ELSEVIER

Insurance: Mathematics and Economics 26 (2000) 91–99

INSURANCE
MATHEMATICS & ECONOMICS

An investigation into stochastic claims reserving models and the chain-ladder technique

R.J. Verrall

Department of Actuarial Science and Statistics, City University, Northampton Square, London EC1V 0HB, UK

Received October 1998; received in revised form May 1999

Abstract

This paper considers a range of stochastic models which give the same reserve estimates as the chain-ladder technique. The relationship between the models described by Renshaw and Verrall (Renshaw, A.E., Verrall, R.J., 1998. British Actuarial Journal 4, 903–923) and Mack (Mack, T., 1993. ASTIN Bulletin 23, 213–225) is explored in more detail than previously. Several new models are suggested and some new ways to allow for negative incremental claims for the chain-ladder technique and other claims reserving methods are put forward. ©2000 Elsevier Science B.V. All rights reserved.

Keywords: Chain-ladder technique; Generalised linear models; Loss run-off triangles

1. Introduction

The chain-ladder technique is an algorithm for producing estimates of outstanding claims, ignoring any tail factors. In accordance with the chain-ladder technique, this paper does not consider projecting outstanding claims beyond the latest development year which has already been observed.

One of the reasons why the chain-ladder technique excites controversy is that there is no definitive source in which it is defined. It has simply emerged as an algorithm which can be used to produce reserve estimates. From the point of view of obtaining reserve estimates, this is of little significance: it is possible to define various models which give the same reserve estimates as the chain-ladder technique. Renshaw and Verrall (1998) and Mack (1993) are examples of papers which explore models which produce the same estimates of outstanding claims as the chain-ladder technique, although the generalised linear modelling approach of Renshaw and Verrall requires certain positivity constraints. These positivity constraints can be overcome by a reparameterisation of the model, and the reserve estimates of the chain-ladder technique reproduced in all cases. As England and Verrall (1999) have commented, the construction of models which merely produce the same reserve estimates as the chain-ladder technique “may seem like a futile exercise” at first sight. However, there are a number of reasons why this is not the case. Firstly, the chain-ladder technique gives only a point estimate of the outstanding claims and no indication of the likely variability of the actual outcome around the reserves. Secondly, in any data analysis exercise, it is important to understand the model

E-mail address: r.j.verrall@city.ac.uk (R.J. Verrall).

PII: S0167-6687(99)00038-4

being used and the data for which it is suitable. In order to go beyond the simple reserve estimates of the chain-ladder technique, it is necessary to specify a stochastic model. Mack (1993) and Renshaw and Verrall (1998) are examples of papers which do this, and this paper will give a number of other specifications of models which reproduce the chain-ladder estimates. The various models make different assumptions and it is hoped that this paper will clarify some of the issues confronting a practitioner when they have to decide which model to use to estimate the variability of the chain-ladder reserve estimates.

In contrast with the chain-ladder technique, the separation technique was defined first as a stochastic model by Verbeek (1972) and only after this was an easy algorithm produced by Taylor (1977), which is akin to the chain-ladder technique. The chain-ladder technique and the separation technique are very closely related (see Verrall, 1996) and the split between a stochastic model and an algorithm for producing point estimates applies to both cases. In the case of the separation technique, a stochastic model was defined first and a simple algorithm derived later. This is a logical way to proceed and it is possible to return to the stochastic model and adjust the assumptions in the light of the data observed. In the days of fast computers, simple algorithms have less importance. This is the philosophy of our approach: the most important thing is the stochastic model, and simple algorithms are useful only if computing facilities are not available. This does not imply that the connection between the stochastic model and the reserve estimates produced by the simple algorithm should be ignored.

We begin with a model which assumes that the data have a Poisson distribution. While this is unrealistic since it implies that the data must be positive integers, it is straightforward to relax this, and it makes the exposition easier to understand. Without loss of generality, we assume that the data consist of a triangle of incremental claims:

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

The cumulative claims are defined by:

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

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

2. The Poisson model

Consider the following model, which assumes a multiplicative structure for expected incremental claims.

$$C_{ij} \sim \text{independent Poisson with } E[C_{ij}]=x_i y_j \quad \text{and} \quad \sum_{k=1}^{n} y_k = 1. \tag{2.1}$$

Clearly $x_i = E[D_{in}]$, which is expected ultimate cumulative claims (up to the latest development year so far observed). The parameters $\left\{y_j:\ j=1,\ldots,n;\ \sum_{j=1}^{n} y_j=1\right\}$ will be referred to as the “column parameters”. The column parameters can be interpreted as the proportions of ultimate claims which emerge in each development year. This model can be reparameterised as follows:

$$C_{ij} \sim \text{Poisson with } E[C_{ij}]=\frac{z_i y_j}{\sum_{k=1}^{n-i+1} y_k} \quad \text{and} \quad \sum_{k=1}^{n} y_k = 1. \tag{2.2}$$

In this case $z_i = E[D_{i,n-i+1}]$, which is the expected value of cumulative claims up to the latest development year observed in accident year i. Using this parameterisation, the likelihood is:

$$L=\prod_{i=1}^{n}\prod_{j=1}^{n-i+1}\left[\frac{(z_i y_j/S_{n-i+1})^{C_{ij}}\,\mathrm{e}^{-z_i y_j/S_{n-i+1}}}{C_{ij}!}\right], \quad \text{where } S_m=\sum_{k=1}^{m} y_k. \tag{2.3}$$

Thus,

$$\begin{aligned} L &= \prod_{i=1}^{n}\left\{ z_i^{\sum_{j=1}^{n-i+1} C_{ij}} \exp\left(-\frac{z_i \sum_{j=1}^{n-i+1} y_j}{S_{n-i+1}} \right) \prod_{j=1}^{n-i+1}\left[\frac{(y_j/S_{n-i+1})^{C_{ij}}}{C_{ij}!} \right] \right\} \\ &= \prod_{i=1}^{n}\left\{ z_i^{D_{i,n-i+1}} \mathrm{e}^{-z_i} \prod_{j=1}^{n-i+1}\left[\frac{1}{C_{ij}!} (y_j/S_{n-i+1})^{C_{ij}} \right] \right\} \\ &= \prod_{i=1}^{n}\left\{ \frac{z_i^{D_{i,n-i+1}} \mathrm{e}^{-z_i}}{D_{i,n-i+1}!} \left[\frac{D_{i,n-i+1}!}{\prod_{i=1}^{n-i+1} C_{ij}!} \prod_{j=1}^{n-i+1} (y_j/S_{n-i+1})^{C_{ij}} \right] \right\}. \end{aligned} \tag{2.4}$$

L can be written as $L = L_D L_C$, where

$$L_D = \prod_{i=1}^{n}\left\{ \frac{z_i^{D_{i,n-i+1}} \mathrm{e}^{-z_i}}{D_{i,n-i+1}!} \right\}, \tag{2.5}$$

and

$$L_C = \prod_{i=1}^{n}\left[\frac{D_{i,n-i+1}!}{\prod_{i=1}^{n-i+1} C_{ij}!} \prod_{j=1}^{n-i+1} \left(\frac{y_j}{S_{n-i+1}} \right)^{C_{ij}} \right]. \tag{2.6}$$

Thus, L is the product of the likelihood of z_i and the likelihood for $\left\{ y_j:\ j=1,\ldots,n;\ \sum_{j=1}^{n} y_j = 1 \right\}$. The first term is the probability for $D_{i,n-i+1}$ and the second term is the probability for $\{C_{ij}: j=1,\ldots,n-i+1\}$, conditional on $D_{i,n-i+1}$. From this, it can be seen that the maximum likelihood estimate of z_i is $D_{i,n-i+1}$. It is also clear that the same estimates of $\left\{ y_j:\ j=1,\ldots,n;\ \sum_{j=1}^{n} y_j = 1 \right\}$ are obtained, whether L or L_C is maximised.

The connection between these estimates and the estimates of the development factors in the chain-ladder technique is as follows. It can be shown that the column parameters of this model, $\left\{ y_j:\ j=1,\ldots,n;\ \sum_{j=1}^{n} y_j = 1 \right\}$, are related to the development factors of the chain-ladder technique via $\lambda_j = S_j/S_{j-1}$ (see Verrall, 1991). Thus, the maximum likelihood estimates of the development factors, $\hat{\lambda}_j$, can be obtained from the maximum likelihood estimates of the column parameters in the multiplicative model, $\hat{y}_j$:

$$\hat{\lambda}_j = \frac{\sum_{k=1}^{j} \hat{y}_k}{\sum_{k=1}^{j-1} \hat{y}_k}.$$

It can be shown that this gives the same development factor estimates as the chain-ladder technique,

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} D_{ij}}{\sum_{i=1}^{n-j+1} D_{i,j-1}}. \tag{2.7}$$

Hence, in this context, the chain-ladder technique can be seen as an algorithm for producing the maximum likelihood estimates of the column parameters for either the unconditional likelihood of $\{C_{ij}: j=1,\ldots,n-i+1; i=1,\ldots,n\}$, L, or for the conditional likelihood, conditioning on $\{D_{i,n-i+1}: i=1,\ldots,n\}$, L_C. A number of points should be noted:

1. It should be emphasised here that this does not imply that it is necessary to view the chain-ladder estimates as arising from this model: there certainly exist other formulations which will give rise to the same estimates of the development factors.

2. It is straightforward to extend the model so that it applies to positive data, which does not necessarily consist solely of positive integers. This can be done by using the quasi-(log)-likelihood, and details are contained in Renshaw and Verrall (1998).
3. This model is not necessarily the one which should be used for all data sets. Again, Renshaw and Verrall (1998) contains some discussion of how to identify a suitable model to use in practice.
4. At this point, it is not clear how negative incremental claims should be dealt with. It is possible to estimate the column parameters, and the development factors when some of the incremental claims are negative, although certain software packages may have difficulty with this. We believe that the treatment of negative incremental claims is a very important subject and we defer further discussion until Section 5.
5. It is crucial to note that the same estimates of the development factors are obtained whether the unconditional likelihood, L, or the conditional likelihood, L_C is used. It is not possible to refer back to the original formulation of the chain-ladder technique to decide which likelihood to use, and hence it is necessary for a decision to be made by the practitioner.

This final point, number 5, is discussed in more detail in Section 4. First, we show how to write this model in recursive form.

3. A recursive model

The chain-ladder technique obtains the estimate of $D_{i,j}$ ($j > n-i+1$), from the observed value of $D_{i,n-i+1}$ in a recursive way:

$$\begin{array}{l}\hat{D}_{i,n-i+2} = \hat{\lambda}_{n-j+2} D_{i,n-i+1} \\ \hat{D}_{i,j} = \hat{\lambda}_j \hat{D}_{i,j-1}\end{array} \qquad (j = n-i+3, \ldots, n).$$

For comparison purposes, and also to inform the choice of model when some of the incremental claim amounts are negative, it is useful to consider the model in Section 2 in recursive form. In order to do this, we use a Bayesian formulation and concentrate on the estimation of z_i. For the purposes of simplicity of exposition, we consider just one row of data:

$$C_{i1}, C_{i2}, \ldots, C_{i,n-i+1}.$$

We drop the i suffix, and write the model for C_j given $z(j)$ as:

$$C_j | z(j) \sim \text{Poisson with mean } \frac{z(j) y_j}{S_j}, \tag{3.1}$$

where $z(j) = E[D_j]$ is the expected value of aggregate claims up to development year j. Note that in the recursive formulation of the model, it is necessary to attach the label j to z, since the definition of z changes as each datum is received. In this notation, the row parameter in Section 2 would be written $z_i(n-i+1)$. $z(j)$ can be related to $z(j-1)$ as follows:

$$z(j) = E\left[D_j\right] = E\left[D_{j-1}\right] + E\left[C_j\right] = z(j-1) + \frac{z(j) y_j}{S_j}.$$

Hence

$$z(j) = \frac{z(j-1)}{1-(y_j/S_j)} = \frac{z(j-1) S_j}{S_{j-1}}. \tag{3.2}$$

Thus, the conditional distribution of C_j given $z(j-1)$ is

$$C_j | z(j-1) \sim \text{ Poisson with mean } \frac{z(j-1) y_j}{S_{j-1}}. \tag{3.3}$$

Now, we use the conjugate prior gamma distribution for $z(j-1)$, and suppose we have the distribution of $z(j-1)$, conditional on the information received up to development year $j-1$:

$$z(j-1)|C_1, C_2, \ldots, C_{j-1} \sim \Gamma(\alpha, \beta), \tag{3.4}$$

for some parameters α and β which will be determined below. Using standard Bayesian analysis, the posterior distribution of $z(j-1)$ is

$$z(j-1)|C_1, C_2, \ldots, C_{j-1}, C_j \sim \Gamma\left(\alpha + C_j, \beta + \frac{y_j}{S_{j-1}}\right). \tag{3.5}$$

Since we have a relationship between $z(j)$ and $z(j-1)$, given in Eq. (3.2), we can obtain the distribution of $z(j)$, conditional on the information received up to development year j by a straightforward transformation as follows. If

$$z(j-1) \sim \Gamma(a, b) \quad \text{then} \quad z(j) \sim \Gamma\left(a, \frac{bS_{j-1}}{S_j}\right). \tag{3.6}$$

From (3.5) and (3.6),

$$z(j)|C_1, C_2, \ldots, C_j \sim \Gamma\left(\alpha + C_j, \left(\beta + \frac{y_j}{S_{j-1}}\right)\frac{S_{j-1}}{S_j}\right). \tag{3.7}$$

This completes a recursive estimation procedure, and we may now derive the distribution of $z(j)\,|\,C_1,C_2,\ldots,C_j$ (for all j) by considering first the case $j=1$. In order to do this, we require a suitable prior distribution for $z(1)$. A non-informative (improper) prior distribution is $f(z(1)) \propto z(1)^{-1}$, and $C_1\,|\,z(1) \sim$ Poisson with mean $z(1)$, since $y_1 = S_1$.

Again, a standard Bayesian analysis yields the posterior distribution of $z(1)$, conditional on $z(1)$:

$$z(1)|C_1 \sim \Gamma(C_1, 1). \tag{3.8}$$

This starts the recursion, and it straightforward to prove by induction, from (3.4) and (3.7) that

$$z(j)|C_1, C_2, \ldots, C_j \sim \Gamma(D_j, 1). \tag{3.9}$$

This is true for $j=1$, as can be seen from (3.8), suppose it is true for $j-1$. Then from (3.7)

$$z(j)|C_1, C_2, \ldots, C_j \sim \Gamma\left(D_{j-1} + C_j, \left(1 + \frac{y_j}{S_{j-1}}\right)\frac{S_{j-1}}{S_j}\right),$$

and hence $z(j)\,|\,C_1,C_2,\ldots,C_j \sim \Gamma(D_j,1)$, as required.

Now consider the distribution of C_j, conditional on the information received up to development year $j-1$. For this, it is necessary to integrate out $z(j-1)$as follows:

$$f(C_j|C_1, C_2, \ldots, C_{j-1}) = \int f(C_j|z(j-1)) f(z(j-1)|C_1, C_2, \ldots, C_{j-1}) \mathrm{d}z(j-1). \tag{3.10}$$

There is a subtle difference in the treatment here from Section 2. Since $z(j)$ is a random variable, C_j are only independent conditional on $z(j)$. It is well known that (3.10) gives a negative binomial distribution, in this case with parameters $p = (S_{j-1}/S_j)$ and $k = D_{j-1}$:

$$f(C_j|C_1, C_2, \ldots, C_{j-1}) = \frac{\Gamma(D_j)}{C_j!\,\Gamma(D_{j-1})}\left(\frac{S_{j-1}}{S_j}\right)^{D_{j-1}}\left(\frac{y_j}{S_j}\right)^{C_j}. \tag{3.11}$$

Thus, in recursive form, the mean and variance of C_j, conditional on the information received up to development year $j-1$ are

$$\frac{D_{j-1}y_j}{S_{j-1}} \quad \text{and} \quad \frac{D_{j-1}y_jS_j}{(S_{j-1})^2}, \quad \text{respectively.} \tag{3.12}$$

Again, this can be generalised by appealing to the quasi-(log)-likelihood, and it represents a further formulation of a stochastic model which will give the same estimates of outstanding claims as the chain-ladder technique.

Finally, we write the recursive model in terms of the development factors, λ_j $(=S_j/S_{j-1})$. $C_j\,|\,C_1,C_2,\ldots,C_{j-1}$ has mean and variance

$$(\lambda_j-1)D_{j-1} \quad \text{and} \quad \lambda_j(\lambda_j-1)D_{j-1}, \quad \text{respectively.} \tag{3.13}$$

Hence, noting that $D_j=D_{j-1}+C_j$, the mean and variance of $D_j\,|\,C_1,C_2,\ldots,C_{j-1}$ are

$$\lambda_jD_{j-1} \quad \text{and} \quad \lambda_j(\lambda_j-1)D_{j-1}, \quad \text{respectively.} \tag{3.14}$$

These moments define a recursive model which will reproduce the reserves given by the chain-ladder technique. Certain positivity constraints exist, and this is discussed further in Section 5. The fact that the model can be reparameterised as in (3.14), with the mean in a particularly simple form, is one of the key reasons why the chain-ladder technique is easy to apply. It happens because the column parameters represent separate factors, and can be replaced by another set of factors, the development factors. This point is discussed further in Section 6. We first consider the conditional and unconditional likelihoods in more detail.

4. The conditional and unconditional likelihoods

Section 2 has shown that the development factors of the chain-ladder technique can be obtained by using either the conditional likelihood, L_C , or the unconditional likelihood, L. If the conditional likelihood is used, then the latest cumulative claims in each row, $D_{i,n-i+1}$, are conditioned on. Thus, the estimates of outstanding claims can be obtained from the development factors, or the column parameters, and the values of $D_{i,n-i+1}$ (which have been conditioned on). If the unconditional likelihood is used, the same estimates of the column parameters (and hence the development factors) are obtained, and the maximum likelihood estimate of z_i is $D_{i,n-i+1}$. Thus, the estimates of the outstanding claims will be the same whichever approach is taken. The fitted values will be the same under the conditional likelihood or the unconditional likelihood. However, the standard errors of the predicted values, the root mean squared prediction errors and the measures of the variability of the reserves will be different in each case. If the conditional likelihood is used, it is only necessary to estimate the $n-1$ column parameters (equivalently the $n-1$ development factors). If the unconditional likelihood is used, there are n more parameters to estimate, making a total of $2n-1$ parameters. The additional parameters are $\{z_i=E[D_{i,n-i+1}]\colon i=1,2,\ldots,n\}$, the likelihood for z_i is $(z_i^{D_{i,n-i+1}}\mathrm{e}^{-z_i}/D_{i,n-i+1}!)$ and the maximum likelihood estimate of z_i is easy to obtain $(\hat{z}_i = D_{i,n-i+1})$. Thus, no additional computational work is required and it will not be clear whether z_i is being estimated or $D_{i,n-i+1}$ conditioned on, if a simple algorithm is used. This is certainly the case with the chain-ladder technique.

The implication of using the conditional likelihood, and conditioning on $\{D_{i,n-i+1}\colon i=1,\ldots,n\}$, is that these values are fixed. No other values would have been possible under the claims process being studied. If the unconditional likelihood is used, then $\{D_{i,n-i+1}\colon i=1,\ldots,n\}$ are treated as the realised values of the random variables which have been obtained in this case. Thus, the standard errors, reserve mean square prediction errors, etc. will be larger in the latter case. Thus, the difference between the two cases is that the former assumes that the row totals are the only ones which could have been attained, while in the latter, the row totals could have been different, but their expected values are estimated by the observed values.

It is possible to take the same approach as was adopted in Section 3 with the (non-recursive) model of Section 2, and integrate out the row parameter, x_i or z_i. Whichever is chosen gives the same distribution for C_{ij}. Using the model formulation of (2.2),

$C_{ij} \sim$ Poisson with

$$E\left[C_{ij}\right] = \frac{z_i y_j}{\sum_{k=1}^{n-i+1} y_k}$$

and from (3.9)

$$z(j)|C_1, C_2, \ldots, C_j \sim \Gamma(D_j, 1).$$

The recursive model used in Section 3 enables us to formulate a model in which the row totals are not conditioned on. Their estimates enter the model implicitly. Thus, integrating out z_i, C_{ij} has a negative binomial distribution, with parameters $p = (1/(1 + (y_j/S_{n-i+1})))$ and $k = D_{i,n-i+1}$. Hence, the mean and variance of C_{ij} are

$$\frac{D_{i,n-i+1} y_j}{\sum_{k=1}^{n-i+1} y_k} \quad \text{and} \quad \frac{D_{i,n-i+1} y_j}{\sum_{k=1}^{n-i+1} y_k}\left(1 + \frac{y_j}{\sum_{k=1}^{n-i+1} y_k}\right), \text{ respectively.} \tag{4.1}$$

Thus, the column parameters (equivalently the development factors) can be estimated from either (3.13), (3.14) or (4.1). These represent the same model written in different forms. Clearly, the recursive formulation has the simplest form. In particular the mean for cumulative claims is straightforward, and this is one of the reasons why the chain-ladder technique appears simple and easy to understand. However, the implications of using the chain-ladder technique, in terms of the variability of the reserve, are not obvious unless an analysis such as that carried out in this paper is done. Also, the chain-ladder technique copes with negative incremental claims, in that estimates of development factors and reserves can still be produced. Again, the implications of using the chain-ladder technique for data containing negative incremental claims requires close examination, and this is the subject of the next section.

5. Treatment of negative incremental claims

We begin this section by restating the recursive model derived in Section 3, and reinstating the notation of Section 2, which applies to data suffixed by i and j. $C_{ij} \mid C_{i1}, C_{i2}, \ldots, C_{i,j-1}$ has mean and variance

$$(\lambda_j - 1) D_{i,j-1} \quad \text{and} \quad \lambda_j(\lambda_j - 1) D_{i,j-1}, \quad \text{respectively,} \tag{5.1}$$

and the mean and variance of $D_{ij} \mid C_{i1}, C_{i2}, \ldots, C_{i,j-1}$ are

$$\lambda_j D_{i,j-1} \quad \text{and} \quad \lambda_j(\lambda_j - 1) D_{i,j-1}, \quad \text{respectively.} \tag{5.2}$$

Notice that if $\lambda_j < 1$, the variance is negative and the model breaks down. $\lambda_j < 1$ implies that incremental claims in column j (or at least some of them) are negative. This is the point at which estimation for data which includes negative incremental claims can break down. Clearly, the assumptions of the stochastic model underlying (5.1) and (5.2) have been violated. The recursive distribution of the data has to be adjusted to allow for the data received, which does not support the stochastic model being used. It is necessary to use a distribution whose support is not restricted to the positive real line, and a suitable candidate is the normal distribution. We can imagine that some refinements of this are likely to be suggested, to allow for the fact that the distribution of the data is unlikely to be symmetrical. However, it is possible to replace (5.2) by a normal distribution, whose mean is unchanged, but whose variance is altered to accommodate the case when $\lambda_j < 1$. Preserving as much of $\lambda_j(\lambda_j - 1)D_{i,j-1}$ as possible, we would expect the variance to be proportional to $D_{i,j-1}$, with the constant of proportionality depending on j. This gives $D_{ij} \mid C_{i1}, C_{i2}, \ldots, C_{i,j-1}$ is approximately normally distributed, with mean and variance

$$\lambda_j D_{i,j-1} \quad \text{and} \quad \phi_j D_{i,j-1}, \quad \text{respectively.} \tag{5.3}$$

This is equivalent to the model of Mack (1993), except that Mack regarded it as a non-parametric model. An important point to notice is that the variability of the estimate of the expected row total is still included in the variance of this estimate. The only time this is not the case is in the conditional model, with likelihood L_C. In other words, $D_{i,j-1}$ is an estimate of $z_i(j-1)$, and is not being conditioned on.

Of the two recursive models, (5.2) and (5.3), the first is preferable because it is not necessary to estimate the parameters in the variance. Of course, the distributional assumptions underlying (5.2) should be checked in the usual way, by, for example, examining appropriate residuals. This is part of the modelling process, and it may lead to the modelling distribution being altered. The model defined by (5.3), and used by Mack (1993) represents an extreme case of this, in which the form of the variance has to be abandoned and the variance estimated from the data. This is necessary when the claims data have been contaminated by, for example, reinsurance recoveries, accounting procedures, or when incurred data (incorporating estimates of claims) have been used.

The non-recursive model (4.1), can also be adjusted to accommodate negative incremental claims by replacing it with the following. C_{ij} is approximately normally distributed, with mean and variance

$$\frac{D_{i,n-i+1}y_j}{\sum_{k=1}^{n-i+1} y_k} \quad \text{and} \quad \phi_j D_{i,n-i+1}, \text{ respectively}. \tag{5.4}$$

In the cases of the models in which the variance is defined in terms of new sets of parameters, (5.3) and (5.4), these parameters must be estimated from the data. This means that these models have more parameters than models such as (2.1), or the models considered in Section 3. Clearly, this is a disadvantage of these models, but it may be a price that must be paid in order to deal with negative incremental claims. If the data satisfy the positivity conditions of Renshaw and Verrall (1998), then that approach is preferable. Alternatively, a model using a (quasi) negative binomial likelihood in either recursive or non-recursive form could be adopted. Mack (1993) provides one method of estimating the extra parameters needed in the variance in (5.3), which is not necessarily the one which we would employ. However, these models are all robust to negative incremental claims and give the same reserve estimates as the chain-ladder technique.

6. Conclusions

This paper has considered a range of stochastic models which give the same reserve estimates as the chain-ladder technique. It has been shown that the model used by Renshaw and Verrall (1998), reparameterised as a multiplicative model, can be written in recursive form. This recursive model has been related to the approach adopted by Mack (1993), and it is shown that the latter paper includes an estimate of the expected cumulative claims, up to the latest accident year so far observed, in the estimation of outstanding claims.

This paper has also shown that the chain-ladder estimates of the development factors may be regarded as the maximum likelihood estimates from the unconditional model, or from the conditional model, conditioning on the latest cumulative claims in each accident year. It is argued that the standard errors and measures of reserve variability from the recursive model, and the approach of Mack (1993) do not condition on the latest cumulative claims in each accident year.

Which approach should be taken in practice? We do not believe that one particular model should be used in all situations. The flexibility of generalised linear models is a great advantage which allows the practitioner to explore different underlying distributions and run-off shapes in order to get a good understanding of the data. This can be done easily with standard statistical software, and it is our belief that this represents the best approach for data which satisfies the positivity requirements. Data sets, such as incurred claims, which contain a significant number of negative incremental claims require careful consideration. It can be seen that the model defined in (5.3) has the simplest structure and will be the easiest model to fit. The non-recursive model may have some advantages: it may be easier, for example, to find a parametric run-off shape for the parameters y_j than for the parameters λ_j. We do not think that a model should be recommended simply because it can be fitted to all data sets: the practitioner should examine the assumptions of the model and consider carefully the implications for the data and the reserves. Nevertheless, the following generalised linear model can be fitted to all data:

$$C_{i,j} \sim N((\lambda_j - 1)D_{i,j-1}, \phi_{ij} D_{i,j-1}), \tag{6.1}$$

$$\phi_{ij} \sim \text{Gamma}, \text{ with log link function and linear predictor } \gamma_j. \tag{6.2}$$

Cumulative claims could be used instead of incremental claims (it makes no difference):

$$D_{i,j} \sim N(\lambda_j D_{i,j-1}, \phi_{ij} D_{i,j-1}). \tag{6.3}$$

It should be noted that any model which has a multiplicative structure can be written in the recursive form given by (4.1). For example, the Hoerl curve, with $E[C_{ij}] = x_i y_j$, where

$$y_j = \alpha^j \exp(\beta(j-1)), \tag{6.4}$$

can be written in recursive form. The link ratios in this case can be written as:

$$\lambda_j = \frac{\sum_{k=1}^{j} y_k}{\sum_{k=1}^{j-1} y_k}, \tag{6.5}$$

where λ_j is now not a factor, but is parameterised according to (6.4). It may be simpler to consider a model where the delay is treated as continuous, and use ratios of integrated gamma functions:

$$\lambda_j = \frac{\int_0^j \alpha^x \exp(\beta(x-1))\mathrm{d}x}{\int_0^{j-1} \alpha^x \exp(\beta(x-1))\mathrm{d}x}. \tag{6.6}$$

The chain-ladder technique is simple to fit because, when using a factor in each column, rather than a parametric form such as the Hoerl curve, the model is simple to reparameterise. In other words, (6.5) can be used to estimated the parameters λ_j and the original column parameters can be reconstructed easily from these. When considering (6.6), the model is non-linear in terms of the parameters α and β, unless some simplified (linear) version of (6.6) can be found. It is clear also that $\lambda_j \geq 1$, when it is defined by (6.6), a restriction which does not apply to (6.3). Thus, the chain-ladder technique allows negative incremental claims. Of course, it is possible that a simple adjustment to (6.6) could be found to allow a Hoerl-type run-off shape to be fitted to data containing negative incremental claims.

It is hoped that this paper has explained why the chain-ladder technique works, and has cast some further light on the advantages and disadvantages of using it. It is also hoped that the model represented in (6.1)–(6.3) will be useful in practice, and that further development of models useful to practitioners (particularly for data containing negative incremental claims) will be forthcoming from the approach used here.

References

England, P.D., Verrall, R.J., 1999. Analytic and bootstrap estimates of prediction error in claims reserving. Insurance: Mathematics and Economics 26, 281–293.

Mack, T., 1993. Distribution-free calculation of the standard error of chain ladder reserve estimates. ASTIN Bulletin 23, 213–225.

Renshaw, A.E., Verrall, R.J., 1998. A stochastic model underlying the chain-ladder technique. British Actuarial Journal 14, 903–923.

Taylor, G.C., 1977. Separation of inflation and other effects from the distribution of non-life insurance claim delays. ASTIN Bulletin 9, 217–230.

Verbeek, H.G., 1972. An approach to the analysis of claims experience in motor liability excess-of-loss reinsurance. ASTIN Bulletin 6, 195–203.

Verrall, R.J., 1991. Chain ladder and maximum likelihood. Journal of the Institute of Actuaries 118, 489–499.

Verrall, R.J., 1996. Claims Reserving. Notes for a Short Course presented at City University.