Cash flow simulation for a model of outstanding liabilities based on claim amounts and claim numbers

María Dolores Martínez Miranda
University of Granada, Campus Fuentenueva, 18071, Granada, Spain
mmiranda@ugr.es
Bent Nielsen
Nuffield College, Oxford OX1 1NF, U.K.
bent.nielsen@nuffield.ox.ac.uk
Jens Perch Nielsen
Cass Business School, City University London, 106 Bunhill Row, London EC1Y 8TZ, U.K.
Jens.Nielsen.1@city.ac.uk
and Richard Verrall
Cass Business School, City University, London
r.j.verrall@city.ac.uk

6 September 2010

Abstract

In this paper we develop a full stochastic cash flow model of outstanding liabilities for the model developed in Verrall, Nielsen and Jessen (2010). This model is based on the simple triangular data available in most non-life insurance companies. By using more data, it is expected that the method will have less volatility than the celebrated chain ladder method. Eventually, our method will lead to lower solvency requirements for those insurance companies that decide to collect counts data and replace their conventional chain ladder method.

1 Introduction

While non-life insurance companies often base their reserves on a simple method, such as a chain-ladder estimate for either paid data or incurred data, Verrall, Nielsen and Jessen (2010) recently pointed out that a model combining a classical paid triangle with another triangle of the same format containing the number of reported claims has the advantage of making a clear split between the RBNS reserve and the IBNR reserve. In this paper we consider the distributional properties of this model and compare it with the chain ladder model. It is to be expected that the additional information of the count triangle should lower the volatility of estimated reserves: when more information is available, the predictions should be better. We also note that the RBNS forecasts use the actual numbers of claims, and can therefore be considered to be based on a conditional model of the claim amounts, given the numbers of claims. On the other hand, the IBNR forecasts require, by definition, forecasts of the numbers of claims. We also examine in more detail some of the estimation issues associated with the model developed by Verrall, Nielsen and Jessen (2010).

The model of Verrall, Nielsen and Jessen (2010) separates the reporting delay from the payment delay from reporting to payment: the reporting delay is observed in the reported counts triangle and the delay observed in the paid triangle is a mixture of the reporting delay and the payment delay. The payment delay and the claims severity are estimated from a conditional model of the paid triangle given the counts triangle. In this paper, we exploit this new method in order to split the full stochastic cash flow into one cash flow related to the RBNS reserve and another related to the IBNR reserve. The RBNS reserve is predicted without modelling the counts but instead exploiting the fact that the reported counts give an advance warning of future payments. The IBNR reserve is constructed by first predicting the reported number of claims and then applying the same conditional model for payments given counts as outlined above. Thus, in this paper we construct bootstrap estimates of the predictive distributions total reserve and its split into RBNS and the IBNR reserves.

Recent related methods are also discussed in Pinheiro, Andrade e Silva and Centeno (2003) and Björkwall, Hössjer and Ohlsson (2009a). It is worth noting though, that these bootstrap methods are aimed at predicting the claims distribution for models which use just the aggregate claims data, and one can therefore expect them to generate higher solvency requirements than the approach of this paper that takes full advantage of the extra information provided by the reported number of claims. Another related idea for combining the reported counts and the paid triangles is the separation method of Taylor (1977). Here the idea is to predict the incurred counts from the count triangle using a chain ladder model and then to apply these predictions in the construction of the reserves. This approach is somewhat similar to the construction of the IBNR reserve in the cash flow but more model dependent than the RBNS reserve of the cash flow model which is constructed conditionally on the counts data. Recently, Björkwall, Hössjer and Ohlsson (2009b) have constructed conditional bootstrap estimators for such a separation method.

Section 6 contains the explanation of the bootstrap method for the model of Verrall, Nielsen and Jessen (2010). In sections 3,4 and 5 we summarise the model and explain some alterations in the set-up and estimation methods. The aim of this is to make the bootstrapping procedure as straightforward as possible to implement. It should be noted that the underlying structure of the model and the basic philosophy of the approach remain the same. In section 7, we consider some simulation studies which are designed to illustrate how the model behaves in general.

2 The data

This paper uses the same motor data as Verrall, Nielsen and Jessen (2010), which originates from the general insurer RSA and is based on a portfolio of motor third party liability policies. These data typically have long settlement delays, and the cashflow model in this paper is aimed at improved stochastic modelling of data of this type. The data available consists of two incremental run-off triangles of dimension $m = 10$, one for reported counts, N_{ij}, and one for aggregated payments, X_{ij}, where $i = 1, \ldots, m$ denotes the accident year and $j = 0, \ldots, m-1$ is the development year. Collectively, we have data $N = \{N_{ij}, (i,j) \in \mathcal{I}\}$ and $X = \{X_{ij}, (i,j) \in \mathcal{I}\}$ where $\mathcal{I}$ is the triangular index set

$$\mathcal{I} = \{i = 1, \ldots, m, j = 0, \ldots, m-1 \text{ with } i + j = 1, \ldots, m\}.$$

The data are shown in Tables 1 and 2, respectively. The data for the aggregate paid claims have been inflation corrected using an inflation index depending on the calendar year $i + j$. This adjustment was carried out using an external economic inflation index before the data were supplied to the authors. It is not known when the payments aggregated as X_{ij} were first reported.

3 The statistical model

As set out in Verrall, Nielsen and Jessen (2010), the statistical model has two ingredients: a conditional model for payments given incurred counts along

$i \setminus j$	0	1	2	3	4	5	6	7	8	9
1	6238	831	49	7	1	1	2	1	2	3
2	7773	1381	23	4	1	3	1	1	3	
3	10306	1093	17	5	2	0	2	2		
4	9639	995	17	6	1	5	4			
5	9511	1386	39	4	6	5				
6	10023	1342	31	16	9					
7	9834	1424	59	24						
8	10899	1503	84							
9	11954	1704								
10	10989									

Table 1: Run-off triangle of number of reported claims, N_{ij}

$i \setminus j$	0	1	2	3	4	5	6	7	8	9
1	451288	339519	333371	144988	93243	45511	25217	20406	31482	1729
2	448627	512882	168467	130674	56044	33397	56071	26522	14346	
3	693574	497737	202272	120753	125046	37154	27608	17864		
4	652043	546406	244474	200896	106802	106753	63688			
5	566082	503970	217838	145181	165519	91313				
6	606606	562543	227374	153551	132743					
7	536976	472525	154205	150564						
8	554833	590880	300964							
9	537238	701111								
10	684944									

Table 2: Run-off triangle of aggregated payments, X_{ij}

with a model for incurred counts. In this section, we give a summary of the model and describe a number of differences in the set-up and estimation which simplify Verrall, Nielsen and Jessen (2010) without making any significant changes to the overall structure.

3.1 The conditional model for payments given counts

The key feature of the cash flow model of Verrall, Nielsen and Jessen (2010) is to identify the payment delay through a conditional model for the payments X given the reported counts N. The reported claims from accident year i and development year j will be paid with some delay, and those paid with delay k are denoted by N_{ijk}^{paid} $(k = 0, 1, \ldots, d)$. Here d denotes the maximum delay in paying a claim which has been reported, which in the context of the estimation in this paper is defined to be such that $d \leq m - 1$. While we recognise that delays greater than $m - 1$ could be envisaged in practice, this would require other estimation methods which are beyond the scope of this paper. Also, it would simplify the model to put $d = m - 1$, but we have found that it is often better to set $d < m - 1$ in practice, as discussed in section 4.2. The overall, but latent, number of payments in development year j is therefore

$$N_{ij}^{paid} = \sum_{k=0}^{\min(j,d)} N_{i,j-k,k}^{paid}. \tag{1}$$

The observed aggregate payment in development year j is then

$$X_{ij} = \sum_{k=1}^{N_{ij}^{paid}} Y_{ij}^{(k)} \tag{2}$$

where $Y_{ij}^{(k)}$ denotes an individual claim payment.

We assume that the delays are independent of the counts, and are assumed to take the values $0, 1, \ldots, d$ with probabilities $p_0, p_1, \ldots, p_d$, where $\sum_{k=0}^{d} p_k = 1$. In principle, these probabilities could depend on accident year and development year, but in this paper we refrain from this further complication.

As in Verrall, Nielsen and Jessen (2010), it is assumed that the individual payments are independent of delays and counts and are identically distributed with expectation μ and variance σ^2. While we recognise that this assumption may be unrealistic, we leave possible extensions to other models for future work. The methodology is quite flexible about the distribution for individual payments. Verrall, Nielsen and Jessen (2010) used a mixed-type distribution which allowed for the possibility of zero claims, and we use this set-up in section 4.2 when examining the delay distribution. In section 6, we illustrate the bootstrap procedures using a simpler assumption that the payments are gamma distributed (with no allowance for the possibility of zero claims). In practice, additional external information may be available on (for instance) the frequency of zero claims and negative claims as well as on the tail behaviour of the claims.

The conditional expectation and variance of claims given the reported counts were computed in Verrall, Nielsen and Jessen (2010):

$$\mathsf{m}_{ij}(N) = \mathsf{E}(X_{ij} \mid N) = \sum_{k=0}^{\min(j,d)} N_{i,j-k}\, \mu p_k, \tag{3}$$

$$\mathsf{v}_{ij}(N) = \mathsf{Var}(X_{ij} \mid N) = \sum_{k=0}^{\min(j,d)} N_{i,j-k} \{\sigma^2 p_k + \mu^2 p_k (1 - p_k)\}. \tag{4}$$

These formulas are used for the estimation of the parameters, and the conditional expectation (3) is used to calculate a point forecasts of the RBNS reserve (using the incurred counts) and of the IBNR reserve (using predicted counts).

3.2 The model for counts

The model for the counts N will be used for predicting the IBNR counts while it has no bearing on the predictions of the RBNS reserve. For the data in Table 1 a standard Poisson chain-ladder model seems reasonable. A generalisation could be to include a calendar effect as in Zehnwirth (1994) and the recent analysis in Kuang, Nielsen and Nielsen (2008a,b, 2010). Bryden and Verrall (2009) also discuss calendar year effects in the context of the chain-ladder technique.

The variables N_{ij} are therefore assumed to be independently Poisson distributed with expectation

$$\log\{\mathsf{E}(N_{ij})\} = \mu_{ij} = \alpha_i + \beta_j,$$

so that α_i is an accident year parameter and β_j is a development year parameter. The maximum likelihood analysis leads to the standard chain ladder analysis as shown by Kremer (1985), for example. Recently Kuang, Nielsen and Nielsen (2009) have revisited the maximum likelihood analysis and shown that the row sums and development factors of the chain ladder analysis are maximum likelihood estimators for the parameters.

4 Estimation

In this section, we summarise the estimation of the parameters, based on the theory of Verrall, Nielsen and Jessen (2010). We start by reviewing chain ladder estimation for the incurred counts. This is followed by a discussion of the estimation for the delay parameters $\psi_k = \mu p_k$. Finally, estimators for the individual payment parameters μ and σ^2 are given.

4.1 Chain ladder estimation of the model for counts

The model for the counts N is a standard Poisson chain ladder model for which maximum likelihood analysis was given by Kremer (1985). The row sums and the development factors,

$$R_i = \sum_{k=0}^{m-i} N_{ik}, \qquad F_\ell = \frac{\sum_{i=1}^{m-\ell} \sum_{j=0}^{\ell} N_{ij}}{\sum_{i=1}^{m-\ell} \sum_{j=0}^{\ell-1} N_{ij}}, \quad 1 \leq \ell \leq m-1, \tag{5}$$

are maximum likelihood estimators for the parameters $\rho_i = \mathsf{E}(R_i)$ and the development factors, Φ_ℓ, see Kuang, Nielsen and Nielsen (2009).

The fitted values are denoted by $\widehat{N}_{ij}$, and for use later we define the ratios

$$\widehat{B}_j = \widehat{N}_{ij}/\widehat{N}_{i0} = \begin{cases} (F_j - 1) \prod_{k=1}^{j-1} F_k & j \geq 2, \\ F_1 - 1 & j = 1, \\ 1 & j = 0, \end{cases} \tag{6}$$

which do not depend on the row index i, see also Kuang, Nielsen and Nielsen (2009, eq. 14).

4.2 Estimating the delay and payment means

Considering next the triangle of paid claims, the delay parameters and the payment mean are estimated from the conditional model for the payments X given incurred counts N as in Verrall, Nielsen and Jessen (2010). The idea is to estimate the parameters through a Poisson regression of payments X on incurred counts N using the conditional mean function $\mathsf{m}_{ij}(N) = \mathsf{E}(X_{ij} \mid N)$ given in (3). The parameters $\psi_k = \mu p_k$ are then estimated by maximising the pseudo likelihood

$$\ell^{pseudo}(\psi; X, N) = \sum_{i,j \in \mathcal{I}} \{X_{ij} \log \mathsf{m}_{ij}(N) - \mathsf{m}_{ij}(N)\}. \tag{7}$$

Based on the estimators $\widehat{\psi}_k$, $k = 0, \ldots, d$ the mean of the claims distribution and the delay probabilities are estimated by

$$\widehat{\mu} = \sum_{k=0}^{d} \widehat{\psi}_k \qquad \widehat{p}_k = \widehat{\psi}_k / \widehat{\mu}. \tag{8}$$

Table 3, first row of first panel, reports the estimates for the data in section 2.

It can immediately be seen that there is a numerical difficulty here, since it is possible that not all values, $\widehat{p}_k$ (or equivalently $\widehat{\psi}_k$), are positive. In

	$\widehat{p}_0$	$\widehat{p}_1$	$\widehat{p}_2$	$\widehat{p}_3$	$\widehat{p}_4$	$\widehat{p}_5$	$\widehat{p}_6$	$\widehat{p}_7$	$\widehat{p}_8$	$\widehat{p}_9$	$\widehat{\mu}$
N	.361	.286	.113	.084	.066	.035	.026	.012	.018	-.001	163.53
	.361	.286	.112	.084	.065	.035	.026	.012	.017		163.69
$\widehat{N}$	.361	.287	.112	.085	.066	.035	.026	.012	.018	-.001	163.43
	.361	.287	.112	.084	.066	.035	.026	.012	.018	0	163.62

Table 3: Pseudo likelihood estimators of the delay probabilities p_k based on data in Tables 1, 2. Panel 1 uses actual counts with maximal delay of $d = m - 1 = 9$ and $d = m - 2 = 8$, respectively. Panel 2 uses predicted counts and the analytic formula (in Row 2 the last entry is replaced by zero).

p_9	1.3e-6	5.4e-3	0.018
$\mathsf{P}(\widehat{p}_9 < 0)$	72%	33%	10%

Table 4: The frequency of zero estimates of p_9 is simulated for different values of p_9 using 1000 repetitions. The delay probabilities in the first column were chosen as p=(0.360,0.288,0.111,0.083,0.066,0.035,0.025,0.013,0.016,1.3e-6). In the second column the last p_9 was substituted by $5.4e-3$ and p_8 was slightly modified so the p_k's sum to one. Finally, the third column considers p=(0.182,0.164,0.145,0.127,0.109,0.091,0.073,0.055,0.036,0.018). The claims distribution considered in the three cases was a gamma (with mean $\mu = 204.91$ and variance $\sigma^2 = 2589440$) mixed with 20% zeros. The counts were kept fixed as in Table 1.

practice, this may arise since the delay probabilities will tend to tail off so that for instance p_{m-1} will be close to zero, with the result that the estimate may be negative in some cases. One response is to impose the restriction that the maximal delay is shorter, for instance $d = m - 2$. Table 3, second row of first panel, reports the estimates for the motor data, and it can be seen that, for this particular data set, there is not much difference between the results. The relative differences are largest, up to 5%, for the longest delays, but in absolute terms the differences are modest, up to 0.1%.

We investigated the chance of negative estimates by simulation, and Table 4 reports the simulated probability that the analytic estimator of the longest delay ψ_9 is negative for different values of ψ_9. The probability of negative estimates was similar using other estimation methods but varies considerably with ψ_9. This indicates that the possibility of negative estimates will typically be an issue in practice.

One approach to deal with negative delay estimates would be to use a constrained optimization routine to ensure that all estimators are non-negative. Having the subsequent bootstrap in mind we suggest a pragmatic estimator, which is numerically less intensive. If the sum of absolute values of negative $\widehat{\psi}_k$ is less than 1% of the sum of absolute values of all $\widehat{\psi}_k$ then the negative estimates are replaced by zero. If the sum of negative estimates is larger than this threshold it may be useful to investigate whether the paid data have special features such as many zeros.

4.3 Analytic estimation of delay parameters

Again keeping the bootstrap procedure in mind, we suggest an analytic estimator of the delay parameters as a numerically less costly alternative to the above iterative procedure. In this, we depart from Verrall, Nielsen and Jessen (2010). While the above estimation procedure conditions on the actual count data the idea of the alternative is to exploit a possible chain ladder structure for the counts data. We expect this to work well as long as the counts data do not deviate much from the chain ladder model, by for instance having a significant calendar effect. This analytic method only works when $d = m-1$, which is what is assumed in this section. In practice, it is necessary to check for negative delay parameter values and set these to zero, as discussed at the end of section 4.2.

Thus, the proposal is to replace the observed counts N in the pseudo likelihood (7) by the fitted counts $\widehat{N}$ from a chain ladder model. In general, information can be lost in a regression model when replacing regressors by predicted regressors. However, this loss will be small when the difference $N-\widehat{N}$ is small, which is not an unreasonable assumption in a Poisson context where expectation equals variance. Moreover, the count data come from aggregation over many policies which should improve their precision. In this paper, we use the analytical method for all our calculations because it makes our extensive simulation study of the numerically complex bootstrapping procedure possible.

Recalling that the ratios $\widehat{B}_j = \widehat{N}_{ij}/\widehat{N}_{i0}$ do not depend on the row index i (see (6)) the conditional expectation evaluated at the predictor has chain ladder structure:

$$\mathsf{m}_{ij}(\widehat{N}) = \sum_{k=0}^{j} \widehat{N}_{i,j-k}\psi_k = \widehat{N}_{i0}\zeta_j \quad \text{where} \quad \zeta_j = \sum_{k=0}^{j} \widehat{B}_{j-k}\psi_k.$$

Evaluating the pseudo log likelihood (7) at $\widehat{N}$ therefore gives

$$\ell^{pseudo}(\psi; X, \widehat{N}) = \sum_{i,j\in\mathcal{I}} X_{ij}\log(\widehat{N}_{i0}) + \sum_{j=1}^{m-1}\{\log(\zeta_j)\sum_{i=1}^{m-j} X_{ij} - \zeta_j \sum_{i=1}^{m-j}\widehat{N}_{i0}\}.$$

This pseudo log likelihood has its maximum at

$$\widehat{\zeta}_j = \frac{\sum_{i=1}^{m-j} X_{ij}}{\sum_{i=1}^{m-j}\widehat{N}_{i0}}. \tag{9}$$

The estimators for the parameters ψ_k then solve the linear system

$$\begin{pmatrix} \widehat{\zeta}_0 \\ \vdots \\ \vdots \\ \widehat{\zeta}_{m-1} \end{pmatrix} = \begin{pmatrix} \widehat{B}_0 & 0 & \cdots & 0 \\ \widehat{B}_1 & \widehat{B}_0 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ \widehat{B}_{m-1} & \cdots & \widehat{B}_1 & \widehat{B}_0 \end{pmatrix} \begin{pmatrix} \widehat{\psi}_0 \\ \vdots \\ \vdots \\ \widehat{\psi}_{m-1} \end{pmatrix}. \tag{10}$$

The second panel of Table 3 shows delay estimates using, first, the analytic estimator as it is, and, secondly, in combination with the pragmatic rule to deal with negative estimates. For this particular data set there is not much difference between any of the reported estimates.

4.4 Estimating the claims variance

The claims variance can be estimated by inserting the estimators $\widehat{\psi}_k$ in the conditional expectation (3) to get $\widehat{\mathsf{m}}_{ij}(N) = \sum_{k=0}^{\min(j,d)} N_{i,j-k}\widehat{\psi}_k$ and computing the over-dispersion statistic

$$\widehat{\varphi} = \frac{1}{df}\sum_{i,j\in\mathcal{I}} \frac{\{X_{ij} - \widehat{\mathsf{m}}_{ij}(N)\}^2}{\widehat{\mathsf{m}}_{ij}(N)}. \tag{11}$$

Here, the degrees of freedom are $df = n - q$ where $n = m(m+1)/2$ is the dimension of X and $q = d + 1$ is the number of estimated delay parameters.

This statistic could be viewed as an estimator of

$$\varphi = \frac{1}{n}\sum_{i,j\in\mathcal{I}} \frac{\mathsf{v}_{ij}(N)}{\mathsf{m}_{ij}(N)} = \frac{\sigma^2+\mu^2}{\mu} - \frac{\mu}{n}\sum_{i,j\in\mathcal{I}} \frac{\sum_{k=0}^{\min(j,d)} N_{i,j-k}p_k^2}{\sum_{k=0}^{\min(j,d)} N_{i,j-k}p_k}, \qquad (12)$$

recalling the expressions for the conditional mean and variance of X_{ij} given N in (3), (4). A consistency argument could possibly be made in which the number of rows was increased in the index set $\mathcal{I}$ while the number of columns is kept fixed. The variance estimator implied by (11), (12) is

$$\widehat{\sigma}^2 = \widehat{\mu}\widehat{\varphi} - \widehat{\mu}^2 + \frac{\widehat{\mu}^2}{n}\sum_{i,j\in\mathcal{I}} \frac{\sum_{k=0}^{\min(j,d)} N_{i,j-k}\widehat{p}_k^2}{\sum_{k=0}^{\min(j,d)} N_{i,j-k}\widehat{p}_k}. \qquad (13)$$

This estimator is slightly different from the variance estimator $\widehat{\sigma}^2_{VNJ} = \widehat{\mu}\widehat{\varphi} - \widehat{\mu}^2$ given in Verrall, Nielsen and Jessen (2010). However, if φ is much larger than μ as for the present data set the difference between the two variance estimators is modest.

4.5 Summary of estimates for motor data

Table 5 gives an overview of the estimates from the motor data. The estimates $\widehat{\psi}_k$ are obtained from the last row of Table 3, that is by the analytic estimator combined with the zero rule of thumb. The sum of these estimators is $\widehat{\mu}$. The estimates $\widehat{p}_k$ are computed as $\widehat{\psi}_k/\widehat{\mu}$, and the variance $\widehat{\sigma}^2$ is obtained using (13).

5 Point forecasts of the reserves

Point forecasts of the reported but not settled (RBNS) reserve and the incurred but not reported (IBNR) reserve can now be constructed along the lines of Verrall, Nielsen and Jessen (2010). As a benchmark for comparison purposes, we also consider the chain ladder reserve, and discuss the construction of this first in section 5.1.

5.1 Point forecasts of the chain ladder reserve

Forecasting using the chain ladder technique assumes that the aggregated payment triangle X indexed by the upper triangle $\mathcal{I}$ has a chain ladder structure, which is extrapolated to the lower triangle of payments indexed by

$$\mathcal{J}_1 = \{i = 1, \ldots, m, j = 0, \ldots, m-1 \text{ where } i + j = m+1, \ldots, 2m-1\}.$$

The index sets $\mathcal{I}$ and $\mathcal{J}_1$ are illustrated in Figure 1.

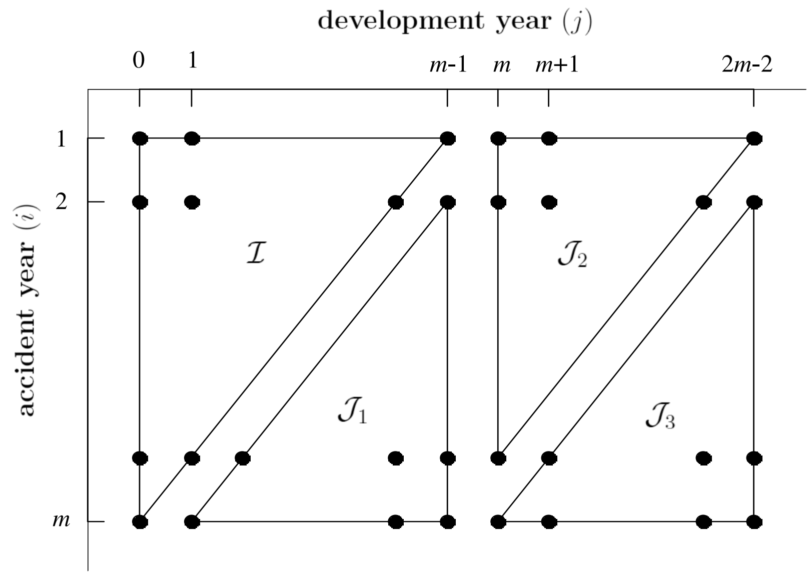

Figure 1: Index sets for reserves.

The cash flow predicted by the chain ladder is shown in the last column of Table 6. The cash flow by calendar year is computed by summing the point forecasts $\tilde{X}_{ij}$ along the diagonals of $\mathcal{J}_1$. Table 6 also shows the RBNS and IBNR forecasts which are discussed in further detail below. Note that while the chain ladder forecast is comparable to that of the sum of the RBNS and IBNR forecasts, it is indeed somewhat smaller. This is in contrast to our simulated results below, where the median of the sum of the RBNS and IBNR reserves estimated by the method of this paper is more or less the same as that of the classical chain ladder reserve. Therefore, it appears that it is the particular sample at hand that causes this difference, and it is not a general feature of the new method. Similar conclusions were noted in Verrall, Nielsen and Jessen (2010).

5.2 Point forecasts of the RBNS reserve

Forecasting the RBNS reserve by the cash flow model assumes that the payments relating to the incurred counts are delayed as described in §3.1. These forecasts vary over the index set $\mathcal{J}_1$ as well as the index set

$$\mathcal{J}_2 = \{i = 1, \ldots, m; j = m, \ldots 2m - 2 \text{ where } i + j = m + 1, \ldots, 2m - 1\},$$

illustrated in Figure 1.

The point forecasts are constructed from the conditional expectation $\mathsf{m}_{ij}(N)$ in (3). Recognising that the counts are only available in the upper triangle $\mathcal{I}$ and inserting the estimates $\widehat{\psi}_\ell$ gives the point forecasts

$$\tilde{\mathsf{m}}_{ij}(N) = \sum_{k=j-m+i}^{\min(j,d)} N_{i,j-k}\widehat{\psi}_k, \tag{14}$$

over the index set $\mathcal{J}_1 \cup \mathcal{J}_2$. The RBNS cash flow by calendar year is computed by summing the point forecasts along the diagonals of $\mathcal{J}_1 \cup \mathcal{J}_2$.

Point forecasts of the RBNS cash flow by calendar year for the motor data are shown in the first column of Table 6. Note that the cash flow for calendar year 19 is zero as the last delay parameter ψ_9 is set to zero.

5.3 Point forecasts of the IBNR reserve

The IBNR forecasts are constructed in two stages. First, predictions of the incurred but not reported counts, $\tilde{N}_{ij}$ are computed over the index set $\mathcal{J}_1$. Secondly, these predictions are inserted in the expression $\tilde{\mathsf{m}}_{ij}(N)$ to get the IBNR point forecasts

$$\tilde{\mathsf{m}}_{ij}(\tilde{N}) = \sum_{k=\max(0,j-m+1)}^{\min(d,j-m+i-1)} \tilde{N}_{i,j-k}\widehat{\psi}_k. \tag{15}$$

Due to the delay, these point forecasts run over the index sets $\mathcal{J}_1$, $\mathcal{J}_2$ as well as the index set

$$\mathcal{J}_3 = \{i = 1, \ldots, m; j = m, \ldots, 2m - 2 \text{ where } i + j = 2m, \ldots, 3m - 2\},$$

illustrated in Figure 1.

The IBNR cash flow is shown in the last column of Table 6. The cash flow by calendar year is computed by summing the point forecasts along the diagonals of $\mathcal{J}_1 \cup \mathcal{J}_2 \cup \mathcal{J}_3$. As the last delay parameter ψ_9 is set to zero the cash flow for calendar year 28 is zero.

6 Bootstrapping the predictive distribution including parameter uncertainty

In this section, we explain the bootstrapping procedure which can be applied to the model set out above. It should be noted that the term 'bootstrapping' can be used to cover a wide range of approaches. For example, it is sometimes used in connection with procedures that just simulate the process distribution. However, it is more common (especially in the actuarial literature) to use it when the estimation error is also included, and this is the context in which it is used in this paper. For completeness, we will also mention the former case in section 6.2, but all the results will include the estimation error. A further distinction in bootstrapping methodology is between parametric and non-parametric bootstrapping. Again, in the actuarial literature, it is more common to encounter non-parametric bootstrapping, where (for example) the residuals of the model are resampled. However, it is possible also to use parametric bootstrapping, and the choice may depend on the particular properties of the model being considered. In the case of the model in this paper, parametric bootstrapping is more appropriate and is used in the remainder of this section. The results of this parametric bootstrapping estimation procedure are compared with non-parametric bootstrapping applied to the chain-ladder technique in section 6.3. In section 7, a simulation study compares the conditional bootstrapping method with the classical unconditional chain ladder method.

6.1 The predictive distribution

We first introduce some notation for the predictive distributions of the RBNS and IBNR reserves, which will be estimated by bootstrapping.

The reported counts N_{ij} are indexed over $\mathcal{I}$. Their distribution is denoted $\mathsf{N}_{\mathcal{I}}(\omega)$ and is Poisson distributed with mean given in terms of the population version ω of the row sums R_i and development parameters F_ℓ.

The distribution of the aggregated claims X_{ij} over $\mathcal{I} \cup \mathcal{J}_1 \cup \mathcal{J}_2$ arising from the incurred counts N_{ij} is denoted $\mathsf{X}_{ij}(\theta, N)$, where $\theta = (p, \mu, \sigma^2)$. This distribution is constructed sequentially. Given the incurred counts, the paid counts N_{ij}^{paid} are defined over the set $\mathcal{I} \cup \mathcal{J}_1 \cup \mathcal{J}_2$ through the formula (1). The individual claims distribution (or the severity distribution) is assumed to be a gamma distribution with mean μ and variance σ^2. Therefore, the shape parameter is $\lambda = \mu^2/\sigma^2$, the scale parameter is $\kappa = \sigma^2/\mu$, and the density is

$$\mathsf{f}(y) = \frac{1}{\gamma(\lambda)\kappa^\lambda} y^{\lambda-1} \exp(-y/\kappa) \qquad \text{for } y > 0.$$

Note that the possibility of zero claims is excluded, in contrast to assumption in Verrall, Nielsen and Jessen (2010). Given the count N_{ij}^{paid}, the aggregate claims X_{ij} are then gamma distributed with shape $N_{ij}^{paid}\lambda$ and scale κ.

The RBNS reserve is the sum over $\mathcal{J}_1 \cup \mathcal{J}_2$ of the aggregate claims arising from the reported counts N, that is $\tilde{\mathsf{m}}_{ij}(N)$ as given in (14).

The IBNR reserve arises from the incurred but not reported counts $\widetilde{N}_{ij}$ over the lower triangle $\mathcal{J}_1$. These are Poisson distributed distributed in a similar way to the reported counts, and in accordance with the notation above, their distribution is denoted $\mathsf{N}_{\mathcal{J}_1}(\rho, \Phi)$. The aggregated claims over $\mathcal{J}_1 \cup \mathcal{J}_2 \cup \mathcal{J}_3$ arising from the predicted counts $\widetilde{N}$, $\tilde{\mathsf{m}}_{ij}(\widetilde{N})$ as given in (15), will then have a mixture distribution $\mathsf{X}_{ij}\{\theta, \mathsf{N}_{\mathcal{J}_1}(\omega)\}$.

The total reserve is found by adding the RBNS and the IBNR reserves.

6.2 Bootstrap predictive distribution of RBNS and IBNR cash flow

The predictive reserve distributions will be estimated using a parametric bootstrapping procedure. As mentioned above, the term 'bootstrapping' is sometimes used to describe the situation where the unknown parameters are simply replaced by the estimated parameters (ignoring the estimation uncertainty). This would give the bootstrap estimators

$$\mathsf{RBNS}(\widehat{\theta}, N), \qquad \mathsf{IBNR}\{\widehat{\theta}, \mathsf{N}_{\mathcal{J}_1}(\widehat{\omega})\}, \qquad \mathsf{Total}\{\widehat{\theta}, N, \mathsf{N}_{\mathcal{J}_1}(\widehat{\omega})\}. \tag{16}$$

The more usual bootstrapping procedure, taking parameter uncertainty into account, is defined as follows.

The delay and severity parameters $\theta = (p, \mu, \sigma^2)$ are estimated using the conditional model of aggregrated payments X given reported counts N, while the chain ladder parameters ω are estimated using the model for the reported counts N. For the bootstrap, this can be replicated by considering these two distributions varying independently in spaces Θ and Ω say. The conditional distribution given N of the estimators of the delay and severity parameters is denoted $\mathsf{D}_\theta(\theta^*; N)$ while the distribution of the estimators of the chain ladder parameters is denoted $\mathsf{C}_\omega(\omega^*)$. Hence the bootstrap distributions of the reserves are the mixtures

$$\mathsf{RBNS}_{mix}(\theta, N) = \int_{\theta^* \in \Theta} \mathsf{RBNS}(\theta^*, N) d\mathsf{D}_\theta(\theta^*; N), \tag{17}$$

$$\mathsf{IBNR}_{mix}\{\theta, \mathsf{N}_{\mathcal{J}_1}(\omega)\} = \int_{(\theta^*, \omega^*) \in (\Theta, \Omega)} \mathsf{IBNR}\{\theta^*, \mathsf{N}_{\mathcal{J}_1}(\omega^*)\} d\mathsf{C}_\omega(\omega^*) d\mathsf{D}_\theta(\theta^*; N), \tag{18}$$

$$\mathsf{Total}_{mix}\{\theta, N, \mathsf{N}_{\mathcal{J}_1}(\omega)\} = \int_{(\theta^*, \omega^*) \in (\Theta, \Omega)} \mathsf{Total}\{\theta^*, N, \mathsf{N}_{\mathcal{J}_1}(\omega^*)\} d\mathsf{C}_\omega(\omega^*) d\mathsf{D}_\theta(\theta^*; N). \tag{19}$$

These bootstrap distributions are evaluated at the estimated parameters giving the bootstrap estimators

$$\mathsf{RBNS}_{mix}(\widehat{\theta}, N),\ \mathsf{IBNR}_{mix}\{\widehat{\theta}, \mathsf{N}_{\mathcal{J}_1}(\widehat{\omega})\},\ \mathsf{Total}_{mix}\{\widehat{\theta}, N, \mathsf{N}_{\mathcal{J}_1}(\widehat{\omega})\}. \tag{20}$$

As the integral (17), (18), (19) cannot be calculated exactly they are approximated by simulation by drawing 999 repetitions of the independent distributions $\mathsf{C}_\omega(\omega^*)$ $\mathsf{D}_\theta(\theta^*; N)$. In each repetition the integrand is evaluated as in (16).

To implement the above bootstrap approximations we define the following bootstrap algorithms for the RBNS and IBNR cash-flows.

Algorithm RBNS

Step 1. *Estimation of the parameters and distributions.* From the original data (N,X) estimate the parameters in the model by $\widehat{\theta} = (\widehat{p}, \widehat{\mu}, \widehat{\sigma}^2)$ through (10) and (8). The delay distribution is estimated by a multinomial distribution with probability parameter $\widehat{p}$. The distribution of the individual payments is estimated by a gamma with shape parameter $\widehat{\lambda} = \widehat{\mu}^2/\widehat{\sigma}^2$ and scale parameter $\widehat{\kappa} = \widehat{\sigma}^2/\widehat{\mu}$.

Step 2. *Bootstrapping the data.* Keep the same counts N but generate new bootstrapped aggregated payments $X^* = \{X^*_{ij},\ (i,j) \in \mathcal{I}\}$ as follows:

- Simulate the delay: from each N_{ij} in $\mathcal{I}$ generate the number of paid claims, N_{ij}^{*paid}, by (1) from the Multinomial distribution estimated at Step 1.
- Get the bootstrapped aggregated payments, X_{ij}^*, from a gamma distribution with shape parameter $N_{ij}^{*paid}\widehat{\lambda}$ and scale parameter $\widehat{\kappa}$, for each $(i,j) \in \mathcal{I}$.

Step 3. *Bootstrapping the parameters.* From the bootstrap data, (N, X^*), get $\theta^* = (p^*, \mu^*, \sigma^{2*})$, calculated in the same way as $\widehat{\theta}$ but with the bootstrap data generated at Step 2.

Step 4. *Bootstrapping the RBNS predictions.*

- Simulate the delay from the Multinomial distribution with bootstrapped probability parameter p^* (as in Step 2). Calculate the number of RBNS claims trough (1) and denote these values by N_{ij}^{*rbns}, with $(i,j) \in \mathcal{J}_1 \cup \mathcal{J}_2$.
- Get the bootstrapped RBNS predictions, $m_{ij}^*(N)$, from a gamma distribution with shape parameter $N_{ij}^{*rbns}\lambda^*$ and scale parameter κ^*. Here $\lambda^* = \mu^{*2}/\sigma^{2*}$ and $\kappa^* = \sigma^{2*}/\mu^*$.

Step 5. *Monte Carlo approximation.* Repeat steps 2-4 B times and get the empirical bootstrap distribution of the RBNS reserve, $\widetilde{m}_{ij}(N)$, from the bootstrapped $\{m_{ij}^{*(b)}(N), b = 1, \ldots, B\}$, for each $(i,j) \in \mathcal{J}_1 \cup \mathcal{J}_2$.

Algorithm IBNR

Step 1. *Estimation of the parameters and distributions.* Estimate θ as in Step 1 of Algorithm RBNS, above. Estimate ω using chain ladder, through (5) and (6).

Step 2. *Bootstrapping the data.* Get new bootstrapped data (N^*, X^*) as follows:

- The counts N^* are simulated from Poisson distributions with mean parameters given in terms of $\widehat{\omega}$ from the original observed reported counts N.

- The bootstrapped aggregated payments X^* are simulated exactly as was described in Step 2 of the algorithm RBNS above.

Step 3. *Bootstrapping the parameters.* From the bootstrap data, (N, X^*), get $\theta^* = (\pi^*, \mu^*, \sigma^{2*})$, and $\omega^* = (\rho^*, \Phi^*)$ calculated in the same way as $(\widehat{\theta}, \widehat{\omega})$, but with the bootstrapped data generated at Step 2. Calculate the bootstrapped count parameters, ω^*, by (5) using N^*, and get the bootstrapped predictions in the lower triangle $N^*_{\mathcal{J}_1}(\omega^*)$.

Step 4. *Bootstrapping the IBNR predictions.*

- For each entry N^*_{ij} in $N^*_{\mathcal{J}_1}(\omega^*)$, simulate the delay from a Multinomial distribution with bootstrapped probability parameter p^*. Calculate the number of IBNR claims trough (1) and denote these values by N^{*ibnr}_{ij}, for each $(i,j) \in \mathcal{J}_1 \cup \mathcal{J}_2 \cup \mathcal{J}_3$.
- Get the bootstrapped IBNR predictions, $m^*_{ij}(N^*_{\mathcal{J}_1}(\omega^*))$, from a gamma distribution with shape parameter $N^{*ibnr}_{ij}\lambda^*$ and scale parameter κ^*, exactly as in algorithm RBNS.

Step 5. *Monte Carlo approximation.* Repeat steps 2-4 B times and get the empirical bootstrap distribution of the IBNR reserve, $\widetilde{m}_{ij}(N)$, from the bootstrapped $\{m^{*(b)}_{ij}(N), b = 1, \ldots, B\}$, for each $(i,j) \in \mathcal{J}_1 \cup \mathcal{J}_2 \cup \mathcal{J}_3$.

An intuitive representation of the above bootstrap algorithms are given in Figures 2 and 3.

Considering the motor data, the summary statistics from the RBNS and IBNR cash-flows, estimated by the just presented bootstrap method are reported in Table 7.

6.3 A comparison with bootstrap estimation for the chain ladder technique

For the chain ladder model, bootstrap methods have been considered in England and Verrall (1999), England (2002) and Pinheiro, Andrade e Silva and Centeno (2003), amongst others. These methods are nonparametric bootstrapping using the residuals in a GLM framework, and a key issue for nonparametric bootstraps is the proper definition of the residuals for bootstrapping. Since the methods consider resampling with replacement, it is

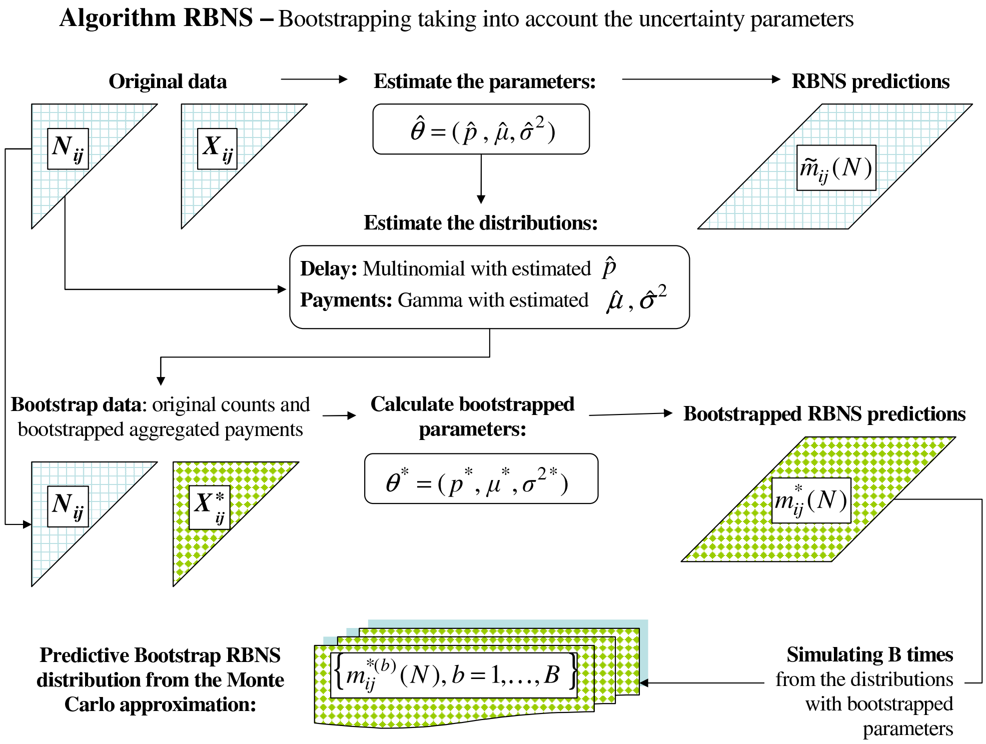

Figure 2: Bootstrapping scheme to approximate the RBNS predictive distribution

necessary to ensure that these residuals are independent and identically distributed. This contrasts with the bootstrap method described above, which is a parametric bootstrap exploiting an assumed distributional form and defining the resampling scheme from the parametric distributions. Other parametric bootstrap methods have been considered recently by Björkwall, Hössjer and Ohlsson (2009a, 2009b).

For comparison purposes with the results from the model described in this paper, we consider the bootstrap approach for estimating derive the predictive distribution of chain ladder forecasts described in England and Verrall (1999) with the modification suggested by England (2002). This constructs the predictive bootstrap distribution by resampling with replacement from the Pearson residuals and then simulating the process distribution using a gamma distribution (or an overdispersed Poisson) with the parameters

Algorithm IBNR – Bootstrapping taking into account the uncertainty parameters

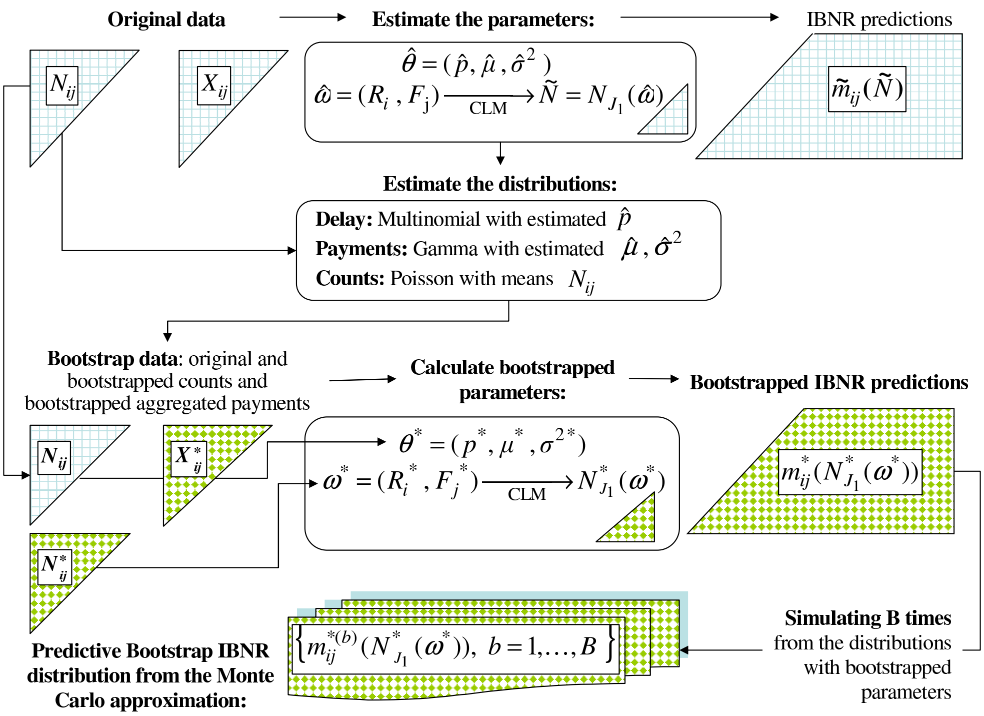

Figure 3: Bootstrapping scheme to approximate the IBNR predictive distribution

estimated from the empirical bootstrap distribution from the first stage.

Table 7 reports results from the England and Verrall (1999) and England (2002) bootstrap method using an R package of Gesmann (2009), together with the results from the new method. The resulting reserves are similar to the chain ladder estimate of outstanding claims (3,315,779). Hence, as also noted in Verrall, Nielsen and Jessen (2010), for this data set the new method does not imply a change in the estimation of the total outstanding claims. We would note, however, that the new model includes a tail, whereas this would have to be added separately for the chain ladder model (thereby increasing the estimate of outstanding claims). The new model is able to generate full cash flows split into two parts, one for the RBNS reserve and one for the IBNR reserve. We also see that the volatility (as measured by the prediction error) is also similar to that from the bootstrap distribution for the chain ladder method. Again, we note that the new model includes the tail and we would expect that the prediction error for the chain ladder model would increase once this is taken into account. Also shown are some percentiles of the bootstrap estimates of the predicitive distributions for the new model and the chain ladder model. In order to assess the performance of the new model, the following section contains a simulation study to compare the results with the classical chain ladder.

7 Simulation study

In order to study the performance of the model described in this paper, in comparison with the standard chain ladder technique, this section considers a simulation study and examines the reserve estimates and capital requirements based on each approach. We generate the data using the assumptions of the new model, but (since the assumptions are deliberately free of any specific structure in terms of the shape of the run-off) we do not believe that this has any affect on the conclusions reached.

7.1 The simulation settings

A scenario for the simulations close to that described in section 6 has been constructed. We consider data triangles with dimension $k = 10$, and generate 999 data sets using the following distribution specifications:

1. The reported counts N_{ij} are defined over a square matrix (with dimension $m = 10$) with the upper triangle being exactly the data entries in Table 1, and the lower triangle completed by generating the entries from a Poisson model with the chain ladder parameters.

2. The delay is generated from a multinomial distribution with probability parameters p_k estimated from the empirical study in section 5.

3. The individual payments are generated from a gamma distribution with first two moments, $\mu = 163.6158$ and $\sigma^2 = 2070821$ (estimated again from the empirical study).

4. A new triangle of aggregated payments is formed from the data generated in step 3, to which the new method and the classical chain ladder method are applied.

7.2 Distribution forecasts

We study the performance of the new bootstrap method (described through algorithms in Section 6.2) in estimating the predictive distribution. Also we make comparisons with the results achieved by applying the standard bootstrap method to the chain ladder method. As in the empirical study in Section 6, we fix the number of bootstrap samples to be $B = 999$ for all the bootstrap methods.

In order to assess the performance of the new method, we require the "actual" predictive distribution, which we simulate using steps 1–4 of the simulations described above. This was done as follows: for each of the 999 simulated data sets we estimate the parameters in the model. These estimates are used to produce the RBNS and IBNR reserves, and from these 999 reserves we calculate the desired quantile of the distribution. This process is repeated 999 times and the simulated "actual" quantiles are defined as by taking the average of the 999 resulting quantiles.

Table 8 shows the distribution forecasts for the total reserve. The bootstrap chain ladder method of England and Verrall (2002) and England (2002), as implemented by Gesmann (2009) gives higher tail quantiles implying higher levels of solvency requirements when using this method. One can consider the trade-off between accepting these extra solvency requirements based on the simple unconditional chain ladder method and rather than collecting the triangle data of reported claims and implementing the more complicated model considered in this paper.

In Table 9 simulations of the medians of the full cash flow are presented for the bootstrap forecast distribution for the new model and for the chain ladder model. The two methods produce almost identical results for the median and it is therefore the distributional properties only that define the difference between these models. They produce almost the same best estimates of reserves. Table 10 gives a breakdown of distribution forecasts for the RBNS and IBNR method and we can see that almost all the upwards bias comes from the RBNS reserve. The chain ladder method does not give such an RBNS/IBNR split and we see the ability to make this split as one of the advances of the new approach. When comparing Table 6 and Table 10 it can be seen that the relative volatility of the IBNR part of the reserve is much bigger than the volatility of the RBNS reserve. This is because of the RBNS is conditional on known counts, whereas the counts of the IBNR contain volatility, and have to be estimated before the final IBNR reserve can be predicted.

8 Conclusions

This paper has examined the properties of the claims reserving method proposed by Verrall, Nielsen and Jessen (2010), and has shown how the full predictive distribution may be obtained using bootstrap methods. In this paper, the structure of the model is identical to Verrall, Nielsen and Jessen (2010) although the detailed assumptions differ in some respects. We believe that this general approach has a great deal to offer: it is essentially as simple to apply as methods such as the chain-ladder technique, but it uses a little more data. We believe that by adding the information regarding the claim counts, much better estimates should be obtained (in general) for the outstanding liabilities and for the predictive distributions. Although the results for the set of data used in this paper did not show any great improvements over the standard chain ladder results, we believe that the general approach has a lot of potential for further development and improvement. We also believe that the coherent approach to the underlying mechanism generating the data, the split between RBNS and IBNR reserves, and the natural and consistent inclusion of the tail in the forecasts are specific advantages of this methods over the ad hoc approach of the chain ladder technique.

Acknowledgement

The computations were done using R (R Development Core Team, 2006). This research was financially supported by a Cass Business School Pump Priming Grant. The first author is also supported by the Project MTM2008-03010/MTM.

References

Björkwall, S., Hössjer, O. & Ohlsson, E. (2009a) Non-parametric and Parametric Bootstrap Techniques for Arbitrary Age-to-Age Development Factor Methods in Stochastic Claims Reserving. *Scandinavian Actuarial Journal* 306-331.

Björkwall, S., Hössjer, O. & Ohlsson, E. (2009b) Bootstrapping the separation method in claims reserving. Stockholm University, Mathematical Statistics, Research Report 2009:2. To appear in *ASTIN Bulletin*

Bryden, D. and Verrall, R.J. (2009) Calendar year effects, claims inflaton and the Chain-Ladder technique. *Annals of Actuarial Science* 4, 287-301.

England, P. (2002) Addendum to "Analytic and Bootstrap Estimates of Prediction Error in Claims Reserving". *Insurance: Mathematics and Economics* 31, 461–466.

England, P. & Verrall, R. (1999) Analytic and Bootstrap Estimates of Prediction Error in Claims Reserving. *Insurance: Mathematics and Economics* 25, 281–293.

Gesmann, M. (2009) R-package 'ChainLadder' version 0.1.2-11 (April, 17, 2009).

Kremer, E. (1985) *Einführung in die Versicherungsmathematik*. Göttingen: Vandenhoek & Ruprecht.

Kuang, D., Nielsen, B. & Nielsen, J.P. (2008a) Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika* 95, 979–986.

Kuang, D., Nielsen, B. & Nielsen, J.P. (2008b) Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika* 95, 987–991.

Kuang, D., Nielsen, B. & Nielsen, J.P. (2009) Chain Ladder as Maximum Likelihood Revisited. *Annals of Actuarial Science* 4, 105–121.

Kuang, D., Nielsen, B. & Nielsen, J.P. (2010) Forecasting in an extended chain-ladder-type model. *Journal of Risk and Insurance* to appear.

Mack, T. (1991) A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *Astin Bulletin*, vol. 39, 35–60.

Pinheiro, P.J.R., Andrade e Silva, J.M. & Centeno, M.d.L. (2003) Bootstrap Methodology in Claim Reserving. *Journal of Risk and Insurance* 4, 701–714.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

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

Verrall, R., Nielsen, J.P. & Jessen, A. (2010) Prediction of RBNS and IBNR claims using claim amounts and claim counts. To appear in *ASTIN Bulletin*.

Zehnwirth, B. (1994) Probabilistic development factor models with applications to loss reserve variability, prediction intervals, and risk based capital. *Casualty Actuarial Society Forum* Spring 1994, 447–605.

k	R_k	F_k	$\widehat{\zeta}_k$	$\widehat{\psi}_k$	$\widehat{p}_k$
0			59.0	59.0	0.361
1	7135	1.135291	54.9	46.9	0.287
2	9190	1.003790	24.9	18.3	0.112
3	11427	1.000917	16.6	13.8	0.084
4	10667	1.000329	12.8	10.7	0.066
5	10951	1.000284	7.27	5.70	0.035
6	11421	1.000234	5.13	4.26	0.026
7	11341	1.000144	2.67	2.02	0.012
8	12486	1.000306	3.21	2.87	0.018
9	13658	1.000421	0.28	0	0
10	10989				

$\widehat{\mu} = 163.62$
$\widehat{\varphi} = 12793.19$
$\widehat{\sigma}^2 = 2070821 = 1439.0^2$
$\widehat{\sigma}^2_{VNJ} = 2066398 = 1437.5^2$

Table 5: Estimates for motor data. R_k and F_k are row sums and development factors for count data in Table 1 computed as in (5). $\widehat{\psi}_k$, $\widehat{p}_k$, $\widehat{\mu}$ are delay parameters estimated as described in §4.2 using the analytic method with a negative $\widehat{\psi}_9$ replaced by zero. $\widehat{\varphi}$ and $\widehat{\sigma}^2$ are computed as in (12), (13). $\widehat{\sigma}^2_{VNJ} = \widehat{\mu}\widehat{\varphi} - \widehat{\mu}^2$.

Future	Calendar Year	RBNS	IBNR	RBNS+IBNR	CL
1	11	1307	93	1399	1354
2	12	720	78	798	754
3	13	494	34	529	489
4	14	323	26	349	318
5	15	188	20	208	185
6	16	117	12	130	115
7	17	65	9	74	63
8	18	37	5	42	36
9	19	0	6	6	2
10	20		1	1	
11	21		0.6	0.6	
12	22		0.4	0.4	
13	23		0.2	0.2	
14	24		0.1	0.1	
15	25		0.07	0.07	
16	26		0.04	0.04	
17	27		0.02	0.02	
18	28		0	0	
Total		3251	287	3538	3316

Table 6: Point forecasts of cashflow by calendar year, in thousands.

	Bootstrap predictive distribution			
	RBNS	IBNR	Total	BCL
mean	3134	274	3408	3314
pe	327	60	340	345
1%	2464	148	2714	2588
5%	2646	183	2895	2780
50%	3105	272	3390	3287
95%	3722	378	4002	3911
99%	3987	435	4275	4061

Table 7: Distribution forecasts of RBNS, IBNR and total reserve, in thousands. The three first column give the summary of the distribution from the proposed bootstrap method which takes into account the uncertainty of the parameters. The last column provides the results for the total reserve for the bootstrap method of England and Verrall (1999) and England (2002).

Future	Actual		Cashflow Bootstrap		CL Bootstrap	
	95%	99%	95%	99%	95%	99%
1	1649	1759	1659	1776	1709	1847
2	992	1085	1001	1094	1018	1115
3	686	765	698	778	707	789
4	482	550	492	562	494	564
5	319	375	326	386	323	381
6	226	276	231	284	224	274
7	157	203	161	210	150	193
8	112	154	117	163	103	140
9	41	76	48	87	24	41
10	9	23	10	25	0	0
11	3	14	3	13	0	0
12	0.8	10	1.4	9	0	0
13	0.2	7	0.4	5	0	0
14	0	3	0.1	3	0	0
15	0	0.8	0	1	0	0
16	0	0.1	0	0.4	0	0
17	0	0	0	0.1	0	0
18	0	0	0	0	0	0

Table 8: Simulation of distribution forecasts of total reserve by calendar year: 95%, 99% quantiles over 999 repetitions. Column 2-3 give actual numbers; Column 4-5 cash flow bootstrap using analytic delay estimation and taking into account the uncertainty parameters; Column 6-7 bootstrap Chain Ladder.

Future	Actual	Cashflow Bootstrap	CL Bootstrap
1	1396	1406	1400
2	793	802	799
3	523	530	528
4	343	350	345
5	203	209	204
6	124	129	124
7	69	74	68
8	38	42	36
9	4	7	3
10	0	0	0

Table 9: Simulation of distribution forecasts of total reserve by calendar year: Medians (50%) over 999 repetitions. Column 2 gives actual numbers; Column 3 cashflow bootstrap using analytic delay estimation and taking into account the uncertainty parameters; Column 4 Chain Ladder bootstrap.

	RBNS				IBNR			
Future	Actual		Bootstrap		Actual		Bootstrap	
	95%	99%	95%	99%	95%	99%	95%	99%
1	1551	1655	1559	1673	155	194	155	192
2	906	993	916	1005	137	175	136	170
3	649	724	660	738	74	102	74	101
4	454	519	463	531	60	86	60	85
5	296	351	303	362	52	76	51	75
6	212	261	216	270	37	60	36	57
7	146	192	150	199	31	54	30	50
8	105	147	110	156	21	41	21	39
9	27	67	35	77	24	45	23	42
10	0	0	0	0	9	23	10	25
11	0	0	0	4	3	14	3	13
12	0	0	0	3	0.8	10	1	9
13	0	0	0	0.5	0.2	7	0.4	6
14	0	0	0	0.1	0	3	0.1	3
15	0	0	0	0	0	0.7	0	1
16	0	0	0	0	0	0.1	0	0.4
17	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0

Table 10: Simulation of distribution forecasts of RBNS/IBNR reserves by calendar year: 95% and 99% quantiles over 999 repetitions. Column 2-5 give RBNS reserve Column 6-9 give IBNR reserve Column 2-3 & 6-7 give actual numbers; Column 4-5 & 8-9 give cashflow bootstrap using analytic delay estimation and taking into account the uncertainty parameters.