quantity defines a smoothness penalty based on

curvature of the spline function $s(x)$. The level of smoothing is controlled by the single parameter $\theta(> 0)$. When θ tends to zero, there is no smoothness penalty, and the model provides a perfect fit; the fitted values are the data points themselves. When θ is large (tends to infinity), the fit is perfectly smooth, and the fitted values fall along a straight line, effectively forcing the relationship to be linear in x. The parameter θ is set between these extremes to produce the desired level of smoothness, and controls the trade-off between goodness of fit and smoothness. For other distributions, such as the Poisson, the fitting procedure is extended using an iterative process, and has some similarity with the procedure used for fitting GLMs (see Hastie & Tibshirani, 1990; Green & Silverman, 1994).

5.2.4 In claims reserving, we simply generalise equation 4.1, and define the predictor (for incremental claims) as:

$$\eta_{ij} = s_{\theta_i}(i) + s_{\theta_j}(\log(j)) + s_{\theta_j}(j).$$

The function $s(i)$ represents a smooth of origin year i, obtained using a smoothing spline with smoothing parameter θ_i. Similarly, the functions $s(j)$ and $s(\log(j))$ represent smoothing splines specifying the shape of the run-off pattern, with smoothing parameter θ_j chosen to be the same for both functions (for simplicity). In practice, it may not be necessary to include smooths in both j and $\log(j)$. It should be noted that both origin year i and development year j are considered as continuous covariates. It can be seen that use of this predictor implicitly assumes the same run-off pattern for all origin years, although the model can be extended using carefully chosen interaction terms.

5.2.5 The extremes of the smoothing parameters are interesting, and provide the link between the chain-ladder model and the Hoerl curve. When θ_i is zero there is no smoothing, and the model is forced to pass through each value of i, which treats origin year i as though it is a factor. The same is true of θ_j; when θ_j is zero, the model is forced to pass through each value of j, and development time is treated as though it is a factor. The result is that the model collapses to the chain-ladder model, and gives the same results as the over-dispersed Poisson model of Section 2.3. When θ_j tends to infinity, the part of the model relating to development time is linear in j and $\log(j)$, giving the Hoerl curve (as in Section 4.2). An example where this is shown can be found in England & Verrall (2001).

5.2.6 These models may be fitted using S-PLUS (2001) (see also Chambers & Hastie, 1992). We believe that a key advantage of using the non-parametric models described in this section is that it is possible to use elegant alternatives to the straightforward deterministic or parametric models. We believe that other methods could also be explored in order to reproduce some of the adjustments which are at present made on an *ad hoc* basis to the deterministic techniques.

6. Bayesian Models and the Bornhuetter-Ferguson Technique

6.1 *Introduction*

The Bornhuetter-Ferguson technique (Bornhuetter & Ferguson, 1972) has proved popular when there is instability in the proportion of ultimate claims paid in the early development years, causing a method such as the chain-ladder technique to produce unsatisfactory results when applied mechanically. The idea is to try to stabilise the results using an external initial estimate of ultimate claims. This is then used with the development factors of the chain-ladder technique, or something similar, to estimate outstanding claims. The recent paper by Mack (2000) provides an excellent summary of the Bornhuetter-Ferguson technique, and subsequent work by other authors, and gives details of an approach which uses credibility theory, first suggested by Benktander (1976). Here, we continue in the spirit of earlier sections, and look at a model within the framework of generalised linear models. Since we use the over-dispersed Poisson model, the theory in this section is not applicable to all sets of data, and can break down in the presence of a sufficient number of negative incremental claims.

6.2 *The Bornhuetter-Ferguson Technique*

6.2.1 For the chain-ladder technique, the estimate of outstanding claims is:

$$D_{i,n-i+1}\left(\lambda_{n-i+2}\lambda_{n-i+3}\ldots\lambda_n - 1\right).$$

That is, the latest cumulative claims are multiplied by the product of the remaining development factors (giving an estimate of the ultimate claims), and the paid to date are subtracted, giving the reserve. This could be written as:

$$U_i^{(CL)}\frac{1}{\lambda_{n-i+2}\lambda_{n-i+3}\ldots\lambda_n}\left(\lambda_{n-i+2}\lambda_{n-i+3}\ldots\lambda_n - 1\right)$$

where $U_i^{(CL)}$ is the chain-ladder estimate of ultimate claims.

6.2.2 The Bornhuetter-Ferguson technique differs from the chain-ladder technique in that the chain-ladder estimate of ultimate claims is replaced by an alternative estimate, $U_i^{(BF)}$, based on external information and expert judgement.

6.2.3 The Bornhuetter-Ferguson estimate of outstanding claims for origin year i is therefore:

$$U_i^{(BF)}\frac{1}{\lambda_{n-i+2}\lambda_{n-i+3}\ldots\lambda_n}\left(\lambda_{n-i+2}\lambda_{n-i+3}\ldots\lambda_n - 1\right)$$

The use of external information to provide the initial estimate $U_i^{(BF)}$ leads naturally to a Bayesian model.

6.3 *Bayesian Models*

6.3.1 In the over-dispersed Poisson chain-ladder model described in Section 2.3, no prior assumptions are made about the row parameters $\{x_i : i = 1, \ldots, n\}$. The Bornhuetter-Ferguson technique assumes that there is prior knowledge about these parameters, making the Bornhuetter-Ferguson technique analogous to a Bayesian approach. The prior information can be summarised using an appropriate prior distribution for the row parameter. Several prior distributions are available for this purpose, although an obvious candidate is the Gamma distribution, giving:

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

such that:

$$\mathrm{E}[x_i] = U_i^{(BF)} = \frac{\alpha_i}{\beta_i}.$$

6.3.2 This Bayesian model can be implemented using a Markov chain Monte Carlo approach (MCMC, see Section 7.11), through the software WinBUGS (Spiegelhalter *et al.*, 1996).

6.3.3 However, it is useful to look at the theoretical predictive distribution for the data (see Verrall, 2001), in order to compare the Bornhuetter-Ferguson technique with the chain-ladder technique. The theoretical predictive distribution of C_{ij} is an over-dispersed negative binomial distribution, with mean:

$$\left(Z_{ij} D_{i,j-1} + (1 - Z_{ij}) \frac{\alpha_i}{\beta_i} \frac{1}{\lambda_j \lambda_{j+1} \ldots \lambda_n}\right)(\lambda_j - 1)$$

where:

$$Z_{ij} = \frac{\dfrac{1}{\lambda_j \lambda_{j+1} \ldots \lambda_n}}{\beta_i \phi + \dfrac{1}{\lambda_j \lambda_{j+1} \ldots \lambda_n}}.$$

6.3.4 It can be seen that this is in the form of an actuarial 'credibility formula'. In modern statistical terms, it uses a natural trade off between two competing estimates:

$$D_{i,j-1} \quad \text{and} \quad \frac{\alpha_i}{\beta_i}\frac{1}{\lambda_j\lambda_{j+1}\ldots\lambda_n} = U_i^{(BF)}\frac{1}{\lambda_j\lambda_{j+1}\ldots\lambda_n}.$$

6.3.5 The Bayesian model has the chain-ladder technique as one extreme (no prior information about the row parameters), and the Bornhuetter-Ferguson technique as the other (perfect prior information about the row parameters). It is interesting to note that the Bornhuetter-Ferguson technique essentially assumes that there is perfect prior information about the row parameters, and does not use the data at all for this part of the estimation. In practice, we would regard this as a rather heroic assumption, and would prefer to use something between the Bornhuetter-Ferguson technique and the chain-ladder technique.

6.3.6 The credibility factor Z_{ij} governs the trade-off between the prior mean and the data. Notice that the further through the development we are, the larger $\dfrac{1}{\lambda_j\lambda_{j+1}\ldots\lambda_n}$ becomes, and the more weight is given to the chain-ladder estimate. The choice of β_i is governed by the prior precision of the initial estimate for ultimate claims, and this should be chosen with due regard given to the over-dispersion parameter (an initial estimate of which could be obtained from the over-dispersed Poisson model of Section 2.3).

7. Predictions and Prediction Errors

7.1 *Introduction*

7.1.1 Claims reserving is a predictive process: given the data, we try to predict future claims. In Sections 2 to 6 different models have been outlined, from which future claims can be predicted. In this context, we use the expected value as the prediction. When considering variability, attention is focused on the root mean squared error of prediction (RMSEP), also known as the prediction error.

7.1.2 Consider a random variable y and a predicted value $\hat{y}$. The mean squared error of prediction (MSEP) is:

$$\mathrm{E}\big[(y-\hat{y})^2\big] = \mathrm{E}\big[((y-\mathrm{E}[y]) - (\hat{y}-\mathrm{E}[y]))^2\big].$$

Plugging in $\hat{y}$ instead of y in the final expectation and expanding gives:

$$\mathrm{E}\big[(y-\hat{y})^2\big] \approx \mathrm{E}\big[(y-\mathrm{E}[y])^2\big] - 2\mathrm{E}[(y-\mathrm{E}[y])(\hat{y}-\mathrm{E}[\hat{y}])] + \mathrm{E}\big[(\hat{y}-\mathrm{E}[\hat{y}])^2\big].$$

Assuming future observations are independent of past observations gives:

$$\mathrm{E}\left[(y-\hat{y})^2\right] \approx \mathrm{E}\left[(y-\mathrm{E}[y])^2\right] + \mathrm{E}\left[(\hat{y}-\mathrm{E}[\hat{y}])^2\right]$$

which, in words, is:

Prediction variance = Process variance + Estimation variance.

7.1.3 When trying to estimate the prediction error of future payments and reserve estimates using classical statistical methods, the problem reduces to estimating the two components: the process variance and the estimation variance. Alternatively, if the full predictive distribution can be found, the RMSEP can be obtained directly by calculating its standard deviation. It is preferable to have the predictive distribution, and, in this respect, Bayesian methods offer the best way forward

7.1.4 It is important to understand the difference between the prediction error and the standard error. Strictly, the standard error is the square root of the estimation variance. The prediction error is concerned with the variability of a forecast, taking account of uncertainty in parameter estimation and also of the inherent variability in the data being forecast. Unfortunately, there is confusion in the literature over terminology, with the RMSEP also being called the standard error of prediction, or simply the standard error.

7.2 *The Over-Dispersed Poisson Chain-Ladder Model*

7.2.1 A general form for the over-dispersed Poisson chain-ladder model can be written as follows:

$$\mathrm{E}\left[C_{ij}\right] = m_{ij} \qquad \mathrm{Var}\left[C_{ij}\right] = \phi m_{ij} \tag{7.1}$$

with the specification being completed by specifying a structural form for m_{ij}. Two structures have been suggested, either:

$$m_{ij} = x_i y_j \tag{7.2}$$

or

$$\log(m_{ij}) = c + \alpha_i + \beta_j. \tag{7.3}$$

7.2.2 With the first structure the model is non-linear in the parameters, and non-linear modelling techniques are required to obtain estimates of the parameters. If maximum likelihood estimates are required, this involves setting up the likelihood and maximising with respect to the parameters, which is not always straightforward.

7.2.3 The second structure defines a generalised linear model, in which

Table 1. Historical loss development study (1991) Automatic Facultative General Liability data (excluding asbestos and environmental)

Claim payments										Reserves
5012	3257	2638	898	1734	2642	1828	599	54	172	0
106	4179	1111	5270	3116	1817	−103	673	535		154
3410	5582	4881	2268	2594	3479	649	603			617
5655	5900	4211	5500	2159	2658	984				1,636
1092	8473	6271	6333	3786	225					2,747
1513	4932	5257	1233	2917						3,649
557	3463	6926	1368							5,435
1351	5596	6165								10,907
3133	2262									10,650
2063										16,339
									Overall	52,135
Development factors										
	2.999	1.624	1.271	1.172	1.113	1.042	1.033	1.017	1.009	

the responses C_{ij} are modelled as Poisson random variables with a logarithmic link function and linear predictor $\eta_{ij} = c + \alpha_i + \beta_j$. Over-dispersion is taken into account by estimating the unknown scale parameter ϕ as part of the fitting procedure. Use of the log link function makes the model linear in the parameters. Standard statistical software which fits GLMs can be used to obtain the parameter estimates by maximum likelihood. However, use of the log link function imposes a limitation, in that interpretation of parameter estimates is not easy. It is possible to relate these parameters back to the chain-ladder link ratios, or the parameters of the non-linear model (y_j), as shown by Verrall (1991b). Nevertheless, the fact that this involves further calculations can be seen as a disadvantage of using the over-dispersed Poisson model. Also, fitted values will always be positive, and equivalent link ratios will always, therefore, be greater than 1.

7.2.4 To illustrate the methodology, consider the claims amounts in Table 1, shown in incremental form. The data relate to Automatic Facultative General Liability (excluding Asbestos and Environmental) from the Historical Loss Development Study (1991), and were used by Mack (1994b). The data show considerable variability between origin years. Notice the negative incremental value in cell (2, 7). Also shown are the standard chain-ladder development factors and reserve estimates, ignoring any tail factor.

7.2.5 Table 2 shows the parameter estimates and their standard errors obtained by fitting the over-dispersed Poisson model. For many of the parameters, the standard errors are large relative to the estimates themselves. This does not provide evidence that those estimates can be set to zero, since doing so may ignore a trend, which itself may be statistically significant. For example, the β parameters clearly show a downward trend, even though the

Table 2. Over-dispersed Poisson model; parameter estimates

	Parameter estimate	Standard error
Constant	7.6551	0.3193
Alpha 2	−0.1108	0.3450
Alpha 3	0.2459	0.3185
Alpha 4	0.4213	0.3100
Alpha 5	0.4291	0.3130
Alpha 6	0.0348	0.3538
Alpha 7	−0.0593	0.3819
Alpha 8	0.2432	0.3786
Alpha 9	−0.1603	0.5143
Alpha 10	−0.0232	0.7816
Beta 2	0.6928	0.2685
Beta 3	0.6260	0.2784
Beta 4	0.2769	0.3115
Beta 5	0.0606	0.3417
Beta 6	−0.1958	0.3885
Beta 7	−1.0831	0.6079
Beta 8	−1.2737	0.7893
Beta 9	−1.9159	1.3617
Beta 10	−2.5076	2.4911
Dispersion	1049.8	

individual estimates are less than twice their standard error for β_4 to β_{10}. Ideally, the strength of that trend should be tested, and modelled directly, but, in this example, we ignore that feature, since the purpose is to fit a model which reproduces chain-ladder estimates.

7.2.6 Estimates of future payments can be obtained from the parameter estimates by inserting them into equation 7.3 and exponentiating, giving:

$$\hat{C}_{ij} = \hat{m}_{ij} = \exp(\hat{\eta}_{ij}). \tag{7.4}$$

Origin year and overall reserve estimates can then be found by summation. We also require prediction errors, and we begin by considering a single incremental payment, C_{ij}.

7.2.7 Consider origin year i and claim payments in development year j (yet to be observed). The mean squared error of prediction is given by:

$$\mathrm{MSEP}\big[\hat{C}_{ij}\big] = \mathrm{E}\big[(C_{ij} - \hat{C}_{ij})^2\big] \approx \mathrm{Var}\big[C_{ij}\big] + \mathrm{Var}\big[\hat{C}_{ij}\big].$$

From equation 7.1, we know that:

$$\mathrm{Var}\big[C_{ij}\big] = \phi m_{ij} \tag{7.5}$$

then, using equation 7.4 and a Taylor series expansion:

$$\mathrm{Var}\left[\hat{C}_{ij}\right] \approx \left|\frac{\partial m_{ij}}{\partial \eta_{ij}}\right|^2 \mathrm{Var}\left[\hat{\eta}_{ij}\right]. \tag{7.6}$$

Combining equations 7.5 and 7.6 gives:

$$\mathrm{MSEP}\left[\hat{C}_{ij}\right] \approx \phi \hat{m}_{ij} + \hat{m}_{ij}^2 \mathrm{Var}\left[\hat{\eta}_{ij}\right]. \tag{7.7}$$

7.2.8 The final component of equation 7.7, the variance of the linear predictor, is usually available directly from statistical software packages, enabling the prediction error to be calculated without difficulty.

7.2.9 The prediction errors for origin year reserve estimates and the total reserve estimates can also be calculated, but require more effort. In those cases, the variance of the sum of predicted values is considered, taking account of any covariances between predicted values. Making certain assumptions (outlined in Renshaw, 1994b), the only covariances which need to be considered arise in the estimation variance. This makes sense intuitively, since predicted values in each row are based on the same parameters, and predicted values in the same column are based on the same parameters, thus introducing dependence.

7.2.10 Denoting the triangle of predicted claims contributing to the reserve estimates by Δ, then the reserve estimate in origin year i is given by summing the predicted values in row i of Δ, that is:

$$\hat{C}_{i+} = \sum_{j \in \Delta_i} \hat{C}_{ij}.$$

From Renshaw (1994b), the squared prediction error of the origin year reserve is given by:

$$\mathrm{MSEP}\left[\hat{C}_{i+}\right] \approx \sum_{j \in \Delta_i} \phi \hat{m}_{ij} + \sum_{j \in \Delta_i} \hat{m}_{ij}^2 \mathrm{Var}\left[\hat{\eta}_{ij}\right] + 2 \sum_{\substack{j_1,\, j_2 \in \Delta_i \\ j_2 > j_1}} \hat{m}_{ij_1} \hat{m}_{ij_2} \mathrm{Cov}\left[\hat{\eta}_{ij_1}, \hat{\eta}_{ij_2}\right]. \tag{7.8}$$

7.2.11 The total reserve estimate is given by:

$$\hat{C}_{++} = \sum_{i,\, j \in \Delta} \hat{C}_{ij}$$

and the squared prediction error of the total reserve is given by:

Table 3. Over-dispersed Poisson model; reserve results

	Reserve	Prediction error	Prediction error %
Year 1	–	–	–
Year 2	154	556	361%
Year 3	617	1,120	181%
Year 4	1,636	1,775	109%
Year 5	2,747	2,231	81%
Year 6	3,649	2,440	67%
Year 7	5,435	3,124	57%
Year 8	10,907	5,032	46%
Year 9	10,650	6,075	57%
Year 10	16,339	12,987	79%
Overall	52,135	18,193	35%

$$\mathrm{MSEP}\left[\hat{C}_{++}\right] \approx \sum_{i,j\in\Delta} \phi\hat{m}_{ij} + \sum_{i,j\in\Delta} \hat{m}_{ij}^2 \mathrm{Var}\left[\hat{\eta}_{ij}\right] + 2 \sum_{\substack{i_1 j_1\in\Delta \\ i_2 j_2\in\Delta \\ i_1 j_1 \neq i_2 j_2}} \hat{m}_{i_1 j_1}\hat{m}_{i_2 j_2}\mathrm{Cov}\left[\hat{\eta}_{i_1 j_1}, \hat{\eta}_{i_2 j_2}\right]. \tag{7.9}$$

7.2.12 Equations 7.8 and 7.9 require considerable care when summing the appropriate elements. The covariance terms are not readily available from statistical software packages. However, provided the *variance-covariance matrix* of the parameter estimates can be extracted from the statistical software package used, a full matrix of the covariance terms can be calculated by forming the *design matrix* of future payments and performing a few simple matrix operations. Indeed, the variances of the linear predictors are simply the diagonal of such a matrix.

7.2.13 Note that the first term in the origin year and overall prediction errors is simply the appropriate sum of the process variances. The remaining terms relate to the estimation variance.

7.2.14 The prediction errors can be calculated using equations 7.8 and 7.9, with the results shown in Table 3, which shows that there is considerable uncertainty in the reserve estimates, particularly in the early development years where the outstanding reserves are small. The prediction errors, as a percentage of the mean, are large in recent years, and, in particular, for the most recent year, due to the estimation error, since more is being predicted and a greater number of parameters are used in the prediction. Notice that the reserve estimates are identical to the chain-ladder results.

7.3 *The Negative Binomial Chain-Ladder Model*

7.3.1 The negative binomial model can be fitted using incremental or cumulative data, and gives the same fitted values irrespective of which

method is used. Unlike the over-dispersed Poisson model, which has origin year parameters and development year parameters, the parameters in the negative binomial model relate to development years only. Considering the model for incremental data:

$$\mathrm{E}\big[C_{ij}\big] = \big(\lambda_j - 1\big)D_{i,j-1} \quad \text{and} \quad \mathrm{Var}\big[C_{ij}\big] = \phi\lambda_j\big(\lambda_j - 1\big)D_{i,j-1} \tag{7.10}$$

where the D_{ij} (where observed) are considered known.
Then, writing:

$$\mathrm{E}\big[C_{ij}\big] = m_{ij} = \big(\lambda_j - 1\big)D_{i,j-1}$$

and taking logs, gives:

$$\log(m_{ij}) = \log\big(\lambda_j - 1\big) + \log\big(D_{i,j-1}\big).$$

Writing:

$$\log\big(\lambda_j - 1\big) = c + \alpha_{j-1} \qquad \text{with:} \quad \alpha_1 = 0, j \geq 2 \tag{7.11}$$

gives:

$$\log(m_{ij}) = c + \alpha_{j-1} + \log\big(D_{i,j-1}\big). \tag{7.12}$$

7.3.2 Equations 7.10, 7.11 and 7.12 specify a generalised linear model with logarithmic link function and negative binomial error structure. The $\log(D_{i,j-1})$ terms are derived from the known values $D_{i,j-1}$, and are specified as offsets in the model.

7.3.3 Notice that the parameters in the model relate to the standard chain-ladder development factors. Like the over-dispersed Poisson model, standard statistical software can be used to obtain maximum likelihood estimates of the parameters. However, not all statistical software packages which fit GLMs include the negative binomial distribution. The authors used S-PLUS (2001) in the example, having recourse to the MASS library, from Venables & Ripley (1999) available on the internet (http://www.stats.ox.ac.uk/Pub/MASS3).

7.3.4 Continuing the example in Section 7.2, parameter estimates and their standard errors are shown in Table 4, using the same data triangle.

7.3.5 Estimates of the development factors can be obtained from the parameter estimates using equation 7.11, and their approximate standard errors can be obtained using:

$$\mathrm{Var}\big[\hat{\lambda}_j\big] = \mathrm{Var}\big[\hat{\lambda}_j - 1\big] \approx \exp\big(\hat{c} + \hat{\alpha}_{j-1}\big)^2\mathrm{Var}\big[\hat{c} + \hat{\alpha}_{j-1}\big] \qquad j \geq 2.$$

Table 4. Negative binomial model; parameter estimates

	Parameter estimate	Standard error
Constant	0.6928	0.2733
Alpha 2	−1.1652	0.3490
Alpha 3	−1.9989	0.3675
Alpha 4	−2.4550	0.3909
Alpha 5	−2.8698	0.4318
Alpha 6	−3.8645	0.6416
Alpha 7	−4.0961	0.8207
Alpha 8	−4.7711	1.3945
Alpha 9	−5.3796	2.5399
Dispersion	1086.76	

The covariance matrix of the parameter estimates is required in the evaluation of the final component, the variance of the predictor.

7.3.6 The development factors and their approximate standard errors are given in Table 5. Notice that the development factors given by the model are exactly the same as the chain-ladder development factors.

7.3.7 For the negative binomial model, and its Normal approximation considered in Section 2.5, the parameters relate to the development factors λ_j, and a similar approach can be adopted when estimating the reserve prediction errors for both models. Again we proceed by considering estimators for the process variance and estimation variance.

7.3.8 For the origin year reserve estimates, the ultimate claims U_i are the cumulative claims in the last development year. That is:

$$U_i = D_{in}.$$

The reserve estimate in origin year i, R_i, is $U_i - D_{i,n-i+1}$, where D_{i-n+1} is the paid to date, which is considered known. Therefore:

Table 5. Negative binomial model; development factors and standard errors

	Estimate	Standard error
Lambda 2	2.999	0.546
Lambda 3	1.624	0.135
Lambda 4	1.271	0.067
Lambda 5	1.172	0.048
Lambda 6	1.113	0.038
Lambda 7	1.042	0.024
Lambda 8	1.033	0.026
Lambda 9	1.017	0.023
Lambda 10	1.009	0.023

$$\mathrm{Var}[R_i] = \mathrm{Var}[U_i] = \mathrm{Var}[D_{in}] \quad \text{and} \quad \mathrm{Var}\big[\hat{R}_i\big] = \mathrm{Var}\big[\hat{U}_i\big] = \mathrm{Var}\big[\hat{D}_{in}\big].$$

The origin year process and estimation variances can be estimated by considering $\mathrm{Var}[D_{in}]$ and $\mathrm{Var}[\hat{D}_{in}]$ respectively.

7.3.9 First consider the process variance $\mathrm{Var}[D_{in}]$. Since the negative binomial model and its Normal approximation, are formulated as recursive models, the calculation of the process variance involves estimating the variance of a k-steps ahead forecast, where $k = i - 1$. It is shown in Appendix 1 that, for the negative binomial model:

$$\mathrm{Var}[D_{in}] \approx \phi D_{i,n-i+1} \prod_{k=n-i+2}^{n} \hat{\lambda}_k \left(\prod_{k=n-i+2}^{n} \hat{\lambda}_k - 1 \right). \tag{7.13}$$

7.3.10 The estimation variance is calculated from:

$$\mathrm{Var}\big[\hat{D}_{in}\big] \approx \mathrm{Var}\left[D_{i,n-i+1} \prod_{k=n-i+2}^{n} \hat{\lambda}_k \right] = D_{i,n-i+1}^2 \mathrm{Var}\left[\prod_{k=n-i+2}^{n} \hat{\lambda}_k \right]. \tag{7.14}$$

Since the parameters in the negative binomial model and its Normal approximation relate to the development factors λ_j, and the covariance matrix of the parameter estimates is readily available, the variance of the product of development factors in equation 7.14 can be calculated. This is straightforward when performed recursively, as described in Appendix 2.

7.3.11 The overall reserve estimation and process variances can be estimated by considering $\mathrm{Var}[R_+]$ and $\mathrm{Var}[\hat{R}_+]$ respectively, where $R_+ = \sum_{i=2}^{n} R_i$. The overall reserve process variance is the sum of the process variances of individual origin year reserves, assuming independence between years. The overall reserve estimation variance is given by:

$$\mathrm{Var}\big[\hat{R}_+\big] \approx \sum_{i=2}^{n} \mathrm{Var}\big[\hat{D}_{in}\big] + 2 \sum_{\substack{i=2 \\ j>i}}^{n} \mathrm{Cov}\big[\hat{D}_{in}, \hat{D}_{jn}\big]. \tag{7.15}$$

That is, the estimation variance of overall reserves is the sum of the estimation variances of individual origin year reserves, with an additional component to take account of the covariance between years induced by dependence on the same parameters. Appendix 2 shows how the covariance terms can be derived.

7.3.12 The reserve estimates can be calculated from the development factors, and the prediction error of the reserves can be calculated using equations 7.13, 7.14 and 7.15. The results are shown in Table 6.

7.3.13 A comparison of the results of the over-dispersed Poisson and

Table 6. Negative binomial model; reserve results

	Reserve	Prediction error	Prediction error %
Year 1	–	–	–
Year 2	154	566	367%
Year 3	617	1,139	185%
Year 4	1,636	1,807	110%
Year 5	2,747	2,271	83%
Year 6	3,649	2,483	68%
Year 7	5,435	3,180	59%
Year 8	10,907	5,122	47%
Year 9	10,650	6,185	58%
Year 10	16,339	13,227	81%
Overall	52,135	18,528	36%

negative binomial models shows that the reserve estimates are identical, and that the prediction errors are very close. The prediction errors of the negative binomial model are 1.75% higher than those from the over-dispersed Poisson model, that figure being the square root of the ratio of the respective dispersion parameters.

7.3.14 Table 7 shows a comparison of the process and estimation variances for the two models, scaled by dividing by the dispersion parameters (to avoid the distortions mentioned above). The sum of the process variance and estimation variance is essentially the same for the two models (the differences being due to numerical accuracy used in the calculations), although the estimation variance is larger for the over-dispersed Poisson model, reflecting the higher number of parameters in that model. Essentially, the two models only differ in the way the models are parameterised. This

Table 7. Scaled prediction variance comparison

	Over-dispersed Poisson			Negative binomial		
	Estimation variance	Process variance	Prediction variance	Estimation variance	Process variance	Prediction variance
Year 2	140	154	294	139	155	294
Year 3	577	617	1,195	561	634	1,195
Year 4	1,366	1,636	3,002	1,268	1,735	3,003
Year 5	1,995	2,747	4,742	1,709	3,035	4,744
Year 6	2,021	3,649	5,671	1,184	4,489	5,673
Year 7	3,863	5,435	9,298	1,470	7,834	9,304
Year 8	13,209	10,907	24,117	4,160	19,980	24,140
Year 9	24,508	10,650	35,158	3,524	31,674	35,197
Year 10	144,330	16,339	160,669	15,234	145,752	160,985
Overall	263,155	52,135	315,290	100,598	215,288	315,886

results in a trade-off between the estimation variance and process variance. The practical implication of this is that it does not matter which model is fitted, the results will be the same. The negative binomial model is interesting, however, in that, through its Normal approximation, it provides insight into Mack's model.

7.4 *The Normal Approximation to the Negative Binomial Chain-Ladder Model*

7.4.1 The Normal approximation to the negative binomial model can also be fitted using incremental or cumulative data. To allow a comparison with Mack's model later, we consider using cumulative data, giving:

$$\mathrm{E}\left[D_{ij}\right] = \lambda_j D_{i,j-1} \quad \text{and} \quad \mathrm{Var}\left[D_{ij}\right] = \phi_j D_{i,j-1} \qquad j \geq 2$$

where the $D_{i,j-1}$ are considered known. Writing $w_{ij} = D_{i,j-1}$ and dividing by w_{ij} gives:

$$\mathrm{E}\left[\frac{D_{ij}}{w_{ij}}\right] = \lambda_j \quad \text{and} \quad \mathrm{Var}\left[\frac{D_{ij}}{w_{ij}}\right] = \frac{\phi_j}{w_{ij}}.$$

7.4.2 We fit the model by focusing on the quantities $f_{ij} = \dfrac{D_{ij}}{w_{ij}}$, which, under the model, are Normally distributed with variance components ϕ_j and weights w_{ij}. Notice that the f_{ij} form a triangle of individual development factors. Notice also that the variance components ϕ_j depend on development year j. This is a complication since, usually, in a generalised linear model the dispersion parameter ϕ is constant for all observations. The variance components are modelled as part of the extended fitting procedure known as joint modelling, for which a thorough description is given in Renshaw (1994a).

7.4.3 The joint modelling process proceeds by providing initial arbitrary positive values for ϕ_j, and fitting a weighted generalised linear model to the f_{ij}, with weights $W_{ij} = \dfrac{w_{ij}}{\phi_j}$ (that is, weights inversely proportional to the variance). It is usual to include a constant in the linear predictor, giving:

$$\mathrm{E}\left[f_{ij}\right] = c + \alpha_{j-1} \qquad \text{with:} \quad \alpha_1 = 0, j \geq 2.$$

Notice that the linear predictor depends on development period j only, and the fitted values $\hat{f}_{ij}$ are estimates of the development factors λ_j. Having obtained the fitted values, the square of the residuals r_{ij} are calculated, where:

Table 8. Normal approximation to negative binomial model; parameter estimates and standard errors

	Parameter estimate	Standard error
Constant	2.9994	1.1781
Alpha 2	−1.3758	1.1865
Alpha 3	−1.7285	1.1818
Alpha 4	−1.8277	1.1784
Alpha 5	−1.8860	1.1786
Alpha 6	−1.9574	1.1783
Alpha 7	−1.9661	1.1781
Alpha 8	−1.9824	1.1782
Alpha 9	−1.9901	1.1782
Dispersion	1.2222	

$$r_{ij}^2 = w_{ij}\left(f_{ij} - \hat{f}_{ij}\right)^2.$$

7.4.4 These residuals squared are used as the response in a second generalised linear model, in which the predictor depends on development year j only. The fitted values $\hat{r}_{ij}^2$ from this second model are used to update the values ϕ_j, and the first model is refitted after updating the weights W_{ij} to reflect the revised estimates of ϕ_j. This completes the joint modelling process, from which estimates of λ_j and ϕ_j can be obtained. Note: since the model for the squared residuals depends only on development year j, the fitted values $\hat{r}_{ij}^2$ will simply be estimated as the average over i of the squared residuals in development year j. The form of the residuals is dictated by the error distribution used in the model for the mean, which, in this case, is a Normal distribution.

7.4.5 Fitting the joint model to the data triangle in Table 1 gives the parameters estimates and standard errors shown in Table 8. Notice that the parameter estimates are now on the scale of the development factors themselves since an identity link function is used. The development factors and their associated standard errors, obtained from these estimates and the associated covariance matrix, are shown in Table 9. The estimates of the variance components ϕ_j are shown in Table 10. Notice that there is insufficient information to estimate ϕ_{10}, since there is only one residual at that point. The estimate of ϕ_{10} has been set to the estimate of the previous value ϕ_9 in this example.

7.4.6 The theory underlying the calculation of prediction errors for this model is the same as for the negative binomial model in Section 7.3. The reserve estimates and their prediction errors, calculated using equations 7.14, 7.15 and the methods shown in Appendices 1 and 2, are shown in Table 11.

7.4.7 Compared to the results of the negative binomial model in Table 6,

Table 9. Normal approximation to negative binomial model; development factors and standard errors

	Estimate	Standard error
Lambda 2	2.999	1.178
Lambda 3	1.624	0.140
Lambda 4	1.271	0.093
Lambda 5	1.172	0.026
Lambda 6	1.113	0.035
Lambda 7	1.042	0.022
Lambda 8	1.033	0.004
Lambda 9	1.017	0.012
Lambda 10	1.009	0.016

Table 10. Normal approximation to negative binomial model; variance components

Phi 2	24785.3
Phi 3	970.0
Phi 4	592.7
Phi 5	51.0
Phi 6	95.6
Phi 7	30.6
Phi 8	0.9
Phi 9	3.9
Phi 10	3.9

Table 11. Normal approximation to negative binomial model; reserve results

	Reserve	Prediction error	Prediction error %
Year 1	–	–	–
Year 2	154	371	241%
Year 3	617	643	104%
Year 4	1,636	764	47%
Year 5	2,747	1,361	50%
Year 6	3,649	1,822	50%
Year 7	5,435	2,017	37%
Year 8	10,907	5,007	46%
Year 9	10,650	5,939	56%
Year 10	16,339	23,375	143%
Overall	52,135	25,706	49%

the reserve prediction errors in Table 11 are lower in the early years, and higher in the later years. Since the overall prediction error is driven mostly by the more recent years, the overall reserve prediction error is higher for the Normal approximation than for the negative binomial model.

7.5 *Mack's Model*

7.5.1 According to Mack's model, the mean and variance of D_{ij} are $\lambda_j D_{i,j-1}$ and $\sigma_j^2 D_{i,j-1}$, respectively. Mack provides estimators for λ_j and σ_j^2 using:

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} w_{ij} f_{ij}}{\sum_{i=1}^{n-j+1} w_{ij}} \qquad \text{where:} \quad w_{ij} = D_{i,j-1} \quad \text{and} \quad f_{ij} = \frac{D_{ij}}{D_{i,j-1}}$$

and

$$\hat{\sigma}_j^2 = \frac{1}{n-j} \sum_{i=1}^{n-j+1} w_{ij} \left(f_{ij} - \hat{\lambda}_j \right)^2.$$

7.5.2 The estimator for the development factors λ_j is the standard volume weighted chain-ladder estimator. Mack shows that this provides unbiased estimates of the development factors. Mack also shows that an unweighted average of individual development factors also provides unbiased estimates, but shows that the weighted average is preferable, since it has a lower variance (it is the minimum variance unbiased estimator).

7.5.3 The variance component σ_j^2 is estimated as an average of weighted residuals, where the divisor is the number of residuals (used in calculating the estimator) minus one. The one is subtracted to provide an unbiased estimator of σ_j^2. These variance components are not used when estimating the development factors, but are required when considering the prediction errors of future payments. Like the Normal approximation to the negative binomial model, there is insufficient information to estimate the final variance component σ_{10}^2. In this example, we first reproduce the results of Mack (1994b), in which the final component $\hat{\sigma}_{10}^2$ is set to $\hat{\sigma}_8^2$, then we fit the model setting $\hat{\sigma}_{10}^2$ to the previous value $\hat{\sigma}_9^2$, which shows that the results in the earlier years are very sensitive to this single parameter.

7.5.4 In Mack (1993), formulae are developed for prediction errors of the origin year and overall reserves in terms of the cumulative claims D_{ij}, the development factors λ_j, and the variance components σ_j^2 only. From Mack (1993), the process variance of the origin year reserves is given by:

$$\text{Var}[R_i] \approx \hat{D}_{in}^2 \sum_{k=n-i+1}^{n-1} \frac{\hat{\sigma}_{k+1}^2}{\hat{\lambda}_{k+1}^2 \hat{D}_{ik}}$$

and the estimation variance is given by:

$$\mathrm{Var}\left[\hat{R}_i\right] \approx \hat{D}_{in}^2 \sum_{k=n-i+1}^{n-1} \frac{\hat{\sigma}_{k+1}^2}{\hat{\lambda}_{k+1}^2 \sum_{q=1}^{n-k} D_{qk}}$$

giving:

$$\mathrm{MSEP}\left[\hat{R}_i\right] \approx \hat{D}_{in}^2 \sum_{k=n-i+1}^{n-1} \frac{\hat{\sigma}_{k+1}^2}{\hat{\lambda}_{k+1}^2} \left(\frac{1}{\hat{D}_{ik}} + \frac{1}{\sum_{q=1}^{n-k} D_{qk}} \right).$$

7.5.5 For the overall reserve prediction error, a covariance adjustment is needed for the estimation variance, giving:

$$\mathrm{MSEP}\left[\hat{R}_+\right] = \sum_{i=2}^{n} \left\{ \mathrm{MSEP}\left[\hat{R}_i\right] + \hat{D}_{in} \left(\sum_{q=i+1}^{n} \hat{D}_{qn} \right) \times \sum_{k=n-i+1}^{n-1} \frac{2\hat{\sigma}_{k+1}^2}{\hat{\lambda}_{k+1}^2 \sum_{q=1}^{n-k} D_{qk}} \right\}$$

where:

$$\hat{R}_+ = \sum_{i=2}^{n} \hat{R}_i.$$

7.5.6 Although the formulae for the prediction errors seem daunting, the calculations can be set up fairly quickly in a spreadsheet.

7.5.7 Table 12 shows the development factors $\hat{\lambda}_j$ and the variance components $\hat{\sigma}_j^2$ obtained by fitting Mack's model. Table 13 shows the reserve

Table 12. Mack's model; development factors and variance components

	Lambda	Sigma squared
$j = 2$	2.999	27883.5
$j = 3$	1.624	1108.5
$j = 4$	1.271	691.4
$j = 5$	1.172	61.2
$j = 6$	1.113	119.4
$j = 7$	1.042	40.8
$j = 8$	1.033	1.3
$j = 9$	1.017	7.9
$j = 10$	1.009	

Table 13. Mack's model; reserve results using $\hat{\sigma}_{10}^2 = \hat{\sigma}_8^2$

	Reserve	Prediction error	Prediction error %
Year 1	–	–	–
Year 2	154	206	134%
Year 3	617	623	101%
Year 4	1,636	747	46%
Year 5	2,747	1,469	53%
Year 6	3,649	2,002	55%
Year 7	5,435	2,209	41%
Year 8	10,907	5,358	49%
Year 9	10,650	6,333	59%
Year 10	16,339	24,566	150%
Overall	52,135	26,909	52%

Table 14. Mack's model; reserve results using $\hat{\sigma}_{10}^2 = \hat{\sigma}_9^2$

	Reserve	Prediction error	Prediction error %
Year 1	–	–	–
Year 2	154	500	324%
Year 3	617	863	140%
Year 4	1,636	1,014	62%
Year 5	2,747	1,623	59%
Year 6	3,649	2,065	57%
Year 7	5,435	2,259	42%
Year 8	10,907	5,391	49%
Year 9	10,650	6,348	60%
Year 10	16,339	24,571	150%
Overall	52,135	27,172	52%

estimates and their prediction errors, first using the value of $\hat{\sigma}_{10}^2$ given in Mack (1994b), and Table 14 shows the effect of using $\hat{\sigma}_{10}^2 = \hat{\sigma}_9^2$.

7.6 *A Comparison of Mack's Model and the Normal Approximation to the Negative Binomial*

7.6.1 A quick glance at the results from Mack's model and the Normal approximation to the negative binomial reveals striking similarities. Estimates of the development factors are identical, as expected. The estimates of ϕ_j, used in the Normal approximation to the negative binomial, and σ_j^2, used in Mack's model, are close, and the prediction errors are also close.

7.6.2 In fact, the similarities run deeper, but are obscured by differences in the way bias corrections are incorporated. In Mack's model, unbiased estimates of σ_j^2 are calculated using an average of weighted residuals, where

Table 15. Mack's model and the normal approximation to the negative binomial; a comparison of variance components, bias correction removed

	Phi	Sigma squared
$j = 2$	24785.3	24785.3
$j = 3$	970.0	970.0
$j = 4$	592.7	592.7
$j = 5$	51.0	51.0
$j = 6$	95.6	95.6
$j = 7$	30.6	30.6
$j = 8$	0.9	0.9
$j = 9$	3.9	3.9
$j = 10$	3.9	3.9

the divisor is the number of data points used minus one. In the Normal approximation to the negative binomial, the estimates of ϕ_j are simply an average of weighted residuals, ignoring the bias correction.

7.6.3 Furthermore, in Mack's model there is no adjustment for the number of parameters used in fitting the model, although, in the Normal approximation to the negative binomial, a bias correction for the number of parameter estimated is incorporated naturally into the dispersion parameter (in Table 8).

7.6.4 It is not immediately clear how bias corrections should be incorporated, and discussion of the issues is beyond the scope of this paper. However, the similarities of the models should be noted, and, to show this, the respective bias corrections have been removed from both models, with the results for the variance components σ_j^2 and ϕ_j shown in Table 15, and the components of the prediction errors shown in Tables 16 and 17.

Table 16. Normal approximation to the negative binomial; reserve prediction errors, bias correction removed

	Normal approximation, bias removed				
	Process variance	Estimation variance	Prediction variance	Prediction error	Prediction error %
Year 2	65,841	58,933	124,773	353	229%
Year 3	188,267	183,859	372,125	610	99%
Year 4	249,915	273,427	523,342	723	44%
Year 5	1,153,186	572,498	1,725,684	1,314	48%
Year 6	2,626,609	567,599	3,194,209	1,787	49%
Year 7	3,341,553	593,708	3,935,261	1,984	36%
Year 8	20,667,184	3,600,498	24,267,682	4,926	45%
Year 9	31,366,047	3,195,596	34,561,643	5,879	55%
Year 10	488,235,825	47,596,937	535,832,762	23,148	142%
Overall	547,894,426	92,394,120	640,288,546	25,304	49%

Table 17. Mack's model; reserve prediction errors, bias correction removed

	Mack, bias removed				
	Process variance	Estimation variance	Prediction variance	Prediction error	Prediction error %
Year 2	65,841	58,933	124,773	353	229%
Year 3	188,267	183,843	372,109	610	99%
Year 4	249,915	273,400	523,314	723	44%
Year 5	1,153,186	572,350	1,725,536	1,314	48%
Year 6	2,626,609	567,276	3,193,885	1,787	49%
Year 7	3,341,553	593,215	3,934,768	1,984	36%
Year 8	20,667,120	3,593,809	24,260,929	4,926	45%
Year 9	31,365,773	3,180,563	34,546,336	5,878	55%
Year 10	488,235,511	46,925,843	535,161,354	23,134	142%
Overall	547,893,774	91,639,423	639,533,197	25,289	49%

7.6.5 Comparison of Tables 15, 16 and 17 reveals that the models are essentially the same, the remaining differences being due to the accuracy of the numerical computations. To establish a proof that the models are identical, it is only necessary to show algebraically that the estimators for the parameters λ, ϕ, and σ^2 are the same. This can be shown, since the parameters depend on development year only, and will, therefore, be a weighted average of the respective quantities used in their estimation; the weights are the same for both models. It is not necessary to show that the formulae for prediction errors are algebraically the same, although they will be if the same approximations are used (and the formulae have been derived correctly), since both models adopt a recursive approach. For uniqueness, it is necessary to show that the Normal distribution is the only one which will result in the same estimates of ϕ and σ^2; this can be shown, since other error distributions will result in different estimators for the residuals in ¶7.4.3. Since we do not provide a proof, the equivalence of Mack's model to the normal approximation to the negative binomial model remains a conjecture.

7.6.6 It is clear that Mack's model was not conceived as a Normal approximation to the negative binomial; that is our interpretation. Mack made no specific distributional assumptions, but simply provided the first two moments. However, it is hoped that some insight has been gained by starting with the over-dispersed Poisson model, moving to the negative binomial and its normal approximation, then comparing with Mack's model.

7.6.7 Although this comparison might seem irrelevant, it becomes important when extending the chain-ladder model to incorporate other predictor structures which allow extrapolation (and therefore the evaluation of tail factors), and also when considering fitting models to smooth the variance factors and impute a value for the missing estimate in the latest development year. Such extensions can occur naturally in the joint modelling framework of Section 7.4.

7.7 *Log-Normal Model*

7.7.1 With the log-Normal models, the first step is to transform the incremental claims by taking their (natural) logarithm. A model is then fitted to the transformed values using ordinary least squares regression analysis. Adopting the chain-ladder type predictor structure, then:

$$Y_{ij} = \log(C_{ij}) = \eta_{ij} + \varepsilon_{ij}$$

where:

$$\eta_{ij} = c + \alpha_i + \beta_j \quad \text{and} \quad \varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2). \tag{7.16}$$

7.7.2 Having obtained estimates for the parameters in the linear predictor, and the process variance σ^2, the fitted values on a log scale are obtained by forming the appropriate sum of estimates. However, to obtain estimates for the mean on the untransformed scale, we cannot simply exponentiate the linear predictor, since that would provide an estimate of the median. The fitted values on the untransformed scale are given by:

$$\hat{C}_{ij} = \exp\left(\hat{\eta}_{ij} + \tfrac{1}{2}\hat{\sigma}^2_{ij}\right) \tag{7.17}$$

where:

$$\hat{\sigma}^2_{ij} = \mathrm{Var}\left[\hat{\eta}_{ij}\right] + \hat{\sigma}^2.$$

7.7.3 The σ^2_{ij} terms are, in fact, the prediction variance of the linear predictor, and are calculated as the sum of the variance of the linear predictor and the underlying process variance. The variance of the linear predictor is usually available from standard statistical software.

7.7.4 It is important to note that the reason why a variance component in equation 7.17 is needed for the log-Normal model, but not for the over-dispersed Poisson, is because the incremental claims, themselves, are used as the response with the over-dispersed Poisson model, but, with the log-Normal model, the logarithm of the incremental claims is used as the response.

7.7.5 Equation 7.17 is in the standard form of the expected value of a log-Normal distribution. Note that the variance component includes the estimation and process error. The prediction variance of future incremental claims C_{ij} is given by:

$$\mathrm{MSEP}\left[\hat{C}_{ij}\right] = \hat{C}^2_{ij}\left(\exp\left(\hat{\sigma}^2_{ij}\right) - 1\right)$$

which is in the standard form of the variance of a log-Normal distribution.

7.7.6 Like the over-dispersed Poisson model, the prediction error of

origin year and overall reserve estimates requires more effort. The variance of the sum of predicted values needs to be considered, taking account of any covariances between predicted values.

7.7.7 Denoting the triangle of predicted claims contributing to the reserve estimates by Δ, then the reserve estimate in origin year i is given by summing the predicted values in row i of Δ, that is:

$$\hat{C}_{i+} = \sum_{j \in \Delta_i} \hat{C}_{ij}.$$

The squared prediction error of the origin year reserve is given by:

$$\text{MSEP}\left[\hat{C}_{i+}\right] \approx \sum_{j \in \Delta_i} \text{MSEP}\left[\hat{C}_{ij}\right] + 2 \sum_{\substack{j_1,\, j_2 \in \Delta_i \\ j_2 > j_1}} \hat{C}_{ij_1} \hat{C}_{ij_2} \left(\exp\left(\text{Cov}\left[\hat{\eta}_{ij_1}, \hat{\eta}_{ij_2}\right]\right) - 1\right).$$

7.7.8 The total reserve estimate is given by:

$$\hat{C}_{++} = \sum_{i,\, j \in \Delta} \hat{C}_{ij}$$

and the squared prediction error of the total reserve is given by:

$$\text{MSEP}\left[\hat{C}_{++}\right] \approx \sum_{i,\, j \in \Delta} \text{MSEP}\left[\hat{C}_{ij}\right] + 2 \sum_{\substack{i_1 j_1 \in \Delta \\ i_2 j_2 \in \Delta \\ i_1 j_1 \neq i_2 j_2}} \hat{C}_{i_1 j_1} \hat{C}_{i_1 j_2} \left(\exp\left(\text{Cov}\left[\hat{\eta}_{i_1 j_1}, \hat{\eta}_{i_1 j_2}\right]\right) - 1\right).$$

7.7.9 The covariance terms are not readily available from statistical software packages. Like the over-dispersed Poisson model, provided the *design matrix* and *variance-covariance matrix* of the parameter estimates can be extracted from the statistical software package used, a full matrix of the covariance terms can be calculated. Christofides (1990) provides an example of how this is achieved.

7.7.10 The data in Table 1 immediately cause a problem, due to the negative value in cell (2, 7). Since it is not possible to take the logarithm of a negative number, it has simply been weighted out of the analysis for the purposes of the example. Fitting the log-Normal model with chain-ladder type predictor gives the parameter estimates and standard errors shown in Table 18. The associated reserve estimates and their prediction errors are given in Table 19.

7.7.11 It is clear from Table 19 that the estimate of the expected reserves given by the log-Normal model is far from the chain-ladder estimate, as is the overall prediction error, which, even as a percentage of the reserve estimate, is about twice the equivalent from the over-dispersed Poisson and

Table 18. Log-Normal model; chain-ladder type; parameter estimates and standard errors

	Parameter estimate	Standard error
Constant	7.1820	0.4200
Alpha 2	−0.1150	0.4263
Alpha 3	0.3180	0.4283
Alpha 4	0.5390	0.4489
Alpha 5	0.1510	0.4737
Alpha 6	0.0780	0.5045
Alpha 7	−0.2120	0.5464
Alpha 8	0.3000	0.6086
Alpha 9	0.1520	0.7155
Alpha 10	0.4490	0.9649
Beta 2	1.1040	0.4095
Beta 3	1.0120	0.4283
Beta 4	0.5380	0.4487
Beta 5	0.5320	0.4731
Beta 6	0.0070	0.5040
Beta 7	−0.5090	0.6077
Beta 8	−0.8140	0.6083
Beta 9	−1.9890	0.7159
Beta 10	−2.0350	0.9650
Dispersion	0.7545	

Table 19. Log-Normal model; chain-ladder type; mean reserve results

	Reserve	Prediction error	Prediction error %
Year 1	–	–	–
Year 2	357	751	210%
Year 3	1,020	1,413	139%
Year 4	3,064	3,291	107%
Year 5	3,753	3,540	94%
Year 6	6,010	5,227	87%
Year 7	7,742	6,678	86%
Year 8	18,806	16,379	87%
Year 9	25,367	24,908	98%
Year 10	56,475	77,519	137%
Overall	122,595	86,312	70%

negative binomial models. This is due to the inclusion of the variance component in equation 7.17, which can have a significant effect when the underlying variability is large. For comparison purposes, median reserve results are shown in Table 20, which are now considerably closer to the chain-ladder estimates, although the overall prediction error as a percentage of the overall reserve estimate, is still large. With the log-Normal model,

Table 20. Log-Normal model; chain-ladder type; median reserve results

	Reserve	Prediction error	Prediction error %
Year 1	–	–	–
Year 2	153	323	211%
Year 3	484	652	135%
Year 4	1,604	1,743	109%
Year 5	2,008	1,916	95%
Year 6	3,300	2,911	88%
Year 7	4,284	3,746	87%
Year 8	10,195	8,963	88%
Year 9	13,004	12,860	99%
Year 10	23,717	32,690	138%
Overall	58,750	38,072	65%

ignoring the variance component is invalid, when focusing on the mean. The interpretation is clarified when considering the predictive distribution of reserves (see Section 8).

7.8 *The Gamma Model*

7.8.1 Obtaining predictions and prediction errors for the Gamma model simply requires a subtle change to the over-dispersed Poisson model. The Gamma model can be written as:

$$\mathrm{E}\left[C_{ij}\right] = m_{ij} \qquad \mathrm{Var}\left[C_{ij}\right] = \phi m_{ij}^2$$

that is, the variance is proportional to the mean squared, rather than proportional to the mean in the case of the over-dispersed Poisson model.

7.8.2 Adopting the chain-ladder type predictor $\log(m_{ij}) = c + \alpha_i + \beta_j$, it is straightforward to obtain parameter estimates and predicted values using generalised linear models. It is also straightforward to estimate the prediction errors of future payments and reserves by making the appropriate change to the process error components of equations 7.7, 7.8, and 7.9. That is:

$$\mathrm{MSEP}\left[\hat{C}_{ij}\right] \approx \phi \hat{m}_{ij}^2 + \hat{m}_{ij}^2 \mathrm{Var}\left[\hat{\eta}_{ij}\right]$$

$$\mathrm{MSEP}\left[\hat{C}_{i+}\right] \approx \sum_{j \in \Delta_i} \phi \hat{m}_{ij}^2 + \sum_{j \in \Delta_i} \hat{m}_{ij}^2 \mathrm{Var}\left[\hat{\eta}_{ij}\right] + 2 \sum_{\substack{j_1,\, j_2 \in \Delta_i \\ j_2 > j_1}} \hat{m}_{ij_1} \hat{m}_{ij_2} \mathrm{Cov}\left[\hat{\eta}_{ij_1}, \hat{\eta}_{ij_2}\right]$$

$$\mathrm{MSEP}\left[\hat{C}_{++}\right] \approx \sum_{i,\, j \in \Delta} \phi \hat{m}_{ij}^2 + \sum_{i,\, j \in \Delta} \hat{m}_{ij}^2 \mathrm{Var}\left[\hat{\eta}_{ij}\right] + 2 \sum_{\substack{i_1 j_1 \in \Delta \\ i_2 j_2 \in \Delta \\ i_1 j_1 \neq i_2 j_2}} \hat{m}_{i_1 j_1} \hat{m}_{i_2 j_2} \mathrm{Cov}\left[\hat{\eta}_{i_1 j_1}, \hat{\eta}_{i_2 j_2}\right]$$

Table 21. Gamma model; chain-ladder type; parameter estimates and standard errors

	Parameter estimate	Standard error
Constant	7.7519	0.4509
Alpha 2	−0.2552	0.4401
Alpha 3	0.0743	0.4603
Alpha 4	0.3056	0.4823
Alpha 5	0.1263	0.5085
Alpha 6	−0.1955	0.5417
Alpha 7	−0.3828	0.5868
Alpha 8	−0.0245	0.6537
Alpha 9	−0.1036	0.7686
Alpha 10	−0.1200	1.0369
Beta 2	0.7650	0.4401
Beta 3	0.7674	0.4603
Beta 4	0.3446	0.4823
Beta 5	0.1766	0.5085
Beta 6	−0.1148	0.5417
Beta 7	−1.1270	0.5868
Beta 8	−1.2352	0.6537
Beta 9	−1.8285	0.7686
Beta 10	−2.6044	1.0369
Dispersion	0.8718	

7.8.3 As before, the variances of the linear predictors can be obtained from the software used to fit the model, and the calculation of the covariances requires a few simple matrix operations.

7.8.4 Parameter estimates and their standard errors obtained by fitting the Gamma model with a chain-ladder type predictor are shown in Table 21, and the reserve results are shown in Table 22. Notice that the expected

Table 22. Gamma model; chain-ladder type; reserve results

	Reserve	Prediction error	Prediction error %
Year 1	–	–	–
Year 2	133	185	139%
Year 3	588	574	98%
Year 4	1,659	1,371	83%
Year 5	2,242	1,650	74%
Year 6	3,330	2,517	76%
Year 7	4,654	3,457	74%
Year 8	9,863	7,480	76%
Year 9	13,630	11,477	84%
Year 10	17,842	19,178	107%
Overall	53,940	25,976	48%

reserves are close, but not identical, to the chain-ladder reserves. Also notice that the overall prediction error is larger than for the over-dispersed Poisson and negative binomial models. Essentially, relatively more weight is being given to smaller incremental values, and less to larger incremental values, giving greater weight to the tail, which is more variable.

7.8.5 Further discussion of the use of the Gamma error structure can be found in England & Verrall (1999, 2001).

7.9 *The Hoerl Curve and Wright's Model*

7.9.1 Implementation of the Hoerl curve is straightforward, and simply involves using a predictor structure with development time and log development time treated as continuous covariates. This has been implemented by the authors for the over-dispersed Poisson model and log-Normal models. The authors have not attempted to use parametric curves with the negative binomial model and its Normal approximation, although such extensions would be straightforward. For illustration purposes we consider the over-dispersed Poisson model only and the predictor structure shown in equation 4.1, using the same parameters β and γ for every origin year. We also normalise the data and adopt the technical adjustments to development time, as recommended in Appendix 1 of Wright (1990), although the effect of these is small. For illustration purposes we extrapolate for a further five development years.

7.9.2 For models in which the linear predictor involves continuous covariates (for example, the Hoerl curve), reserve prediction errors can be calculated in exactly the same way as their chain-ladder type counterparts, by calculating the appropriate variances and covariances of the linear predictors and including them in the relevant formula (for example, using equations 7.7, 7.8, and 7.9 for the over-dispersed Poisson model). Using standard software packages, and matrix manipulations, generic routines can be set up which perform the appropriate calculations, regardless of the form of the linear predictor.

7.9.3 Table 23 shows the parameter estimates and their standard errors from this model, and Table 24 shows the equivalent development factors, including extrapolation. Reserve results, without and with extrapolating into the tail, are shown in Tables 25 and 26 respectively.

7.9.4 Although the fit is poorer overall with the Hoerl curve, the dispersion parameter is reduced, because of the drop in number of parameters in the model. The equivalent development factors, calculated by simply fitting a standard chain-ladder model to the fitted values, show how model fitting and extrapolation can occur concurrently. The rigid form of the Hoerl curve means that the fit is a compromise. This can clearly be seen, since the earliest development factor is high relative to the chain-ladder model, and low later in the development.

7.9.5 In this example, the Hoerl curve has been fitted across the entire

Table 23. Over-dispersed Poisson model; Hoerl curve; parameter estimates and standard errors

	Parameter estimate	Standard error
Constant	8.8769	0.2920
Alpha 2	−0.1056	0.3197
Alpha 3	0.2569	0.2946
Alpha 4	0.4351	0.2865
Alpha 5	0.4576	0.2897
Alpha 6	0.0450	0.3280
Alpha 7	−0.0617	0.3542
Alpha 8	0.2508	0.3505
Alpha 9	−0.1654	0.4755
Alpha 10	0.0458	0.7179
Log(time)	0.5097	0.2864
Time	−0.4887	0.1330
Dispersion	904.5	

Table 24. Over-dispersed Poisson model; Hoerl curve; equivalent development factors

	Estimate
Lambda 2	3.230
Lambda 3	1.603
Lambda 4	1.284
Lambda 5	1.157
Lambda 6	1.093
Lambda 7	1.057
Lambda 8	1.036
Lambda 9	1.023
Lambda 10	1.015
Lambda 11	1.009
Lambda 12	1.006
Lambda 13	1.004
Lambda 14	1.002
Lambda 15	1.002

range of development time. In practice, it is likely that the fit would be improved by allowing the first two or three development years to have their own parameter, then adopting the form of the Hoerl curve from that point on.

7.9.6 There are some inherent dangers of using a fixed parametric curve like the Hoerl curve. For some combinations of parameter estimates, the curve can appear to be following a downward trend, but can start increasing when extrapolating. If used in combination with log-normal models, the overall trend can be downwards, but the bias correction can result in the fitted values showing an increasing trend when extrapolating.

Table 25. Over-dispersed Poisson model; Hoerl curve; reserve results without extrapolation

	Reserve	Prediction error	Prediction error %
Year 1	–	–	–
Year 2	243	486	200%
Year 3	885	984	111%
Year 4	2,033	1,589	78%
Year 5	3,582	2,216	62%
Year 6	3,849	2,301	60%
Year 7	5,393	2,873	53%
Year 8	11,091	4,686	42%
Year 9	10,568	5,563	53%
Year 10	17,654	12,801	73%
Overall	55,297	17,357	31%

Table 26. Over-dispersed Poisson model; Hoerl curve; reserve results with extrapolation

	Reserve	Prediction error	Prediction error %
Year 1	435	702	162%
Year 2	634	864	136%
Year 3	1,447	1,406	97%
Year 4	2,704	2,043	76%
Year 5	4,269	2,656	62%
Year 6	4,304	2,582	60%
Year 7	5,801	3,109	54%
Year 8	11,650	4,971	43%
Year 9	10,936	5,750	53%
Year 10	18,109	13,101	72%
Overall	60,291	19,241	32%

7.10 *Smoothing Models*

7.10.1 The smoothing models described in Section 5 are very appealing, since the rigid form of the Hoerl curve is not imposed, yet extrapolation is still possible. However, the availability of statistical software which implements smoothing models is currently limited. The authors used S-PLUS for the following example, which includes generalised additive models as standard. In S-PLUS, it is straightforward to adopt a smoothing approach by simply invoking the '*gam*' function, specifying the covariates to smooth over, the amount of smoothing and the underlying error distribution. In this section we consider using an over-dispersed Poisson error distribution, with a parameter for each origin year, smoothing over development time and log development time, with the same smoothing parameter for each component. That is:

$$\mathrm{E}[C_{ij}] = m_{ij} \qquad \mathrm{Var}[C_{ij}] = \phi m_{ij} \tag{7.18}$$

with:

$$\log(m_{ij}) = c + \alpha_i + s_\theta(j) + s_\theta(\log(j)). \tag{7.19}$$

7.10.2 Since this is treated as a non-parametric model, column parameter estimates are not available, because the mathematical form of the underlying spline is hidden from the user, with all calculations being performed behind the scenes. However, fitted values and their approximate standard errors are readily available. It is also possible to obtain approximate covariances of the predictors, enabling the prediction errors to be approximated by plugging-in the appropriate quantities to the relevant formula.

7.10.3 Table 27 shows implied development factors obtained after fitting the model described by equations 7.18 and 7.19, using a moderate amount of smoothing, and extrapolating for a further five years. Tables 28 and 29 show reserve results without and with extrapolation respectively.

7.10.4 It can be seen that the development factors adhere closely to the chain-ladder development factors across the range of development time. An alternative way of assessing the models is to view the component relating to development time graphically. That is, draw a graph of the predictor ignoring the parameters relating to origin year (which simply sets the level for each origin year). Figure 1 shows the development time component for the chain-ladder model, the Hoerl curve, and the smoothing model, and shows the effect of extrapolation. Note that this represents the shape of the run-off of incremental claims.

Table 27. Over-dispersed Poisson model; smoothing model; equivalent development factors

	Estimate
Lambda 2	3.006
Lambda 3	1.618
Lambda 4	1.279
Lambda 5	1.171
Lambda 6	1.100
Lambda 7	1.054
Lambda 8	1.031
Lambda 9	1.019
Lambda 10	1.013
Lambda 11	1.009
Lambda 12	1.006
Lambda 13	1.005
Lambda 14	1.003
Lambda 15	1.003

Table 28. Over-dispersed Poisson model; smoothing model; reserve results without extrapolation

	Reserve	Prediction error	Prediction error %
Year 1	–	–	–
Year 2	243	453	186%
Year 3	885	873	99%
Year 4	2,033	1,370	67%
Year 5	3,582	1,943	54%
Year 6	3,849	2,201	57%
Year 7	5,393	2,936	54%
Year 8	11,091	4,928	44%
Year 9	10,568	5,945	56%
Year 10	17,654	12,378	70%
Overall	55,297	18,124	33%

Table 29. Over-dispersed Poisson model; smoothing model; reserve results with extrapolation

	Reserve	Prediction error	Prediction error %
Year 1	435	700	161%
Year 2	634	814	128%
Year 3	1,447	1,217	84%
Year 4	2,704	1,688	62%
Year 5	4,269	2,226	52%
Year 6	4,304	2,400	56%
Year 7	5,801	3,120	54%
Year 8	11,650	5,168	44%
Year 9	10,936	6,159	56%
Year 10	18,109	12,739	70%
Overall	60,291	19,448	32%

7.10.5 Changing the amount of smoothing changes the shape of the run-off. If there is no smoothing, the chain-ladder model is reproduced. If the maximum level of smoothing is chosen, the model becomes linear in development time and log development time, giving the Hoerl curve. Having two well-known models as a result of choosing the extremes of the smoothing parameter is a very appealing characteristic of the smoothing models.

7.10.6 Although this simple example shows some benefits of smoothing, the benefits are more obvious in Figure 2, which shows the run-off pattern of incremental claims from a motor portfolio in which data were collected quarterly. The class was fairly long-tailed, and showed considerable variability in the tail. Fitting a smoothing model shows that the run-off is essentially bi-modal, with one peak due predominantly to property damage claims being paid fairly quickly, and a second peak due to the longer settlement of bodily injury claims. Smoothed development factors can be

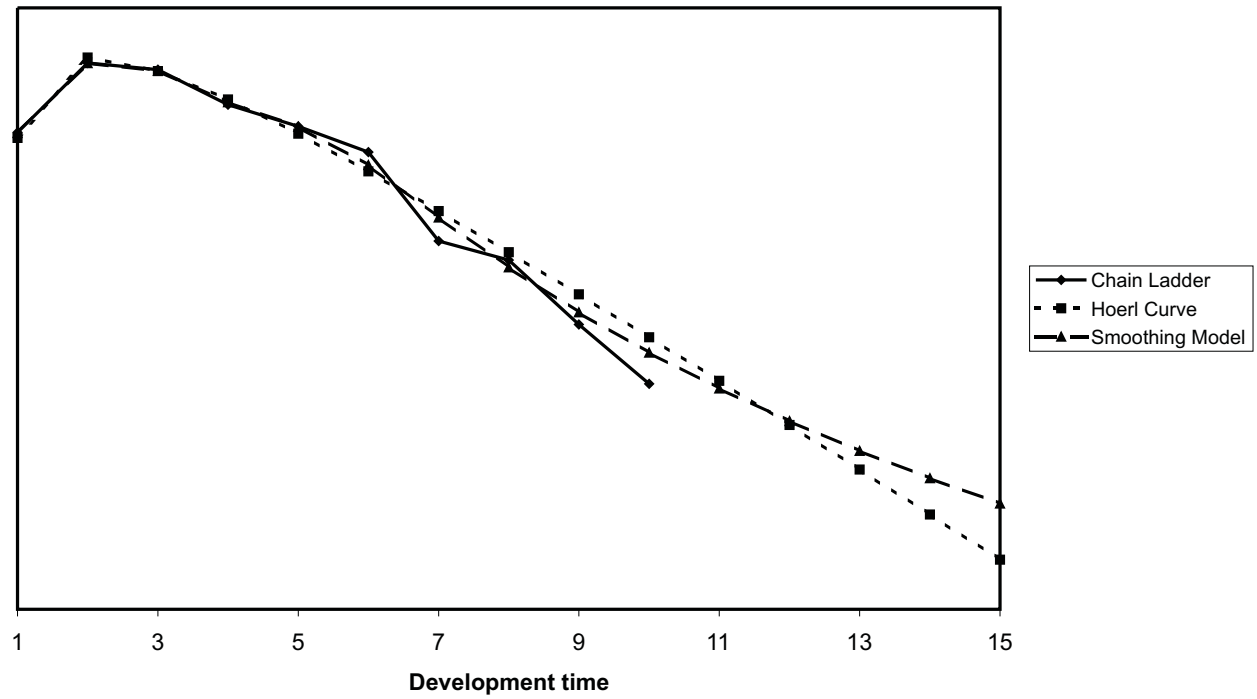

Figure 1. Run-off pattern of incremental claims

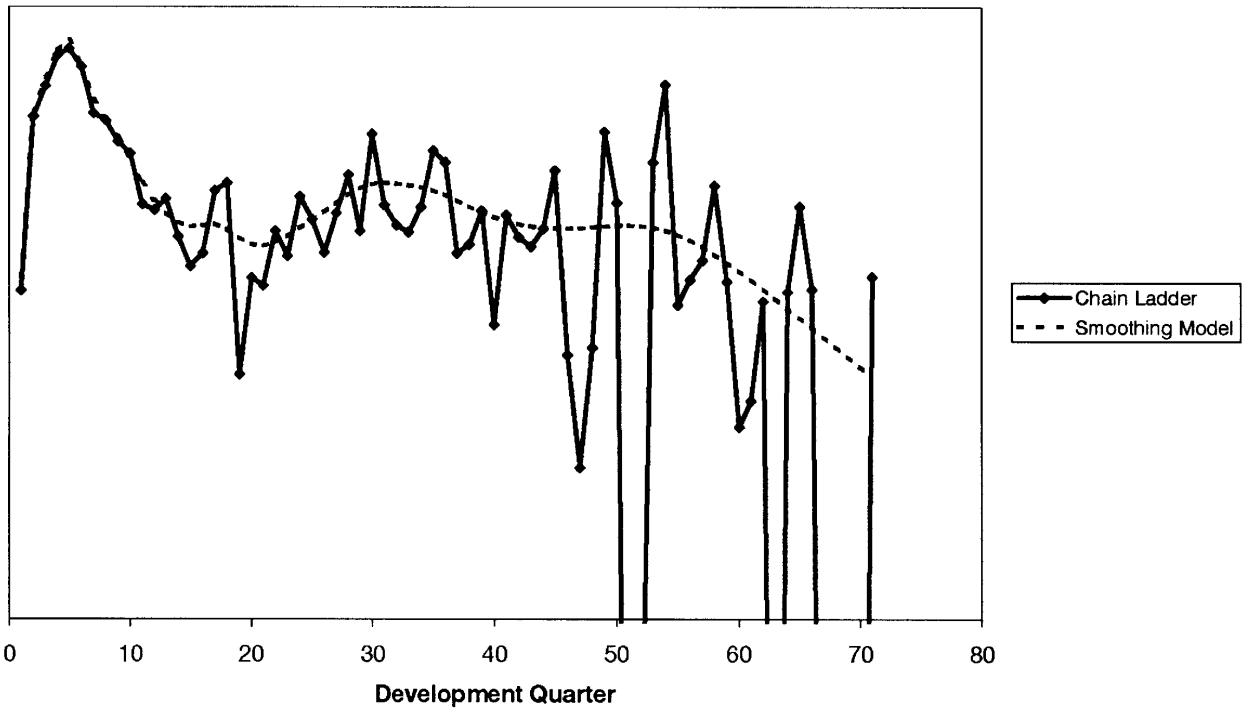

Figure 2. Run-off pattern of incremental claims; motor example with quarterly development

obtained from the model, with assurance that a consistent level of smoothing has been adopted. Obtaining the equivalent results by hand would be a tedious process.

7.11 *Bayesian Models and the Bornhuetter-Ferguson Technique*

7.11.1 The implementation of Bayesian models has been simplified enormously in recent years, due to the availability of WinBUGS (Spiegelhalter *et al.*, 1996, http://www.mrc-bsu.cam.ac.uk/bugs), a freely available software package specifically designed to fit Bayesian models using Gibbs sampling, a technique for Markov chain Monte Carlo analysis. The term 'Markov chain Monte Carlo' simply refers to the way in which simulation techniques are implemented to obtain the simulated posterior distribution of the parameters: 'Monte Carlo' refers to sampling from a given distribution; and 'Markov chain' refers to the way in which model parameters are treated sequentially using an iterative procedure. In fact, the main hurdle in adopting Bayesian methods is understanding the terminology, and making the mental shift from a purely mathematical solution to a solution based on simulation techniques. Once that shift has been made, the techniques make available solutions to problems which otherwise would be intractable to solve mathematically. An excellent overview of MCMC methods with applications in actuarial science is provided by Scollnik (2001).

7.11.2 To illustrate the techniques, we consider the over-dispersed Poisson model of Section 3.2, parameterised as a multiplicative model. That is:

$$\mathrm{E}[C_{ij}] = m_{ij} \qquad \mathrm{Var}[C_{ij}] = \phi m_{ij}$$

and

$$m_{ij} = x_i y_j.$$

7.11.3 When specified in this way, x_i is the expected ultimate claims and y_j is the proportion of ultimate claims to emerge in each development year. Notice that the model is non-linear in the parameters, which would usually cause complications, with traditional modelling techniques. Even using Bayesian simulation techniques, convergence is not guaranteed, and the model is less stable than using a log link and a linear predictor.

7.11.4 The model is set up by specifying the underlying distribution of the data (over-dispersed Poisson), the structure relating the parameters of the model to the mean, and prior distributions for the parameters. In the case of the standard chain-ladder model, 'vague' priors are used, that is, distributions with a given mean, but with large variance. As explained in Section 6.3, a Bornhuetter-Ferguson type methodology can be implemented by providing a 'proper' prior in which the mean of x_i is specified with greater precision.

7.11.5 Using Bayesian methods, a posterior distribution of the parameters is obtained, conditional on the data observed. Various summary statistics can then be obtained, if desired, such as the mean, median or mode of the posterior distribution. The posterior mode is analogous to the maximum likelihood estimate of classical statistics.

7.11.6 The predictive distribution of unobserved observations can be obtained naturally in WinBUGS by setting unobserved observations as 'missing values', which are then imputed by the software.

7.11.7 Since MCMC methods provide a predictive distribution of unobserved values using simulation, it is straightforward to calculate the prediction error, which is simply the standard deviation of the predictive distribution. This highlights the power and simplicity of simulation techniques, since the need to derive and evaluate complicated formulae involving covariances of predictor values, or even to separate the prediction error into its process and estimation components, is made redundant.

7.11.8 Since the focus is on the distribution of (the sum of) unobserved future claims, we can ignore the simulated posterior distribution of parameter estimates. Using 'vague priors' for the parameters, that is, prior distributions with very large variance, the mean and standard deviation of the simulated predictive distribution of the reserves are shown in Table 30. Note that the standard deviation of the predictive distribution is an estimate of the prediction error. Note also that the results are close to the results of the over-dispersed Poisson model obtained analytically in Table 3. It would be desirable if the results were closer still. The differences are due to a number of factors, including: simulation error, choice of prior for the parameters, and choice of the form of the model linking the parameters to the mean. Using a log link and linear predictor gives superior results with this example.

7.11.9 Suppose that we believe the predicted ultimate for year ten is too high, due to the payment in the first development period being unusually high. We can put an informative prior on x_{10}. Initially, suppose the prior on x_{10} has mean 16,000 and standard deviation 1, that is, we use a very precise prior, since the standard deviation is so low. The reserve results are shown in

Table 30. Over-dispersed Poisson model; Bayesian chain-ladder model with vague priors; mean and prediction error of reserves

	Bayesian mean reserve	Bayesian prediction error	Bayesian prediction error %	Analytic mean reserve	Analytic prediction error	Analytic prediction error %
Year 2	159	553	349%	154	556	361%
Year 3	636	1,107	174%	617	1,120	181%
Year 4	1,649	1,730	105%	1,636	1,775	109%
Year 5	2,808	2,190	78%	2,747	2,231	81%
Year 6	3,731	2,388	64%	3,649	2,440	67%
Year 7	5,539	3,104	56%	5,435	3,124	57%
Year 8	11,030	4,876	44%	10,907	5,032	46%
Year 9	10,840	5,975	55%	10,650	6,075	57%
Year 10	16,990	12,990	76%	16,339	12,987	79%
Overall	53,380	18,010	34%	52,135	18,193	35%

Table 31. Over-dispersed Poisson model; Bayesian chain-ladder model with precise priors; mean and prediction error of reserves

	Bayesian mean reserve	Bayesian prediction error	Bayesian prediction error %	Bornhuetter-Ferguson reserve
Year 2	155	545	352%	154
Year 3	625	1,104	177%	617
Year 4	1,659	1,766	106%	1,636
Year 5	2,811	2,214	79%	2,747
Year 6	3,717	2,399	65%	3,649
Year 7	5,511	3,065	56%	5,435
Year 8	11,000	4,918	45%	10,907
Year 9	10,900	6,016	55%	10,650
Year 10	14,440	4,924	34%	14,206
Overall	50,820	12,880	25%	50,002

Table 31, together with the standard chain-ladder estimates and the Bornhuetter-Ferguson estimates. Notice that the reserve estimate for year ten is similar to the result of the Bornhuetter-Ferguson technique, as might be expected with a precise prior. Notice also that the prediction errors have reduced substantially, reflecting the degree of precision of the prior: it could be questioned whether use of such a strong prior is appropriate.

7.11.10 Suppose now we use an informative prior for x_{10}, with mean 16,000 and standard deviation 5000. We are incorporating prior belief about the ultimate claims for year ten, but allowing for uncertainty. The associated reserve results are shown in Table 32. Notice that the reserves for year ten are between the chain-ladder and Bornhuetter-Ferguson results. Notice, also, that the precision of the prior has influenced the prediction errors, but to a lesser extent.

Table 32. Over-dispersed Poisson model; Bayesian chain-ladder model with informative priors; mean and prediction error of reserves

	Bayesian mean reserve	Bayesian prediction error	Bayesian prediction error %
Year 2	166	566	340%
Year 3	635	1,115	176%
Year 4	1,675	1,759	105%
Year 5	2,820	2,223	79%
Year 6	3,713	2,391	64%
Year 7	5,523	3,057	55%
Year 8	11,090	4,932	44%
Year 9	10,870	5,923	54%
Year 10	14,640	6,324	43%
Overall	51,130	13,390	26%

7.11.11 It should be noted that, in general, an informative prior on a row parameter will affect the results in all rows, to some extent, through its affect on the joint posterior distribution of the parameters. An exception is when a prior is put on the row parameter in the most recent year only (as in the example), since the observed single data point in that year does not contribute to the fit. This can be seen intuitively by considering the chain-ladder technique, in which the value of incremental claims in the most recent origin year does not contribute to the calculation of development factors.

8. Predictive Distributions and Simulation

8.1 *Introduction*

8.1.1 So far in this paper the focus has been on obtaining an estimate of the mean reserves and an estimate of the prediction error of the reserves, given an underlying statistical model. The prediction error is useful as a measure of precision of the reserve estimates, which might be considered useful for setting prudent reserves. For example, Wright (1990) suggested allowing for prudence in the reserves by adding a multiple of the prediction error to the best estimate. Practitioners will differ in their opinions of the merits of this suggestion, but it is eminently sensible to consider the possible downside, as well as the best estimate.

8.1.2 Although calculating the prediction error is a good starting point, it is still only a measure of the second moment of the full predictive distribution of possible reserve outcomes. Other summary measures, such as measures of skewness or extreme percentiles of the predictive distribution, are also of interest. In fact, the 'holy grail' is to obtain a full predictive distribution from which summary measures can be calculated, if desired. Notice that we say *a* predictive distribution, not *the* predictive distribution, since it is possible to have more than one distribution giving the same mean and variance, while having a different distribution overall.

8.1.3 There is little in the actuarial literature which considers the predictive distribution of reserve outcomes; to date the primary focus has been on estimating variability using prediction errors. That is because the predictive distribution of reserve estimates, which are, themselves, the sum of random variables, is difficult to obtain analytically, taking account of the variability due to the underlying statistical process and the variability due to the estimation of parameters. In this context, a paper which does consider a predictive distribution is Wright (1997), which describes a way of fitting models to individual claims data from which a predictive distribution of reserve outcomes can be obtained. The paper also describes an interesting way of calculating a prudential margin using the proportional hazards principle, introduced by Wang (1995).

8.1.4 If the distribution of the underlying data is unspecified, it is not possible to obtain a predictive distribution without further assumptions. For example, Mack (1994b) suggested calculating a 90% confidence interval of reserve estimates, given the mean and prediction error of the reserves, making the additional assumptions that the reserves are approximately Normally or log-Normally distributed.

8.1.5 With the advent of fast computers, and an increasing acceptance of simulation techniques, it is now possible to devise methods of obtaining a predictive distribution of reserve estimates using simulation methods. A description of some of the methods appears in Sections 8.2 to 8.4. Variations on these methods, and possibly some new ones, are likely to be suggested in the future.

8.2 *Bootstrapping*

8.2.1 Bootstrapping (Efron & Tibshirani, 1993) is a powerful, yet simple, technique for obtaining information from a single sample of data, which would usually be obtained using analytic techniques. The methodology revolves around sampling with replacement from the observed data sample, to create a large number of sets of pseudo-data, which are consistent with the same underlying distribution. Statistics of interest can then be obtained for each set of pseudo-data, and the distribution of those statistics investigated to obtain further insight. For example, the mean of each set of pseudo-data can be calculated, and the standard deviation of the set of means calculated as an estimate of the standard error of the mean.

8.2.2 In England & Verrall (1999), bootstrapping was used to obtain the estimation error of reserve estimates from the chain-ladder model. An analytic adjustment was suggested to calculate the process error, and the two components combined to form the prediction error. In England (2001), the method was extended to simulate the process error, in addition to using bootstrapping, to obtain the estimation error, thereby allowing a predictive distribution to be obtained.

8.2.3 The advantage of this two-stage bootstrapping/simulation approach is that it is very easy to set up in a spreadsheet, and does not require sophisticated statistical software or the calculation of complex formulae.

8.2.4 In a standard application of the bootstrap, where data are assumed to be independent and identically distributed, resampling with replacement takes place from the data themselves. 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 problems, it is common to bootstrap residuals, rather than the data themselves, since the residuals are approximately independent and identically distributed, or can be made so. However, it is important to use an appropriate residual definition for the

problem at hand. For generalised linear models, a range of extended definitions of residuals exists, the precise form being dictated by the underlying modelling distribution (see McCullagh & Nelder, 1989). For the over-dispersed Poisson chain-ladder model, we use the Pearson residuals for bootstrapping.

8.2.5 Dropping the suffices indicating the origin and development year, the Pearson residuals r_P are defined as:

$$r_P = \frac{C - \hat{m}}{\sqrt{\hat{m}}} \tag{8.1}$$

where $\hat{m}$ is the fitted incremental claims given by equation 7.2 or 7.3.

8.2.6 The bootstrap process involves resampling, with replacement, from the residuals. A bootstrap data sample is then created by inverting equation 8.1, using the resampled residuals, together with the fitted values. Given a resampled Pearson residual r_P^*, together with the fitted value m, the associated bootstrap incremental claims amount C^* is given by:

$$C^* = r_P^* \sqrt{\hat{m}} + \hat{m}.$$

8.2.7 Resampling the residuals (with replacement) gives rise to a new triangle of past claims payments. Having obtained the bootstrap sample, the model is refitted and the statistic of interest calculated. Strictly, we ought to fit an over-dispersed Poisson GLM to the bootstrap sample to obtain a bootstrap reserve estimate. However, we can obtain identical reserve estimates using standard chain-ladder methodology. It is at this point that the usefulness of the bootstrap process becomes apparent; we do not need sophisticated software to fit the model, a spreadsheet will suffice.

8.2.8 Having fitted the chain-ladder model to the bootstrap sample, and obtained forecast incremental claims payments, the second stage of the procedure is invoked which replicates the process variance. This is achieved by simulating an observed claims payment for each future cell in the run-off triangle, using the bootstrap value as the mean, and using the process distribution assumed in the underlying model, which, in this case, is over-dispersed Poisson, using equation 7.1.

8.2.9 The procedure is repeated a large number of times, each time providing a new bootstrap value and simulated forecast payment. For each iteration, the reserves are calculated by summing the simulated forecast payments. The set of reserves obtained in this way forms the predictive distribution, from which summary statistics, such as the prediction error (which is simply the standard deviation of the distribution of reserve estimates), can be obtained.

8.2.10 A complete list of the steps required appears in Appendix 3. A

Table 33. Chain-ladder model; bootstrap results

	Chain ladder reserve	Bootstrap mean	Bootstrap prediction error	Bootstrap prediction error %	Analytic prediction error	Analytic prediction error %
Year 2	154	177	695	392%	556	361%
Year 3	617	639	1,343	210%	1,120	181%
Year 4	1,636	1,655	1,992	120%	1,775	109%
Year 5	2,747	2,770	2,377	86%	2,231	81%
Year 6	3,649	3,769	2,563	68%	2,440	67%
Year 7	5,435	5,459	3,093	57%	3,124	57%
Year 8	10,907	11,259	5,135	46%	5,032	46%
Year 9	10,650	10,902	6,018	55%	6,075	57%
Year 10	16,339	16,580	13,644	82%	12,987	79%
Overall	52,135	53,210	19,267	36%	18,193	35%

more complete description together with a worked example and a discussion of various practical considerations can be obtained by referring to the original papers.

8.2.11 The bootstrap procedure, outlined in Appendix 3, was applied to the data in Table 1. The mean and standard deviation of 1000 bootstrap iterations is shown in Table 33, together with the chain-ladder reserve estimates and analytic prediction errors for the over-dispersed Poisson chain-ladder model. The mean of the bootstrap process is reassuringly close to the chain-ladder results. The standard deviation of the bootstrap sample is also close to the prediction error calculated analytically, providing further comfort. An advantage of the bootstrap procedure is that it also provides a predictive distribution of reserves. Various percentiles of that distribution are shown in Table 34, with a histogram of the overall reserves shown in Figure 3.

Table 34. Percentiles of predictive distributions; bootstrap chain-ladder, log-Normal, and Bayesian over-dispersed Poisson

	Bootstrap reserve	Log-normal reserve	Bayes ODP reserve
1st percentile	16,258	41,375	22,290
5th percentile	25,951	55,171	29,070
10th percentile	30,694	61,810	32,950
25th percentile	39,282	77,049	40,700
50th percentile	51,059	107,295	50,390
75th percentile	64,428	148,078	62,990
90th percentile	78,491	222,397	76,550
95th percentile	87,668	278,500	87,210
99th percentile	109,445	489,267	108,500

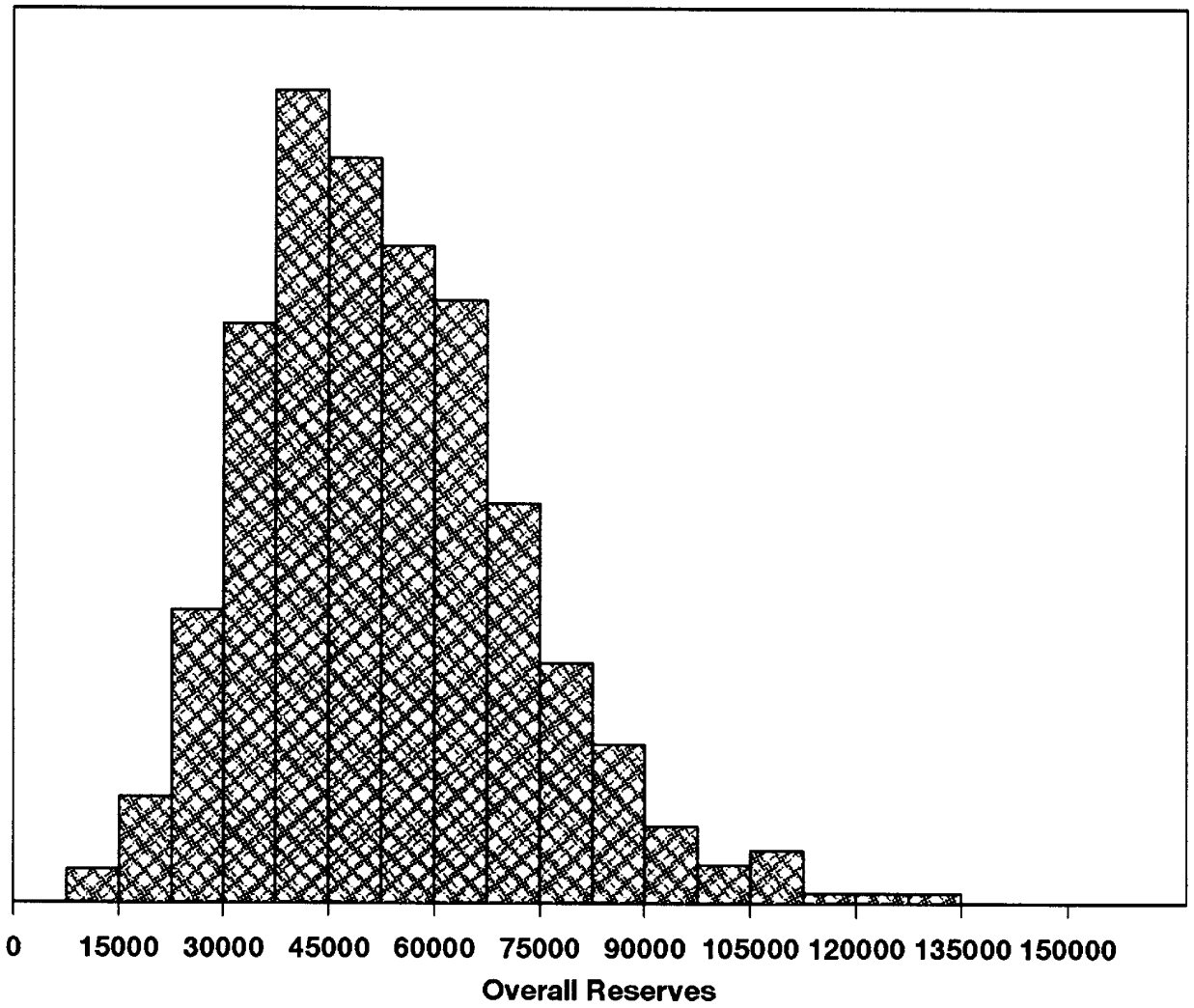

Figure 3. Bootstrap chain-ladder model; predictive distribution of overall reserves

8.2.12 Bootstrapping is simple and quick to implement for the over-dispersed Poisson version of the chain-ladder model, since the standard chain-ladder model can be fitted at each iteration of the process. For other models, such as the log-Normal models and models with a parametric predictor structure, it is also possible to use bootstrapping, but the implementation is less straightforward.

8.3 *Simulation from the Parameters*

8.3.1 An alternative to bootstrapping is to simulate the estimation error by simulating the parameters from an appropriate joint distribution, using information from the fitted model. The model predictor can then be formed, giving a distribution for the linear predictor. The process error can then be incorporated by drawing a random observation from the process distribution, given the simulated linear predictor. As with the bootstrap procedure, this is a two-stage process.

8.3.2 The complication with this procedure is identifying an appropriate joint distribution for the parameters from which to sample and obtaining the appropriate parameters. For models based on GLM methodology using a Normal error distribution and identity link function, the parameter estimates and the covariance matrix of the parameter estimates are all that is required, and are usually readily available from standard statistical software packages. It is often assumed that the distribution of the linear predictor, in this case, is approximately multivariate Normal. Therefore, given the parameter estimates and their associated covariance matrix, a distribution of parameters can be simulated by sampling from a multivariate Normal distribution. Given the joint distribution of parameter estimates, the appropriate sum of parameter estimates can be formed (incorporating continuous covariates if necessary, as in the Hoerl curve) to create a distribution of linear predictors. A random observation from the process distribution can then be drawn for each simulated predictor value, whose mean is the predictor value simulated at the first stage, and whose variance is dictated by the form of the process distribution.

8.3.3 The result of this two-stage process is a predictive distribution for each future payment in the run-off triangle, having the correct correlation characteristics. The appropriate sum of simulated payments can then be formed to give a predictive distribution of reserve estimates. The mean and standard deviation of that distribution should be the same as the mean and prediction error calculated analytically, if the procedure has been carried out correctly.

8.3.4 To illustrate the methodology, consider the log-Normal model of Section 7.7. Taking the parameter estimates in Table 18 as the means, together with the associated covariance matrix, a set of parameter estimates was simulated from a multivariate Normal distribution, and the predictor in equation 7.16 formed for each future cell in the run-off triangle. For each simulated predictor value, an observation from a Normal distribution was simulated with the predictor value as the mean, and the dispersion parameter in Table 18 as the variance. That simulated observation was then exponentiated to form the forecast payment. Those forecast payments were then summed to give origin year and overall reserves. The mean and standard deviation of the simulated reserves were taken as the expected reserves and prediction error, respectively. The results are shown in Table 35, together with the equivalent results obtained analytically, and again the results are reassuringly close. Various percentiles of the predictive distribution are shown in Table 34, alongside the bootstrap results of the over-dispersed Poisson model. It can be seen that the distribution for the log-Normal model has a higher level generally, and a larger variance, reflecting the impact of the bias correction. A histogram of the overall reserves is shown in Figure 4. It is clear from the histogram that the predictive distribution, when using the log-Normal model, is heavily skewed, with extreme observations influencing

Table 35. Log-Normal model; chain-ladder type predictor; simulation results

	Simulated mean reserve	Simulated prediction error	Simulated prediction error %	Analytic mean reserve	Analytic prediction error	Analytic prediction error %
Year 2	378	719	190%	357	751	210%
Year 3	1,021	1,346	132%	1,020	1,413	139%
Year 4	3,334	3,528	106%	3,064	3,291	107%
Year 5	3,586	2,919	81%	3,753	3,540	94%
Year 6	6,132	5,184	85%	6,010	5,227	87%
Year 7	7,565	5,810	77%	7,742	6,678	86%
Year 8	18,526	13,987	76%	18,806	16,379	87%
Year 9	27,510	27,997	102%	25,367	24,908	98%
Year 10	60,772	77,984	128%	56,475	77,519	137%
Overall	128,823	84,842	66%	122,595	86,312	70%

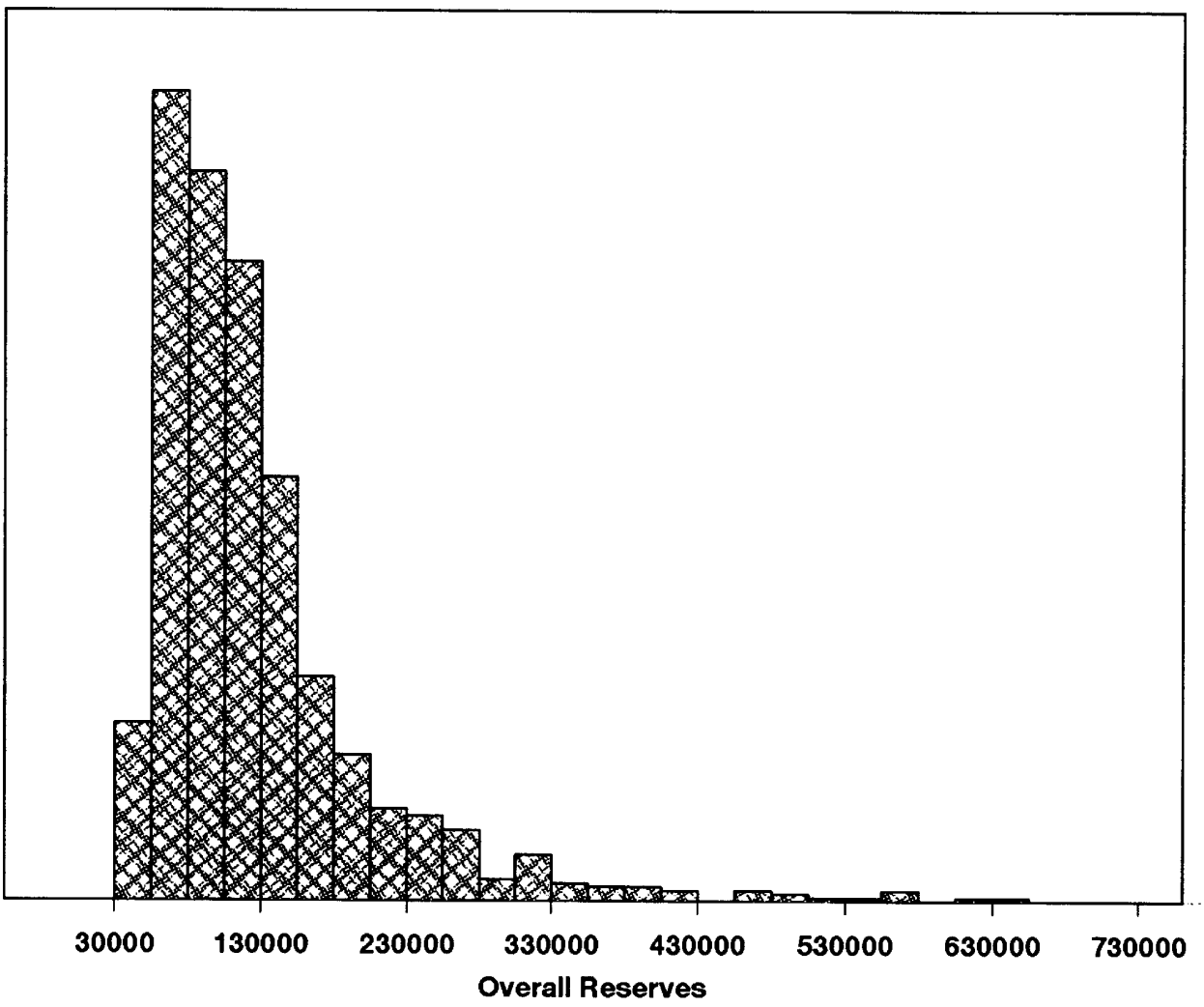

Figure 4. Log-Normal model; chain-ladder type predictor; predictive distribution of overall reserves

the expected value. Comparison of Figures 3 and 4 shows that the modes of the distributions (the most likely values) are similar. Inference can therefore be very different, depending on whether the means of the predictive distributions are being compared, or their 'maximum likelihood' values.

8.3.5 Barnett & Zehnwirth (1998) also suggest simulating from a multivariate (log) Normal distribution, but the mechanics of the process are not described, and it is not clear whether the two-stage procedure outlined in this section is adopted, or the estimation error alone is considered.

8.4 *Bayesian Methods*

8.4.1 The method of simulation from the parameters outlined in the previous section is reasonable if the joint distribution of the parameters can be identified. For GLMs with a non-Normal error structure and non-identity link, this may not be straightforward. In fact, identifying such a distribution might only be possible by working in a Bayesian world, making assumptions about the prior distribution of the parameters, and finding the posterior joint distribution, given the data. Even then, the distribution may not be recognisable as a standard one. Markov chain Monte Carlo methods obviate this problem using simulation and numerical methods, and consider estimation and prediction at the same time within the same framework.

8.4.2 MCMC methods do not provide parameter estimates per se, but a simulated joint distribution of parameter estimates. In the claims reserving context, a distribution of future payments in the run-off triangle is produced automatically, where that distribution is the predictive distribution. The appropriate sums of the simulated predicted values can then be formed to provide predictive distributions of origin year and total reserves. The means of those distributions may be used as the best estimates. Other summary statistics can also be investigated, since the full predictive distribution is available.

8.4.3 Unlike the methods based on GLMs (and Mack's method), which focus on modelling the mean, then investigate predictive distributions, MCMC methods provide the predictive distribution, from which summary statistics, such as the mean, can be calculated. As such, there are no extra steps to perform or approximations to be made when considering the predictive distribution.

8.4.4 The methodology is illustrated using results from the Bayesian version of the over-dispersed Poisson chain-ladder model described in Section 7.10. Vague priors are used to enable a comparison with the bootstrap approach. The means and standard deviations of the predictive distributions have already been shown in Table 30. Various percentiles of the predictive distribution are shown in Table 34. It can be seen that the Bayesian over-dispersed Poisson model and the bootstrap procedure give very similar distributions, as might be expected. A histogram of the overall reserves is

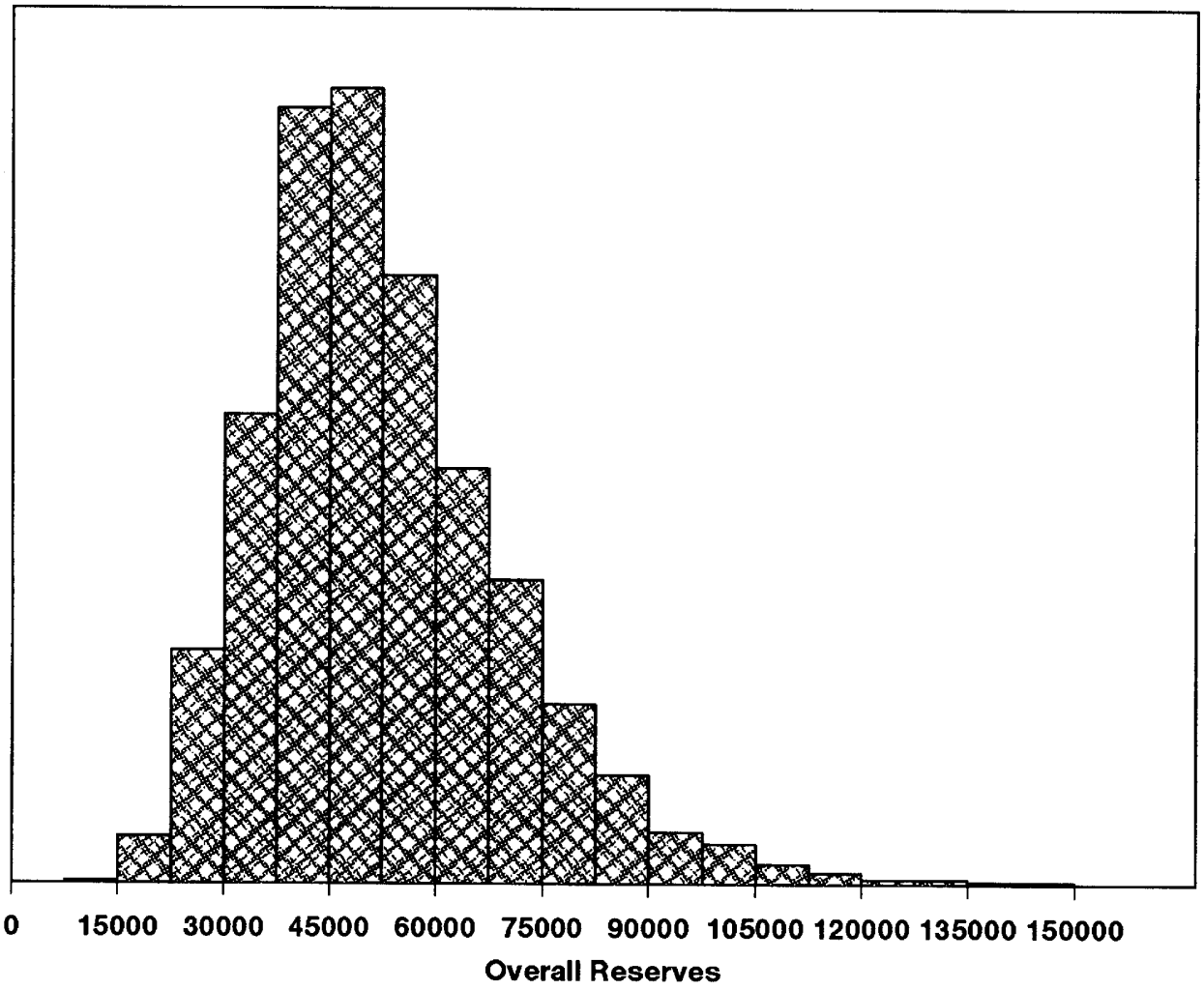

Figure 5. Bayesian over-dispersed Poisson model; predictive distribution of overall reserves

shown in Figure 5, again highlighting the similarity between the Bayesian results and the bootstrap results.

9. Dynamic Financial Analysis

9.1 In a dynamic financial analysis (DFA) exercise, cash flows of an insurance enterprise are simulated to help with business planning, pricing, capital modelling, and risk profiling generally. One component of risk carried by an insurance operation lies within its outstanding claims reserves, and simulating reserve movements is an ingredient of a full DFA exercise. Methods are therefore required for DFA which enable a predictive distribution of future claim payments to be obtained. Clearly, it is desirable if this can be performed in a consistent manner, such that, when the

simulated future claim payments are summed to give origin period reserves, or total reserves, the mean and variance of those sums matches their analytic equivalents. For example, it is inappropriate to simulate future payments as independent random variables, even if the mean and prediction error of those payments have been estimated correctly. Although the mean of simulated reserves would be correct, the standard deviation of the simulated reserves would be underestimated. That is because the underlying model relies on the same parameters, so dependence is introduced through estimation.

9.2 In this paper we have proposed various models for estimating reserves using stochastic methods. We have also shown how measures of variability can be calculated, and shown, for some models, alternative ways in which predictive distributions can be simulated. It is the predictive distributions which are of interest in a DFA exercise, and interest in stochastic reserving models is likely to increase as DFA becomes more popular. As simulation techniques are understood and accepted, and computing technology continues to improve, business planning on a best estimate basis, or including a few adverse scenarios whose probability can only be guessed, will appear increasingly inadequate.

9.3 However, simply obtaining the predictive distribution is not enough for a DFA exercise. Although the methods can be used to simulate calendar year reserve movements which are consistent with a traditional reserving model, a company would be expected to re-evaluate its reserves on a regular basis, taking account of the reserve movements in the intervening period. Such dynamic behaviour should be incorporated in a DFA exercise.

9.4 Furthermore, it is likely that claims inflation would be considered in a DFA exercise, perhaps linked to retail prices inflation, or other economic measures. In that case, the component of the DFA model generating outstanding reserves would need to be linked to an economic scenario generator in a sensible way.

9.5 It is also possible that reserve movements in different classes of business are dependent in some way. Although stochastic models could be fitted to several run-off triangles at the same time, and dependencies explored, the models would quickly become complex and involve numerous interaction terms. Even so, it might be desirable to consider such dependencies in a DFA exercise, and, again, an appropriate methodology for incorporating that is required. It is likely that judgement would be used in setting dependence parameters, and unlikely that sufficient data would be available to fully justify or reject such parameters.

9.6 It is not our intention to suggest here how those features might be incorporated, and, indeed, that could be the subject of a further paper.

10. Discussion

10.1 The subject of stochastic models and claims reserving has given rise, in recent years, to a number of controversies, most notably, discussing which stochastic model underlies the chain-ladder technique (Mack & Venter, 2000). Focusing on models which reproduce chain-ladder estimates is, in many ways, a futile exercise, although considerable insight can be gained by starting with those models, since they provide a link between traditional deterministic methods and stochastic methods. However, the chain-ladder technique is simply part of a spectrum of models. In this paper other models within that spectrum have been described, and an attempt has been made to explain the similarities and connections between them, where such similarities exist.

10.2 Somebody new to the topic might be inclined to ask the question: "Which model is best?", to which there is no straightforward answer. It is unlikely that an answer could be given to the same question aimed at traditional methods, since a particular model will suit a particular problem or data set; the data should be examined in detail in order to find an appropriate model, rather than using the same modelling approach in all circumstances. This is true for traditional and stochastic reserving models. In this context, there is little to choose between the over-dispersed Poisson model and the negative binomial model, since predictions and prediction errors are essentially the same, and the same restrictions regarding negative incremental claims apply, although the way in which the negative binomial model is parameterised is more consistent with the traditional chain-ladder model. Implementation of the negative binomial model is less straightforward, since the availability of software to fit the model is limited. The Normal approximation to the negative binomial offers greater flexibility in the presence of negative incremental claims, and its similarity to Mack's model offers greater insight.

10.3 Once within a stochastic framework, there is considerable flexibility in the choice of predictor structures, with predictors including continuous covariates offering the possibility of extrapolation beyond the range of data observed. This offers an alternative to the usual practice of fitting a standard model to cumulative claims, then fitting another model to the development factors so derived. Smoothing models offer even greater flexibility, and remove reliance on a predictor with a rigid parametric form. The disadvantage of the smoothing models is that obtaining a predictive distribution is not straightforward. Combining predictors with continuous covariates, either in a parametric or non-parametric context, with the Normal approximation to the negative binomial, offers extensions to Mack's model.

10.4 Bayesian models have much to recommend them, since estimation and prediction occur in the same model at the same time. Since model fitting

and prediction is performed using simulation, the methods automatically provide a predictive distribution of reserve estimates, from which the prediction error, if required, can be estimated by calculating the standard deviation of the simulated results; there is no need to evaluate complicated formulae. Bayesian models also have the advantage that actuarial judgement can be incorporated through the choice of informative prior distributions. This is also a major disadvantage, since it leaves the methods open to abuse. Practical difficulties associated with Bayesian models include choice of prior distribution and assurance that the software has converged on the optimum solution. Although such reassurance can be gained by knowing what the results should be (using analytic methods), a pragmatic alternative is to repeat the analysis several times, starting the simulations from very different initial values, and checking convergence.

10.5 Recently, the concept of a *best estimate* of reserves, and a range of best estimates, has received attention (see, for example, Gibson, 2000). It is far from clear what this means. For example, the median of the predictive distribution of reserves provides a measure of location of the underlying distribution, but a 50% probability of outstanding liabilities being greater than the best estimate is unlikely to be acceptable. Furthermore, the median provides no information concerning the downside potential. The mean of the predictive distribution is closer to the intended interpretation of a best estimate (in the United Kingdom), although, again, outstanding liabilities set on this basis being sufficient 'on average' does not leave much room for comfort.

10.6 It could be argued that several different deterministic reserving methods applied to different data sets provide a range of best estimates. Although that range is unlikely to be wide, prudence dictates that setting reserves at the lowest end of the range, without good reason, is unwise. The standard error of reserves (estimation error) also provides a measure of variability, but the interpretation is very different. The standard error represents the standard deviation of the mean reserves that would be obtained if we could repeat an experiment many times, in which we go back in time and repeat the loss experience, each time estimating the mean reserves. Although that may be of interest, the prediction error is of more interest, representing, not the standard deviation of the expected reserves, but the standard deviation of the actual outstanding liabilities. Clearly, the prediction error will be larger than the standard error.

10.7 In fact, as has been shown in the examples, the prediction error could be disappointingly large. This is entirely consistent with the small sample of data usually available in run-off triangles, and the limitations of the models available for estimation. However, the prediction error could be exacerbated using a model with a poor fit. For example, it is possible to fit a model with a single parameter, resulting in the estimated incremental payments being the mean of the observed incrementals. Such a model would provide a large prediction error, not as a result of inherent variability, but

due to a poor choice of model. Similarly, it is possible to construct a data set in which each origin year has a different payment pattern, but with no variability around that pattern within a year. Fitting a model in which each year is assumed to have the same payment pattern will result in a positive prediction error, whereas fitting the same model used to create the data would result in a prediction error of zero.

10.8 This is an important point; the prediction error reflects the underlying assumptions of the model. If the underlying assumptions are incorrect, the prediction error will also be incorrect (as will the expected value). In particular, the models assume that variability is due to random statistical variation, according to a particular statistical distribution. If it is believed that the observed variability could equally occur in the future, the assumption is acceptable. In general, the underlying assumptions will not hold in reality, to a greater or lesser extent. Some causes of variability, such as future changes in legislation, are not easy to predict, therefore cannot easily be incorporated. This does not provide a reason to dismiss attempting modelling, neither does it excuse poor modelling. Rather, it provides good reason to understand the modelling assumptions, and to test different models. Like traditional methods, different stochastic methods will give different results.

10.9 Interest in the concept of a best estimate (in the U.K.) stems primarily from actuarial sign-off of Lloyd's reserves, where an actuary is required to sign that the reserves are: "at least as large as those implied by a 'best estimate' basis without precautionary margins" (Guidance Notes 20 and 33, *Manual of Actuarial Practice*, Institute of Actuaries). Not only does this raise the question of the definition of a best estimate, it also raises the issue of prudential margins held in reserves. In this context, the term 'best estimate' is intended to represent: "the expected value of the distribution of possible outcomes of the unpaid liabilities". Essentially, this is the mean of the predictive distribution. The guidance notes, quite rightly, do not prescribe what that distribution should be, or how it might be obtained. Nonetheless, the wording is interesting. Even if the reserves held are at least as large as those implied by a best estimate, it would be of considerable value to know the estimated probability that outstanding liabilities might exceed the reserves held, which could be obtained if the predictive distribution can be estimated.

10.10 Concerning prudential margins, the Inland Revenue, regulators and rating agencies are at odds. Put simply, the Inland Revenue objects to possible profits (and hence revenue from taxes) being held back, whereas regulators and rating agencies are concerned with protecting the policyholder. Shareholders are torn between the desire for profits, and the security of their shareholding. It could be argued that reserves set on an undiscounted basis include an implicit margin for prudence, although, in the current climate of low interest rates, that margin is very much reduced. If

reserves are set on a discounted basis, there is a strong case for including an explicit prudential margin. As such, a risk margin based on a risk measure from a predictive distribution of claims reserves is a strong contender.

10.11 The reporting of variability of claims reserves also needs careful consideration. It could be seen as desirable to display a predictive distribution of reserves graphically in a report, as in Figures 3, 4 and 5, which highlights the levels of uncertainty. The disadvantage of providing a full distribution is that it highlights the potential upside, as well as the downside, which requires careful interpretation. If the aim of the report is to inform management for the purposes of setting reserves, highlighting the upside might be inadvisable. However, if the reserves are being reviewed for the purposes of a merger or acquisition, the potential upside will be of interest, as well as the downside. In any event, the distribution may need careful interpretation when presented to management.

10.12 Although the example data set used throughout this paper is based on incurred data (that is, paid losses and aggregate case estimates combined), in general, the methods are better suited to paid data. That is because case estimates are set individually and often a little conservatively, resulting in over-estimation when considered in aggregate, leading to negative incremental amounts in the later stages of development. The problem is more with the data than the methods, since, clearly, it is the estimation of aggregate case reserves which is at fault. Ideally, methods need to be found which help provide better estimates of aggregate case reserves. In this respect, models based on individual claims, rather than data aggregated into triangles, are likely to be of benefit. Aggregate triangles are useful for management information, and have the advantage that simple deterministic methods can be used to analyse them. However, it has to be borne in mind that traditional techniques were developed before the advent of desktop computers, using methods which could be evaluated using pencil and paper. With the continuing increase in computer power, it has to be questioned whether it would not be better to examine individual claims rather than use aggregated data. Databases of individual claims are routinely used for pricing purposes, so the provision of individual claims databases for reserving is feasible. Models could be developed, investigating, for example, the time taken to report claims, the sizes and timings of partial payments, the delay between occurrence and reporting of claims, reinsurance recoveries, and so on. In effect, this allows much closer modelling of the process, including the individual case reserve amounts. This introduces several complexities for model fitting, with a Bayesian approach offering help in this context, which is more amenable to tackling problems of this nature. Indeed, some progress has already been made by Haastrup & Arjas (1996), Haastrup (1997) and Pereira (2000), using work pioneered by Norberg (1993, 1999). Those papers and theses show that claims can be treated individually, with Bayesian inference being used to update the distributions of all

quantities of interest. Using these techniques, it is possible to obtain predictive distributions which are likely to be better than those obtained from data triangles. However, the implementation is far from straightforward, and further development is required before firm conclusions can be reached. Even so, several of the techniques outlined in this paper can equally be used with individual claims data, from which extensions can be developed.

10.13 In practice, it is often the case that triangles of gross, net, and ceded data are analysed separately. Fitting stochastic models to the data sets independently is likely to lead to difficulty in making inferences from the results. It is not obvious how to solve this problem without recourse to individual claims data, combined with suitable modelling of the reinsurance process. Reinsurance programmes are often complex, and obtaining the relevant information for older years, even if the process can be modelled, is not easy. Even if there is the desire to improve data standards, and obtain individual claims data, it is likely to be some time before the data sets are large enough to reap the benefits. Companies with good data standards will already be able to construct the required databases.

10.14 There are many areas where further research could be conducted. Of considerable value would be an empirical study of a large number of data triangles from various classes of business and companies, making reserve forecasts, and following the development over time. In this way, the performance of stochastic claims reserving models could be assessed, in particular by monitoring the frequency with which estimated percentiles are breached.

10.15 Principal omissions from this paper concern the use of exposure measures, modelling of claims inflation, and model testing and comparison. It is straightforward to extend the methods to incorporate exposure measures as weights in the models, allowing different weightings to be attached to different origin years. It is also straightforward to extend the models to estimate an inflationary trend by including a calendar year covariate. In the case of models using a logarithmic link, the coefficient of a term which is linear in calendar time provides an estimate of the force of claims inflation. Attempting to model claims inflation in each year individually is usually problematical, due to the number of parameters in the model and dependencies with the origin period and development period. If an element of claims inflation is believed to be known, it is better to strip out its effect before modelling.

10.16 A thorough analysis will include formal model testing using residuals to identify systematic and isolated departures from the fitted model, and goodness-of-fit measures to compare models and assess statistical significance of model terms. Such model checking techniques are entirely possible for the models outlined in this paper. Since there are many texts which describe model checking methods in detail, the topic has been omitted, as a complete exposition would add to the length of this paper.

10.17 In conclusion, the assessment of the financial strength of a general insurance enterprise includes a thorough analysis of the outstanding claims reserves, including an assessment of the possible variability in the reserves. Failure to do so will result in the insolvency of some insurers, as witnessed in recent months. This inevitably puts the spotlight on advisors, including auditors and actuaries. Including estimates of the variability of claims reserves in actuarial reports would change the emphasis on a best estimate, and might prove useful should litigation arise. Methods of analysis which help with reserve estimation, as well as providing insight into the variability of those reserves, are to be welcomed. It is important that the characteristics of the methods are explored, and their assumptions understood, and it is hoped that this paper contributes to that process.

Acknowledgements

The authors acknowledge the many useful discussions with Dr Arthur Renshaw which have occurred over time, and also acknowledge Dr Andrzej Czernuszewicz for the idea of simulating from the parameters of a generalised linear model.

References

Barnett, G. & Zehnwirth, B. (1998). Best estimates for reserves. Casualty Actuarial Society, Fall Forum.

Benktander, G. (1976). An approach to credibility in calculating IBNR for casualty excess reinsurance. *Actuarial Review*, April 1976, 7.

Bornheutter, R. L. & Ferguson, R. E. (1972). The actuary and IBNR. *Proceedings of the Casualty Actuarial Society*, **LIX**, 181-195.

Chambers, J. M. & Hastie, T. J. (1992). *Statistical models in S*. Chapman and Hall, London.

Christofides, S. (1990). Regression models based on log-incremental payments. *Claims Reserving Manual*, **2**, Institute of Actuaries, London.

De Jong, P. & Zehnwirth, B. (1983). Claims reserving, state-space models and the Kalman filter. *J.I.A.* **110**, 157-182.

Doray, L. G. (1996). UMVUE of the IBNR reserve in a lognormal linear regression model. *Insurance: Mathematics and Economics*, **18**, 43-57.

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

England, P. D. (2001). Addendum to 'Analytic and bootstrap estimates of prediction errors in claims reserving'. Actuarial Research Paper No. 138, Department of Actuarial Science and Statistics, City University, London, EC1V 0HB.

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. (2001). A flexible framework for stochastic claims reserving. *Proceedings of the Casualty Actuarial Society* (to appear).

Gibson, L. (2000). Reserving best estimates and ranges. *Proceedings of the General Insurance Convention*, **2**,185-187, Institute of Actuaries.

Green, P. J. & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*. Chapman and Hall, London.

HAASTRUP, S. (1997). Some fully Bayesian micro models for claims reserving. PhD thesis, University of Copenhagen.

HAASTRUP, S. & ARJAS, E. (1996). Claims reserving in continuous time: a nonparametric Bayesian approach. *ASTIN Bulletin*, **26**, 139-164.

HASTIE, T. J. & TIBSHIRANI, R. J. (1990). *Generalized additive models*. Chapman and Hall, London.

HISTORICAL LOSS DEVELOPMENT STUDY (1991). Reinsurance Association of America. Washington D.C.

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

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.

MACK, T. (1994a). Which stochastic model is underlying the chain-ladder model? *Insurance: Mathematics and Economics*, **15**, 133-138.

MACK, T. (1994b). Measuring the variability of chain-ladder reserve estimates. Casualty Actuarial Society, Spring Forum.

MACK, T. (2000). Credible claims reserves: the Benktander method. *ASTIN Bulletin*, **30**, 333-347.

MACK, T. & VENTER, G. (2000). A comparison of stochastic models that reproduce chain-ladder reserve estimates. *Insurance: Mathematics and Economics*, **26**, 101-107.

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

MURPHY, D. M. (1994). Unbiased loss development factors. *Proceedings of the Casualty Actuarial Society*, **LXXXI**, 154-222.

NORBERG, R. (1993). Prediction of outstanding liabilities in non-life insurance. *ASTIN Bulletin*, **23**, 95-115.

NORBERG, R. (1999). Prediction of outstanding liabilities II. Model variations and extensions. *ASTIN Bulletin*, **29**, 5-25

PEREIRA, F. C. (2000). Bayesian Markov chain Monte Carlo methods in general insurance. PhD thesis, City University, London.

RENSHAW, A. E. (1989). Chain-ladder and interactive modelling (claims reserving and GLIM). *J.I.A.* **116**, 559-587.

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

RENSHAW, A. E. (1994b). On the second moment properties and the implementation of certain GLIM based stochastic claims reserving models. Actuarial Research Paper No. 65, Department of Actuarial Science and Statistics, City University, London.

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

SCHMIDT, K. D. & SCHNAUS, A. (1996). An extension of Mack's model for the chain-ladder method. *ASTIN Bulletin*, **26**, 247-262.

SCOLLNIK, D. P. M. (2001). Actuarial modeling with MCMC and BUGS. *North American Actuarial Journal*, **5** (2), 96-124.

SHERMAN, R. E. (1984). Extrapolating, smoothing and interpolating development factors. *Proceedings of the Casualty Actuarial Society*, **LXXI**, 122-192.

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

S-PLUS (2001). *S-PLUS 6 for Windows user's guide*. Insightful Corporation, Seattle, WA.

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

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

VENABLES, W. N. & RIPLEY, B. D. (1999). *Modern applied statistics with S-Plus. 3rd edition.* New-York: Springer-Verlag.

VERRALL, R. J. (1989). A state space representation of the chain-ladder linear model. *J.I.A.* **116**, 589-610.

VERRALL, R. J. (1990). Bayes and empirical Bayes estimation for the chain-ladder model. *ASTIN Bulletin*, **20**, 2, 217-243.

VERRALL, R. J. (1991a). On the unbiased estimation of reserves from loglinear models. *Insurance: Mathematics and Economics*, **10**, 1, 75-80.

VERRALL, R. J. (1991b). Chain-ladder and maximum likelihood. *J.I.A.* **118**, 489-499.

VERRALL, R. J. (1994). A method for modelling varying run-off evolutions in claims reserving. *ASTIN Bulletin*, **24**, 2, 325-332.

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

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. (2001). A Bayesian generalised linear model for the Bornhuetter-Ferguson method of claims reserving. Actuarial Research Paper No. 139, Department of Actuarial Science and Statistics, City University, London.

VERRALL, R. J. & ENGLAND, P. D. (2000). Comments on: 'A comparison of stochastic models that reproduce chain-ladder reserve estimates', by Mack & Venter. *Insurance: Mathematics And Economics*, **26**, 109-111.

WANG, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics And Economics*, **17**, 43-54.

WHITTAKER, E. T. (1923). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, **41**, 63-75.

WRIGHT, T. S. (1990). A stochastic method for claims reserving in general insurance. *J.I.A.* **117**, 677-731.

WRIGHT, T. S. (1997). Probability distribution of outstanding liability from individual payments data. *Claims Reserving Manual*, **2**, Institute of Actuaries, London.

ZEHNWIRTH, B. (1985). *ICRFS version 4 manual and users guide*. Benhar Nominees Pty Ltd, Turramurra, NSW, Australia.

ZEHNWIRTH, B. (1989). The chain-ladder technique — a stochastic model. *Claims Reserving Manual*, **2**, Institute of Actuaries, London.

APPENDIX 1

RECURSIVE MODELS: PROCESS ERROR

A1.1 *General case*

A1.1.1 The negative binomial model and its Normal approximation are formulated as recursive models, so the calculation of the process variance involves estimating the variance of a k-steps-ahead forecast, using standard results from the analysis of conditional distributions. First, the general case is considered, followed by specific formulae for the negative binomial model and its Normal approximation.

A1.1.2 The simplest way to proceed is to start with the 2-steps-ahead moments, followed by the 3-steps-ahead, and spot a pattern from which it is straightforward to obtain the k-steps-ahead moments. In order to simplify the notation, we look at a single row, dropping the suffix i, and consider both the expected value and variance.

2 steps-ahead:

$$\begin{aligned}
\mathrm{E}\left[D_{j+1}|C_1, C_2, \ldots, C_{j-1}\right] &= \mathrm{E}\left[\mathrm{E}\left[D_{j+1}|C_1, C_2, \ldots, C_{j-1}, C_j\right]|C_1, C_2, \ldots, C_{j-1}\right] \\
&= \mathrm{E}\left[\lambda_j D_j|C_1, C_2, \ldots, C_{j-1}\right] \\
&= \lambda_{j+1}\lambda_j D_{j-1}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}\left[D_{j+1}|C_1, C_2, \ldots, C_{j-1}\right] &= \mathrm{E}\left[\mathrm{Var}\left[D_{j+1}|C_1, C_2, \ldots, C_{j-1}, C_j\right]|C_1, C_2, \ldots, C_{j-1}\right] \\
&\quad + \mathrm{Var}\left[E\left[D_{j+1}|C_1, C_2, \ldots, C_{j-1}, C_j\right]|C_1, C_2, \ldots, C_{j-1}\right] \\
&= \mathrm{E}\left[\mathrm{Var}\left[D_{j+1}|C_1, C_2, \ldots, C_j\right]|C_1, C_2, \ldots, C_{j-1}\right] \\
&\quad + \mathrm{Var}\left[\lambda_{j+1} D_j|C_1, C_2, \ldots, C_{j-1}\right] \\
&= \mathrm{E}\left[\mathrm{Var}\left[D_{j+1}|C_1, C_2, \ldots, C_j\right]|C_1, C_2, \ldots, C_{j-1}\right] \\
&\quad + \lambda_{j+1}^2 \mathrm{Var}\left[D_j|C_1, C_2, \ldots, C_{j-1}\right]
\end{aligned}$$

which is now in terms of the 1-step-ahead variance, allowing a recursive procedure to be implemented.

3-steps-ahead:

$$\begin{aligned}
\mathrm{E}\big[D_{j+2}|C_1, C_2, \ldots, C_{j-1}\big] &= \mathrm{E}\big[\mathrm{E}\big[D_{j+2}|C_1, C_2, \ldots, C_{j-1}, C_j\big]|C_1, C_2, \ldots, C_{j-1}\big] \\
&= \mathrm{E}\big[\lambda_{j+2}\lambda_{j+1}D_j|C_1, C_2, \ldots, C_{j-1}\big] \\
&= \lambda_{j+2}\lambda_{j+1}\lambda_j D_{j-1}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}\big[D_{j+2}|C_1, C_2, \ldots, C_{j-1}\big] &= \mathrm{E}\big[\mathrm{Var}\big[D_{j+2}|C_1, C_2, \ldots, C_{j-1}, C_j\big]|C_1, C_2, \ldots, C_{j-1}\big] \\
&\quad + \mathrm{Var}\big[\mathrm{E}\big[D_{j+2}|C_1, C_2, \ldots, C_{j-1}, C_j\big]|C_1, C_2, \ldots, C_{j-1}\big] \\
&= \mathrm{E}\big[\mathrm{Var}\big[D_{j+1}|C_1, C_2, \ldots, C_j\big]|C_1, C_2, \ldots, C_{j-1}\big] \\
&\quad + \lambda_{j+2}^2\lambda_{j+1}^2\mathrm{Var}\big[D_j|C_1, C_2, \ldots, C_{j-1}\big]
\end{aligned}$$

which is now in terms of the 1-step-ahead and 2-steps-ahead variances.

A1.1.3 Note that we have used the 2-steps-ahead formula for the mean in the derivations of the 3-steps-ahead moments.

A1.1.4 It is easy to see that this process can be continued to produce the k-steps-ahead moments.

A1.2 *Negative Binomial*

A1.2.1 From the negative binomial model in Section 2.4, we know that the variance of the 1-step-ahead forecast is given by:

$$\mathrm{Var}\big[D_j|C_1, C_2, \ldots, C_{j-1}\big] = \phi\lambda_j\big(\lambda_j - 1\big)D_{j-1}.$$

A1.2.2 Then the variance of the 2-steps-ahead forecast, from Section A1.1, is given by:

$$\begin{aligned}
\mathrm{Var}\big[D_{j+1}|C_1, C_2, \ldots, C_{j-1}\big] &= \mathrm{E}\big[\phi\lambda_{j+1}\big(\lambda_{j+1} - 1\big)D_j|C_1, C_2, \ldots, C_{j-1}\big] \\
&\quad + \lambda_{j+1}^2\phi\lambda_j\big(\lambda_j - 1\big)D_{j-1} \\
&= \phi\lambda_{j+1}\big(\lambda_{j+1} - 1\big)\lambda_j D_{j-1} + \lambda_{j+1}^2\phi\lambda_j\big(\lambda_j - 1\big)D_{j-1} \\
&= \phi\lambda_j\lambda_{j+1}\big(\lambda_j\lambda_{j+1} - 1\big)D_{j-1}
\end{aligned}$$

and the variance of the 3-steps-ahead forecast is given by:

$$\begin{aligned}\mathrm{Var}\left[D_{j+2}|C_1, C_2, \ldots, C_{j-1}\right] &= \mathrm{E}\left[\phi\lambda_{j+1}\lambda_{j+2}\left(\lambda_{j+1}\lambda_{j+2}-1\right)D_j\right]+\lambda_{j+2}^2\lambda_{j+1}^2\phi\lambda_j\left(\lambda_j-1\right)D_{j-1}\\ &= \phi\lambda_{j+1}\lambda_{j+2}\left(\lambda_{j+1}\lambda_{j+2}-1\right)\lambda_j D_{j-1}+\lambda_{j+2}^2\lambda_{j+1}^2\phi\lambda_j\left(\lambda_j-1\right)D_{j-1}\\ &= \phi\lambda_j\lambda_{j+1}\lambda_{j+2}\left(\lambda_j\lambda_{j+1}\lambda_{j+2}-1\right)D_{j-1}.\end{aligned}$$

A1.2.3 It is straightforward to see that the variance of a k-steps-ahead forecast is given by:

$$\mathrm{Var}\left[D_{j+k-1}|C_1, C_2, \ldots, C_{j-1}\right] = \phi\lambda_j\lambda_{j+1}\ldots\lambda_{j+k-1}\left(\lambda_j\lambda_{j+1}\ldots\lambda_{j+k-1}-1\right)D_{j-1}.$$

A1.3 *Normal Approximation to the Negative Binomial*

A1.3.1 From the Normal approximation to the negative binomial model in Section 2.5, we know that the variance of the 1-step-ahead forecast is given by:

$$\mathrm{Var}\left[D_j|C_1, C_2, \ldots, C_{j-1}\right] = \phi_j D_{j-1}.$$

A1.3.2 Then the variance of the 2-steps-ahead forecast, from Section A1.1, is given by:

$$\begin{aligned}\mathrm{Var}\left[D_{j+1}|C_1, C_2, \ldots, C_{j-1}\right] &= \mathrm{E}\left[\phi_{j+1}D_j|C_1, C_2, \ldots, C_{j-1}\right]+\lambda_{j+1}^2\phi_j D_{j-1}\\ &= \left(\phi_{j+1}\lambda_j+\lambda_{j+1}^2\phi_j\right)D_{j-1}\end{aligned}$$

and the variance of the 3-steps-ahead forecast is given by:

$$\begin{aligned}\mathrm{Var}\left[D_{j+2}|C_1, C_2, \ldots, C_{j-1}\right] &= \mathrm{E}\left[\phi_{j+2}\lambda_{j+1}D_j+\lambda_{j+2}^2\phi_{j+1}D_j|C_1, C_2, \ldots, C_{j-1}\right]\\ &\quad +\lambda_{j+2}^2\lambda_{j+1}^2\phi_j D_{j-1}\\ &= \left(\lambda_{j+1}\lambda_j\phi_{j+2}+\lambda_{j+2}^2\lambda_j\phi_{j+1}+\lambda_{j+2}^2\lambda_{j+1}^2\phi_j\right)D_{j-1}.\end{aligned}$$

A1.3.3 It is straightforward to continue this process to consider further steps ahead. We can recover the moments of the negative binomial model by substituting $\phi_k = \phi\lambda_k(\lambda_k - 1)$.

APPENDIX 2

RECURSIVE MODELS: ESTIMATION ERROR

A2.1 For the negative binomial model and its Normal approximation, the estimation error can also be calculated recursively. Again looking at a single row, and dropping the suffix i, we require $\mathrm{Var}[\hat{D}_n|D_{n-i+1}]$, which is the variance of the predicted row total, given the latest cumulative claims. Note that this is the same as the variance of the sum of incremental predicted values, which provides an alternative way of calculating the estimation error, but arrives at the same result:

$$\mathrm{Var}\left[\hat{D}_n|D_{n-i+1}\right] = \mathrm{Var}\left[\hat{\lambda}_{n-i+2}\ldots\hat{\lambda}_n D_{n-i+1}|D_{n-i+1}\right] = D_{n-i+1}^2\mathrm{Var}\left[\hat{\lambda}_{n-i+2}\ldots\hat{\lambda}_n|D_{n-i+1}\right].$$

A2.2 It can be seen that we require the variance of a product of estimates of development factors, which can be obtained recursively, multiplied by the square of the latest cumulative claims. To simplify the notation, the explicit conditioning notation is dropped in what follows.

A2.3 For row 2, the estimation variance is simply $D_{n-i+1}^2\mathrm{Var}[\hat{\lambda}_n]$.

A2.4 For row 3, we require $D_{n-i+1}^2\mathrm{Var}[\hat{\lambda}_{n-1}\hat{\lambda}_n]$, where:

$$\mathrm{Var}\left[\hat{\lambda}_{n-1}\hat{\lambda}_n\right] = \left(\mathrm{E}\left[\hat{\lambda}_{n-1}\right]\right)^2\mathrm{Var}\left[\hat{\lambda}_n\right] + \left(\mathrm{E}\left[\hat{\lambda}_n\right]\right)^2\mathrm{Var}\left[\hat{\lambda}_{n-1}\right] + \mathrm{Var}\left[\hat{\lambda}_n\right]\mathrm{Var}\left[\hat{\lambda}_{n-1}\right].$$

A2.5 Note that this assumes that the estimates of development factors are independent, or at least, uncorrelated (see Mack, 1994, Appendix G). In practice, it is straightforward to show that the development factors are uncorrelated under the model by calculating their covariance matrix, which has zero for all values except the leading diagonal. The variances of the estimates of the development factors can be obtained from the results of the model fitting, and we replace the expectations by their observed values, giving:

$$\mathrm{Var}\left[\hat{\lambda}_{n-1}\hat{\lambda}_n\right] \approx \left(\hat{\lambda}_{n-1}\right)^2\mathrm{Var}\left[\hat{\lambda}_n\right] + \left(\hat{\lambda}_n\right)^2\mathrm{Var}\left[\hat{\lambda}_{n-1}\right] + \mathrm{Var}\left[\hat{\lambda}_n\right]\mathrm{Var}\left[\hat{\lambda}_{n-1}\right]$$

A2.6 For the next row we use:

$$\begin{aligned}\mathrm{Var}\left[\hat{\lambda}_{n-2}\left(\hat{\lambda}_{n-1}\hat{\lambda}_n\right)\right] = {} & \left(\mathrm{E}\left[\hat{\lambda}_{n-2}\right]\right)^2\mathrm{Var}\left[\hat{\lambda}_{n-1}\hat{\lambda}_n\right] + \left(\mathrm{E}\left[\hat{\lambda}_{n-1}\hat{\lambda}_n\right]\right)^2\mathrm{Var}\left[\hat{\lambda}_{n-2}\right] \\ & + \mathrm{Var}\left[\hat{\lambda}_{n-1}\hat{\lambda}_n\right]\mathrm{Var}\left[\hat{\lambda}_{n-2}\right]\end{aligned}$$

We substitute for $\mathrm{Var}[\hat{\lambda}_{n-1}\hat{\lambda}_n]$ using the value calculated at the previous step to obtain the required variance, and this sets up the recursive procedure.

For any step, one additional development factor is required, and the variance of the product of two 'terms' is all that is ever required.

A2.7 Given the standard errors of the development factors, the calculations can be performed in a single line of a spreadsheet.

A2.8 For the overall reserve, the estimation variance is given by:

$$\mathrm{Var}\left[\hat{R}_{+}\right] \approx \sum_{i=2}^{n} \mathrm{Var}\left[\hat{D}_{in}\right] + 2 \sum_{\substack{i=2 \\ j>i}}^{n} \mathrm{Cov}\left[\hat{D}_{in}, \hat{D}_{jn}\right]$$

and we require the calculation of the covariance terms. Again this is straightforward, requiring the variances of the products of development factors considered above.

Consider $i = 2$, then:

$$\begin{aligned} \mathrm{Cov}\left[\hat{D}_{2n}, \hat{D}_{3n}\right] &= \mathrm{Cov}\left[D_{2,n-1}\hat{\lambda}_n, D_{3,n-2}\hat{\lambda}_{n-1}\hat{\lambda}_n\right] \\ &= D_{2,n-1}D_{3,n-2}\hat{\lambda}_{n-1}\mathrm{Var}\left[\hat{\lambda}_n\right] \end{aligned}$$

under independence of development factors. Similarly:

$$\begin{aligned} \mathrm{Cov}\left[\hat{D}_{2n}, \hat{D}_{4n}\right] &= \mathrm{Cov}\left[D_{2,n-1}\hat{\lambda}_n, D_{4,n-3}\hat{\lambda}_{n-2}\hat{\lambda}_{n-1}\hat{\lambda}_n\right] \\ &= D_{2,n-1}D_{4,n-3}\hat{\lambda}_{n-2}\hat{\lambda}_{n-1}\mathrm{Var}\left[\hat{\lambda}_n\right] \end{aligned}$$

and so on, considering the covariance of all rows > 2 with row 2.

Consider $i = 3$, then:

$$\begin{aligned} \mathrm{Cov}\left[\hat{D}_{3n}, \hat{D}_{4n}\right] &= \mathrm{Cov}\left[D_{3,n-2}\hat{\lambda}_{n-1}\hat{\lambda}_n, D_{4,n-3}\hat{\lambda}_{n-2}\hat{\lambda}_{n-1}\hat{\lambda}_n\right] \\ &= D_{3,n-2}D_{4,n-3}\hat{\lambda}_{n-2}\mathrm{Var}\left[\hat{\lambda}_{n-1}\hat{\lambda}_n\right]. \end{aligned}$$

We substitute for $\mathrm{Var}[\hat{\lambda}_{n-1}\hat{\lambda}_n]$ using the value calculated above, and continue in the same way for the covariance of all rows > 3 with row 3.

A2.9 This procedure continues up to row $n - 1$, where:

$$\begin{aligned} \mathrm{Cov}\left[\hat{D}_{n-1,n}, \hat{D}_{n,n}\right] &= \mathrm{Cov}\left[D_{n-1,2}\hat{\lambda}_3\hat{\lambda}_4 \ldots \hat{\lambda}_n, D_{n,1}\hat{\lambda}_2\hat{\lambda}_3 \ldots \hat{\lambda}_n\right] \\ &= D_{n-1,2}D_{n,1}\hat{\lambda}_2\mathrm{Var}\left[\hat{\lambda}_3\hat{\lambda}_4 \ldots \hat{\lambda}_n\right] \end{aligned}$$

each time substituting for the variance of the product of development factors calculated earlier.

APPENDIX 3

THE BOOTSTRAP PROCEDURE

A3.1 The bootstrap procedure is performed by completing the following steps, which can be performed without difficulty in a spreadsheet:

— Obtain the standard chain-ladder development factors from cumulative data.
— Obtain cumulative fitted values for the past triangle by backwards recursion, starting with the observed cumulative paid to date in the latest diagonal, using $\hat{D}_{i,n-i+1} = D_{i,n-i+1}$, and $\hat{D}_{i,k-1} = \hat{D}_{i,k}\lambda_k^{-1}$.
— Obtain incremental fitted values, $\hat{m}_{ij}$, for the past triangle by differencing.
— Calculate the unscaled Pearson residuals for the past triangle using:

$$r_{ij}^{(P)} = \frac{C_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}}}. \tag{A3.1}$$

— Calculate the Pearson scale parameter ϕ, where:

$$\phi = \frac{\sum_{i,\, jn-i+1} \left(r_{ij}^{(P)}\right)^2}{\frac{1}{2}n(n+1) - 2n + 1}$$

that is, the sum of the Pearson residuals squared divided by the degrees of freedom, where the degrees of freedom is the number of observations minus the number of parameters estimated.
— Adjust the Pearson residuals using:

$$r_{ij}^{adj} = \sqrt{\frac{n}{\frac{1}{2}n(n+1) - 2n + 1}} \times r_{ij}^{(P)}$$

to replicate the bias correction using an analytic approach.
— Begin iterative loop, to be repeated N times ($N = 1000$, say):
  — Resample the adjusted residuals with replacement, creating a new past triangle of residuals.
  — For each cell in the past triangle, solve equation A3.1 for C, giving a set of pseudo-incremental data for the past triangle.
  — Create the associated set of pseudo-cumulative data.
  — Fit the standard chain-ladder model to the pseudo-cumulative data.
  — Project to form a future triangle of cumulative payments.

- Obtain the corresponding future triangle of incremental payments by differencing, to be used as the mean when simulating from the process distribution.
- For each cell (i, j) in the future triangle, simulate a payment from the process distribution with mean $\tilde{m}_{ij}$ (obtained at the previous step), and variance $\phi\tilde{m}_{ij}$, using the value of ϕ calculated previously.
- Sum the simulated payments in the future triangle by origin year and overall to give the origin year and total reserve estimates respectively.
- Store the results, and return to start of iterative loop.

A3.2 The set of stored results forms the predictive distribution. The mean of the stored results should be compared to the standard chain-ladder reserve estimates to check for errors. The standard deviation of the stored results gives an estimate of the prediction error.

Printed by Bell & Bain Ltd., Glasgow