COMPARISON OF STOCHASTIC RESERVING METHODS

Jackie Li BCom, MFin, FIAA

Centre for Actuarial Studies, University of Melbourne

Email: k.li3@pgrad.unimelb.edu.au

ABSTRACT

This paper compares several stochastic reserving methods on both qualitative and quantitative aspects in dealing with the outstanding claims liabilities. These methods include Bayesian estimation with Markov chain Monte Carlo (MCMC) simulation, the chain ladder method with bootstrapping, generalised linear models (GLMs) with bootstrapping, the Kalman filter on state-space models, the Mack model, and the stochastic chain ladder method. To start with, the outline of this paper and different types of uncertainty are set forth in Sections 1 and 2. The notation and terminology are stated in Section 3. The strengths and limitations of the methods are examined by considering the underlying structures, assumptions, and estimation mechanics, in Sections 4 to 10. The application of each method is then tested on a particular claims data set in Section 11, similar to the analysis in Mack (1993a). Conclusions are presented in Section 12. This paper is an excerpt of the author's PhD thesis.

All the calculations were done on Excel spreadsheets with VBA (Visual Basic for Applications) coding and the software BUGS (Bayesian Inference Using Gibbs Sampling). When large amounts of simulation were carried out in the analysis, approximation formulae were used to provide reasonable checks. In addition, some proofs and derivations are stated in the appendices for reference purposes. Details regarding the use of VBA and BUGS codings can be provided upon request to the author.

Keywords: outstanding claims liability, uncertainty, risk margin, 75th percentile, coefficient of variation, model structure, Bayesian, MCMC, chain ladder, bootstrapping, GLM, Kalman filter, state-space model, Mack model, stochastic chain ladder

1 Outline

Many of the methods mentioned above have been applied and discussed widely in the academic literature over the past twenty years. With the recent legislative changes in Australia and the upcoming international accounting reform, there is a tendency to shift the focus from creating new methodologies to testing effectiveness of existing ones in practical situations. With regard to the new regulatory requirements, this paper aims to follow a more application-oriented approach in its analysis. England and Verrall (2002) provide a comprehensive review of numerous stochastic reserving techniques. A similar study has been conducted for this paper, but with the work expanded specifically to include elaborate comparison between different reserving methods in light of new legislation and practical application. The estimation procedures of various reserving methods for the expected value, the 75^{th} percentile, and the standard deviation of the outstanding claims liabilities, together with the application results, are listed and compared in detail.

This paper does not aim to offer a panacea for measuring uncertainty of the outstanding claims liabilities, or to provide an exhaustive list of reserving methods. By and large, the reserving methods considered are applied mechanically without exercising actuarial judgement. It is important to realise that each method is appropriate only under certain conditions, and that a thorough understanding of a variety of models and their disparities leads to more flexibility in handling different situations. It is also important to note that the methods are merely tools to help make decisions and cannot substitute for judgement in practice.

2 Types of Uncertainty

Under GPS 210, the outstanding claims liabilities must be valued at the 75th percentile. The risk margin is the difference between the 75th percentile and the mean, subject to a minimum of one half of the coefficient of variation when the risk margin is expressed as a percentage of the mean. An example is given in Appendix 1 to illustrate the necessity of the minimum in some cases where the probability distribution is highly skewed. In general, the purpose of setting up the risk margin is to provide a 'cushion' effect for the uncertainty beneath the measurement of the liabilities. Before embarking on assessing the risk margin, different types of uncertainty underlying the liability calculation need to be identified.

The following gives an account of different types of uncertainty inherent in the assessment of the outstanding claims liabilities. Hart et al (1996) and PS 300 provide similar lists of types of uncertainty.

Parameter error – Due to sampling variation of the past claims data, the exact values of the parameters of the claims run-off process are unknown. Thus, there is uncertainty in parameter estimation.

Process error (or random error) – Since the outstanding claims liabilities relate to the claims that are not paid by the valuation date, the exact amounts of the future payments are random and unknown.

Model specification error – The reserving methods adopted may not reasonably reflect the underlying claims development mechanism.

Data error – There may be mistakes or omissions in the data provided.

Future trends variability – The future trends of various aspects such as parameter evolution over time, inflation, interest rate, exchange rate, claims run-off patterns, claims

management process, reinsurance arrangements, policy conditions, exposure, business mix, staff departure, legislation, insurance market cycle, and technology may change over time. The past does not necessarily predict the future, and judgement about future trends may be inappropriate.

Reinsurance risk – There is a probability of default that the reinsured portion of a large claim cannot be fully recovered due to insufficient capital held by the reinsurer.

This paper focuses on measuring parameter error and process error. Furthermore, consideration is given to model specification error when applying different reserving methods to the same set of claims data.

Another dimension to differentiate the types of uncertainty is to identify systematic and non-systematic risks. GN 353 and Houltram (2003) state some definitions of the two types of risk. When the risk margin is expressed as a percentage of the mean of the outstanding claims liabilities, systematic risk is the portion of the risk margin percentage that does not decrease as the portfolio size or the number of lines of business increases, and non-systematic risk is the portion that decreases as the portfolio size or the number of lines increases.

The types of uncertainty mentioned previously may be subdivided into systematic and non-systematic risks. For example, for a larger portfolio with more past claims data, there is less parameter error as there are more data for estimation (i.e. non-systematic risk), but a certain part of model specification error may not reduce because no reserving method can completely model the claims development of a portfolio (i.e. systematic risk). To cite another example, as the number of lines increases, provided these lines are less than perfectly correlated and have similar volatility levels, part of process error decreases due to increased diversification benefits between the different lines (i.e. non-systematic risk), but part of the process error remains unchanged (i.e. systematic risk).

This paper does not adopt the concept of systematic and non-systematic risks.

3 Notation and Terminology

This section introduces some notation and terminology used throughout the paper.

3.1 Common Notation

It is assumed that all claims will be settled in n periods of time and there are n periods of past claims data. Typically the unit of time is one year, but it can also be say one quarter or one month.

Let $X_{i,j}$, $C_{i,j}$, R_i, and R be random variables (for $1 \le i$, $j \le n$), where $X_{i,j}$ represents the incremental claim amount of accident period i and development period j , $C_{i,j}$ represents the cumulative claim amount of accident period i and development period j , R_i represents the outstanding claims liability of accident period i , and R represents the total outstanding claims liability. Accordingly, the relationships of the variables are $C_{i,j} = X_{i,1} + X_{i,2} + ... + X_{i,j}$, $R_i = X_{i,n+2-i} + X_{i,n+3-i} + ... + X_{i,n} = C_{i,n} - C_{i,n+1-i}$, and $R = R_2 + R_3 + ... + R_n$.

For example, when n = 10, the claims run-off pattern is shown below:

i \ j	1	2	3	4	5	6	7	8	9	10
1	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$	$X_{1,9}$	$X_{1,10}$
2	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$	$X_{2,8}$	$X_{2,9}$	
3	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	$X_{3,6}$	$X_{3,7}$	$X_{3,8}$		
4	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,6}$	$X_{4,7}$			
5	$X_{5,1}$	$X_{5,2}$	$X_{5,3}$	$X_{5,4}$	$X_{5,5}$	$X_{5,6}$				
6	$X_{6,1}$	$X_{6,2}$	$X_{6,3}$	$X_{6,4}$	$X_{6,5}$					
7	$X_{7,1}$	$X_{7,2}$	$X_{7,3}$	$X_{7,4}$						
8	$X_{8,1}$	$X_{8,2}$	$X_{8,3}$							
9	$X_{9,1}$	$X_{9,2}$								
10	$X_{10,1}$									

The $X_{i,j}$'s shown in the upper left triangle of cells (for $i + j \leq n + 1 = 11$) represent the past claims data. The missing lower right triangle (for $i + j > n + 1 = 11$) relates to the outstanding claims liability to be assessed.

We use aggregated claim payments as $X_{i,j}$'s in this paper. Claim counts, incurred claim amounts, and individual claim amounts are not dealt with.

3.2 75th Percentile and Standard Deviation

The reserving methods under consideration provide different routes to estimate the 75th percentile and the standard deviation required for the coefficient of variation. Under some reserving methods, the 'standard error of prediction' will be estimated instead of the simple standard deviation, in order to allow for both parameter error and process error explicitly. The standard error of prediction is the square root of the 'mean square error of prediction'. Verrall (1999) and England and Verrall (2002) state the definitions of the mean square error of prediction and the standard error of prediction.

The mean square error of prediction of R_i is defined for $2 \leq i \leq n$ as:

$$MSEP_i = \mathrm{E}\left[\left(R_i - \hat{R}_i\right)^2\right] \cong \mathrm{Var}\left(R_i\right) + \mathrm{Var}\left(\hat{R}_i\right), \qquad (1)$$

where $\hat{R}_i$ is the estimator of R_i.

Being a measure of the discrepancy between the future outcome and its estimator, the mean square error of prediction provides for both parameter error and process error. In addition, since R_i is related to the future claims and $\hat{R}_i$ is an estimator based on the past claims data, the approximate equality of the relationship on the right hand side of (1) is an exact equality under some reserving methods with the assumptions that the cells in the upper left triangle are independent of the cells in the lower right triangle and $\hat{R}_i$ is unbiased.

We prove the exact equality as follows:

$$MSEP_i = \mathrm{E}\left[\left(R_i - \hat{R}_i\right)^2\right] = \mathrm{E}\left[\left(R_i - \mathrm{E}(R_i) - \hat{R}_i + \mathrm{E}(R_i)\right)^2\right]$$

$$= \mathrm{E}\left[(R_i - \mathrm{E}(R_i))^2\right] - 2\mathrm{E}\left[(R_i - \mathrm{E}(R_i))(\hat{R}_i - \mathrm{E}(R_i))\right] + \mathrm{E}\left[\left(\hat{R}_i - \mathrm{E}(R_i)\right)^2\right]$$

$$= \mathrm{Var}(R_i) - 2\mathrm{E}(R_i - \mathrm{E}(R_i))\mathrm{E}\left(\hat{R}_i - \mathrm{E}(R_i)\right) + \mathrm{Var}\left(\hat{R}_i\right)$$

($\because$ R_i and $\hat{R}_i$ are independent and $\hat{R}_i$ is unbiased)

$$= \mathrm{Var}(R_i) + \mathrm{Var}\left(\hat{R}_i\right)$$

In practice, the relationship above is an approximation when the conditions are not met. The first component of the relationship allows for process error and the second component allows for parameter error.

Likewise, the mean square error of prediction of *R* is defined as:

$$MSEP = \mathrm{E}\left[\left(R - \hat{R}\right)^2\right] \cong \mathrm{Var}(R) + \mathrm{Var}\left(\hat{R}\right), \qquad (2)$$

where $\hat{R}$ is the estimator of *R*.

Similarly, under some reserving methods, the 75th percentiles of $R_i - \hat{R}_i$ and $R - \hat{R}$ will be estimated instead of the 75th percentiles of R_i and *R*, in order to allow for both parameter error and process error explicitly.

3.3 Terminology

The terms ‘claims development’ and ‘claims run-off’ are used interchangeably. The term ‘reserving’ generally refers to ‘measuring the outstanding claims liability’ or ‘setting up a reserve for the outstanding claims liability’. ‘OCL’ is used as an abbreviation for the ‘outstanding claims liability’.

4 Bayesian Estimation with Markov Chain Monte Carlo Simulation

Bayesian estimation is a technique to derive a probability distribution for a random variable given certain past data. The derivation procedure is capable of incorporating expert, judgemental, or prior information in a formal manner. Parameter error and process error are automatically allowed for in the derivation. Verrall (1990) provides an insight into applying the technique in general insurance. In this paper, we adopt Gibbs sampling, a type of MCMC simulation, to carry out Bayesian analysis.

There are some recent papers demonstrating the application of MCMC simulation. Scollnik (2001) provides a clear presentation on adopting the software called BUGS (Bayesian Inference Using Gibbs Sampling) to carry out MCMC simulation in actuarial modelling. Ntzoufras and Dellaportas (2002) present four Bayesian models (including lognormal and state-space models) exploiting MCMC simulation to predict the OCL of a portfolio of motor claims.

4.1 Bayesian Estimation

Bayesian estimation requires pre-specifying two types of probability distribution – process distribution and prior distribution. The probability distributions assumed for $X_{i,j}$ are the process distributions, which allow for process error. Since certain parameters of the process distributions are unknown and need to be estimated, they are conceptually assumed to be random variables following other probability distributions, namely the prior distributions. The prior distributions, which allow for parameter error, offer a direct and formal way to incorporate objective prior information.

Given the past claims data, the Bayesian mechanism generates two types of probability distribution – posterior distributions for the parameters treated as random variables and predictive distributions for $X_{i,j}$ (for $i + j > n + 1$). Both parameter error and process error

are automatically incorporated in the predictive distributions. Inferences are then drawn from these two types of distribution.

Considering a simple case with only one parameter treated as a random variable (and so only one prior distribution) for a process distribution, the rationale of the Bayesian mechanism is shown below.

The posterior distribution is derived from the process distribution, prior distribution, and past claims data. The posterior density is proportional to the likelihood function multiplied by the prior density:

$$f_{\text{posterior}}\left(\theta \middle| X_{g,h:g+h\le n+1}\right) \propto L\left(\theta \middle| X_{g,h:g+h\le n+1}\right) \times f_{\text{prior}}\left(\theta\right) ,$$

i.e. $$f_{\text{posterior}}\left(\theta \middle| X_{g,h:g+h\le n+1}\right) \propto \prod_{i+j\le n+1} f_{\text{process}}\left(x_{i,j} \middle| \theta\right) \times f_{\text{prior}}\left(\theta\right) ,$$

where f denotes a density function, and θ, assumed to be a random variable, is the unknown parameter of the process distribution, and $x_{i,j}$ is the observed value of $X_{i,j}$.

The predictive density is derived by integrating the product of the process density and posterior density over the unknown parameter for $i + j > n + 1$:

$$f_{\text{predictive}}\left(x_{i,j} \middle| X_{g,h:g+h\le n+1}\right) = \int f_{\text{process}}\left(x_{i,j} \middle| \theta\right) f_{\text{posterior}}\left(\theta \middle| X_{g,h:g+h\le n+1}\right) d\theta .$$

The mechanism shown above is mathematically straightforward and the prior distribution is useful for taking prior information into account. Nevertheless, actual implementation can be cumbersome when there is more than one parameter treated as a random variable and so more than one prior distribution, or when the combination of the process distribution and prior distribution is complicated. MCMC simulation can be used to alleviate this problem, which is described below.

4.2 Gibbs Sampling

A statistical MCMC simulation method called Gibbs sampling is applied in this paper. Theoretically, the 'full conditional posterior distribution' – the probability distribution of a parameter given all the other parameters and the past claims data – is obtained for each parameter, and then samples are simulated from a Markov chain that has its stationary distribution equal to the joint probability distribution of the parameters and $X_{i,j}$. Under weak regularity conditions as stated in Geman and Geman (1984) and when the number of iterations tends to infinity, the samples are simulated as if they were simulated from the true joint probability distribution.

We make use of the general process of Gibbs sampling described in Scollnik (2001). To start Gibbs sampling on the OCL assessment, the full conditional posterior density is derived for each parameter as follows:

$$f_{\text{full conditional posterior } \theta_s}\left(\theta_s \middle| \theta_{t:t\neq s}, X_{g,h:g+h\leq n+1}\right)$$

$$\propto \left[\prod_{i+j\leq n+1} f_{X_{i,j}(\theta_s)}\left(x_{i,j}(\theta_s)\middle| p\left(X_{i,j}(\theta_s)\right)\right)\right] f_{\theta_s}\left(\theta_s \middle| p(\theta_s)\right),$$

where θ_s (s = 1, 2, … , k ; k = number of parameters treated as random variables) is one of the parameters, $X_{i,j}(\theta_s)$ is the incremental claim amount of accident period i and development period j having a process distribution with θ_s as its parameter or one of its parameters, and $p(Y)$ represents all the parameters of the probability distribution of a random variable Y.

Given the corresponding parameters, $X_{i,j}$'s are independent random variables over all values of i and j , i.e. they are conditionally independent. Conditional independence also applies to θ_s's for all values of s.

The determination of the initial value of each θ_s can be based on certain prior information. Starting with the initial values, samples of each θ_s are simulated from its full conditional posterior density as derived above:

a sample of $\theta_1^{(1)}$ is drawn from $f_{\text{full conditional posterior } \theta_1}\left(\theta_1 \middle| \theta_2^{(0)}, \theta_3^{(0)}, \theta_4^{(0)}, ..., X_{g,h:g+h\le n+1}\right)$,

a sample of $\theta_2^{(1)}$ is drawn from $f_{\text{full conditional posterior } \theta_2}\left(\theta_2 \middle| \theta_1^{(1)}, \theta_3^{(0)}, \theta_4^{(0)}, ..., X_{g,h:g+h\le n+1}\right)$,

a sample of $\theta_3^{(1)}$ is drawn from $f_{\text{full conditional posterior } \theta_3}\left(\theta_3 \middle| \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)}, ..., X_{g,h:g+h\le n+1}\right)$,

and so on, where the superscript (m) indicates the m$^{\text{th}}$ iteration, e.g. $m = 0$ for the initial values and $m = 1$ for the first iteration.

Each iteration generates a sample for each θ_s. The simulation process is repeated until the simulated samples of each θ_s show effective convergence. To assess effective convergence, the simulation process above is carried out with different sets of initial values. Effective convergence is achieved when the probability inferences of the simulated samples of each θ_s do not depend on the initial values. The different sets of initial values should be adequately dispersed to ensure the assessment is meaningful.

One way to identify effective convergence more formally is to use the modified Gelman-Rubin convergence diagnostic, as stated in Scollnik (2001). For a parameter θ_s, the diagnostic is the ratio of the x% interval of the pooled samples (pooling the simulated samples from all sets of initial values) to the average x% interval (averaging over different simulated samples from different sets of initial values). This ratio tends to one as the simulated samples of θ_s effectively converge.

The simulation process is repeated further for say 100,000 times after reaching effective convergence. For each iteration, samples of $X_{i,j}$ (for $i + j > n + 1$) are then simulated from the process density $f_{X_{i,j}}\left(x_{i,j} \middle| \theta_1^{(m)}, \theta_2^{(m)}, \theta_3^{(m)}, ..., \theta_k^{(m)}\right)$, where m is larger than the number of iterations for reaching effective convergence.

4.3 75th Percentile and Standard Deviation

By summing up the relevant simulated $X_{i,j}$'s (for $i + j > n + 1$), 100,000 samples of R_i ($2 \leq i \leq n$) and R are formed. Since both parameter error and process error are automatically incorporated, we use the sample 75th percentile and the sample standard deviation as the estimates of the 75th percentile and the standard deviation of the OCL.

In addition, the sample mean is used as an estimate of the expected OCL.

4.4 Software

The software BUGS is exploited in this paper. It is specialised software for carrying out MCMC simulation using Gibbs sampling. It has both a graphical interface for simple modelling and a programming interface for more sophisticated modelling. It is convenient to use and contains several built-in functions, such as checking the specified model, generating the initial values, calculating various statistics for analysis and validation of the model chosen, and plotting different types of graphs. (In particular, BUGS uses an 80% interval for the modified Gelman-Rubin convergence diagnostic.) It also offers great flexibility in handling the results, e.g. by selecting say every mth iteration of the simulated samples to contribute to the required statistics, the autocorrelations within the samples are reduced.

Scollnik (2001) gives a detailed description of how to adopt the software in general actuarial modelling of claims development.

5 Chain Ladder Method with Bootstrapping

The chain ladder method is the most widely applied reserving method. Taylor (2000) provides a description of the standard chain ladder method. While it is a deterministic and mechanical method, some authors have tried to link it with different stochastic models, such as Mack (1993b) and Verrall (1999, 2000). Considerable arguments have been made on which stochastic model is truly underlying the chain ladder structure.

The chain ladder method can be amalgamated with bootstrapping. The development factors calculated from applying the chain ladder method to the past claims data are used to project the cumulative claim amounts in the last calendar period back to development period zero. This backward calculation produces a 'mean' for the past claims data, and the residuals are computed from the difference between this mean and the past claims data. The residuals are then used in the resampling process of bootstrapping. Thereafter, pseudo past data samples, bootstrap forecast samples, and pseudo future data samples are simulated. This process is repeated many times to obtain the required empirical distribution. Some bootstrapping examples are stated in Verrall (1999), Pinheiro et al (2000), Taylor (2000), Bateup and Reed (2001) and Houltram (2003). The ways to apply this combination are many and varied, and the approach in this paper is stated in the following subsections.

5.1 Chain Ladder Method

Taylor (2000) provides a detailed discussion about the chain ladder method. First, the estimated development factor from development period j to $j + 1$ for $1 \le j \le n - 1$ is:

$$\hat{f}_j = \frac{\sum_{r=1}^{n-j} C_{r,j+1}}{\sum_{r=1}^{n-j} C_{r,j}} \quad , \tag{3}$$

and using the estimated development factors, the estimators of $\mathrm{E}(C_{i,j})$ and $\mathrm{E}(X_{i,j})$ are:

$\hat{C}_{i,j} = C_{i,n+1-i}\,\hat{f}_{n+1-i}\,\hat{f}_{n+2-i}\cdots\hat{f}_{j-1}$ (for $i + j > n + 1$), $\hat{C}_{i,j} = C_{i,j}$ (for $i + j = n + 1$),

$$\hat{C}_{i,j} = \frac{C_{i,n+1-i}}{\hat{f}_j \hat{f}_{j+1} \cdots \hat{f}_{n-i}} \text{ (for } i + j < n + 1) \text{ , } \hat{X}_{i,j} = \hat{C}_{i,j} - \hat{C}_{i,j-1} \text{ (for } 2 \le j \le n) \text{ ,}$$

and $\hat{X}_{i,1} = \hat{C}_{i,1}$. (4)

Accordingly, the estimators of $\mathrm{E}(R_i)$ (for $2 \le i \le n$) and $\mathrm{E}(R)$ are:

$$\hat{R}_i = \hat{C}_{i,n} - C_{i,n+1-i} = C_{i,n+1-i} \hat{f}_{n+1-i} \hat{f}_{n+2-i} \cdots \hat{f}_{n-1} - C_{i,n+1-i} \text{ , and} \quad (5)$$

$$\hat{R} = \hat{R}_2 + \hat{R}_3 + \ldots + \hat{R}_n \text{ ,} \quad (6)$$

which produce the estimates of the expected OCL.

(As stated in Taylor (2000), it can be assumed that $\hat{f}_j = 1 + ab^{j-r}$ for $j > r$ with a and b being the parameters and with r being the development period beyond which the ratios are smoothed. This type of smoothing is useful for achieving parameter parsimony and for extrapolation beyond the development period n of the past claims data.)

5.2 Residuals

With the $\hat{X}_{i,j}$'s calculated above, the estimated standard deviation of $\left(X_{i,j} - \hat{X}_{i,j}\right)$ for development period j is calculated as follows:

$$\hat{\sigma}_j = \sqrt{\frac{\sum_{r=1}^{n+1-j} \left(X_{r,j} - \hat{X}_{r,j}\right)^2}{n-j}} \text{ (for } 1 \le j \le n-1) \text{ and } \hat{\sigma}_n = \min\left(\frac{\hat{\sigma}_{n-1}^2}{\hat{\sigma}_{n-2}}, \hat{\sigma}_{n-2}\right) . \quad (7)$$

In practice, one may smooth the $\hat{\sigma}_j$'s for higher values of j to obtain more stable results. Consequently, the residual for each cell (i , j) in the run-off triangle is, for $i + j \le n + 1$:

$$residual_{i,j} = \frac{X_{i,j} - \hat{X}_{i,j}}{\hat{\sigma}_j} \text{ .} \quad (8)$$

Effectively, these residuals are adjusted according to the variability of $\left(X_{i,j} - \hat{X}_{i,j}\right)$ for each development period. The approach we adopt here is similar to that in Houltram

(2003). Ignoring the zero residuals for the cells (1 , n) and (n , 1), there are in total $\left(n(n+1)/2-2\right)$ residuals for resampling. Moreover, to justify the bootstrapping process, the residuals are assumed to be independently and identically distributed random variables. Other forms of residuals may also be used in some situations, such as those stated in Subsection 6.2.

5.3 Resampling Process of Bootstrapping

We exploit the bootstrapping approaches stated in Pinheiro et al (2000). The residuals calculated in the previous subsection are resampled with replacement. Pseudo past data samples of $X_{i,j}^{\text{pseudo past}}$ are then formed for $i + j \le n + 1$ by:

$$X_{i,j}^{\text{pseudo past}} = residual_{i,j}^{\text{resampled}}\hat{\sigma}_j + \hat{X}_{i,j} \ . \qquad (9)$$

Pseudo past data samples of $C_{i,j}^{\text{pseudo past}}$ (for $i + j \le n + 1$) are then formed by adding up the relevant $X_{i,j}^{\text{pseudo past}}$'s as in Subsection 3.1. Applying (3), (5), and then (6) to these samples leads to the estimates of the expected OCL of the pseudo past data samples of $C_{i,j}^{\text{pseudo past}}$. The estimates become bootstrap forecast samples of $\hat{R}_i^{\text{bootstrap}}$ (for $2 \le i \le n$) and $\hat{R}^{\text{bootstrap}}$, which allows for parameter error.

Furthermore, the residuals are resampled with replacement again to form pseudo future data samples of $X_{i,j}^{\text{pseudo future}}$ for $i + j > n + 1$ by:

$$X_{i,j}^{\text{pseudo future}} = residual_{i,j}^{\text{resampled}}\hat{\sigma}_j + \hat{X}_{i,j} \ . \qquad (10)$$

Pseudo future data samples of $R_i^{\text{pseudo future}}$ (for $2 \le i \le n$) and $R^{\text{pseudo future}}$ are then formed by summing up the relevant $X_{i,j}^{\text{pseudo future}}$'s as in Subsection 3.1. This computation allows for process error.

The resampling process is repeated for say 1,000 times, forming 1,000 samples of $\left(R_i^{\text{pseudo future}} - \hat{R}_i^{\text{bootstrap}}\right)$, $\left(R^{\text{pseudo future}} - \hat{R}^{\text{bootstrap}}\right)$, $\left(R_i^{\text{pseudo future}} - \hat{R}_i^{\text{bootstrap}}\right)^2$, and $\left(R^{\text{pseudo future}} - \hat{R}^{\text{bootstrap}}\right)^2$. Inferences are then drawn from these samples, as discussed in the next subsection.

5.4 75th Percentile and Standard Error of Prediction

Since both parameter error and process error are explicitly allowed for, we add the sample 75th percentiles of $\left(R_i^{\text{pseudo future}} - \hat{R}_i^{\text{bootstrap}}\right)$ and $\left(R^{\text{pseudo future}} - \hat{R}^{\text{bootstrap}}\right)$ to $\hat{R}_i$ and $\hat{R}$ respectively to estimate the 75th percentile of the OCL. Similarly, we use the square root of the sample means of $\left(R_i^{\text{pseudo future}} - \hat{R}_i^{\text{bootstrap}}\right)^2$ and $\left(R^{\text{pseudo future}} - \hat{R}^{\text{bootstrap}}\right)^2$ to estimate the standard error of prediction of the OCL.

Verrall (1999) argues that in the bootstrapping process the residuals are resampled without considering the number of parameters estimated. His paper hence suggests an adjustment for its bootstrapping results. This adjustment may be considered as a process of studentising the residuals (refer to Subsection 6.2). A similar adjustment is made here: the sample 75th percentiles of $\left(R_i^{\text{pseudo future}} - \hat{R}_i^{\text{bootstrap}}\right)$ and $\left(R^{\text{pseudo future}} - \hat{R}^{\text{bootstrap}}\right)$, together with the square root of the sample means of $\left(R_i^{\text{pseudo future}} - \hat{R}_i^{\text{bootstrap}}\right)^2$ and $\left(R^{\text{pseudo future}} - \hat{R}^{\text{bootstrap}}\right)^2$, are multiplied by a factor $\sqrt{n_d/(n_d - n_p)}$, in which n_d is the number of cells or data points in the upper left run-off triangle and n_p is the number of parameters to be estimated. For example, when n = 10, the factor is equal to $\sqrt{55/36}$ as there are nineteen parameters in total estimated by using (3) and (7). One final note is that in general the sample mean of $R^{\text{bootstrap}}$ is not significantly different from $\hat{R}$. Otherwise, it becomes necessary to adjust the uncertainty estimates with the ratio of the two figures.

6 Generalised Linear Models with Bootstrapping

Generalised linear models (GLMs) are exploited mainly to perform regression to identify underlying relationships between different variables of interest. They offer a rigorous, sophisticated, and flexible framework. Not only do they help explore the data more efficiently, but also they can be validated in a formal way.

GLMs comprise a set of statistical models that share a number of analogous properties. Essentially, a linear relationship is assumed to exist between certain variables of interest or their transformations. There is also a choice of probability distribution, from the exponential family, for the uncertainty involved. Maximum likelihood is then used in parameter estimation. After selecting the model structure – the linear relationship and the probability distribution – and obtaining the parameter estimates based on the past data, future unobserved values are predicted accordingly. McCullagh and Nelder (1989) provide a detailed description of various technical aspects of GLMs. A few examples of applying GLMs in general insurance are Renshaw (1989), Verrall (1999), and Taylor and McGuire (2004).

Mixing GLMs with bootstrapping is similar to combining the chain ladder method with bootstrapping. The residuals obtained from applying a GLM to the past claims data are used in the resampling process of bootstrapping. Thereafter, an empirical distribution is formed, from which the required inferences are drawn.

6.1 GLM Framework

The basic framework of the GLMs used in this paper is based on McCullagh and Nelder (1989) and Verrall (1999). The linear relationship and the probability distribution stated below constitute the basic assumptions of the GLM framework.

The linear predictor $\eta_{i,j}$ is expressed as $\eta_{i,j} = \nu + \alpha_i + \beta_j + \gamma_{i+j-1} = h\left(\mu_{i,j}\right)$. The parameter ν represents the overall average amount of $h\left(\mu_{i,j}\right)$. The parameters α_i, β_j, and γ_{i+j-1} accommodate the accident period effect, the development period effect, and the calendar period effect. The function h is called the link function, and $\mu_{i,j} = \mathrm{E}\left(X_{i,j}\right)$. Some examples of h are $\mu_{i,j}$ (identity), $\ln\left(\mu_{i,j}\right)$ (log), $\ln\left(\mu_{i,j}/\left(1-\mu_{i,j}\right)\right)$ (logit), $\Phi^{-1}\left(\mu_{i,j}\right)$ (probit), and $\ln\left(-\ln\left(1-\mu_{i,j}\right)\right)$ (complementary log-log). Effectively, the linear relationship is between the linear predictor and the parameters.

The density function of $X_{i,j}$ is (from the exponential family):

$$f_{X_{i,j}}\left(x_{i,j}\right) = \exp\left(\frac{x_{i,j}\theta\left(\mu_{i,j}\right) - b\left(\theta\left(\mu_{i,j}\right)\right)}{a(\phi)} + c\left(x_{i,j}, \phi\right)\right),$$

where a, b, c, and θ are functions depending on the probability distribution assumed, ϕ is the dispersion parameter, $X_{i,j}$ and $X_{g,h}$ are independent for $(i, j) \neq (g, h)$, and $x_{i,j}$ is an observation of $X_{i,j}$.

The disparities of a, b, c, θ, ϕ, and other characteristics between different probability distributions, including normal, Poisson, gamma, and inverse Gaussian, are shown in Appendix 2. The subscript (i, j) of $X_{i,j}$ is hidden in this appendix for convenience.

The linear relationship (i.e. the parameters ν, α_i, β_j, and γ_{i+j-1}, for $i + j \leq n + 1$) and the dispersion parameter (i.e. ϕ) are estimated by applying maximum likelihood estimation to the past claims data, based on the selected probability distribution. We adopt the iteration procedure described in McCullagh and Nelder (1989) to carry out maximum likelihood estimation. The parameter estimates are then used to find the inverse of the link function for all i and j:

$$\hat{X}_{i,j} = h^{-1}\left(\hat{\eta}_{i,j}\right) = h^{-1}\left(\hat{\nu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{i+j-1}\right), \qquad (11)$$

where $\hat{\nu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$, $\hat{\gamma}_{i+j-1}$, $\hat{\eta}_{i,j}$, and $\hat{X}_{i,j}$ are the estimators of ν, α_i, β_j, γ_{i+j-1}, $\eta_{i,j}$, and $\mu_{i,j}$ (the significance of the parameters can be tested as stated in Appendix 3).

For $i + j > n + 1$, $\hat{\gamma}_{i+j-1}$ cannot be estimated from the past claims data, as the parameter is related to future calendar period effect. It may then be assumed that $\gamma_{i+j-1} = \gamma_n$ or $\gamma_{i+j-1} = \frac{1}{n}\sum_{r=1}^{n}\gamma_r$, or actuarial judgement may be made regarding the future trends of the calendar period effect. As argued in England and Verrall (2002), if the past claims inflation can be readily computed and the future claims inflation can be reasonably determined, it is more desirable to remove the calendar period effect before modelling.

It is important to avert over-parameterisation in order to obtain valid solutions from maximum likelihood estimation. To exemplify, for n = 10, the maximum number of parameters to be estimated (excluding the dispersion parameter) is $2n-1=19$. On the other hand, parsimony of parameter use can be achieved by adjusting the linear predictor. For example, as noted in Taylor (2000) and England and Verrall (2002), the Hoerl curve can be applied as $\ln\left(\mu_{i,j}\right) = \alpha_i + \beta\ln\left(j\right) + \lambda j$ with β and λ being the only parameters for the development period effect and with j treated as a covariate, for certain development periods. (These arguments apply similarly to Bayesian estimation with MCMC simulation.)

Accordingly, the estimators of $\mathrm{E}\left(R_i\right)$ (for $2 \le i \le n$) and $\mathrm{E}\left(R\right)$ are:

$$\hat{R}_i = \hat{X}_{i,n+2-i} + \hat{X}_{i,n+3-i} + ... + \hat{X}_{i,n} \text{ , and} \qquad (12)$$

$$\hat{R} = \hat{R}_2 + \hat{R}_3 + ... + \hat{R}_n \text{ ,} \qquad (13)$$

which produce the estimates of the expected OCL.

In the GLM estimation, while the probability distributions of $X_{i,j}$ and $\hat{X}_{i,j}$ (for identity link) can be determined approximately (refer to Appendix 3 for details), the probability

distributions of R_i , R , $\hat{R}_i$, and $\hat{R}$ cannot be estimated readily as the corresponding convolution is intricate. Under this situation, bootstrapping is expedient in computing the desired empirical distribution. The next subsection exhibits the different types of residuals that can be input into the bootstrapping process.

6.2 Residuals

Making use of the $\hat{X}_{i,j}$'s calculated previously, according to McCullagh and Nelder (1989), there are three types of residuals for each cell (i , j) in the upper left run-off triangle defined for $i + j \leq n + 1$ as:

$$residual_{i,j}^{\text{Pearson}} = \frac{X_{i,j} - \hat{X}_{i,j}}{\sqrt{b''\left(\theta\left(\hat{X}_{i,j}\right)\right)}} ,$$

$$residual_{i,j}^{\text{Deviance}} = Sign(X_{i,j} - \hat{X}_{i,j})\sqrt{2\left(X_{i,j}\left(\theta\left(X_{i,j}\right) - \theta\left(\hat{X}_{i,j}\right)\right) - b\left(\theta\left(X_{i,j}\right)\right) + b\left(\theta\left(\hat{X}_{i,j}\right)\right)\right)} ,$$

and $$residual_{i,j}^{\text{Anscombe}} = \frac{A\left(X_{i,j}\right) - A\left(\hat{X}_{i,j}\right)}{A'\left(\hat{X}_{i,j}\right)\sqrt{b''\left(\theta\left(\hat{X}_{i,j}\right)\right)}} , \qquad (14)$$

where $b''(\theta)$ is the second derivative of $b(\theta)$ with respect to θ, $A = \int \left(b''\left(\theta(t)\right)\right)^{-1/3} dt$ (ignoring the constant term), and A' is the first derivative of A with respect to t.

Pearson residuals and deviance residuals become 'studentised standardised' residuals if divided by $\sqrt{1 - s_r}$ with s_r being the r$^{\text{th}}$ diagonal component of the 'hat matrix', and by $\sqrt{\hat{\phi}}$ with $\hat{\phi}$ being the estimator of ϕ (refer to Appendix 3 for details). In addition, weights varying across different cells of the run-off triangle can be incorporated into the model structure and hence into the residual formulae (by multiplying the residual of a cell with the square root of the weight of that cell). Accommodating varying weights is useful when it is believed that some data points carry more credibility than the others and when there is heteroscedasticity between different cells.

In general, deviance and Anscombe residuals are closer to being normally distributed than Pearson residuals. Pearson residuals are normally distributed only when the probability distribution of $X_{i,j}$ is normal. The distribution of studentised standardised residuals tends to standard normal.

6.3 Resampling Process of Bootstrapping

Again, we adopt the bootstrapping approaches stated in Pinheiro et al (2000). The residuals computed in the previous subsection are resampled with replacement. Using the resampled residuals and the $\hat{X}_{i,j}$'s calculated, pseudo past data samples of $X_{i,j}^{\text{pseudo past}}$ (for $i + j \leq n + 1$) are formed by using the inverse of (14). Maximum likelihood, (11), (12), and (13) are then applied to these pseudo past data samples of $X_{i,j}^{\text{pseudo past}}$. This computation forms bootstrap forecast samples of $\hat{R}_i^{\text{bootstrap}}$ (for $2 \leq i \leq n$) and $\hat{R}^{\text{bootstrap}}$, which allows for parameter error.

The residuals are then resampled with replacement again. Using the resampled residuals and the $\hat{X}_{i,j}$'s calculated, pseudo future data samples of $X_{i,j}^{\text{pseudo future}}$ (for $i + j > n + 1$) are formed by using the inverse of (14). Thereafter, pseudo future data samples of $R_i^{\text{pseudo future}}$ (for $2 \leq i \leq n$) and $R^{\text{pseudo future}}$ are then formed by adding up the relevant $X_{i,j}^{\text{pseudo future}}$'s as in Subsection 3.1. This computation allows for process error.

The resampling process is repeated for say 1,000 times, forming 1,000 samples of $\left(R_i^{\text{pseudo future}} - \hat{R}_i^{\text{bootstrap}}\right)$, $\left(R^{\text{pseudo future}} - \hat{R}^{\text{bootstrap}}\right)$, $\left(R_i^{\text{pseudo future}} - \hat{R}_i^{\text{bootstrap}}\right)^2$, and $\left(R^{\text{pseudo future}} - \hat{R}^{\text{bootstrap}}\right)^2$. Inferences are then drawn from these samples.

6.4 75th Percentile and Standard Error of Prediction

The 75th percentile and the standard error of prediction of the OCL are estimated in the same way as described in Subsection 5.4. The adjustment of the factor $\sqrt{n_d / (n_d - n_p)}$ also applies here, where n_p is the number of parameters to be estimated (including ν, α_i, β_j, and γ_{i+j-1}).

6.5 Alternative Approach

As mentioned in England and Verrall (2002), there is an alternative means to assess the OCL apart from bootstrapping. Provided the parameters and the covariance matrix of the parameter estimators are obtained from maximum likelihood estimation, and assuming the parameter estimators are approximately multinormal, samples can then be simulated from the multinormal distribution. This computation provides for parameter error. Accordingly, these samples are used as the parameters for the model structure, and samples of $X_{i,j}$ (for $i + j > n + 1$) are simulated from the selected probability distribution. This computation allows for process error. The simulation process is then repeated many times, and the sample 75th percentile and the sample standard deviation are used as the estimates of the 75th percentile and the standard deviation of the OCL.

6.6 Validation of Model Structure

The residuals can be further analysed by plotting them against different dimensions, i.e. the accident period, the development period, the calendar period, and $\hat{X}_{i,j}$. In addition, the empirical distribution of the residuals can also be formed. If the residuals are evenly spread in the scatter plot, and if there is no skewness in the empirical density of the residuals, the model structure can be regarded as being able to provide a reasonable fit over the past claims data. The residuals may then be assumed to be independently and identically distributed normal random variables. McCullagh and Nelder (1989) and Taylor and McGuire (2004) elaborate on examination of residuals.

To further rank different arrangements of the parameters in the linear relationship, the following statistic can be used, which is stated in McCullagh and Nelder (1989). It is mentioned that the range $2 \le \alpha \le 6$ below may provide a set of reasonable initial models for further investigation. The statistic is defined as:

$$Q = D + \alpha \cdot n_p \cdot \hat{\phi} \ , \tag{15}$$

where $D = \sum_{i+j \le n+1} \left(residual_{i,j}^{\text{Deviance}} \right)^2$ and $2 \le \alpha \le 6$.

This statistic penalises both poor fitting and excessive use of parameters. Lower value of Q is preferred in determining how the parameters are associated in the linear relationship.

6.7 Relationship with Chain Ladder Bootstrapping

As suggested in Verrall (1999) and proved in Taylor (2000), applying maximum likelihood to the GLM structure $\eta_{i,j} = \alpha_i + \beta_j = \ln\left(\mu_{i,j}\right)$ and $X_{i,j} \sim \text{Pn}\left(\mu_{i,j}\right)$ gives the same estimates of the expected OCL as the chain ladder method produces. If the type of residuals adopted is the same, there is no difference between mixing this GLM structure and mixing the chain ladder method with bootstrapping.

6.8 Approximation Formulae

We now extend the ideas in Verrall (1999) by constructing some general formulae to estimate the standard error of prediction of the OCL with the use of (1) and (2).

Since $\text{Var}\left(R_i\right) = \sum_{j=n+2-i}^{n} \text{Var}\left(X_{i,j}\right)$ and $\text{Var}\left(R\right) = \sum_{i+j>n+1} \text{Var}\left(X_{i,j}\right)$ due to independence of $X_{i,j}$, the estimators of $\text{Var}\left(R_i\right)$ and $\text{Var}\left(R\right)$ are respectively:

$$\sum_{j=n+2-i}^{n} \hat{\phi}\, \hat{X}_{i,j}^{\,p} \ , \text{ and} \tag{16}$$

$$\sum_{i+j>n+1} \hat{\phi}\, \hat{X}_{i,j}^{\,p} \ , \tag{17}$$

where $p = 0$ for normal distribution, $p = 1$ for (over-dispersed) Poisson distribution, $p = 2$ for gamma distribution, and $p = 3$ for inverse Gaussian distribution (refer to Appendix 2 for details).

Furthermore, the variance of $\hat{R}_i$ and $\hat{R}$ are respectively:

$$\mathrm{Var}\left(\hat{R}_i\right) = \sum_{j=n+2-i}^{\mathrm{n}} \mathrm{Var}\left(\hat{X}_{i,j}\right) + \sum_{\substack{j>n+1-i \\ h>n+1-i \\ j\neq h}} \mathrm{Cov}\left(\hat{X}_{i,j}, \hat{X}_{i,h}\right) \text{ , and} \tag{18}$$

$$\mathrm{Var}\left(\hat{R}\right) = \sum_{i+j>n+1} \mathrm{Var}\left(\hat{X}_{i,j}\right) + \sum_{\substack{i+j>n+1 \\ g+h>n+1 \\ (i,j)\neq(g,h)}} \mathrm{Cov}\left(\hat{X}_{i,j}, \hat{X}_{g,h}\right) \ . \tag{19}$$

By using Taylor series expansion, $\hat{X}_{i,j}$ is formulated as (up to only the first order):

$$\hat{X}_{i,j} = h^{-1}\left(\hat{\eta}_{i,j}\right) \approx h^{-1}\left(\eta_{i,j}\right) + \frac{dh^{-1}\left(\eta_{i,j}\right)}{d\eta_{i,j}}\left(\hat{\eta}_{i,j} - \eta_{i,j}\right) = \mu_{i,j} + \frac{d\mu_{i,j}}{d\eta_{i,j}}\left(\hat{\eta}_{i,j} - \eta_{i,j}\right) \ .$$

The variance and covariance of $\hat{X}_{i,j}$ are then:

$$\mathrm{Var}\left(\hat{X}_{i,j}\right) \approx \left(\frac{d\mu_{i,j}}{d\eta_{i,j}}\right)^2 \mathrm{Var}\left(\hat{\eta}_{i,j}\right) \text{ , and}$$

$$\mathrm{Cov}\left(\hat{X}_{i,j}, \hat{X}_{g,h}\right) \approx \frac{d\mu_{i,j}}{d\eta_{i,j}} \frac{d\mu_{g,h}}{d\eta_{g,h}} \mathrm{Cov}\left(\hat{\eta}_{i,j}, \hat{\eta}_{g,h}\right) \text{ (for } (i\ ,\ j) \neq (g\ ,\ h)) \ .$$

Accordingly, the estimators of $\mathrm{Var}\left(\hat{X}_{i,j}\right)$ and $\mathrm{Cov}\left(\hat{X}_{i,j}, \hat{X}_{g,h}\right)$ (for $(i\ ,\ j) \neq (g\ ,\ h)$) are respectively:

$$\left(\frac{d\mu_{i,j}}{d\eta_{i,j}}\right)^2 \Bigg|_{\eta_{i,j}=\hat{\eta}_{i,j}} \times \text{ Estimator of } \mathrm{Var}\left(\hat{\eta}_{i,j}\right) \text{ , and} \tag{20}$$

$$\left.\frac{d\mu_{i,j}}{d\eta_{i,j}}\right|_{\eta_{i,j}=\hat{\eta}_{i,j}} \left.\frac{d\mu_{g,h}}{d\eta_{g,h}}\right|_{\eta_{g,h}=\hat{\eta}_{g,h}} \times \text{Estimator of Cov}\left(\hat{\eta}_{i,j},\hat{\eta}_{g,h}\right), \tag{21}$$

where the estimators of $\text{Var}\left(\hat{\eta}_{i,j}\right)$ and $\text{Cov}\left(\hat{\eta}_{i,j},\hat{\eta}_{g,h}\right)$ are computed from the covariance matrix of the parameter estimators calculated from maximum likelihood estimation.

The estimators of $\text{Var}\left(\hat{R}_i\right)$ and $\text{Var}\left(\hat{R}\right)$ are then found by using (20) and (21) in (18) and (19).

Making use of the estimated $\text{Var}\left(R_i\right)$, $\text{Var}\left(R\right)$, $\text{Var}\left(\hat{R}_i\right)$, and $\text{Var}\left(\hat{R}\right)$ as above, $MSEP_i$ and $MSEP$ are computed approximately using (1) and (2). Finally, the estimate of the standard error of prediction is simply the square root of the estimate of the mean square error of prediction.

7 Kalman Filter

The Kalman filter is a recursive algorithm. Kalman (1960) introduces this algorithm as a different means to make statistical prediction. It has been used mainly in the engineering area, but it has also been brought to the finance area, e.g. Wells (1996). The algorithm generates estimates for the unknown parameters of a state-space model, and successively updates the estimates as new data are collected over time. The estimators it provides are least squares estimators if the error terms are normally distributed, and are least squares linear estimators otherwise.

Applying the Kalman filter to a state-space model offers a flexible framework for analysing the past claims data. The parameters can be assumed to evolve over time, instead of being assumed constant. The volatility of the forecasts can be computed directly, and prior information can be readily incorporated into the calculation. The application of the Kalman filter has been introduced into general insurance by De Jong and Zehnwirth (1983). Zehnwirth (1996) exhibits an application of the algorithm to reserving.

The following subsections give an explanation of the structure of a state-space model and the Kalman filter algorithm adopted in this paper.

7.1 State-Space Model

As stated in Zehnwirth (1996), a state-space model comprises two sets of equations. The first set of equations is called observation equation, in which we choose to take the logarithm of the claims data for modelling for $1 \le t \le 2n-1$:

$$Y(t) = A(t)\theta(t) + \varepsilon(t) ,$$

where $Y(t)$ is a vector of $\ln(X_{i,j})$ at time t with $t = i + j - 1$, $\theta(t)$ is a vector of the parameters at time t to be estimated, $A(t)$ is the design matrix at time t that states how

$X_{i,j}$ is related to the parameters, and $\varepsilon(t)$ is a vector of error terms (for the volatility of $\ln(X_{i,j})$) at time *t*.

The second set of equations is called state equation, which describes the evolution of the parameters over time for $1 \le t \le 2n-1$:

$\theta(t) = H(t)\theta(t-1) + \delta(t)$,

where $H(t)$ is a matrix at time *t* that states how the parameters of successive periods are related, and $\delta(t)$ is a vector of error terms (for the volatility of the parameters) at time *t*.

The two matrices $A(t)$ and $H(t)$ are the basic assumptions that determine the relationships between $X_{i,j}$ and the parameters. For $t \ne t^*$, $\varepsilon(t)$ and $\varepsilon(t^*)$ are assumed to be uncorrelated with each other, and similarly for $\delta(t)$ and $\delta(t^*)$, $\varepsilon(t)$ and $\delta(t)$, and $\varepsilon(t)$ and $\delta(t^*)$, in which $\varepsilon(t) \sim \mathrm{N}(0, \Sigma(t))$ and $\delta(t) \sim \mathrm{N}(0, \Gamma(t))$, where $\Sigma(t)$ and $\Gamma(t)$ are covariance matrices of $\varepsilon(t)$ and $\delta(t)$ respectively.

7.2 Kalman Filter

An informal derivation of the Kalman filter is introduced in De Jong and Zehnwirth (1983). Based on Zehnwirth (1996) and Wells (1996), the following algorithm can be operated sequentially on the past claims data for $1 \le t \le n$:

$$P(t|t-1) = H(t)P(t-1|t-1)H'(t) + \Gamma(t) \text{ (for } 1 \le t \le n+1), \quad (22)$$

$$K(t) = P(t|t-1)A'(t)\left(A(t)P(t|t-1)A'(t) + \Sigma(t)\right)^{-1}, \quad (23)$$

$$\hat{\theta}(t|t-1) = H(t)\hat{\theta}(t-1|t-1) \text{ (for } 1 \le t \le n+1), \quad (24)$$

$$\hat{Y}(t|t-1) = A(t)\hat{\theta}(t|t-1), \quad (25)$$

$$\hat{\theta}(t|t) = \hat{\theta}(t|t-1) + K(t)\left(Y(t) - \hat{Y}(t|t-1)\right), \text{ and} \quad (26)$$

$$P(t|t) = \left(I - K(t)A(t)\right)P(t|t-1), \quad (27)$$

where

$\hat{\theta}(t|t)$ is the estimator of $\theta(t)$ given the past claims data to time t,

$\hat{\theta}(t|t-1)$ is the estimator of $\theta(t)$ given the past claims data to time $t-1$,

$P(t|t)$ is the covariance matrix of $\theta(t)-\hat{\theta}(t|t)$,

$P(t|t-1)$ is the covariance matrix of $\theta(t)-\hat{\theta}(t|t-1)$,

$K(t)$ is the Kalman gain matrix,

$\hat{Y}(t|t-1)$ is the one-step-ahead forecast of $Y(t)$ given the past claims data to time $t-1$,

and $A(t)P(t|t-1)A'(t)+\Sigma(t)$ is the covariance matrix of $Y(t)-\hat{Y}(t|t-1)$.

The calculation embarks on setting the initial values of $P(0|0)$ and $\hat{\theta}(0|0)$, which are put into (22) and (24) respectively. Other prior information relating to the volatilities of $\ln(X_{i,j})$ and the parameters (i.e. the covariance matrices $\Sigma(t)$ and $\Gamma(t)$ for $1 \le t \le 2n-1$) also need to be incorporated. De Jong and Zehnwirth (1983) describe an approximate method to determine the initial values and the volatilities.

For each time period's claims data, appearing as a diagonal of the claims run-off triangle, the algorithm is run from (22) to (27) to update the parameter estimates, which are given by (26). This computation is effectively a recursive process that runs the algorithm once for each time period from 1 to *n*.

The parameter estimates obtained after using *n* time periods of the past claims data may be used as new initial values of $\hat{\theta}(0|0)$, and the whole recursive process is repeated. Again, the final parameter estimates may be used as $\hat{\theta}(0|0)$ to repeat the whole recursive process. We find that the parameter estimates converge in the calculation in Section 11.

The Kalman gain matrix is similar to the credibility factor in credibility theory. For example, if $\Sigma(t)$ is large (i.e. the volatility of $\ln(X_{i,j})$ is large), $K(t)$ is small, and so less weight is put on $X_{i,j}$ when updating the parameter estimates using (26).

7.3 Mean and Covariance Matrix

Verrall (1983) depicts the use of *m*-step-ahead forecasts. Regarding the measurement of the OCL (for $n + 1 \le t \le 2n - 1$), *m*-step-ahead forecasts (for $m = 1, 2, \ldots, n - 1$) given the past claims data to time *n* are calculated by using an analogy of (24) and (25):

$$\hat{Y}(n+m|n) = A(n+m)\hat{\theta}(n+m|n) = \ldots$$

$$= A(n+m)H(n+m)H(n+m-1)\ldots H(n+2)\hat{\theta}(n+1|n) \quad . \qquad (28)$$

To correct the bias approximately, we take the exponential of (28) plus half of the diagonal of $\Sigma(n+m)$ to estimate the expected OCL. Moreover, we derive some covariance matrices as follows.

Since

$$Y(n+m) = A(n+m)\theta(n+m) + \varepsilon(n+m)$$

$$= A(n+m)H(n+m)H(n+m-1)\ldots H(n+2)\theta(n+1) + A(n+m)\delta(n+m)$$

$$+ A(n+m)H(n+m)\delta(n+m-1) + \ldots + A(n+m)H(n+m)H(n+m-1)\ldots H(n+3)\delta(n+2)$$

$$+ \varepsilon(n+m) \ ,$$

(from repeated use of the state equation)

the covariance matrix of $\left(Y(n+m) - \hat{Y}(n+m|n)\right)$ is

$$\mathrm{Cov}\left(Y(n+m) - \hat{Y}(n+m|n)\right)$$

$$= A(n+m)H(n+m)\ldots H(n+2)P(n+1|n)H'(n+2)\ldots H'(n+m)A'(n+m)$$

$$+ A(n+m)\Gamma(n+m)A'(n+m) + A(n+m)H(n+m)\Gamma(n+m-1)H'(n+m)A'(n+m) + \ldots$$

$$+ A(n+m)H(n+m)\ldots H(n+3)\Gamma(n+2)H'(n+3)\ldots H'(n+m)A'(n+m) + \Sigma(n+m) \quad .$$

(29)

Furthermore, let $Y = \begin{bmatrix} Y(n+1) \\ Y(n+2) \\ \vdots \\ Y(2n-1) \end{bmatrix}$, $\hat{Y} = \begin{bmatrix} \hat{Y}(n+1|n) \\ \hat{Y}(n+2|n) \\ \vdots \\ \hat{Y}(2n-1|n) \end{bmatrix}$, $\varepsilon = \begin{bmatrix} \varepsilon(n+1) \\ \varepsilon(n+2) \\ \vdots \\ \varepsilon(2n-1) \end{bmatrix}$, and

$\delta = \begin{bmatrix} \delta(n+1) \\ \delta(n+2) \\ \vdots \\ \delta(2n-1) \end{bmatrix}$, where $\varepsilon \sim \mathrm{N}(0, \Sigma)$ and $\delta \sim \mathrm{N}(0, \Gamma)$.

Since $Y = A\theta(n+1) + B\delta + \varepsilon$ and $\hat{Y} = A\hat{\theta}(n+1|n)$, where

$A = \begin{bmatrix} A(n+1) \\ A(n+2)H(n+2) \\ \vdots \\ A(2n-1)H(2n-1)H(2n-2)...H(n+2) \end{bmatrix}$, and

$B = \begin{bmatrix} 0 & ... & & & \\ 0 & A(n+2) & 0 & ... & \\ 0 & A(n+3)H(n+3) & A(n+3) & 0 & ... \\ \vdots & & & & \\ 0 & A(2n-1)H(2n-1)...H(n+3) & A(2n-1)H(2n-1)...H(n+4) & ... & A(2n-1) \end{bmatrix}$,

the covariance matrix of $\left(Y - \hat{Y}\right)$ is:

$$\mathrm{Cov}\left(Y - \hat{Y}\right) = AP(n+1|n)A' + B\Gamma B' + \Sigma \quad . \qquad (30)$$

The covariance matrices above allow for both parameter error and process error explicitly.

An example of the observation equation is $\ln\left(X_{i,j}\right) = \alpha_i + \beta_j + \varepsilon_{i,j}$ with $\varepsilon_{i,j} \sim \mathrm{N}\left(0, \sigma_\varepsilon^2\right)$, in which α_i and β_j are constant parameters depicting the accident period effect and the development period effect, and $\varepsilon_{i,j}$ is an error term.

In this example, for $n + 1 \le t \le 2n - 1$, $\Sigma(t) = \sigma_\varepsilon^2 I$, $H(t) = I$, and $\Gamma(t) = 0$, and so $\Sigma = \sigma_\varepsilon^2 I$, A is in terms of $A(t)$ only, and $\Gamma = 0$.

7.4 75th Percentile and Standard Deviation

To incorporate parameter error and process error approximately, we choose to simulate 1,000 samples of $\ln(X_{i,j})$ (for $i + j > n + 1$) from a multinormal distribution. The mean vector of the multinormal distribution is set equal to the forecasts from (28) and the corresponding covariance matrix is given by (30). The Cholesky decomposition process is exploited to transform independently drawn normal random variables into multinormal random variables. Golan (2004) demonstrates the algorithm of the Cholesky decomposition process.

Accordingly, 1,000 samples of R_i (for $2 \le i \le n$) and R are formed by summing up the relevant simulated $X_{i,j}$'s (for $i + j > n + 1$). Since both parameter error and process error are incorporated in the estimated covariance matrix, we use the sample 75th percentile and the sample standard deviation to estimate the 75th percentile and the standard deviation of the OCL.

7.5 Relationship with General Linear Model

Applying the Kalman filter to a state-space model reduces to applying maximum likelihood estimation to a general linear model (a type of GLM with identity link function and normal distribution) under certain circumstances (for $1 \le t \le 2n - 1$). These are: (1) $\Sigma(t) = \hat{\sigma}^2 I$ (the term $\hat{\sigma}^2$ is the estimator of the GLM dispersion parameter); (2) $H(t) = I$; (3) $\Gamma(t) = 0$; (4) $P(0|0)$ is very large (i.e. non-informative prior). These conditions are briefly mentioned in Verrall (1989). Some further details are set forth in Appendix 4.

8 Mack Model

This reserving model has been proposed by Mack (1993a). The structure of the model embraces the mechanics of the chain ladder method. It is essentially distribution-free, i.e. it does not require an assumption of a probability distribution, and it provides closed-form formulae. It calculates the expected OCL in the same way as the chain ladder method does. It also estimates the mean square error of prediction of the OCL. Both parameter error and process error are accommodated explicitly in the formulae.

In order to obtain the 75th percentile of the OCL approximately, however, a probability distribution is needed, and a lognormal distribution is assumed for the OCL in this paper. The parameters of the lognormal distribution are calculated by setting the estimates of the expected OCL and the mean square error of prediction as the mean and variance of the lognormal distribution. Other probability distributions may be used to reflect different right-tail properties.

8.1 Model Structure

As stated in Mack (1993a), the three underlying assumptions of the model are as follows:

(1) $\mathrm{E}\left(C_{i,j+1}\middle|C_{i,1},C_{i,2},...,C_{i,j}\right)=C_{i,j}f_j$ (for $1\le j\le n-1$),

(2) $C_{i,j}$ and $C_{g,h}$ are independent (for $i\ne g$), and

(3) $\mathrm{Var}\left(C_{i,j+1}\middle|C_{i,1},C_{i,2},...,C_{i,j}\right)=C_{i,j}\sigma_j^2$ (for $1\le j\le n-1$).

There are two types of parameters, f_j and σ_j^2, where f_j is the development factor from development period j to $j+1$, and σ_j^2 is related to the variance of $C_{i,j+1}$. Under the three underlying assumptions, Mack (1993a) shows that the expected values of $C_{i,n}$ and R_i are $\mathrm{E}(C_{i,n})=\mathrm{E}(C_{i,n+1-i})f_{n+1-i}f_{n+2-i}...f_{n-1}$ and $\mathrm{E}(R_i)=\mathrm{E}\left(C_{i,n+1-i}\right)f_{n+1-i}f_{n+2-i}...f_{n-1}-\mathrm{E}\left(C_{i,n+1-i}\right)$ (for $2\le i\le n$).

8.2 Fitting the Model

The Mack model is based on the chain ladder method and so their estimators are the same. Mack (1993a) describes the following estimators.

The estimator of f_j for $1 \le j \le n-1$ is:

$$\hat{f}_j = \frac{\sum_{r=1}^{n-j} C_{r,j+1}}{\sum_{r=1}^{n-j} C_{r,j}} \quad . \tag{31}$$

The estimator of $\mathrm{E}(C_{i,j})$ for $2 \le i \le n$ and $n+1-i \le j \le n$ is:

$$\hat{C}_{i,j} = C_{i,n+1-i}\,\hat{f}_{n+1-i}\,\hat{f}_{n+2-i} \cdots \hat{f}_{j-1} \quad . \tag{32}$$

The estimators of $\mathrm{E}(R_i)$ (for $2 \le i \le n$) and $\mathrm{E}(R)$ are:

$$\hat{R}_i = \hat{C}_{i,n} - C_{i,n+1-i} = C_{i,n+1-i}\,\hat{f}_{n+1-i}\,\hat{f}_{n+2-i} \cdots \hat{f}_{n-1} - C_{i,n+1-i} \quad \text{, and} \tag{33}$$

$$\hat{R} = \hat{R}_2 + \hat{R}_3 + \ldots + \hat{R}_n \quad , \tag{34}$$

which produce the estimates of the expected OCL.

Taylor (2000) demonstrates that (31) can be derived by applying weighted least squares estimation. Mack (1993a) shows that the estimators given by (31) to (34) are unbiased, and that $\hat{f}_j$ and $\hat{f}_k$ are uncorrelated for $j > k$, under the model assumptions.

8.3 Standard Error of Prediction

Mack (1993a) gives the estimator of σ_j^2 as:

$$\hat{\sigma}_j^2 = \frac{1}{n-j-1} \sum_{r=1}^{n-j} C_{r,j} \left(\frac{C_{r,j+1}}{C_{r,j}} - \hat{f}_j \right)^2 \text{ (for } 1 \le j \le n-2\text{) and } \hat{\sigma}_{n-1}^2 = \min\left(\frac{\hat{\sigma}_{n-2}^4}{\hat{\sigma}_{n-3}^2}, \hat{\sigma}_{n-3}^2 \right) \quad . \tag{35}$$

In addition, the mean square error of prediction of R_i given the past claims data is derived for $2 \le i \le n$ as:

$$MSEP_i \mid C_{g,h:g+h\le n+1} = \mathrm{Var}\left(C_{i,n} \middle| C_{g,h:g+h\le n+1}\right) + \left(\mathrm{E}\left(C_{i,n} \middle| C_{g,h:g+h\le n+1}\right) - \hat{C}_{i,n}\right)^2 ,$$

in which the first component allows for process error and the second component allows for parameter error.

The mean square error of prediction of R_i is then estimated for $2 \le i \le n$ by:

$$\hat{C}_{i,n}^2 \sum_{j=n+1-i}^{n-1} \left(\frac{\hat{\sigma}_j^2}{\hat{f}_j^2}\right)\left(\frac{1}{\hat{C}_{i,j}} + \frac{1}{\sum_{r=1}^{n-j} C_{r,j}}\right) . \qquad (36)$$

Likewise, the mean square error of prediction of $\hat{R}$ given the past claims data is derived as:

$$MSEP \mid C_{g,h:g+h\le n+1} = \mathrm{Var}\left(\sum_{i=2}^{n} C_{i,n} \middle| C_{g,h:g+h\le n+1}\right) + \left(\mathrm{E}\left(\sum_{i=2}^{n} C_{i,n} \middle| C_{g,h:g+h\le n+1}\right) - \sum_{i=2}^{n} \hat{C}_{i,n}\right)^2 ,$$

in which the first component allows for process error and the second component allows for parameter error.

The mean square error of prediction of $\hat{R}$ is then estimated by:

$$\sum_{i=2}^{n} \text{Estimator of } MSEP_i \mid C_{g,h:g+h\le n+1} + \sum_{i=2}^{n-1} \sum_{l=i+1}^{n} 2\hat{C}_{i,n}\hat{C}_{l,n} \sum_{j=n+1-i}^{n-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2 \sum_{r=1}^{n-j} C_{r,j}} . \qquad (37)$$

The square root of (36) and of (37) are used to estimate the standard error of prediction of the OCL. Mack (1993a) demonstrates the derivation of the two closed-form formulae (36) and (37).

8.4 Probability Distribution for 75th Percentile

The Mack model estimates the mean square error of prediction but not the 75th percentile as the model is distribution-free. In this paper, we assume the lognormal distribution to be the underlying probability distribution of the OCL in order to provide an approximation of the 75th percentile. Probability distributions other than lognormal may also be used. (Strictly speaking, R_i and R cannot be both lognormally distributed, though the assumption may be approximately correct.)

The parameters of the lognormal distribution are then computed by moment matching – equating the estimates of the expected OCL and the corresponding mean square error of prediction to the mean and variance of the lognormal distribution.

To cite an example, let $R \sim \text{Ln}(a,b)$. Then a and b are found by equating the following:

$\text{E}(R) = \exp\left(a + \frac{1}{2}b^2\right) = \hat{R}$, and

$\text{Var}(R) = \left(\text{E}(R)\right)^2 \left(\exp\left(b^2\right) - 1\right) = \text{Estimator of } MSEP \mid C_{g,h:g+h\le n+1}$.

Finally, the 75th percentile of the total OCL is estimated as $\exp\left(a + b\ \Phi^{-1}\left(0.75\right)\right)$, where $\Phi^{-1}\left(0.75\right)$ is the 75th percentile of the standard normal distribution.

8.5 Justification of Assumptions

Houltram (2003) sets forth three tests for justifying the underlying assumptions of the Mack model. Likewise, they may be used for the chain ladder method with bootstrapping.

9 Stochastic Chain Ladder Method

The stochastic chain ladder method has been documented and applied in some references such as Hertig (1985), Taylor (2000), Collings and White (2001) and De Jong (2003). It is effectively a stochastic version of the chain ladder method, which uses simulation. The development factors are assumed to follow a normal distribution, the mean and variance of which are computed from the past claims data, and the factors may be adjusted or smoothed based on experience or judgement. Samples of the ratios are then simulated from the normal distribution with the estimated parameters. Samples of future unobserved aggregated claim amounts are formed accordingly.

This method is flexible. Correlations between claim amounts of different accident periods, development periods, or calendar periods can be accommodated. While the correlations can be estimated from the past claims data, appropriate judgement is often needed due to data limitations. Development beyond the latest development period in the past claims data can also be included by extrapolating the past observed run-off pattern. De Jong (2003) describes a variety of these flexible features.

The ways to structure the stochastic chain ladder method are many and varied, and the approach in this paper is detailed in the following subsections.

9.1 Model Structure

The stochastic chain ladder method is a stochastic method, as the development factors are assumed to be normal random variables. We make the following assumptions:

(1) $C_{i,j+1} = C_{i,j} \exp\left(F_{i,j}\right)$ (for $1 \le j \le n-1$) ,

(2) $F_{i,j} \sim \mathrm{N}\left(f_j, \sigma_j^2\right)$ (for $1 \le j \le n-1$) ,

(3) $F_{i,j}$ and $F_{g,h}$ are independent (for $(i\ ,\ j) \ne (g\ ,\ h)$) , and

(4) $C_{i,j}$ and $C_{g,h}$ are independent (for $i \ne g$) .

$F_{i,j}$ is the development factor from development period j to $j + 1$ for accident period i, with f_j and σ_j^2 being the mean and variance of $F_{i,j}$ (different to the Mack model's f_j and σ_j^2). We then derive the expected value and variance of $C_{i,n}$, R_i, and R (given certain past claims data) as follows.

Using the assumptions (1), (2), and (3), for $2 \le i \le n$,

$$C_{i,n} = C_{i,n+1-i} \exp\left(F_{i,n+1-i} + F_{i,n+2-i} + ... + F_{i,n-1}\right) ,$$

$$R_i = C_{i,n+1-i} \exp\left(F_{i,n+1-i} + F_{i,n+2-i} + ... + F_{i,n-1}\right) - C_{i,n+1-i} \text{ , and}$$

$$F_{i,n+1-i} + F_{i,n+2-i} + ... + F_{i,n-1} \sim \mathrm{N}\left(f_{n+1-i} + f_{n+2-i} + ... + f_{n-1}, \sigma_{n+1-i}^2 + \sigma_{n+2-i}^2 + ... + \sigma_{n-1}^2\right) .$$

Given the latest calendar period's cumulative claim amount of accident period i in the past claims data, the expected values of $C_{i,n}$ and R_i for $2 \le i \le n$ are:

$$\mathrm{E}\left(C_{i,n} \middle| C_{i,n+1-i}\right) = C_{i,n+1-i} \exp\left(f_{n+1-i} + f_{n+2-i} + ... + f_{n-1} + \frac{1}{2}\left(\sigma_{n+1-i}^2 + \sigma_{n+2-i}^2 + ... + \sigma_{n-1}^2\right)\right) \text{ , and}$$

$$\mathrm{E}\left(R_i \middle| C_{i,n+1-i}\right) = C_{i,n+1-i} \exp\left(f_{n+1-i} + f_{n+2-i} + ... + f_{n-1} + \frac{1}{2}\left(\sigma_{n+1-i}^2 + \sigma_{n+2-i}^2 + ... + \sigma_{n-1}^2\right)\right) - C_{i,n+1-i} .$$

Given the latest calendar period's cumulative claim amounts in the past claims data, the expected value of R is, based on the assumption (4):

$$\mathrm{E}\left(R \middle| C_{g,h:g+h=n+1}\right) = \mathrm{E}\left(R_2 \middle| C_{2,n-1}\right) + \mathrm{E}\left(R_3 \middle| C_{3,n-2}\right) + ... + \mathrm{E}\left(R_n \middle| C_{n,1}\right) .$$

Given the latest calendar period's cumulative claim amount of accident period i in the past claims data, the variances of $C_{i,n}$ and R_i for $2 \le i \le n$ are:

$$\mathrm{Var}\left(C_{i,n} \middle| C_{i,n+1-i}\right) = \mathrm{Var}\left(R_i \middle| C_{i,n+1-i}\right) = \left(\mathrm{E}\left(C_{i,n} \middle| C_{i,n+1-i}\right)\right)^2 \left(\exp\left(\sigma_{n+1-i}^2 + \sigma_{n+2-i}^2 + ... + \sigma_{n-1}^2\right) - 1\right) .$$

9.2 Fitting the Model

As stated in De Jong (2003), the estimator of f_j is defined for $1 \le j \le n-1$ as:

$$\hat{f}_j = \frac{1}{n-j}\sum_{r=1}^{n-j} F_{r,j} = \frac{1}{n-j}\sum_{r=1}^{n-j} \ln\left(\frac{C_{r,j+1}}{C_{r,j}}\right) . \tag{38}$$

Under the assumptions (2) and (3), $\hat{f}_j$ is unbiased, in which $\hat{f}_j \sim \mathrm{N}\left(f_j, \frac{\sigma_j^2}{n-j}\right)$, and $\hat{f}_j$ and $\hat{f}_h$ are independent for $j \ne h$. Consequently, for $2 \le i \le n$,

$$\hat{f}_{n+1-i} + \hat{f}_{n+2-i} + \ldots + \hat{f}_{n-1} \sim \mathrm{N}\left(f_{n+1-i} + f_{n+2-i} + \ldots + f_{n-1}, \frac{\sigma_{n+1-i}^2}{i-1} + \frac{\sigma_{n+2-i}^2}{i-2} + \ldots + \frac{\sigma_{n-1}^2}{1}\right) .$$

In order to provide an unbiased estimate, we define the estimator of σ_j^2 as follows, which is different to that in De Jong (2003):

$$\hat{\sigma}_j^2 = \frac{1}{n-j-1}\sum_{r=1}^{n-j}\left(F_{r,j} - \hat{f}_j\right)^2 = \frac{1}{n-j-1}\sum_{r=1}^{n-j}\left(\ln\left(\frac{C_{r,j+1}}{C_{r,j}}\right) - \hat{f}_j\right)^2 \text{ (for } 1 \le j \le n-2\text{), and}$$

$$\hat{\sigma}_{n-1}^2 = \min\left(\frac{\hat{\sigma}_{n-2}^4}{\hat{\sigma}_{n-3}^2}, \hat{\sigma}_{n-3}^2\right) , \tag{39}$$

in which $\hat{\sigma}_j^2$ is unbiased for $1 \le j \le n-2$ under the assumptions (2) and (3).

Hence, we derive the estimators of $\mathrm{E}\left(C_{i,n}\middle|C_{i,n+1-i}\right)$ and $\mathrm{E}\left(R_i\middle|C_{i,n+1-i}\right)$ for $2 \le i \le n$ as:

$$\hat{C}_{i,n} = C_{i,n+1-i}\exp\left(\hat{f}_{n+1-i} + \hat{f}_{n+2-i} + \ldots + \hat{f}_{n-1} + \frac{1}{2}\left(\hat{\sigma}_{n+1-i}^2 + \hat{\sigma}_{n+2-i}^2 + \ldots + \hat{\sigma}_{n-1}^2\right)\right) \text{, and} \tag{40}$$

$$\hat{R}_i = C_{i,n+1-i}\exp\left(\hat{f}_{n+1-i} + \hat{f}_{n+2-i} + \ldots + \hat{f}_{n-1} + \frac{1}{2}\left(\hat{\sigma}_{n+1-i}^2 + \hat{\sigma}_{n+2-i}^2 + \ldots + \hat{\sigma}_{n-1}^2\right)\right) - C_{i,n+1-i} . \tag{41}$$

The estimator of $\mathrm{E}\left(R\middle|C_{g,h:g+h=n+1}\right)$ is then defined as:

$$\hat{R} = \hat{R}_2 + \hat{R}_3 + \ldots + \hat{R}_n . \tag{42}$$

Equations (41) and (42) provide the estimates of the expected OCL.

9.3 75th Percentile and Standard Deviation

From the previous subsections, it is noted that for $2 \le i \le n$:

$F_{i,n+1-i} + F_{i,n+2-i} + ... + F_{i,n-1} \sim \mathrm{N}\left(f_{n+1-i} + f_{n+2-i} + ... + f_{n-1}, \sigma^2_{n+1-i} + \sigma^2_{n+2-i} + ... + \sigma^2_{n-1}\right)$ and

$\hat{f}_{n+1-i} + \hat{f}_{n+2-i} + ... + \hat{f}_{n-1} \sim \mathrm{N}\left(f_{n+1-i} + f_{n+2-i} + ... + f_{n-1}, \frac{\sigma^2_{n+1-i}}{i-1} + \frac{\sigma^2_{n+2-i}}{i-2} + ... + \frac{\sigma^2_{n-1}}{1} \right)$, where

$\left(\sigma^2_{n+1-i} + \sigma^2_{n+2-i} + ... + \sigma^2_{n-1}\right)$ allows for process error and $\left(\frac{\sigma^2_{n+1-i}}{i-1} + \frac{\sigma^2_{n+2-i}}{i-2} + ... + \frac{\sigma^2_{n-1}}{1} \right)$ is related to parameter error.

Furthermore, it follows that for $2 \le i < r \le n$:

$\mathrm{Cov}\left(\hat{f}_{n+1-i} + \hat{f}_{n+2-i} + ... + \hat{f}_{n-1}, \hat{f}_{n+1-r} + \hat{f}_{n+2-r} + ... + \hat{f}_{n-1}\right)$

$= \mathrm{Var}\left(\hat{f}_{n+1-i} + \hat{f}_{n+2-i} + ... + \hat{f}_{n-1}\right) + \mathrm{Cov}\left(\hat{f}_{n+1-i} + \hat{f}_{n+2-i} + ... + \hat{f}_{n-1}, \hat{f}_{n+1-r} + \hat{f}_{n+2-r} + ... + \hat{f}_{n-i}\right)$

$= \left(\frac{\sigma^2_{n+1-i}}{i-1} + \frac{\sigma^2_{n+2-i}}{i-2} + ... + \frac{\sigma^2_{n-1}}{1} \right)$ (because of independence of $\hat{f}_j$),

which is estimated by $\left(\frac{\hat{\sigma}^2_{n+1-i}}{i-1} + \frac{\hat{\sigma}^2_{n+2-i}}{i-2} + ... + \frac{\hat{\sigma}^2_{n-1}}{1} \right)$. (43)

Hence, $\left(\frac{\sigma^2_{n+1-i}}{i-1} + \frac{\sigma^2_{n+2-i}}{i-2} + ... + \frac{\sigma^2_{n-1}}{1} \right)$ also represents the covariance between the sums of the development factor estimators of accident periods i and r for $2 \le i < r \le n$.

Accordingly, to incorporate parameter error and process error approximately, we choose to simulate 1,000 samples of $\left(F_{i,n+1-i} + F_{i,n+2-i} + ... + F_{i,n-1}\right)$ (for $2 \le i \le n$) from a multinormal distribution. The mean vector of the multinormal distribution is set equal to $\left(\hat{f}_{n+1-i} + \hat{f}_{n+2-i} + ... + \hat{f}_{n-1}\right)$ computed from (38), the variance of the distribution is given by

$\left(\hat{\sigma}^2_{n+1-i} + \hat{\sigma}^2_{n+2-i} + ... + \hat{\sigma}^2_{n-1} + \frac{\hat{\sigma}^2_{n+1-i}}{i-1} + \frac{\hat{\sigma}^2_{n+2-i}}{i-2} + ... + \frac{\hat{\sigma}^2_{n-1}}{1} \right)$ computed from (39), and the covariance of the distribution is set as $\left(\frac{\hat{\sigma}^2_{n+1-i}}{i-1} + \frac{\hat{\sigma}^2_{n+2-i}}{i-2} + ... + \frac{\hat{\sigma}^2_{n-1}}{1} \right)$ from (43). Golan (2004) demonstrates the algorithm of the Cholesky decomposition process for generating multinormal random variables.

Finally, 1,000 samples of R_i (for $2 \le i \le n$) and so R are generated by using $R_i = C_{i,n+1-i} \exp\left(F_{i,n+1-i} + F_{i,n+2-i} + ... + F_{i,n-1} \right) - C_{i,n+1-i}$. Since both parameter error and process error are approximately accommodated, we use the sample 75th percentile and the sample standard deviation to estimate the 75th percentile and the standard deviation of the OCL.

9.4 Approximation Formulae

In this section, we introduce our derivation of some approximation formulae. The standard error of prediction of the OCL can be estimated by using (1) and (2).

The estimator of $\mathrm{Var}\left(C_{i,n} \middle| C_{i,n+1-i} \right)$ is defined for $2 \le i \le n$ as:

$$\hat{C}^2_{i,n} \left(\exp\left(\hat{\sigma}^2_{n+1-i} + \hat{\sigma}^2_{n+2-i} + ... + \hat{\sigma}^2_{n-1} \right) - 1 \right) . \qquad (44)$$

Consider $\hat{C}^*_{i,n} = C_{i,n+1-i} \exp\left(\hat{f}_{n+1-i} + \hat{f}_{n+2-i} + ... + \hat{f}_{n-1} \right)$, in which, for $2 \le i < r \le n$:

$$\mathrm{E}\left(\hat{C}^*_{i,n} \middle| C_{i,n+1-i} \right) = C_{i,n+1-i} \exp\left(f_{n+1-i} + f_{n+2-i} + ... + f_{n-1} + \frac{1}{2} \left(\frac{\sigma^2_{n+1-i}}{i-1} + \frac{\sigma^2_{n+2-i}}{i-2} + ... + \frac{\sigma^2_{n-1}}{1} \right) \right) ,$$

$$\mathrm{Var}\left(\hat{C}^*_{i,n} \mid C_{i,n+1-i} \right) = \left(\mathrm{E}\left(\hat{C}^*_{i,n} \middle| C_{i,n+1-i} \right) \right)^2 \left(\exp\left(\frac{\sigma^2_{n+1-i}}{i-1} + \frac{\sigma^2_{n+2-i}}{i-2} + ... + \frac{\sigma^2_{n-1}}{1} \right) - 1 \right) , \text{ and}$$

$$\mathrm{E}\left(\hat{C}_{i,n}^{*}\hat{C}_{r,n}^{*} \mid C_{i,n+1-i}, C_{r,n+1-r}\right)$$

$$= C_{i,n+1-i}C_{r,n+1-r}\exp\left(\sum_{l=n+1-r}^{n-i} f_l + 2\sum_{l=n+1-i}^{n-1} f_l + \frac{1}{2}\sum_{l=n+1-r}^{n-i}\frac{\sigma_l^2}{n-l} + 2\sum_{l=n+1-i}^{n-1}\frac{\sigma_l^2}{n-l}\right).$$

Accordingly, we define the following three estimators:

Estimator of $\mathrm{E}\left(\hat{C}_{i,n}^{*}\middle| C_{i,n+1-i}\right)$

$$= C_{i,n+1-i}\exp\left(\hat{f}_{n+1-i} + \hat{f}_{n+2-i} + \ldots + \hat{f}_{n-1} + \frac{1}{2}\left(\frac{\hat{\sigma}_{n+1-i}^2}{i-1} + \frac{\hat{\sigma}_{n+2-i}^2}{i-2} + \ldots + \frac{\hat{\sigma}_{n-1}^2}{1}\right)\right), \tag{45}$$

Estimator of $\mathrm{Var}\left(\hat{C}_{i,n}^{*}\middle| C_{i,n+1-i}\right)$

$$= C_{i,n+1-i}^2\exp\left(2\sum_{l=n+1-i}^{n-1}\hat{f}_l + \sum_{l=n+1-i}^{n-1}\frac{\hat{\sigma}_l^2}{n-l}\right)\left(\exp\left(\sum_{l=n+1-i}^{n-1}\frac{\hat{\sigma}_l^2}{n-l}\right) - 1\right), \tag{46}$$

and Estimator of $\mathrm{E}\left(\hat{C}_{i,n}^{*}\hat{C}_{r,n}^{*}\middle| C_{i,n+1-i}, C_{r,n+1-r}\right)$

$$C_{i,n+1-i}C_{r,n+1-r}\exp\left(\sum_{l=n+1-r}^{n-i}\hat{f}_l + 2\sum_{l=n+1-i}^{n-1}\hat{f}_l + \frac{1}{2}\sum_{l=n+1-r}^{n-i}\frac{\hat{\sigma}_l^2}{n-l} + 2\sum_{l=n+1-i}^{n-1}\frac{\hat{\sigma}_l^2}{n-l}\right). \tag{47}$$

The three estimators above are used to approximate $\mathrm{E}\left(\hat{C}_{i,n}\middle| C_{i,n+1-i}\right)$, $\mathrm{Var}\left(\hat{C}_{i,n}\middle| C_{i,n+1-i}\right)$, and $\mathrm{E}\left(\hat{C}_{i,n}\hat{C}_{r,n}\middle| C_{i,n+1-i}, C_{r,n+1-r}\right)$.

From (1), we derive the mean square error of prediction of R_i for $2 \le i \le n$ as:

$$MSEP_i = \mathrm{E}\left(\left(R_i - \hat{R}_i\right)^2\right) \approx \mathrm{Var}\left(R_i\right) + \mathrm{Var}\left(\hat{R}_i\right) \approx \mathrm{Var}\left(R_i\middle| C_{i,n+1-i}\right) + \mathrm{Var}\left(\hat{R}_i\middle| C_{i,n+1-i}\right)$$

$$= \mathrm{Var}\left(C_{i,n}\middle| C_{i,n+1-i}\right) + \mathrm{Var}\left(\hat{C}_{i,n}\middle| C_{i,n+1-i}\right),$$

which is then approximated by using (44) and (46).

From (2), we derive the mean square error of prediction of R as:

$$MSEP = \mathrm{E}\left(\left(R - \hat{R}\right)^2\right) \approx \mathrm{Var}(R) + \mathrm{Var}\left(\hat{R}\right) = \mathrm{Var}\left(\sum_{i=2}^{n} R_i\right) + \mathrm{Var}\left(\sum_{i=2}^{n} \hat{R}_i\right)$$

$$= \sum_{i=2}^{n} \mathrm{Var}\left(R_i\right) + \sum_{i=2}^{n} \mathrm{Var}\left(\hat{R}_i\right) + \sum_{2 \le i < r \le n} 2\mathrm{Cov}\left(\hat{R}_i, \hat{R}_r\right) \text{ (from the assumption (4))}$$

$$\approx \sum_{i=2}^{n} \mathrm{Var}\left(R_i \middle| C_{i,n+1-i}\right) + \sum_{i=2}^{n} \mathrm{Var}\left(\hat{R}_i \middle| C_{i,n+1-i}\right) + \sum_{2 \le i < r \le n} 2\mathrm{Cov}\left(\hat{R}_i, \hat{R}_r \middle| C_{i,n+1-i}, C_{r,n+1-r}\right)$$

$$= \sum_{i=2}^{n} \mathrm{Var}\left(C_{i,n} \middle| C_{i,n+1-i}\right) + \sum_{i=2}^{n} \mathrm{Var}\left(\hat{C}_{i,n} \middle| C_{i,n+1-i}\right) + \sum_{2 \le i < r \le n} 2\mathrm{Cov}\left(\hat{C}_{i,n}, \hat{C}_{r,n} \middle| C_{i,n+1-i}, C_{r,n+1-r}\right),$$

which is then approximated by using (44) to (47).

Finally, the estimate of the standard error of prediction is simply the square root of the estimate of the mean square error of prediction from above.

10 Comparison of Reserving Methods

This section presents our comparison of the reserving methods described previously. From an application perspective, the comparison is focused on the qualitative aspects, in which the underlying structures, assumptions, and estimation mechanics are examined thoroughly.

Each of the following subsections discusses one particular aspect of the reserving methods. 'BMCMC', 'CLB', 'GLMB', 'KF', 'MACK', and 'SCL' are abbreviations for Bayesian estimation with MCMC simulation, the chain ladder method with bootstrapping, GLMs with bootstrapping, the Kalman filter on state-space models, the Mack model, and the stochastic chain ladder method. (These abbreviations are also used in the rest of this paper.) The mathematical notation in this section is the same as that in the corresponding section about each reserving method.

10.1 Probability Distribution

The following compares the probability distributions under the reserving methods.

BMCMC: The incremental claim amount $X_{i,j}$ follows the process distribution.

CLB: No assumption is made for the chain ladder method as it is a mechanical method, but the residuals for bootstrapping are assumed to be identically distributed.

GLMB: The incremental claim amount $X_{i,j}$ follows a probability distribution from the exponential family. The residuals are normally distributed if the model structure provides a reasonable fit over the past claims data.

KF: The logarithm of the incremental claim amount $\ln\left(X_{i,j}\right)$ is normally distributed, i.e. $X_{i,j}$ is lognormally distributed.

MACK: No assumption is made under the basic model structure. To obtain an approximate value of the 75^{th} percentile, however, the lognormal distribution is assumed for R_i or R.

SCL: The development factor $F_{i,j}$ is normally distributed, i.e. $C_{i,j+1}/C_{i,j}$ is lognormally distributed.

The majority of the methods require an assumption of a probability distribution. Since general insurance liabilities are highly uncertain and typically skewed to the right with various degrees, flexibility in modelling the underlying probability distribution is important. In this regard, BMCMC and GLMB offer a range of choices for $X_{i,j}$. MACK also allows assumptions other than lognormal for R_i or R. For KF, if $\ln\left(X_{i,j}\right)$ is not normally distributed, the estimator of the algorithm is a least squares linear estimator (not necessarily a least squares estimator). For SCL, if $F_{i,j}$ is not normally distributed, the convolution of $\sum F_{i,j}$ may be intractable under our model assumptions. Finally, CLB does not need such an assumption, in which an empirical distribution is formed from bootstrapping.

10.2 Parameters

The following compares the parameters of the reserving methods.

BMCMC: Some parameters are fixed but some may follow the prior distributions. They are input as prior information or estimated through MCMC simulation.

CLB: The parameters are estimated mechanically.

GLMB: The parameters are fixed and estimated through maximum likelihood.

KF: Some parameters are fixed but some may evolve over time. They are estimated through the Kalman filter algorithm, and prior information is needed to start the algorithm.

MACK: The parameters are fixed and estimated mechanically.

SCL: The parameters are fixed and estimated mechanically.

The parameters computed from the past claims data are used or modified to predict future unobserved values, so appropriate parameter estimation is one of the paramount issues in

modelling claims development. While CLB, GLMB, MACK, and SCL assume constant parameters and estimate them rather mechanically, BMCMC and KF offer more room for coping with the change of parameters over time. BMCMC conceptually assumes the unknown parameters themselves are random variables following the prior distributions; KF provides a mechanism for the parameters to evolve over time. Both methods recognise that the parameters are not constant in reality and do change as time passes, especially for longer periods of time.

In addition, BMCMC, GLMB, and KF are more flexible with the arrangement of parameters compared to CLB, MACK and SCL. The next subsection provides further details regarding parameter arrangement.

10.3 Parameter Arrangement

The following compares the arrangement of parameters under the reserving methods.

BMCMC: This method can accommodate the accident period effect, the development period effect, and the calendar period effect explicitly. It can also allow the parameters to evolve over time and to be treated as random variables.

CLB: Explicit allowance is possible for the development period effect only. The accident period effect is implicitly accommodated by using the most recent cumulative claim amount of each accident period in the estimation. The future calendar period effects are implicitly assumed as an average of the past calendar period effects.

GLMB: This method can accommodate the accident period effect, the development period effect, and the calendar period effect explicitly.

KF: This method can provide for the accident period effect, the development period effect, and the calendar period effect explicitly. It can also allow the parameters to evolve over time.

MACK: Explicit allowance can be made for the development period effect only. The accident period effect is implicitly allowed for by using the most recent cumulative claim amount of each accident period in the estimation.

The future calendar period effects are implicitly assumed as an average of the past calendar period effects.

SCL: This method can incorporate the development period effect explicitly. Correlations across the accident periods, development periods, or calendar periods can be accommodated.

BMCMC, GLMB, and KF can take the three dimensions' effects into account explicitly. In particular, BMCMC and KF allow non-constant parameters. SCL is also quite flexible and can assume correlations across the three dimensions by adjusting the approach in this paper. In contrast, CLB and MACK can make an explicit allowance for the development period effect only.

In addition, due to their flexibility of parameter arrangement, BMCMC, GLMB, and KF can sometimes lead to a smaller number of necessary parameters compared to the other methods. To cite some examples for GLMB, the Hoerl curve can be applied as stated in Subsection 6.1, or for the most recent accident periods with less data, the same α_i in (11) can be used where appropriate in order to improve the stability of the estimation, as suggested in Renshaw (1989).

10.4 Independence

The following considers the independence of the random variables under the reserving methods.

BMCMC: Both $X_{i,j}$'s and the parameters are conditionally independent.

CLB: No assumption is made for the chain ladder method as it is a mechanical method, but the residuals for bootstrapping are assumed to be independent.

GLMB: Both $X_{i,j}$'s and the residuals for bootstrapping are independent.

KF: The error terms are uncorrelated across the time periods. The error terms for the volatility of $\ln\left(X_{i,j}\right)$ are uncorrelated with the error terms for the

volatility of the parameters. The independence of the error terms of the same time period depends on $\Sigma(t)$ and $\Gamma(t)$.

MACK: The incremental claim amounts $X_{i,j}$'s are independent across the accident periods only. The expected value of $X_{i,j}$ depends on $C_{i,j-1}$ given $C_{i,1}$, $C_{i,2}$, ... , and $C_{i,j-1}$.

SCL: The development factors $F_{i,j}$'s are independent. The incremental claim amounts $X_{i,j}$'s are independent across the accident periods only, in which $X_{i,j}$ depends on $C_{i,j-1}$.

The methods exhibit a variety of independence structures. BMCMC, KF, and SCL offer comparatively more room for adjusting the independence structures. The others are rather restricted in this aspect.

10.5 Volatility

The following compares the volatilities of the random variables under the reserving methods.

BMCMC: The volatilities of $X_{i,j}$ and the parameters depend on the prior and process distributions.

CLB: The variance of $\left(X_{i,j} - \hat{X}_{i,j}\right)$ is different for each development period.

GLMB: The variance of $X_{i,j}$ is equal to $\mu_{i,j}^{p}\phi / w_{i,j}$ (refer to Appendix 2).

KF: The covariance matrices of the error terms are $\Sigma(t)$ and $\Gamma(t)$.

MACK: The variance of $X_{i,j}$ depends on $C_{i,j-1}$ given $C_{i,1}$, $C_{i,2}$, ... , and $C_{i,j-1}$.

SCL: The variance of $F_{i,j}$ is different for each development period.

The methods demonstrate various volatility assumptions. BMCMC, GLMB, and KF can readily accommodate the weight of each cell (i , j) of the run-off triangle, while the others

cannot. This feature is important for dealing with any heteroscedasticity between different cells. Furthermore, correlations across the three dimensions within the run-off triangle may be incorporated in BMCMC, KF, and SCL, while the others are relatively restrictive in allowing for correlations.

10.6 Estimation Procedures

The following compares the estimation procedures of the reserving methods.

BMCMC: BUGS is used to perform Gibbs sampling of $X_{i,j}$.

CLB: The parameters are estimated mechanically, and the residuals are resampled to carry out bootstrapping to sample say $\left(R^{\text{pseudo future}} - \hat{R}^{\text{bootstrap}}\right)$, i.e. the discrepancy between the total OCL and its estimator.

GLMB: The parameters are estimated through maximum likelihood (using Excel VBA coding or standard GLM software such as SAS), and the residuals are resampled to carry out bootstrapping to sample say $\left(R^{\text{pseudo future}} - \hat{R}^{\text{bootstrap}}\right)$, i.e. the discrepancy between the total OCL and its estimator.

KF: The parameters are estimated by running the Kalman filter algorithm, starting with certain prior information. The parameter estimates are then used to generate multinormal random variables of $\ln\left(X_{i,j}\right)$ through the Cholesky decomposition process.

MACK: The parameters are estimated mechanically, and the mean square error of prediction is computed from the closed-form formulae. Inferences are then drawn by assuming R_i or R is lognormally distributed.

SCL: The parameters are estimated mechanically and then used to generate multinormal random variables of $\left(F_{i,n+1-i} + F_{i,n+2-i} + ... + F_{i,n}\right)$ through the Cholesky decomposition process.

All methods involve simulation except MACK. BUGS is needed for Gibbs sampling in BMCMC. Excel with VBA coding suffices for simulation purposes of the other methods. Standard GLM software such as SAS can also be used. CLB and GLMB involve the bootstrapping process in which a final adjustment is necessary to take the number of estimated parameters into account. The Cholesky decomposition process is adopted in KF and SCL to generate multinormal random variables.

BUGS provides a fast and convenient way for carrying out Gibbs sampling. Its functions are flexible and user-friendly, and it can be freely downloaded from the Internet. Nevertheless, it may have difficulty when dealing with too weak a prior, and if there is a hidden problem with certain types of models that have not been featured in the past, it may not crash and may just continue to produce results that are incorrect indeed. As such, extreme care is necessary when applying it to certain sophisticated models that have not been fully tested before.

Excel VBA coding is more complicated for GLMB and KF than for CLB and SCL, as the former two methods involve some recursive procedures with plenty of matrix operations. GLMB and KF also require longer computation time for their simulation. SAS offers a faster means to carry out maximum likelihood estimation for GLMs.

10.7 Unbiasedness

The following considers whether the estimators of the reserving methods are unbiased or not.

BMCMC: Under weak regularity conditions and when the number of iterations tends to infinity, $X_{i,j}$ and the parameters are simulated as if they were simulated from their true joint probability distribution. The sample means are then theoretically unbiased.

CLB: Unbiasedness is undefined as the chain ladder method is a mechanical method.

GLMB: Parameter estimators are maximum likelihood estimators, which may be biased. These estimators are asymptotically unbiased for large samples.

KF: Parameter estimators are least squares estimators, which may be biased.

MACK: The estimators of f_j, $\mathrm{E}(R_i)$, and $\mathrm{E}(R)$ are unbiased. (As shown in Taylor (2002), however, the OCL estimator is biased if $X_{i,j}$'s are assumed to be independent instead.)

SCL: The estimator of f_j is unbiased.

MACK is the only method that clearly shows unbiasedness of the estimator of the expected OCL under its own model structure. The other methods are more likely to provide biased estimators of the expected OCL under their own assumptions and structures. Certain adjustments may be made to correct the bias approximately, e.g. the adjustment in Subsection 7.3.

10.8 Prior Information

The following considers how prior information regarding the parameters is incorporated under the reserving methods.

BMCMC: Prior information is incorporated formally through the prior distributions.

CLB: Prior information cannot be accommodated formally.

GLMB: Prior information cannot be accommodated formally. (As shown in Taylor et al (2003), however, the exponential dispersion family (EDF) filter implemented with GLM estimation can incorporate prior information.)

KF: Prior information is incorporated formally by setting the initial values of $P(0|0)$, $\hat{\theta}(0|0)$, $\Sigma(t)$, and $\Gamma(t)$.

MACK: Prior information cannot be accommodated formally.

SCL: Prior information cannot be accommodated formally.

Only BMCMC and KF offer the option to input objective prior information or judgemental opinions in a formal way. On the other hand, if the judgement is

inappropriate or the information is incorrect, this flexibility may lead to undesirable results.

10.9 Other Characteristics

Some other characteristics of the reserving methods are listed in the following:

- BMCMC, GLMB, and KF can be applied to both individual and group claims data. The other methods can only be applied to aggregated claims data.
- Some approximation formulae are available for BMCMC, GLMB, and SCL to check the simulation results.
- Standardised residuals are readily defined for CLB and GLMB. Checking these residuals is an effective means to examine the reasonableness of fitting a particular model over the past claims data.
- GLMB offers a statistical measure for validating the model structure chosen. For BMCMC, BUGS also provides a similar measure for model validation in some cases.
- Some GLMs can be incorporated into BMCMC's mechanism.

11 Application of Reserving Methods to Claims Data

This section applies the reserving methods under discussion to a particular claims data set, which has been analysed in Taylor and Ashe (1983), Renshaw (1989), Verrall (1989, 1990, 1991b, 1999), and Mack (1993a). The expected value, the standard deviation or standard error of prediction, and the 75th percentile of the OCL of the data set are computed by applying these reserving methods. The risk margin is hence calculated and accommodates parameter error and process error. Consideration is also given to model specification error. The mathematical notation in this section is the same as that in the corresponding section about each reserving method.

11.1 Claims Data

The claims data set of incremental claim amounts used in Taylor and Ashe (1983) is shown below:

i \ j	1	2	3	4	5	6	7	8	9	10
1	357,848	766,940	610,542	482,940	527,326	574,398	146,342	139,950	227,229	67,948
2	352,118	884,021	933,894	1,183,289	445,745	320,996	527,804	266,172	425,046	
3	290,507	1,001,799	926,219	1,016,654	750,816	146,923	495,992	280,405		
4	310,608	1,108,250	776,189	1,562,400	272,482	352,053	206,286			
5	443,160	693,190	991,983	769,488	504,851	470,639				
6	396,132	937,085	847,498	805,037	705,960					
7	440,832	847,631	1,131,398	1,063,269						
8	359,480	1,061,648	1,443,370							
9	376,686	986,608								
10	344,014									

The reserving methods are used to estimate the missing lower right triangle, whose sum is the total OCL, and to calculate the volatility or uncertainty beneath the liability measurement. It is assumed that all claims will be settled in 10 periods of time and there are 10 periods of past claims data (i.e. $n = 10$).

11.2 Model Structures

For BMCMC, we introduce two model structures as stated below, based on Ntzoufras and Dellaportas (2002).

Log Link + Gamma

It is assumed that:

$X_{i,j} \sim \gamma\left(\mu_{i,j}, r_{i,j}\right)$, $\ln\left(\mu_{i,j}\right) = \nu + \alpha_i + \beta_j$, $r_{i,j} = \mu_{i,j}^2 \sigma^{-2}$, $\sigma^{-2} \sim \gamma\left(1,10^{-9}\right)$,

$\nu \sim \mathrm{N}\left(0, \sigma_\nu^2\right)$, $\sigma_\nu^{-2} \sim \gamma\left(1,10^{-2}\right)$, $\alpha_i \sim \mathrm{N}\left(0, \sigma_\alpha^2\right)$, $\sigma_\alpha^{-2} \sim \gamma\left(1,10^{-2}\right)$, $\alpha_1 = 0$,

$\beta_j \sim \mathrm{N}\left(0, \sigma_\beta^2\right)$, $\sigma_\beta^{-2} \sim \gamma\left(1,10^{-2}\right)$, and $\beta_1 = 0$.

Smoothing + Lognormal

It is assumed that:

$\ln\left(X_{i,j}\right) \sim \mathrm{N}\left(\mu_{i,j}, \sigma^2\right)$, $\mu_{i,j} = \nu + \alpha_i + \beta_j$, $\sigma^{-2} \sim \gamma\left(1,10^{-6}\right)$,

$\nu \sim \mathrm{N}\left(0, \sigma_\nu^2\right)$, $\sigma_\nu^{-2} \sim \gamma\left(1,10^{-2}\right)$, $\alpha_i = \alpha_{i-1} + \Delta_i$, $\Delta_i \sim \mathrm{N}\left(0, \sigma_\Delta^2\right)$, $\sigma_\Delta^{-2} \sim \gamma\left(1,10^{-2}\right)$,

$\alpha_1 = 0$, $\Delta_1 = 0$, $\beta_j \sim \mathrm{N}\left(0, \sigma_\beta^2\right)$, $\sigma_\beta^{-2} \sim \gamma\left(1,10^{-2}\right)$, and $\beta_1 = 0$.

In the two BMCMC model structures, $\mu_{i,j} = \mathrm{E}\left(X_{i,j}\right)$ ($\mu_{i,j} = \mathrm{E}\left(\ln\left(X_{i,j}\right)\right)$ in the second structure), ν is the parameter representing the overall average amount of $\ln\left(X_{i,j}\right)$, α_i and β_j are the parameters allowing for the accident period effect and the development period effect, and σ^2 , σ_ν^2 , σ_α^2 , σ_Δ^2, and σ_β^2 are the volatility terms.

The first model structure is similar to a GLM with log link function and gamma distribution. The second one is similar to a state-space model allowing α_i to evolve over time by smoothing α_i. Both model structures employ vague non-informative priors, with the fixed parameters chosen to ensure BUGS runs smoothly. Informative priors may be used instead if extra prior information is available.

For GLMB, the following three model structures are tested, as stated in Verrall (1991b, 1999).

Identity Link + Lognormal (from Verrall (1991b))

It is assumed that the linear relationship is $\eta_{i,j} = \nu + \alpha_i + \beta_j = \mu_{i,j}$ with $\alpha_1 = 0$ and $\beta_1 = 0$ and that the probability distribution is normal for $\ln\left(X_{i,j}\right)$. This model structure is effectively a general linear model on $\ln\left(X_{i,j}\right)$. $X_{i,j}$ is replaced with $\ln\left(X_{i,j}\right)$ in the maximum likelihood estimation and the bootstrapping process. The difference between the maximum likelihood estimator and the unbiased estimator of $X_{i,j}$ is explained in Appendix 5. The latter is applied here.

Log Link + Poisson (from Verrall (1999))

It is assumed that the linear relationship is $\eta_{i,j} = \nu + \alpha_i + \beta_j = \ln\left(\mu_{i,j}\right)$ with $\alpha_1 = 0$ and $\beta_1 = 0$ and that the probability distribution is (over-dispersed) Poisson for $X_{i,j}$.

Log Link + Gamma (from in Verrall (1999))

It is assumed that the linear relationship is $\eta_{i,j} = \nu + \alpha_i + \beta_j = \ln\left(\mu_{i,j}\right)$ with $\alpha_1 = 0$ and $\beta_1 = 0$ and that the probability distribution is gamma for $X_{i,j}$.

In the three GLMB model structures, $\eta_{i,j}$ is the linear predictor, ν is the parameter representing the overall average amount of $\ln\left(X_{i,j}\right)$, α_i and β_j are the parameters allowing for the accident period effect and the development period effect, and $\mu_{i,j} = \mathrm{E}\left(X_{i,j}\right)$ ($\mu_{i,j} = \mathrm{E}\left(\ln\left(X_{i,j}\right)\right)$ in the first structure). In addition, due to computational convenience, only Pearson residuals are used for the bootstrapping process of GLMB.

For KF, we introduce the following model structure.

Smoothing + Lognormal

It is assumed that:

$$\mathrm{E}\left(\ln\left(X_{i,j}\right)\right)=\nu+\alpha_i+\beta_j\ ,\ \alpha_1=0\ ,\ \beta_1=0\ ,$$

$$\theta(t)=\begin{bmatrix}\nu\\ \alpha_2\\ \alpha_3\\ \vdots\\ \alpha_t\\ \beta_2\\ \beta_3\\ \vdots\\ \beta_n\end{bmatrix},\ H(t)=\begin{bmatrix}1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & 1 & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1\end{bmatrix},$$

$$\Sigma(t)=\sigma_\varepsilon^2 I\ \text{, and }\Gamma(t)=\sigma_\gamma^2\begin{bmatrix}0 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & & & \\ & & & 1 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0\end{bmatrix}.$$

The parameters in $\theta(t)$ have the same meaning as above, the blank areas of $H(t)$ and $\Gamma(t)$ are zeros, the $(t-1)^{\text{th}}$ column of $H(t)$ has two 1's, the t^{th} column of $\Gamma(t)$ is σ_γ^2, and σ_ε^2 and σ_γ^2 are the volatility terms. This model structure leads to the smoothing of α_i, i.e. $\alpha_t=\alpha_{t-1}+\delta_t$, where δ_t is the t^{th} component of $\delta(t)$. For example, when $t = 4$,

$$Y(4)=\begin{bmatrix}\ln\left(X_{1,4}\right)\\ \ln\left(X_{2,3}\right)\\ \ln\left(X_{3,2}\right)\\ \ln\left(X_{4,1}\right)\end{bmatrix}\ \text{and}\ A(4)=\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdots\\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & \cdots\\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & \cdots\\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \cdots\end{bmatrix}.$$

11.3 Results

Expected Outstanding Claims Liability

Table 1 at the end of this section presents our estimation results of the expected OCL. Our results of CLB, GLMB (log link with Poisson and log link with gamma), and MACK are the same as those shown in Mack (1993a) and Verrall (1999). Our results of CLB, GLMB (log link with Poisson), and MACK are also the same as each other, as reasoned in Subsections 6.7 and 8.2.

The average of the nine calculated estimates of the expected total OCL is 18,717,555. The largest difference between the expected total OCL estimates is 1,260,328, which is between BMCMC (log link with gamma) and GLMB (log link with gamma). It is nearly 7% of the average. In addition, the pattern of the estimates across the accident periods does not vary significantly for different reserving methods.

Standard Deviation or Standard Error of Prediction

Table 2 shows our estimation results of the standard deviation or standard error of prediction of the OCL. Both parameter error and process error are incorporated in each method.

The estimates vary significantly across different methods. For the standard deviation or standard error of prediction of the total OCL, the average estimate is $2,620,292, and the maximum difference is 832,354 between CLB and GLMB (log link with Poisson). The difference is around 4% of the average expected total OCL estimate.

Furthermore, the pattern of the estimates across the accident periods changes considerably for different reserving methods. The pace of increase of the estimates down the accident periods of BMCMC (log link with gamma) and CLB is slower than that of

the other methods. Particularly, the three GLMB model structures provide very large figures for the most recent accident periods.

Table 3 expresses our estimation results of the standard deviation or standard error of prediction in Table 2 as a percentage of the corresponding estimates in Table 1 (strictly speaking, the former is the coefficient of variation). Our results under GLMB (log link with Poisson) are very close to those in Verrall (1999). Our results under MACK are the same as those in Mack (1993a).

Again, the percentages vary considerably for different methods, with the average being around 14% and the largest difference being nearly 5% between BMCMC (log link with gamma) and GLMB (log link with Poisson) for the standard deviation or standard error of prediction of the total OCL.

Risk Margin from Half of Standard Deviation or Standard Error of Prediction

Half of the coefficient of variation acts as a minimum in the risk margin percentage calculation, and our estimates are shown in Table 5. Effectively, half of the standard deviation or standard error of prediction represents the minimum risk margin level, of which our estimates are shown in Table 4. The aggregate risk margin percentages range approximately from 6% to 8%.

Table 6 shows the sums of the figures from Tables 1 and 4. For the expected total OCL estimate plus the aggregate risk margin, the average is 20,027,701, and the largest difference is 1,003,095 between BMCMC (smoothing with lognormal) and GLMB (log link with gamma). The difference is related to model specification error, which is around 5% of the average expected total OCL estimate.

Risk Margin from 75th Percentile

Table 7 exhibits our estimation results of the risk margin computed from the estimates of the 75th percentile of the OCL. Both parameter error and process error are accommodated under each method. The average aggregate risk margin is 1,709,136.

The maximum difference for the aggregate risk margin is 934,543 between BMCMC (log link with gamma) and the KF structure. It is around 5% of the average expected total OCL estimate. Moreover, the pace of increase of the estimates down the accident periods of the two BMCMC model structures is slower than that of the other methods. In particular, the three GLMB model structures and the KF structure provide very large figures for the most recent accident periods. Overall, these features are more obvious than those reflected in the risk margin estimates computed from the estimates of the standard deviation or standard error of prediction.

For all the methods, the aggregate risk margin from the 75th percentile is larger than the aggregate risk margin from the standard deviation or standard error of prediction, implying that the distribution of the total OCL is not excessively skewed. Hence the minimum risk margin level does not apply here.

Table 8 shows our estimation results of the risk margin computed from the estimates of the 75th percentile in Table 7 expressed as a percentage of the corresponding estimates in Table 1. The aggregate risk margin percentages range from about 8% to 13%.

Table 9 exhibits the sums of the figures from Table 1 and 7. For the expected total OCL estimate plus the aggregate risk margin, the average is 20,426,691, and the largest difference is 1,382,615 between GLMB (identity link with lognormal) and the KF structure. This difference is around 7% of the average expected total OCL estimate. This difference is related to model specification error and is larger than the 5% model specification error calculated previously. As such, assuming the true underlying 75th percentile lies within the aggregate estimates (19,675,636 – 21,058,252) in Table 9, an

additional risk margin of 3-4% (7% ÷ 2 = 3.5%) may then be appropriate if an allowance for model specification error is necessary when using the average aggregate estimate (20,426,691) from Table 9.

In addition, some methods produce larger expected OCL estimates than the others but smaller risk margin estimates, and vice versa. The effects offset each other to some extent and so the final reserving levels are not too different between these methods. For example, BMCMC (log link with gamma) produces the largest expected total OCL estimate but the smallest aggregate risk margin, while the KF structure provides the third lowest expected total OCL estimate but the largest aggregate risk margin. The final reserving levels of the two methods differ only by about 1%.

11.4 Approximation Formulae

Subsections 6.8 and 9.4 suggest some approximation formulae for GLMB and SCL to calculate the standard error of prediction. Tables 10 and 11 present a comparison of our estimates of the standard error of prediction between using the bootstrapping or simulation methods and the approximation formulae.

The results produced by the approximation formulae are fairly close to the results computed by the bootstrapping or simulation methods for two of the GLMB model structures and SCL.

Table 1 Expected Outstanding Claims Liability

i \ Methods	BMCMC (Log Link + Gamma)	BMCMC (Smoothing + Lognormal)	CLB	GLMB (Identity Link + Lognormal)	GLMB (Log Link + Poisson)	GLMB (Log Link + Gamma)	KF (Smoothing + Lognormal)	MACK	SCL
2	264,300	118,700	94,634	97,489	94,634	93,316	101,374	94,634	94,923
3	651,400	490,600	469,511	443,122	469,511	446,505	457,788	469,511	461,112
4	930,400	723,100	709,638	616,470	709,638	611,145	651,123	709,638	695,763
5	1,252,000	1,113,000	984,889	1,029,604	984,889	992,023	1,035,739	984,889	966,753
6	1,665,000	1,567,000	1,419,459	1,448,127	1,419,459	1,453,085	1,473,338	1,419,459	1,436,311
7	2,324,000	2,251,000	2,177,641	2,175,705	2,177,641	2,186,161	2,190,410	2,177,641	2,232,543
8	3,508,000	3,371,000	3,920,301	3,559,622	3,920,301	3,665,066	3,442,432	3,920,301	3,964,480
9	4,063,000	4,334,000	4,278,972	4,183,833	4,278,972	4,122,398	4,269,816	4,278,972	4,313,194
10	4,688,000	5,285,000	4,625,811	4,586,268	4,625,811	4,516,073	5,027,791	4,625,811	4,775,023
Total	**19,346,100**	**19,253,400**	**18,680,856**	**18,140,241**	**18,680,856**	**18,085,772**	**18,649,809**	**18,680,856**	**18,940,103**

Table 2 Standard Deviation (or Standard Error of Prediction*)

i \ Methods	BMCMC (Log Link + Gamma)	BMCMC (Smoothing + Lognormal)	CLB*	GLMB* (Identity Link + Lognormal)	GLMB* (Log Link + Poisson)	GLMB* (Log Link + Gamma)	KF (Smoothing + Lognormal)	MACK*	SCL
2	216,500	61,330	75,226	53,605	105,692	47,243	54,755	75,535	76,550
3	317,700	173,200	134,129	174,655	224,920	167,334	178,242	121,699	127,551
4	379,200	210,400	159,292	201,238	255,164	181,864	198,744	133,549	136,992
5	442,700	276,400	272,327	285,880	313,920	266,631	271,135	261,406	255,286
6	507,900	345,000	398,078	365,223	362,927	359,143	360,715	411,010	446,121
7	570,600	474,600	504,574	572,805	527,029	565,084	522,967	558,317	577,543
8	709,200	717,100	662,788	984,998	807,214	965,656	808,061	875,328	961,796
9	792,200	934,100	793,498	1,281,939	1,069,103	1,221,480	1,054,731	971,258	1,036,933
10	953,000	1,228,000	1,302,403	1,841,354	2,130,225	1,824,191	1,425,522	1,363,155	1,459,151
Total	**2,276,000**	**2,526,000**	**2,206,235**	**2,870,044**	**3,038,589**	**2,855,065**	**2,809,220**	**2,447,095**	**2,554,383**

Table 3 Coefficient of Variation

i \ Methods	BMCMC (Log Link + Gamma)	BMCMC (Smoothing + Lognormal)	CLB	GLMB (Identity Link + Lognormal)	GLMB (Log Link + Poisson)	GLMB (Log Link + Gamma)	KF (Smoothing + Lognormal)	MACK	SCL
2	81.9%	51.7%	79.5%	55.0%	111.7%	50.6%	54.0%	79.8%	80.6%
3	48.8%	35.3%	28.6%	39.4%	47.9%	37.5%	38.9%	25.9%	27.7%
4	40.8%	29.1%	22.4%	32.6%	36.0%	29.8%	30.5%	18.8%	19.7%
5	35.4%	24.8%	27.7%	27.8%	31.9%	26.9%	26.2%	26.5%	26.4%
6	30.5%	22.0%	28.0%	25.2%	25.6%	24.7%	24.5%	29.0%	31.1%
7	24.6%	21.1%	23.2%	26.3%	24.2%	25.8%	23.9%	25.6%	25.9%
8	20.2%	21.3%	16.9%	27.7%	20.6%	26.3%	23.5%	22.3%	24.3%
9	19.5%	21.6%	18.5%	30.6%	25.0%	29.6%	24.7%	22.7%	24.0%
10	20.3%	23.2%	28.2%	40.1%	46.1%	40.4%	28.4%	29.5%	30.6%
Total	**11.8%**	**13.1%**	**11.8%**	**15.8%**	**16.3%**	**15.8%**	**15.1%**	**13.1%**	**13.5%**

Table 4 Risk Margin from Half of Standard Deviation (or Standard Error of Prediction*)

i \ Methods	BMCMC (Log Link + Gamma)	BMCMC (Smoothing + Lognormal)	CLB*	GLMB* (Identity Link + Lognormal)	GLMB* (Log Link + Poisson)	GLMB* (Log Link + Gamma)	KF (Smoothing + Lognormal)	MACK*	SCL
2	108,250	30,665	37,613	26,802	52,846	23,621	27,377	37,768	38,275
3	158,850	86,600	67,064	87,327	112,460	83,667	89,121	60,849	63,775
4	189,600	105,200	79,646	100,619	127,582	90,932	99,372	66,774	68,496
5	221,350	138,200	136,163	142,940	156,960	133,316	135,568	130,703	127,643
6	253,950	172,500	199,039	182,611	181,463	179,571	180,357	205,505	223,060
7	285,300	237,300	252,287	286,402	263,514	282,542	261,484	279,158	288,772
8	354,600	358,550	331,394	492,499	403,607	482,828	404,030	437,664	480,898
9	396,100	467,050	396,749	640,969	534,552	610,740	527,366	485,629	518,466
10	476,500	614,000	651,201	920,677	1,065,113	912,095	712,761	681,577	729,575
Total	**1,138,000**	**1,263,000**	**1,103,117**	**1,435,022**	**1,519,294**	**1,427,533**	**1,404,610**	**1,223,547**	**1,277,192**

Table 5 Risk Margin Percentage from Half of Coefficient of Variation

i \ Methods	BMCMC (Log Link + Gamma)	BMCMC (Smoothing + Lognormal)	CLB	GLMB (Identity Link + Lognormal)	GLMB (Log Link + Poisson)	GLMB (Log Link + Gamma)	KF (Smoothing + Lognormal)	MACK	SCL
2	41.0%	25.8%	39.7%	27.5%	55.8%	25.3%	27.0%	39.9%	40.3%
3	24.4%	17.7%	14.3%	19.7%	24.0%	18.7%	19.5%	13.0%	13.8%
4	20.4%	14.5%	11.2%	16.3%	18.0%	14.9%	15.3%	9.4%	9.8%
5	17.7%	12.4%	13.8%	13.9%	15.9%	13.4%	13.1%	13.3%	13.2%
6	15.3%	11.0%	14.0%	12.6%	12.8%	12.4%	12.2%	14.5%	15.5%
7	12.3%	10.5%	11.6%	13.2%	12.1%	12.9%	11.9%	12.8%	12.9%
8	10.1%	10.6%	8.5%	13.8%	10.3%	13.2%	11.7%	11.2%	12.1%
9	9.7%	10.8%	9.3%	15.3%	12.5%	14.8%	12.4%	11.3%	12.0%
10	10.2%	11.6%	14.1%	20.1%	23.0%	20.2%	14.2%	14.7%	15.3%
Total	**5.9%**	**6.6%**	**5.9%**	**7.9%**	**8.1%**	**7.9%**	**7.5%**	**6.5%**	**6.7%**

Table 6 Expected Outstanding Claims Liability plus Risk Margin (from Half of Coefficient of Variation)

i \ Methods	BMCMC (Log Link + Gamma)	BMCMC (Smoothing + Lognormal)	CLB	GLMB (Identity Link + Lognormal)	GLMB (Log Link + Poisson)	GLMB (Log Link + Gamma)	KF (Smoothing + Lognormal)	MACK	SCL
2	372,550	149,365	132,247	124,292	147,480	116,937	128,751	132,401	133,198
3	810,250	577,200	536,576	530,449	581,971	530,172	546,909	530,361	524,887
4	1,120,000	828,300	789,284	717,089	837,220	702,077	750,495	776,412	764,259
5	1,473,350	1,251,200	1,121,052	1,172,544	1,141,848	1,125,339	1,171,306	1,115,592	1,094,396
6	1,918,950	1,739,500	1,618,498	1,630,739	1,600,923	1,632,657	1,653,695	1,624,964	1,659,372
7	2,609,300	2,488,300	2,429,927	2,462,108	2,441,155	2,468,703	2,451,893	2,456,799	2,521,315
8	3,862,600	3,729,550	4,251,695	4,052,121	4,323,908	4,147,894	3,846,462	4,357,965	4,445,378
9	4,459,100	4,801,050	4,675,721	4,824,802	4,813,524	4,733,138	4,797,182	4,764,601	4,831,660
10	5,164,500	5,899,000	5,277,012	5,506,945	5,690,923	5,428,169	5,740,552	5,307,388	5,504,599
Total	**20,484,100**	**20,516,400**	**19,783,973**	**19,575,263**	**20,200,150**	**19,513,305**	**20,054,419**	**19,904,403**	**20,217,294**

Table 7 Risk Margin from 75th Percentile

i \ Methods	BMCMC (Log Link + Gamma)	BMCMC (Smoothing + Lognormal)	CLB	GLMB (Identity Link + Lognormal)	GLMB (Log Link + Poisson)	GLMB (Log Link + Gamma)	KF (Smoothing + Lognormal)	MACK	SCL
2	111,300	25,800	51,394	32,336	57,137	28,006	31,465	24,126	50,818
3	185,600	89,500	94,083	103,668	122,246	110,548	117,897	70,276	76,873
4	227,600	112,000	89,297	122,072	157,136	121,187	147,370	81,272	87,213
5	260,000	157,000	181,591	165,595	203,295	157,577	187,673	150,209	173,062
6	309,000	192,000	243,276	231,819	218,842	229,313	226,635	231,578	280,574
7	360,000	287,000	335,706	383,015	338,252	331,342	393,798	323,120	317,838
8	450,000	399,000	430,564	612,912	565,611	588,921	574,051	519,525	691,775
9	492,000	529,000	514,480	671,100	691,025	825,031	705,978	574,882	657,398
10	607,000	686,000	943,503	1,008,558	1,337,353	1,118,220	1,135,970	764,771	821,473
Total	**1,473,900**	**1,576,600**	**1,502,161**	**1,535,396**	**2,001,318**	**1,626,050**	**2,408,443**	**1,545,192**	**1,713,164**

Table 8 Risk Margin Percentage from 75th Percentile

i \ Methods	BMCMC (Log Link + Gamma)	BMCMC (Smoothing + Lognormal)	CLB	GLMB (Identity Link + Lognormal)	GLMB (Log Link + Poisson)	GLMB (Log Link + Gamma)	KF (Smoothing + Lognormal)	MACK	SCL
2	42.1%	21.7%	54.3%	33.2%	60.4%	30.0%	31.0%	25.5%	53.5%
3	28.5%	18.2%	20.0%	23.4%	26.0%	24.8%	25.8%	15.0%	16.7%
4	24.5%	15.5%	12.6%	19.8%	22.1%	19.8%	22.6%	11.5%	12.5%
5	20.8%	14.1%	18.4%	16.1%	20.6%	15.9%	18.1%	15.3%	17.9%
6	18.6%	12.3%	17.1%	16.0%	15.4%	15.8%	15.4%	16.3%	19.5%
7	15.5%	12.7%	15.4%	17.6%	15.5%	15.2%	18.0%	14.8%	14.2%
8	12.8%	11.8%	11.0%	17.2%	14.4%	16.1%	16.7%	13.3%	17.4%
9	12.1%	12.2%	12.0%	16.0%	16.1%	20.0%	16.5%	13.4%	15.2%
10	12.9%	13.0%	20.4%	22.0%	28.9%	24.8%	22.6%	16.5%	17.2%
Total	**7.6%**	**8.2%**	**8.0%**	**8.5%**	**10.7%**	**9.0%**	**12.9%**	**8.3%**	**9.0%**

Table 9 Expected Outstanding Claims Liability plus Risk Margin (from 75th Percentile)

i \ Methods	BMCMC (Log Link + Gamma)	BMCMC (Smoothing + Lognormal)	CLB	GLMB (Identity Link + Lognormal)	GLMB (Log Link + Poisson)	GLMB (Log Link + Gamma)	KF (Smoothing + Lognormal)	MACK	SCL
2	375,600	144,500	146,027	129,825	151,771	121,322	132,838	118,760	145,741
3	837,000	580,100	563,594	546,790	591,758	557,053	575,685	539,787	537,985
4	1,158,000	835,100	798,935	738,542	866,774	732,333	798,493	790,910	782,976
5	1,512,000	1,270,000	1,166,480	1,195,199	1,188,184	1,149,601	1,223,411	1,135,098	1,139,815
6	1,974,000	1,759,000	1,662,736	1,679,947	1,638,301	1,682,398	1,699,973	1,651,037	1,716,885
7	2,684,000	2,538,000	2,513,346	2,558,721	2,515,892	2,517,503	2,584,208	2,500,761	2,550,382
8	3,958,000	3,770,000	4,350,865	4,172,534	4,485,912	4,253,987	4,016,483	4,439,826	4,656,255
9	4,555,000	4,863,000	4,793,453	4,854,933	4,969,997	4,947,429	4,975,794	4,853,855	4,970,592
10	5,295,000	5,971,000	5,569,314	5,594,826	5,963,164	5,634,293	6,163,761	5,390,582	5,596,497
Total	**20,820,000**	**20,830,000**	**20,183,016**	**19,675,636**	**20,682,174**	**19,711,823**	**21,058,252**	**20,226,048**	**20,653,266**

Table 10 Standard Error of Prediction (Bootstrapping or Simulation Methods vs Approximation Formulae)

	GLMB (Log Link + Poisson)		GLMB (Log Link + Gamma)		SCL	
i \ Methods	**Bootstrap**	**Approximate**	**Bootstrap**	**Approximate**	**Simulated[#]**	**Approximate**
2	105,692	110,099	47,243	45,166	76,550	79,316
3	224,920	216,042	167,334	160,556	127,551	128,364
4	255,164	260,871	181,864	177,624	136,992	141,130
5	313,920	303,549	266,631	254,470	255,286	261,081
6	362,927	375,012	359,143	351,334	446,121	428,767
7	527,029	495,376	565,084	526,287	577,543	599,632
8	807,214	789,957	965,656	941,319	961,796	1,002,091
9	1,069,103	1,046,508	1,221,480	1,175,943	1,036,933	1,018,366
10	2,130,225	1,980,091	1,824,191	1,667,387	1,459,151	1,386,330
Total	**3,038,589**	**2,945,646**	**2,855,065**	**2,702,701**	**2,554,383**	**2,543,762**

[#] The figures are sample standard deviation.

Table 11 Standard Error of Prediction in Percentage of Expected Outstanding Claims Liability (Bootstrapping or Simulation Methods vs Approximation Formulae)

	GLMB (Log Link + Poisson)		GLMB (Log Link + Gamma)		SCL	
i \ Methods	**Bootstrap**	**Approximate**	**Bootstrap**	**Approximate**	**Simulated[##]**	**Approximate**
2	111.7%	116.3%	50.6%	48.4%	80.6%	83.6%
3	47.9%	46.0%	37.5%	36.0%	27.7%	27.8%
4	36.0%	36.8%	29.8%	29.1%	19.7%	20.3%
5	31.9%	30.8%	26.9%	25.7%	26.4%	27.0%
6	25.6%	26.4%	24.7%	24.2%	31.1%	29.9%
7	24.2%	22.7%	25.8%	24.1%	25.9%	26.9%
8	20.6%	20.2%	26.3%	25.7%	24.3%	25.3%
9	25.0%	24.5%	29.6%	28.5%	24.0%	23.6%
10	46.1%	42.8%	40.4%	36.9%	30.6%	29.0%
Total	**16.3%**	**15.8%**	**15.8%**	**14.9%**	**13.5%**	**13.4%**

[##] The figures are sample standard deviation.

12 Conclusions

Overall, BMCMC offers a sophisticated mechanism for analysing the past claims data. Its rationale is straightforward and it allows various types of model structures. In addition, it is easy to implement with the help of BUGS, despite the few limitations mentioned previously. GLMB and KF also provide rigorous frameworks for claims analysis, which offer great flexibility in setting the parameters, the probability distributions, and many other aspects of the model structures. As these two methods involve many matrix operations, simulating with Excel VBA programming is relatively slower. Standard software such as SAS can be adopted for GLM modelling as a faster means of calculation. Adjustments can also be made for SCL to accommodate numerous desirable features, with the potential cost of making the model more intricate and intractable. In contrast, CLB and MACK are rather mechanical and less flexible, and they can be exploited to generate simple results for checking purposes.

As suggested by Friedman (1953), 'a hypothesis is important if it "explains" much by little, that is, if it abstracts the common and crucial elements from the mass of complex and detailed circumstances…'. The real world of general insurance is extremely complicated and the paramount function of a reserving method is to model the essential features, but not every aspect, of the liabilities. BMCMC, GLMB, KF, and SCL are versatile methods to depict and highlight the most important properties under various situations, especially in identifying the underlying trends along the accident period, the development period, and the calendar period. Nevertheless, the real trends could sometimes be covered up or distorted by noise in the data, and extreme care is necessary in confirming any past trends discovered and in assuming any future trends. CLB and MACK are rather restrictive in this regard, and they can be deemed as fast means to provide a rough picture of the overall situation.

There are several important issues when applying these methods in practice. First, the accuracy of the data needs thorough checking. 'Garbage in, garbage out' is always the first thing to avoid prior to embarking any calculations. No matter how sophisticated and

adaptable the method is, the final output is meaningless if the data are incorrect or misinterpreted. Moreover, when prior information such as expert opinion and industry experience are input, especially when there is insufficient data (say, beyond the latest development period of the past claims data), the reasonableness and relevance of the information require radical examination. One such situation is that if certain catastrophic events or large claims do not happen in the past, these types of potential liabilities are not reflected in the data, and so an explicit and reasonable allowance has to be made. In addition, there may be a tendency for practitioners to choose the 'most likely value' for certain figures or parameters. Strictly speaking, this value is indeed the mode, but not the statistical mean as required under the current regulations.

The past incremental claim amounts are generally adjusted for economic inflation (e.g. CPI and AWE) to current values before valuation. Sometimes it may be feasible to incorporate superimposed inflation (or even economic inflation) into modelling. As stated in Subsection 6.1, however, it is better to remove the inflation effects from the model, if these inflation effects can be reasonably determined, in order to reduce intricacy.

Decisions about the final parameter values require more than mechanically applying the methods. Practitioners need to assess the reasonableness of their final choices of the parameters. The reserving method applied can be stress-tested by changing the parameter values based on different scenarios, so as to check the sensitivity of the results to those parameters.

Caution is necessary when dealing with the tail of the claims run-off, i.e. the latest development periods of the earlier accident years or the development periods beyond what the past data has reached. The residuals of the tail are often volatile, and adjustment is hence required if they are used for say the bootstrapping process. Extrapolation beyond the latest development period of the data and determination of the corresponding volatility remain largely judgemental, which require caution and reasonableness.

Experience analysis between the estimates and the actual future payments has to be carried out in the long run, in order to validate the functionality of a reserving method and identify any necessary modifications. Afterall, as suggested in England and Verrall (2002), the effectiveness of a particular reserving method and modelling can be completely tested only if it is applied on a multitude of data triangles from various lines of business and companies and the estimated results are then compared with how the claims develop over time.

The techniques described in this paper can be similarly adopted for higher percentiles, in which more samples need to be generated by the simulation process. Moreover, the claim amounts can be adjusted for exposures, and the development period can be modified to allow for operational time if one believes the finalisation rate has changed. Other modifications are feasible as long as they do not contradict the underlying assumptions. Moreover, when the run-off data are not in a triangular form (e.g. trapezium), certain methods like BMCMC, GLMB, and KF can be readily adopted to model the data.

Finally, the reserving methods and their applications in this paper so far involve only parameter error and process error, and touch model specification error lightly. Calculating the final risk margin also requires contemplating the other types of uncertainty including data error, future trends variability, and reinsurance risk. As such, the way that the risk margin is calculated in this paper at best provides a minimum for the actual risk margin to be adopted in practice. (In addition, a chapter of the author's thesis provides further examination of model specification error based on hypothetical examples.)

As noted at the outset, this paper does not intend to provide an exhaustive list of reserving methods. Some omissions include credibility theory and modelling negative incremental claim amounts. Taylor (2000) provides an elaborate description of credibility theory in reserving. Verrall (1991a) and De Alba (2002) suggest some techniques to handle negative claim amounts.

Appendix 1

The following is an example of a probability distribution with its 75th percentile being insufficient for setting the risk margin when the skewness increases by a large extent. The coefficient of variation is used instead to obtain a more sensible result.

Assume $X_{i,j} \sim \gamma(\mu, r)$, in which the density function of $X_{i,j}$ is:

$$f_{X_{i,j}}\left(x_{i,j}\right) = \frac{r^r \mu^{-r} e^{-r\mu^{-1}x_{i,j}} x_{i,j}^{r-1}}{\Gamma(r)} \ .$$

The mean is μ, the standard deviation is $\mu/\sqrt{r}$, the coefficient of variation is $1/\sqrt{r}$, and the coefficient of skewness is $2/\sqrt{r}$. As r decreases, the coefficient of skewness increases.

Assume $\mu/r = 10{,}000$. The table below shows the change of (75th percentile – μ) as a percentage of μ and the change of one half of the coefficient of variation, as r decreases and the coefficient of skewness increases.

r	**Coefficient of Skewness**	**(75th Percentile – μ) / μ**	**$\frac{1}{2}$ Coefficient of Variation**	**Difference**
3.5	1.1	29%	27%	2%
3.0	1.2	31%	29%	2%
2.5	1.3	33%	32%	1%
2.0	1.4	35%	35%	0%
1.5	1.6	37%	41%	-4%
1.0	2.0	39%	50%	-11%

In this example, the 75th percentile method and the coefficient of variation method produce risk margin percentages of similar magnitude when the coefficient of skewness is small. When the coefficient of skewness becomes larger, however, the former method produces a much lower risk margin percentage comparatively.

Under some other situations, the 75th percentile can even be lower than the mean. As such, the new regulation has the minimum of one half of the coefficient of variation in place, to avoid the potential insufficiency of including only the 75th percentile in the risk margin calculation procedure.

Appendix 2*

	Normal	Poisson	Gamma	Inverse Gaussian
Notation	$\mathrm{N}(\mu,\sigma^2)$	$\mathrm{Pn}(\mu)$	$\gamma(\mu,r)$	$\mathrm{IG}(\mu,\sigma^2)$
Density Function	$\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2\sigma^2}(x-\mu)^2}$	$\frac{e^{-\mu}\mu^x}{x!}$	$\frac{r^r\mu^{-r}e^{-r\mu^{-1}x}x^{r-1}}{\Gamma(r)}$	$\frac{1}{\sigma\sqrt{2\pi}}x^{-\frac{3}{2}}e^{\frac{-1}{2x\mu^2\sigma^2}(x-\mu)^2}$
$a(\phi)$**	$\phi=\sigma^2$	$\phi=1$	$\phi=r^{-1}$	$\phi=\sigma^2$
$b(\theta)$	$\frac{\theta^2}{2}$	e^{θ}	$-\ln(-\theta)$	$-(-2\theta)^{\frac{1}{2}}$
$c(x,\phi)$	$-\frac{1}{2}\left(\frac{x^2}{\phi}+\ln(2\pi\phi)\right)$	$-\ln(x!)$	$\frac{1}{\phi}\ln\left(\frac{x}{\phi}\right)-\ln x-\ln\left(\Gamma\left(\frac{1}{\phi}\right)\right)$	$-\frac{1}{2}\left(\ln(2\pi\, x^3\phi)+\frac{1}{x\phi}\right)$
$\theta(\mu)$	μ	$\ln\mu$	$-\frac{1}{\mu}$	$-\frac{1}{2\mu^2}$
$b''(\theta)$***	1	$e^{\theta}=\mu$	$\frac{1}{\theta^2}=\mu^2$	$(-2\theta)^{-\frac{3}{2}}=\mu^3$

* This appendix is partly extracted from McCullagh and Nelder (1989).

** The weight *w* of a cell can be put into $a(\phi)$ as $a(\phi)=\phi/w$.

*** $\mathrm{Var}(X)=b''(\theta)a(\phi)$

Appendix 3

Asymptotically, the maximum likelihood estimators of the parameters are distributed as $\hat{v} \sim \mathrm{N}\left(v, \sigma_v^2\right)$, $\hat{\alpha}_i \sim \mathrm{N}\left(\alpha_i, \sigma_{\alpha_i}^2\right)$, $\hat{\beta}_j \sim \mathrm{N}\left(\beta_j, \sigma_{\beta_j}^2\right)$, and $\hat{\gamma}_{i+j-1} \sim \mathrm{N}\left(\gamma_{i+j-1}, \sigma_{\gamma_{i+j-1}}^2\right)$, in which σ_v^2, $\sigma_{\alpha_i}^2$, $\sigma_{\beta_j}^2$, and $\sigma_{\gamma_{i+j-1}}^2$ are the variance terms of the estimators. These variance terms and the covariance between the parameter estimators are computed from the covariance matrix estimated from maximum likelihood.

The hypothesis $v = v^*$ can be tested approximately by computing the test statistic $\left(\hat{v} - v^*\right)^2 / \hat{\sigma}_v^2 \sim \chi_1^2$, in which $\hat{\sigma}_v^2$ is the estimator of σ_v^2. The hypothesis is rejected if the test statistic is larger than the selected critical value (one-sided). The significance of *v* can be tested with the hypothesis $v = 0$. This test statistic applies equally to the other parameters. Likewise, the hypothesis $\alpha_1 = \alpha_2$ can be tested approximately by computing the test statistic $\left(\hat{\alpha}_1 - \hat{\alpha}_2\right)^2 / \hat{\sigma}_{\hat{\alpha}_1, \hat{\alpha}_2}^2 \sim \chi_1^2$, where $\hat{\sigma}_{\hat{\alpha}_1, \hat{\alpha}_2}^2 = \hat{\sigma}_{\alpha_1}^2 - 2\mathrm{Cov}\left(\hat{\alpha}_1, \hat{\alpha}_2\right) + \hat{\sigma}_{\alpha_2}^2$, $\hat{\sigma}_{\alpha_1}^2$ and $\hat{\sigma}_{\alpha_2}^2$ are the estimators of $\sigma_{\alpha_1}^2$ and $\sigma_{\alpha_2}^2$ respectively, and $\mathrm{Cov}\left(\hat{\alpha}_1, \hat{\alpha}_2\right)$ can be estimated from the covariance matrix computed.

The estimator of the dispersion parameter ϕ is defined as:

$$\hat{\phi} = \sum_{i+j \le n+1} \frac{w_{i,j}\left(X_{i,j} - \hat{X}_{i,j}\right)^2}{b''\left(\theta\left(\hat{X}_{i,j}\right)\right)\left(n_d - n_p\right)} = \frac{\sum_{i+j \le n+1} \left(residual_{i,j}^{\text{Pearson}}\right)^2}{n_d - n_p} \ .$$

The means to determine the hat matrix and the covariance matrix of the parameter estimators are stated in McCullagh and Nelder (1989).

Appendix 4

Under the circumstances described in Subsection 7.5, we find that the formulae (22) to (27) reduce to:

$P\left(t\middle|t-1\right)=P\left(t-1\middle|t-1\right)$,

$K\left(t\right)=P\left(t-1\middle|t-1\right)A'\left(t\right)\left(A\left(t\right)P\left(t-1\middle|t-1\right)A'\left(t\right)+\hat{\sigma}^{2}I\right)^{-1}$,

$\hat{\theta}\left(t\middle|t-1\right)=\hat{\theta}\left(t-1\middle|t-1\right)$,

$\hat{Y}\left(t\middle|t-1\right)=A\left(t\right)\hat{\theta}\left(t-1\middle|t-1\right)$,

$\hat{\theta}\left(t\middle|t\right)=\hat{\theta}\left(t-1\middle|t-1\right)+K\left(t\right)\left(Y\left(t\right)-\hat{Y}\left(t\middle|t-1\right)\right)$, and

$P\left(t\middle|t\right)=\left(I-K\left(t\right)A\left(t\right)\right)P\left(t-1\middle|t-1\right)$.

The formulae (28), (29) and (30) also reduce to:

$\hat{Y}\left(n+m\middle|n\right)=A\left(n+m\right)\hat{\theta}\left(n+1\middle|n\right)$,

$\mathrm{Cov}\left(Y\left(n+m\right)-\hat{Y}\left(n+m\middle|n\right)\right)=A\left(n+m\right)P\left(n+1\middle|n\right)A'\left(n+m\right)+\hat{\sigma}^{2}I$, and

$$\mathrm{Cov}\left(Y-\hat{Y}\right)=\begin{bmatrix}A\left(n+1\right)\\A\left(n+2\right)\\\vdots\\A\left(2n-1\right)\end{bmatrix}P\left(n+1\middle|n\right)\begin{bmatrix}A\left(n+1\right)\\A\left(n+2\right)\\\vdots\\A\left(2n-1\right)\end{bmatrix}'+\hat{\sigma}^{2}I\ .$$

Now suppose $A_{past}=\begin{bmatrix}A\left(1\right)\\A\left(2\right)\\\vdots\\A\left(n\right)\end{bmatrix}$.

For the Kalman filter algorithm to produce the maximum likelihood parameter estimates for a state-space model reduced to a general linear model, $P\left(0\middle|0\right)$ has to be set as

$q\hat{\sigma}^2\left(A'_{past}A_{past}\right)^{-1}$. The term $\hat{\sigma}^2\left(A'_{past}A_{past}\right)^{-1}$ is indeed the estimator of the covariance matrix of the parameter estimators in the maximum likelihood estimation procedure, and q is a number larger than zero.

Furthermore, we now show in the following that $P\left(n+1\middle|n\right)$ tends to $\hat{\sigma}^2\left(A'_{past}A_{past}\right)^{-1}$ if q is very large.

Since $I - K(t)A(t)$

$= I - P\left(t-1\middle|t-1\right)A'(t)\left(A(t)P\left(t-1\middle|t-1\right)A'(t)+\hat{\sigma}^2 I\right)^{-1}A(t)$

$= I - \left(P\left(t-1\middle|t-1\right)A'(t)A(t)+\hat{\sigma}^2 I\right)^{-1}P\left(t-1\middle|t-1\right)A'(t)A(t)$

(from the matrix formula $D\left(I+ED\right)^{-1}E = \left(I+DE\right)^{-1}DE$)

$= \left(\hat{\sigma}^{-2}P\left(t-1\middle|t-1\right)A'(t)A(t)+I\right)^{-1}$,

we can see that

$P\left(t\middle|t\right) = \left(I-K(t)A(t)\right)P\left(t-1\middle|t-1\right) = \left(\hat{\sigma}^{-2}P\left(t-1\middle|t-1\right)A'(t)A(t)+I\right)^{-1}P\left(t-1\middle|t-1\right)$.

If $P\left(0\middle|0\right) = q\hat{\sigma}^2\left(A'_{past}A_{past}\right)^{-1}$, then

$P\left(1\middle|1\right) = \left(q\left(A'_{past}A_{past}\right)^{-1}A'(1)A(1)+I\right)^{-1}q\hat{\sigma}^2\left(A'_{past}A_{past}\right)^{-1}$

$= q\hat{\sigma}^2\left(qA'(1)A(1)+A'_{past}A_{past}\right)^{-1}$,

$P\left(2\middle|2\right) = \left(q\left(qA'(1)A(1)+A'_{past}A_{past}\right)^{-1}A'(2)A(2)+I\right)^{-1}q\hat{\sigma}^2\left(qA'(1)A(1)+A'_{past}A_{past}\right)^{-1}$

$= q\hat{\sigma}^2\left(qA'(2)A(2)+qA'(1)A(1)+A'_{past}A_{past}\right)^{-1}$, … , and

$P\left(n\middle|n\right)$

$= \left(q\left(qA'(n-1)A(n-1)+qA'(n-2)A(n-2)+...+qA'(1)A(1)+A'_{past}A_{past}\right)^{-1}A'(n)A(n)+I\right)^{-1}$

$q\hat{\sigma}^2\left(qA'(n-1)A(n-1)+qA'(n-2)A(n-2)+...+qA'(1)A(1)+A'_{past}A_{past}\right)^{-1}$

$$= q\hat{\sigma}^2 \left(qA'(n) A(n) + qA'(n-1) A(n-1) + ... + qA'(1) A(1) + A'_{past} A_{past} \right)^{-1}$$

$$= \frac{q}{q+1} \hat{\sigma}^2 \left(A'_{past} A_{past} \right)^{-1} .$$

As such, $P(n+1|n) = \frac{q}{q+1} \hat{\sigma}^2 \left(A'_{past} A_{past} \right)^{-1}$, which tends to $\hat{\sigma}^2 \left(A'_{past} A_{past} \right)^{-1}$ when q is very large.

Appendix 5

Let $K_{i,j} = \ln\left(X_{i,j}\right)$, i.e. take the logarithm of the past claims data. For a general linear model on $K_{i,j}$ instead of $X_{i,j}$, the maximum likelihood estimator is $\hat{K}_{i,j}$ instead of $\hat{X}_{i,j}$.

Accordingly, the maximum likelihood estimator of $\mu_{i,j} = \mathrm{E}\left(X_{i,j}\right)$ for $i + j > n + 1$ is:

$$\hat{X}_{i,j} = \exp\left(\hat{K}_{i,j} + \frac{1}{2}\frac{n_d - n_p}{n_d}\hat{\phi}\right),$$

where n_d is the number of cells or data points in the upper left run-off triangle, n_p is the number of parameters to be estimated in the linear relationship, and $\hat{\phi}$ is the estimator of the dispersion parameter for $K_{i,j}$.

Verrall (1991b) proposes an unbiased estimator of $\mu_{i,j} = \mathrm{E}\left(X_{i,j}\right)$ for $i + j > n + 1$:

$$\tilde{X}_{i,j} = \exp\left(\hat{K}_{i,j}\right) g_{n_d - n_p}\left(\frac{1}{2}\left(1 - \vec{F}_{i,j}\left(\vec{F}'\vec{F}\right)^{-1}\vec{F}'_{i,j}\right)\hat{\phi}\right),$$

where $g_q\left(x\right) = \sum_{i=0}^{\infty}\frac{q^i\left(q+2i\right)}{q\left(q+2\right)\left(q+4\right)...\left(q+2i\right)}\frac{x^i}{i!}$, $\vec{F}$ is the matrix that describes how the future claims are related to the parameters, and $\vec{F}_{i,j}$ is the row of $\vec{F}$ that determines the parameters for $K_{i,j}$.

References

Australian Prudential Regulation Authority (ARPA 2002). *Prudential Standard GPS 210 and Guidance Note GGN 210.1*.

Bateup R. and Reed I. (2001). *Research and data analysis relevant to the development of standards and guidelines on liability valuation for general insurance*. The Institute of Actuaries of Australia and Tillinghast – Towers Perrin.

Collings S. and White G. (2001). *APRA risk margin analysis*. The Institute of Actuaries of Australia XIIIth General Insurance Seminar, Trowbridge Consulting.

De Alba E. (2002). *Claims reserving when there are negative values in the development triangle*. 37th Actuarial Research Conference, The University of Waterloo.

De Jong P. and Zehnwirth B. (1983). *Claims reserving, state-space models and the Kalman filter*. Journal of The Institute of Actuaries 110: 157-181.

De Jong P. (2003). *Forecasting general insurance liabilities*. Department of Actuarial Studies, Macquarie University.

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

Friedman M. (1953). *The methodology of positive economics*. Essays in Positive Economics. University of Chicago.

Geman S. and Geman D. (1984). *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*. IEE Transactions on Pattern Analysis and Machine Intelligence 6: 721-741.

Golan J. S. (2004). *The linear algebra a beginning graduate student ought to know*. Kluwer Academic Publishers.

Hart D. G., Buchanan R. A., and Howe B. A. (1996). *The actuarial practice of general insurance*. The Institute of Actuaries of Australia.

Hertig J. (1985). *A statistical approach to the IBNR-reserves in marine reinsurance*. ASTIN Bulletin 15: 171-183.

Houltram A. (2003). *Reserving judgement*. The Institute of Actuaries of Australia XIVth General Insurance Seminar.

Kalman R. E. (1960). *A new approach to linear filtering and prediction problems*. Journal of Basic Engineering 82: 340-345.

Mack T. (1993a). *Distribution-free calculation of the standard error of chain ladder reserve estimates*. ASTIN Bulletin 23(2): 213-225.

Mack T. (1993b). *Which stochastic model is underlying the chain ladder method?* XXIVth ASTIN Colloquim, Cambridge.

McCullagh P. and Nelder J. A. (1989). *Generalised linear models (3rd edition)*. Chapman and Hall, New York.

Ntzoufras I. and Dellaportas P. (2002). *Bayesian Modelling of outstanding liabilities incorporating claim count uncertainty*. North American Actuarial Journal 6(1): 113-128.

Pinheiro P. J. R., Andrade e Silva J. M., and de Lourdes Centeno M. (2000). *Bootstrap methodology in claim reserving*. Working Paper No. 11, Centre for Applied Maths to Forecasting & Economic Decision.

Renshaw A. E. (1989). *Chain ladder and interactive modelling (claims reserving and GLIM)*. Journal of The Institute of Actuaries 116: 559.

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

Taylor G. C. and Ashe F. R. (1983). *Second moments of estimates of outstanding claims*. Journal of Econometrics 23: 37-61.

Taylor G. (2000). *Loss reserving – an actuarial perspective*. Kluwer Academic Publishers.

Taylor G. (2002). *Chain ladder bias*. Research Paper Number 92, Centre for Actuarial Studies, The University of Melbourne.

Taylor G., McGuire G., and Greenfield A. (2003). *Loss reserving: past, present, and future*. Research Paper Number 109, Centre for Actuarial Studies, The University of Melbourne.

Taylor G. and McGuire G. (2004). *Loss reserving with GLMs: a case study*. Research Paper Number 113, Centre for Actuarial Studies, The University of Melbourne.

The Institute of Actuaries of Australia (IAA 2002). *Professional Standard PS 300 and Guidance Note GN 353*.

Verrall R. J. (1983). *Forecasting – the Bayesian approach*. Journal of The Institute of Actuaries 110: 183-203.

Verrall R. J. (1989). *A state space representation of the chain ladder linear model*. Journal of The Institute of Actuaries 116: 589-610.

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

Verrall R. J. (1991a). *Negative incremental claims: chain ladder and linear models*. Journal of The Institute of Actuaries 120: 171-183.

Verrall R. J. (1991b). *On the estimation of reserves from loglinear models*. Insurance: Mathematics and Economics 10: 75-80.

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

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

Wells C. (1996). *The Kalman filter in finance*. Kluwer Academic Publishers.

Zehnwirth B. (1996). *Kalman filters with applications to loss reserving*. Research Paper Number 35, Centre for Actuarial Studies, The University of Melbourne.