Cass Business School
CITY UNIVERSITY LONDON

Faculty of Actuarial Science and Insurance

Actuarial Research Paper
No. 192

Exploration of a Novel Bootstrap Technique for Estimating the Distribution of Outstanding Claims Reserves in General Insurance.

Robert Cowell

October 2009

Cass Business School
106 Bunhill Row
London EC1Y 8TZ
Tel +44 (0)20 7040 8470
www.cass.city.ac.uk

ISBN 978-1-905752-24-9

Actuarial Research Report

Exploration of a novel bootstrap technique for estimating the distribution of outstanding claims reserves in general insurance

Robert Cowell

Faculty of Actuarial Science and Insurance
Cass Business School
City University London
106 Bunhill Row
London EC1Y 8TZ

October 13, 2009

Contents

Abstract

This is a report on an exploration of the effectiveness of a novel non-parametric bootstrap method for estimating claims reserves, which we call the *local chain ladder bootstrap* technique. The method is simple and can readily be implemented in a spreadsheet. In addition analytic estimates of the first few moments of reserves are shown to the readily evaluated, obviating the need for simulation if desired. The behaviour of the method on three datasets is presented and compared to published predictions of some other stochastic methods. In addition, a small study of the method using simulated claims triangles is presented and compared with other stochastic models.

In summary, this report presents:

- A simple mathematical proof of a symmetry property of the standard chain ladder technique.
- A description of the novel local chain ladder bootstrap technique, with three variants of it.
- A comparison of the local chain ladder bootstrap technique to published predictions of reserve estimates for several other claims reserving techniques, using three sets of historical data.
- The results of a simulation study in which claims triangles were simulated, and for which the true distribution of reserves is thus known, comparing in an absolute sense the predictive performance of the local chain ladder bootstrap technique and two other standard techniques in the literature.
- A weblink to a program with a simple graphical user interface that implements the local chain ladder bootstrap technique, and two other techniques, on an input claims triangle. The program runs under the Microsoft Windows operating system, and is freely available to download.

Keywords

Stochastic claims reserving; chain ladder; non parametric; bootstrap; distribution of reserves.

1 Introduction

For many years the chain-ladder technique for estimating claims reserves has been widely used. Two reasons can be put forward for this: (i) a chain-ladder analysis is simple for practitioners to implement; (ii) the results from a chain-ladder analysis usually accord reasonably well with the expectations of experienced practitioners. One basic restriction of the chain-ladder technique is that it can only provide a point estimate of the reserves. In recent years there have been a number of proposals to overcome this limitation in order to model the distribution of reserves. Many of these proposals use stochastic models that either are based upon the chain-ladder technique, or are constructed in order to reproduce the chain-ladder estimates in expectation. Interest in these models has grown among academic actuaries because of the realization that the variability in the reserves can be more informative that only a simple point estimate. These concerns have also prompted interest from practising actuaries, however the use of such models in practise appears limited. Two reasons could be put forward for this. Firstly, such models can be quite complex mathematically, and so be difficult to implement. Secondly, there is no agreement on which of the stochastic models is best to use. Even a single stochastic model may have several variants distinguished, for example, by choice of distributions (gamma, over-dispersed Poisson, normal, log-normal, etc.). In addition, some models based upon positive distributions might not be applicable to data in which incremental claims are negative in one or more development years.

This variety of stochastic models, and the lack of consensus on which model is appropriate for particular data (eg: why use a gamma distribution instead of a log-normal distribution?), means that a practising actuary could have difficulty in justifying the use of a particular model to a regulator. One approach would be to estimate the distribution of reserves using several models, and see if they approximately agree in their predictions: the one providing the most conservative estimates might then used. This would not be a problem if the models were simple to implement, however already mentioned they are quite complex and so such a strategy could be beyond the resources available to many practising actuaries.

In this report I present a new method of estimating reserves based on a simple bootstrap simulation method. The result of the simulation is a sample from which the distribution of reserves may be constructed and analysed. The method is non-parametric in nature—no distributional assumptions, for example, about individual claim sizes or their number are made. The model can cope with negative incremental claims.

The plan of this paper is as follows. The next section summarises the standard chain-ladder technique. I then present the new bootstrap method, which I call the *local chain ladder bootstrap* model. Three variants of the method are presented. They are first applied to some historical datasets, and then used in a simulation study in which claims triangles are simulated and whose distribution of true completions are known.

2 The chain ladder technique

It is assumed that the reader is familiar with the standard chain ladder technique: expositions may be found in (Taylor 2000) and (Wüthrich and Merz 2008). This section introduces the notation used, and also discusses a non-standard formulation of the chain ladder technique, and presents a simple mathematical proof of the identity of the two methods.

2.1 Standard formalism

Rows label the period of origin of the claim. The columns represent the development year. The data in the upper triangle represents the amount paid out on claims. Inflation etc. is not modelled.

Following England and Verrall (2002), we use C_{ij} to represent the incremental claim amount in origin year (row) i and development year (column) j. With this notation, the claims are laid out as in Table 1.

Table 1: Incremental claims-triangle format

	Development year						
Origin year	1	2	$\cdots$	j	$\cdots$	$n-1$	n
1	$C_{1,1}$	$C_{1,2}$	$\cdots$	$C_{1,j}$	$\cdots$	$C_{1,n-1}$	$C_{1,n}$
2	$C_{2,1}$	$C_{2,2}$	$\cdots$	$C_{2,j}$	$\cdots$	$C_{2,n-1}$	
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$		
i	$C_{i,1}$	$C_{i,2}$	$\cdots$				
$\cdots$	$\cdots$	$\cdots$	$\cdots$				
$n-1$	$C_{n-1,1}$	$C_{n-1,2}$					
n	$C_{n,1}$						

Summing a row year up to a certain development period leads to *cumulative claims*,

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

that are shown laid out in Table 2.

Table 2: Cumulative claims triangle

	Development year						
Origin year	1	2	$\cdots$	j	$\cdots$	$n-1$	n
1	$D_{1,1}$	$D_{1,2}$	$\cdots$	$D_{1,j}$	$\cdots$	$D_{1,n-1}$	$D_{1,n}$
2	$D_{2,1}$	$D_{2,2}$	$\cdots$	$D_{2,j}$	$\cdots$	$D_{2,n-1}$	
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$		
i	$D_{i,1}$	$D_{i,2}$	$\cdots$				
$\cdots$	$\cdots$	$\cdots$	$\cdots$				
$n-1$	$D_{n-1,1}$	$D_{n-1,2}$					
n	$D_{n,1}$						

The purpose of analysing these data is to try and fill in the lower triangular parts of these squares, and so estimate how much e.g. insurance companies will be ultimately liable for, and hence how much they should retain in reserves. The chain ladder technique consists of using the values in the development triangle to construct so-called *development factors* $\lambda_j, \quad j = 2, \ldots, n$ for each development year,

$$\lambda_j = \frac{\sum_{i=1}^{n-j+1} D_{i,j}}{\sum_{i=1}^{n-j+1} D_{i,j-1}}.$$

These are used as multiplicative factors to fill-in the lower half of the cumulative claims triangle, in a recursive manner according to the formula:

$$\hat{D}_{i,j+1} = \lambda_j \hat{D}_{i,j}, \text{ for } j > n-i-1$$

where we define the diagonal entries:

$$\hat{D}_{i,n-i+1} = D_{i,n-i+1} \text{ for } i = 2, \ldots, n.$$

The values in the final column of the completed square give the estimates of the total claims for each year—also known as the *ultimate claims*. Subtracting from the ultimate of each row the corresponding diagonal entry in the triangle yields the estimate of the reserves required for each year to meet the expected claim. This is illustrated in Table 3. Adding up the expected reserves for each gives the total estimate of reserves required to meet the claims.

Table 3: Cumulative development triangle

Origin year	1	2	$\cdots$	j	$\cdots$	$n-1$	n	R_i
1	$D_{1,1}$	$D_{1,2}$	$\cdots$	$D_{1,j}$	$\cdots$	$D_{1,n-1}$	$\hat{D}_{1,n}$	0
2	$D_{2,1}$	$D_{2,2}$	$\cdots$	$D_{2,j}$	$\cdots$	$D_{2,n-1}$	$\hat{D}_{2,n}$	$\hat{D}_{2,n} - D_{2,n-1}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
i	$D_{i,1}$	$D_{i,2}$	$\cdots$	$D_{i,j}$	$\cdots$	$\hat{D}_{i,n-1}$	$\hat{D}_{i,n}$	$\hat{D}_{i,n} - D_{i,j}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$n-1$	$D_{n-1,1}$	$D_{n-1,2}$	$\cdots$	$\cdots$	$\cdots$	$\hat{D}_{n-1,n-1}$	$\hat{D}_{n-1,n}$	$\hat{D}_{n-1,n} - D_{n-1,2}$
n	$D_{n,1}$	$\hat{D}_{n,2}$	$\cdots$	$\hat{D}_{n,j}$	$\cdots$	$\hat{D}_{n-1,n}$	$\hat{D}_{n,n}$	$\hat{D}_{n,n} - D_{n,1}$

2.2 Non-standard formalism

Consider the following small triangle of *incremental* claims, based on four development years.

Origin	Development year			
year	1	2	3	4
1	5	8	3	4
2	2	7	1	
3	6	5		
4	3			

Cumulating across columns, we get the *cumulative* claims triangle

Origin	Development year			
year	1	2	3	4
1	5	13	16	20
2	2	9	10	
3	6	11		
4	3			

From these, we find the usual chain-ladder development factors $\bar{\lambda} = (2.54, 1.18, 1.25)$ where $\lambda_2 = (13+9+11)/(5+2+6) = 33/13 = 2.54$ etc.

Applying these in the usual way, we get the chain-ladder estimates of *cumulative* italicized in the following table:

		λ_2	λ_3	λ_4
	1	2	3	4
1	5	13	16	20
2	2	9	10	*12.5*
3	6	11	*13*	*16.25*
4	3	*7.62*	*9*	*11.25*

From this the estimates of incremental claim amounts can be found, here shown in the following table in italics together with the observed incremental claims.

	1	2	3	4
1	5	8	3	4
2	2	7	1	*2.5*
3	6	5	*2*	*3.25*
4	3	*4.62*	*1.38*	*2.25*

Adding together the italicised values yields the reserve estimate value of *16*. All this is an example of the standard chain ladder technique described in Section 2.1.

Now it turns out that the same results can be obtained from the incremental claims triangle, by cumulating across rows instead of columns. we call this the non-standard chain ladder technique. Thus for example, taking again the small example,

	1	2	3	4
1	5	8	3	4
2	2	7	1	
3	6	5		
4	3			

and cumulating across rows, we get:

		1	2	3	4
	1	5	8	3	4
ρ_2	2	7	15	4	
ρ_3	3	13	20		
ρ_4	4	16			

From these, row "origin factors" $\bar{\rho} = (\rho_2, \rho_3, \rho_4) = (1.63, 1.5, 1.23)$ may be found, where

$$\begin{aligned}\rho_2 &= (7+15+4)/(5+8+3) = 26/16 = 1.63 \\ \rho_3 &= (13+20)/(7+15) = 33/22 = 1.5 \\ \rho_4 &= 16/13 = 1.23\end{aligned}$$

Applying these factors to the previous triangle, we get cumulative row estimates of

	1	2	3	4
1	5	8	3	4
2	7	15	4	*6.5*
3	13	20	*6*	*9.75*
4	16	*24.62*	*7.38*	*12*

From which, subtracting adjacent row entries in the lower triangle, we get the incremental estimates,

	1	2	3	4
1	5	8	3	4
2	2	7	1	*2.5*
3	6	5	*2*	*3.25*
4	3	*4.62*	*1.38*	*2.25*

which is the same as before using the usual column development factors, and hence the same reserve estimate.

This equality is no coincidence, as is now shown.

2.3 Equivalence of the two approaches

Suppose we have an *incremental* claims triangle of size n, with incremental claim in year i and development year j denoted by $C_{i,j}$. Then the entries in the cumulative claims triangle are $D_{i,j}$ where $D_{i,1} = C_{i,1}$, and $D_{i,j} = C_{i,j} + D_{i,j-1}$ with $i + j \leq n$. Alternatively, $D_{i,j} = \sum_{k=1}^{j} C_{i,k}$.

Now the development factor for column $j, (j = 2, \ldots, n)$ and denoted by λ_j, is given by:

$$\lambda_j = \frac{\sum_{i=1}^{n-j+1} D_{i,j}}{\sum_{i=1}^{n-j+1} D_{i,j-1}}$$

However, this may be expressed in terms of the incremental claims as the ratio of two double summations:

$$\lambda_j = \frac{\sum_{i=1}^{n-j+1} \sum_{k=1}^{j} C_{i,k}}{\sum_{i=1}^{n-j+1} \sum_{k=1}^{j-1} C_{i,k}}$$

which we can write as:

$$\begin{aligned}
\lambda_j &= \frac{\left(\sum_{i=1}^{n-j+1} \sum_{k=1}^{j-1} C_{i,k}\right) + \left(\sum_{i=1}^{n-j+1} C_{i,j}\right)}{\sum_{i=1}^{n-j+1} \sum_{k=1}^{j-1} C_{i,k}} \\
&= 1 + \frac{\sum_{i=1}^{n-j+1} C_{i,j}}{\sum_{i=1}^{n-j+1} \sum_{k=1}^{j-1} C_{i,k}}
\end{aligned}$$

Now, we have, applying the chain-ladder technique, that

$$\hat{D}_{i,j} = \lambda_j \hat{D}_{i,j-1}$$

hence the incremental claim estimate $\hat{C}_{i,j}$ is given by

$$\hat{C}_{i,j} = \hat{D}_{i,j} - \hat{D}_{i,j-1} = (\lambda_j - 1)\hat{D}_{i,j-1} = \frac{\sum_{r=1}^{n-j+1} C_{r,j}}{\sum_{r=1}^{n-j+1} \sum_{k=1}^{j-1} C_{r,k}} \hat{D}_{i,j-1}$$

Now suppose that $j = n - i + 2$, that is one year ahead from the latest observed incremental claim in the i-th row. Then we have that $\hat{D}_{i,j-1} = D_{i,n-i+1} = \sum_{k=1}^{n-i+1} C_{i,k}$, so that

$$\hat{C}_{i,n-i+2} = \frac{\left(\sum_{r=1}^{i-1} C_{r,n-i+2}\right)\left(\sum_{k=1}^{n-i+1} C_{i,k}\right)}{\sum_{r=1}^{i-1} \sum_{k=1}^{n-i+1} C_{r,k}}$$

This formula is illustrated for the estimate of the entry $C_{5,4}$ in the 7×7 table. The two terms in the numerator are sums over the rectangles in row 5 and column 4, whilst the denominator is the sum of terms in the rectangular are spanned by the first four rows and the first three columns.

	1	2	3	4	5	6	7
1	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$	$C_{1,4}$	$C_{1,5}$	$C_{1,6}$	$C_{1,7}$
2	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$	$C_{2,4}$	$C_{2,5}$	$C_{2,6}$	
3	$C_{3,1}$	$C_{3,2}$	$C_{3,3}$	$C_{3,4}$	$C_{3,5}$		
4	$C_{4,1}$	$C_{4,2}$	$C_{4,3}$	$C_{4,4}$			
5	$C_{5,1}$	$C_{5,2}$	$C_{5,3}$	$\hat{C}_{5,4}$			
6	$C_{6,1}$	$C_{6,2}$					
7	$C_{7,1}$						

The table illustrates the symmetry involved in evaluating $\hat{C}_{5,4}$

$$\begin{aligned}\hat{C}_{5,4} &= \frac{\left(\sum_{k=1}^{4} C_{k,4}\right)\left(\sum_{k=1}^{3} C_{5,k}\right)}{\sum_{r=1}^{4}\sum_{k=1}^{3} C_{r,k}} \\ &= \frac{\left(\sum_{k=1}^{3} C_{5,k}\right)\left(\sum_{k=1}^{4} C_{k,4}\right)}{\sum_{k=1}^{3}\sum_{r=1}^{4} C_{r,k}}\end{aligned}$$

This is what would be obtained by making the "origin factors" vertically. More formally, let $\Upsilon_{i,j}$ denote "cumulative sums" over the j-th column entries in the first i rows, with row origin factor denoted by ρ_i. That is, $\Upsilon_{1,j} = C_{1,j}$, and $\Upsilon_{i,j} = C_{i,j} + \Upsilon_{i-1,j}$ with $i + j \leq n$. Alternatively, $\Upsilon_{i,j} = \sum_{k=1}^{i} C_{k,j}$. The row origin factor for row i ($i = 2, \ldots, n$) is given by:

$$\rho_i = \frac{\sum_{k=1}^{n-i+1} \Upsilon_{i,k}}{\sum_{k=1}^{n-i+1} \Upsilon_{i-1,k}} = 1 + \frac{\sum_{k=1}^{n-i+1} C_{i,k}}{\sum_{k=1}^{n-i+1} \Upsilon_{i-1,k}}$$

This is illustrated for the 7×7 table:

		1	2	3	4	5	6	7
	1	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$	$C_{1,4}$	$C_{1,5}$	$C_{1,6}$	$C_{1,7}$
ρ_2	2	$\Upsilon_{2,1}$	$\Upsilon_{2,2}$	$\Upsilon_{2,3}$	$\Upsilon_{2,4}$	$\Upsilon_{2,5}$	$\Upsilon_{2,6}$	
ρ_3	3	$\Upsilon_{3,1}$	$\Upsilon_{3,2}$	$\Upsilon_{3,3}$	$\Upsilon_{3,4}$	$\Upsilon_{3,5}$		
ρ_4	4	$\Upsilon_{4,1}$	$\Upsilon_{4,2}$	$\Upsilon_{4,3}$	$\Upsilon_{4,4}$			
ρ_5	5	$\Upsilon_{5,1}$	$\Upsilon_{5,2}$	$\Upsilon_{5,3}$				
ρ_6	6	$\Upsilon_{6,1}$	$\Upsilon_{6,2}$					
ρ_7	7	$\Upsilon_{7,1}$						

Now we have the estimate:

$$\hat{\Upsilon}_{i,j} = \rho_i \hat{\Upsilon}_{i-1,j}$$

from which the incremental claim estimate is

$$\begin{aligned}\hat{C}_{i,j} &= (\rho_i - 1)\hat{\Upsilon}_{i-1,j} \\ &= \frac{\sum_{k=1}^{n-i+1} C_{i,k}}{\sum_{k=1}^{n-i+1} \Upsilon_{i-1,k}} \hat{\Upsilon}_{i-1,j} \\ &= \frac{\sum_{k=1}^{n-i+1} C_{i,k}}{\sum_{r=1}^{i-1} \sum_{k=1}^{n-i+1} C_{r,k}} \hat{\Upsilon}_{i-1,j}\end{aligned}$$

Now suppose that $j = n - i + 2$, that is just below the observed diagonal values. Then we have that $\hat{\Upsilon}_{i-1,n-i+2} = \Upsilon_{i-1,n-i+2} = \sum_{r=1}^{i-1} C_{r,n-i+2}$, so that

$$\hat{C}_{i,n-i+2} = \frac{\left(\sum_{k=1}^{n-i+1} C_{i,k}\right)\left(\sum_{r=1}^{i-1} C_{r,n-i+2}\right)}{\sum_{r=1}^{i-1} \sum_{k=1}^{n-i+1} C_{r,k}}$$

which is the same as obtained via the usual chain-ladder formulation.

Looking at the 7×7 example, filling in the entries just below the diagonal, we have:

			λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
		1	2	3	4	5	6	7
	1	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$	$C_{1,4}$	$C_{1,5}$	$C_{1,6}$	$C_{1,7}$
ρ_2	2	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$	$C_{2,4}$	$C_{2,5}$	$C_{2,6}$	$\hat{C}_{2,7}$
ρ_3	3	$C_{3,1}$	$C_{3,2}$	$C_{3,3}$	$C_{3,4}$	$C_{3,5}$	$\hat{C}_{3,6}$	
ρ_4	4	$C_{4,1}$	$C_{4,2}$	$C_{4,3}$	$C_{4,4}$	$\hat{C}_{4,5}$		
ρ_5	5	$C_{5,1}$	$C_{5,2}$	$C_{5,3}$	$\hat{C}_{5,4}$			
ρ_6	6	$C_{6,1}$	$C_{6,2}$	$\hat{C}_{6,3}$				
ρ_7	7	$C_{7,1}$	$\hat{C}_{7,2}$					

Now consider the estimates for the incremental claims just below the diagonal of this table. We begin with the estimate $\hat{C}_{3,7}$. Using the usual development factors, we have

$$
\begin{aligned}
\hat{C}_{3,7} &= (\lambda_7 - 1)\hat{D}_{3,6} \\
&= (\lambda_7 - 1)\left(\sum_{j=1}^{5} C_{3,j} + \hat{C}_{3,6}\right) \\
&= (\lambda_7 - 1)\left(\sum_{j=1}^{5} C_{3,j} + (\lambda_6 - 1)\sum_{j=1}^{5} C_{3,j}\right) \\
&= \lambda_6(\lambda_7 - 1)(\sum_{j=1}^{5} C_{3,j}) \\
&= \frac{\sum_{r=1}^{2}\sum_{k=1}^{6} C_{r,k}}{\sum_{r=1}^{2}\sum_{k=1}^{5} C_{r,k}} \times \frac{C_{1,7}}{\sum_{k=1}^{6} C_{1,k}} \times \sum_{j=1}^{5} C_{3,j}
\end{aligned}
$$

Using the row origin factors we get the same result:

$$
\begin{aligned}
\hat{C}_{3,7} &= (\rho_3 - 1)\hat{\Upsilon}_{2,7} \\
&= (\rho_3 - 1)(C_{1,7} + \hat{C}_{2,7}) \\
&= (\rho_3 - 1)(C_{1,7} + (\rho_2 - 1)C_{1,7}) \\
&= (\rho_3 - 1)\rho_2 C_{1,7} \\
&= \frac{\sum_{j=1}^{5} C_{3,j}}{\sum_{r=1}^{2}\sum_{k=1}^{5} C_{r,k}} \times \frac{\sum_{r=1}^{2}\sum_{k=1}^{6} C_{r,k}}{\sum_{k=1}^{6} C_{1,k}} \times C_{1,7}
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\hat{C}_{4,6} &= \lambda_5(\lambda_6 - 1)(\sum_{j=1}^{4} C_{4,j}) \\
&= \frac{\sum_{r=1}^{3}\sum_{k=1}^{5} C_{r,k}}{\sum_{r=1}^{3}\sum_{k=1}^{4} C_{r,k}} \times \frac{\sum_{r=1}^{2} C_{r,6}}{\sum_{r=1}^{2}\sum_{k=1}^{5} C_{r,k}} \times (\sum_{j=1}^{4} C_{4,j}) \\
&= \frac{\sum_{r=1}^{3}\sum_{k=1}^{5} C_{r,k}}{\sum_{r=1}^{2}\sum_{k=1}^{5} C_{r,k}} \times \frac{\sum_{j=1}^{4} C_{4,j}}{\sum_{r=1}^{3}\sum_{k=1}^{4} C_{r,k}} \times \sum_{r=1}^{2} C_{r,6} \\
&= \rho_3(\rho_4 - 1) \times \sum_{r=1}^{2} C_{r,6} \\
&= (\rho_4 - 1)\hat{\Upsilon}_{3,6}
\end{aligned}
$$

Similarly, the remaining estimates $\hat{C}_{5,5}, \hat{C}_{6,4}\hat{C}_{7,3}$ can be obtained using the standard chain-ladder technique, and can be shown to equal to the equal to the respective entries is calculated using the row origin factors. However we wish to show equivalence of both methods for all missing entries in arbitrary sized claims triangle, so we need some notation to help in the proof.

The notation we shall use is an implied summation, indicated by raised indices. Thus if $Y_{i,j}$ is some set of numbers defined over a square array of indices $i = 1, \ldots, n$ and $j = 1, \ldots, n$, then we define:

$$\begin{aligned} Y^{i}{}_{j} &= \sum_{r=1}^{i} Y_{y,j} \\ Y_{i}{}^{j} &= \sum_{k=1}^{j} Y_{i,k} \\ Y^{i:j} &= \sum_{r=1}^{i} \sum_{k=1}^{j} Y_{r,k} \end{aligned}$$

Using this summation notation we have, for a claims triangle of size n:

$$\begin{aligned} \lambda_j &= \frac{C^{n-j+1:j}}{C^{n-j+1:j-1}} \\ \lambda_j - 1 &= \frac{C^{n-j+1}{}_{j}}{C^{n-j+1:j-1}} \\ \rho_i &= \frac{C^{i:n-i+1}}{C^{i-i:n-i+1}} \\ \rho_i - 1 &= \frac{C_{i}{}^{n-i+1}}{C^{i-i:n-i+1}} \end{aligned}$$

Now consider the chain ladder estimate of the missing entry $C_{i,n-i+1+m}$ for $m > 0$.

With $m = 1$ this is given by

$$
\begin{aligned}
\hat{C}_{i,n-i+2} &= (\lambda_{n-i+2} - 1)\hat{D}_{i,n-i+1} \\
&= \frac{C^{i-1}{}_{n-i+2}}{C^{i-1:n-i+1}} C_i{}^{n-i+1} \\
&= \frac{C_i{}^{n-i+1}}{C^{i-1:n-i+1}} C^{i-1}{}_{n-i+2} \\
&= (\rho_i - 1)\Upsilon_{i-1,n-i+1}
\end{aligned}
$$

For $m > 1$ this is given by

$$
\begin{aligned}
\hat{C}_{i,n-i+1+m} &= \lambda_{n-i+2} \dots \lambda_{n-i+m}(\lambda_{n-i+m+1} - 1)\hat{D}_{i,n-i+1} \\
&= \frac{C^{i-1:n-i+2}}{C^{i-1:n-i+1}} \frac{C^{i-2:n-i+3}}{C^{i-2:n-i+2}} \dots \frac{C^{i-m+1:n-i+m}}{C^{i-m+1:n-i+m-1}} \frac{C^{i-m}{}_{n-i+m+1}}{C^{i-m:n-i+m}} C_i{}^{n-i+1} \\
&= \frac{C_i{}^{n-i+1}}{C^{i-1:n-i+1}} \frac{C^{i-1:n-i+2}}{C^{i-2:n-i+2}} \dots \frac{C^{i-m+1:n-i+m}}{C^{i-m:n-i+m}} C^{i-m}{}_{n-i+m+1} \\
&= (\rho_i - 1)\rho_{i-1}\rho_{i-2} \dots \rho_{i-m+1}\Upsilon_{i-m,n-i+m+1}
\end{aligned}
$$

which proves the identity of the usual chain-ladder development and the row origin factor estimates.

3 The Local Chain Ladder Bootstrap method

We now describe our technique for estimating the distribution of reserves from a non-parametric bootstrap procedure. The technique consists of two parts. First we construct a set of local ratios. Then we form bootstrap samples of these to form completed cumulative claims triangles, which on repetition yields a simulation of the reserves. If the construction of the development factors proceeds in a similar manner to the standard chain ladder technique, we call this the *horizontal* method. An alternative method following the non-standard technique we call the *vertical* method. Finally a blend of the two methods is presented that we call the *mixed method*.

3.1 The horizontal method

In the standard chain ladder technique, development factors are formed by taking the ratios of column sums in the cumulative claims triangle. For the local chain ladder bootstrap method we form just the ratios of neighbouring values in the cumulative claims distribution table. That is, we calculate the ratios

$$\lambda_{i,j+1} = D_{i,j+1}/D_{i,j},$$

for each pair of neighbouring values in the cumulative claims distribution table, this is illustrated in Table 4, in which the first column gives the initial claims amounts, and the remaining entries show the local development factors. This part is only performed once.

Table 4: Claims triangle

	Development year						
Origin year	1	2	$\cdots$	j	$\cdots$	$n-1$	n
1	$D_{1,1}$	$\lambda_{1,2}$	$\cdots$	$\lambda_{1,j}$	$\cdots$	$\lambda_{1,n-1}$	$\lambda_{1,n}$
2	$D_{2,1}$	$\lambda_{2,2}$	$\cdots$	$\lambda_{2,j}$	$\cdots$	$\lambda_{2,n-1}$	
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$		
i	$D_{i,1}$	$\lambda_{i,2}$	$\cdots$				
$\cdots$	$\cdots$	$\cdots$	$\cdots$				
$n-1$	$D_{n-1,1}$	$\lambda_{n-1,2}$					
n	$D_{n,1}$						

The second step is to construct a bootstrap sample, which we do as follows: we fill in empty cell Table 4 below the lower diagonal with one of the local development factors calculated in step 1, where we choose the local development factor by sampling with replacement from those in the column in which the empty cell is located. Thus for example, the empty cell located at $(n, 2)$ will be randomly filled with one of the $n-1$ local development factors $\lambda_{j,2} : j = 1, \ldots, n-1$. The final column will be filled with $\lambda_{1,n}$ because there is only one choice of local development factors to choose from.

Having filled the cumulative claims distribution table with sampled local development factors, the ultimate claims for each row can be found by multiplying together all entries in the row, that is, the initial claim amount and the remaining $n-1$ local development factors. This yields one bootstrap sample of ultimate claims, from which a sample of reserves can be found.

By repeatedly carrying out the second step on the lower diagonal, we obtain as big a bootstrap sample as desired, from which the distribution of the reserves for individual years and the distribution of total reserves may be empirically estimated, together with summary statistics such as mean and variance of reserves.

3.1.1 Analytic results

As an alternative to simulations for estimation lower order moments such as the mean and variance of reserves, one can use the local development factors to calculate explicitly these quantities as they would be found from an sample bootstrap of infinite size.

Consider the i-th row of the cumulative claims triangle. The diagonal element will be $D_{i,n-i+1}$. Let us denote the local development factor sampled in column j for this row by $\Lambda_{i,j}$, this will be sampled from the set of values $\{\lambda_{1,j}, \lambda_{2,j}, \ldots, \lambda_{n-j-1,j}\}$. The ultimate for row i will then take the value

$$D_{i,n-i+1} \prod_{j=n-i+2}^{n} \Lambda_{i,j}.$$

Now the $\Lambda_{i,j}$ in the various columns are sampled independently, hence the expected value of the bootstrap sample cumulative value in the final column of row i will be

$$E\left[D_{i,n-i+1} \prod_{j=n-i+2}^{n} \Lambda_{i,j}\right] = D_{i,n-i+1} \prod_{j=n-i+2}^{n} E[\Lambda_{i,j}]. \tag{1}$$

The quantity $E[\Lambda_{i,j}]$ for column j is simply found by averaging the local development factors in column j. Hence the mean in (1) may easily be found without simulation.

The same applies for higher moments, thus for any integer $k \geq 1$:

$$E\left[\left(D_{i,n-i+1} \prod_{j=n-i+2}^{n} \Lambda_{i,j}\right)^k\right] = D_{i,n-i+1}^k \prod_{j=n-i+2}^{n} E[\Lambda_{i,j}^k], \tag{2}$$

where the factors $E[\Lambda_{i,j}^k]$ may be found by averaging the k-th powers of the local development factors in column j. From these moments, the mean, variance and skewness of the reserves according to the local chain ladder bootstrap method may be calculated.

3.1.2 Nature of the assumptions

The basic assumption in the bootstrap algorithm as presented above is that the local development factors within a column are independent and identically distributed. Many stochastic models for claims triangles assume that incremental claims in different origin years are independent, and so the assumption of independence of local development factors within a column is not too unreasonable. The assumption of being identically distributed is used when generating the bootstrap-replicates, because of the sampling with replacement.

However it is one assumption that could be challenged, and indeed as we analyse the three datasets in the Section 4 we shall see how a judgement can be made and suggest ways in which the bootstrap may be modified if the identically distributed assumption is found to be lacking.

3.2 The vertical method

In Section 2.3 we showed the equivalence of the standard and non-standard chain ladder techniques. The horizonal local chain ladder bootstrap of the previous section is based on the standard chain ladder; an alternative, but non-equivalent method local chain ladder bootstrap method based on the non-standard bootstrap mey be developed. In this local origin factors are formed and boostrap samples made using these. This is illustrated for the 7×7 table:

	1	2	3	4	5	6	7
1	$D_{1,1}$	$D_{1,2}$	$D_{1,3}$	$D_{1,4}$	$D_{1,5}$	$D_{1,6}$	$D_{1,7}$
2	$\rho_{2,1}$	$\rho_{2,2}$	$\rho_{2,3}$	$\rho_{2,4}$	$\rho_{2,5}$	$\rho_{2,6}$	
3	$\rho_{3,1}$	$\rho_{3,2}$	$\rho_{3,3}$	$\rho_{3,4}$	$\rho_{3,5}$		
4	$\rho_{4,1}$	$\rho_{4,2}$	$\rho_{4,3}$	$\rho_{4,4}$			
5	$\rho_{5,1}$	$\rho_{5,2}$	$\rho_{5,3}$				
6	$\rho_{6,1}$	$\rho_{6,2}$					
7	$\rho_{7,1}$						

Thus to find a bootstrap value of $\hat{\Upsilon}_{2,7}$, we take $D_{1,7}$ and multiply it by one of the six $\rho_{2,j}$ which are sampled uniformly at random. Other entries are found similarly. Then by row-wise subtraction, a sample of incremental values on the lower diagonal is obtained, from which ultimates may be found.

I have not found explicit expressions for the expectations of the reserves, the formulae appear much more complicated, but it seems feasible that they could be found with a computer algorithm.

3.3 The mixed method

The mixed method combines both the standard and non-standard chain ladder technique into a local chain ladder bootstrap technique. This is illustrated in finding the $(4, 5)$ entry in the 7×7 table.

	1	2	3	4	5	6	7
1	$D_{1,1}$	$D_{1,2}$	$D_{1,3}$	$D_{1,4}$	$D_{1,5}$	$D_{1,6}$	$D_{1,7}$
2	-	-	-	-	$\lambda_{2,5}$	-	
3	-	-	-	-	$\lambda_{3,5}$		
4	$\rho_{4,1}$	$\rho_{4,2}$	$\rho_{4,3}$	$\rho_{4,4}$			
5	-	-					
6	-						
7	-						

Here we find the sampled incremental value of the $(4, 5)$ entry by using either the local development factors $\lambda_{2,5}$ or $\lambda_{3,5}$, or the local origin factors $\rho_{4,1}$, $\rho_{4,2}$, $\rho_{4,3}$ or $\rho_{4,4}$, one of which is sampled uniformly.

Note that in this method, the square is completed by filling in the lower subdiagonals in turn, starting from the largest and ending with the (n, n) entry.

Again, we have not found explicit expressions for the expectations of the reserves, the formulae appear even more complicated than in the case of the vertical method.

4 Application to historical data sets

In this section we consider three datasets that have been analysed by other methods in the literature. This enables a comparison of the local chain ladder bootstrap method to results from the analysis of other other stochastic claims reserving models. For each of the three datasets we shall only compare the horizontal bootstrap method to the other methods, as moments can be found explicitly without recourse to bootstrap sampling.

4.1 Data set 1

Our first dataset is taken from England and Verrall (2001), and refers to incremental paid losses from an aggregation of classes of business. The data appears in Appendix A, in both incremental and cumulative format. The incremental values have a negative value at position $(3, 3)$. England and Verrall (2001) analysed this data using an over-dispersed Poisson model with a logarithmic link function. They considered three choices of the predictor structure:

1. The model of Renshaw and Verrall (1998)

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

2. The Hoerl curve model (ignoring inflation):

$$\eta_{ij} = \mu_j + c + \alpha_i + \beta \log(j) + \gamma j$$

3. A generalised additive model smoothed over log development time:

$$\eta_{ij} = \mu_j + c + \alpha_i + s_{\theta_j}(\log(j)).$$

We shall refer to these models as ODP1, ODP2 and ODP3 respectively. They also considered gamma models using the same three predictor structures above, which we shall refer to as G1, G2 and G3 respectively. The reader is referred to the original paper for more details about the models.

Table 5: Local development factors for Dataset 1. The first column shows the initial claims amounts in the first development year, the remainder of the cells in the triangle show the local development factors. the column means of the local development factors are also show, together with development factors calculated using the chain-ladder technique. The final row shows the mean squares of the local development factors that are used in estimating the standard deviation of the reserve estimates.

Initial	Development year								
Claim	2	3	4	5	6	7	8	9	10
45630	1.51172	1.04239	1.02501	1.02723	1.01590	1.01688	1.00720	1.00986	1.00781
53025	1.49912	1.03559	1.02123	1.00871	1.01679	1.00463	1.00620	1.00390	
67318	1.62885	0.98309	1.02948	1.02744	1.02878	1.02480	1.02174		
93489	1.40083	1.05674	1.04804	1.02901	1.03861	1.01219			
80517	1.41061	1.06043	1.03593	1.03208	1.01825				
68690	1.49397	1.05501	1.05706	1.03040					
63091	1.51034	1.09380	1.06600						
64430	1.50428	1.08681							
68548	1.51593								
76013									
Mean	1.49730	1.05173	1.04039	1.02581	1.02367	1.01462	1.01171	1.00688	1.00781
CL	1.49057	1.05165	1.04194	1.02676	1.02538	1.01492	1.01303	1.00673	1.00781
Mean Sq	2.24578	0.98414	1.08267	1.05235	1.04797	1.02952	1.02361	1.01382	1.01567

The first step in the bootstrap process is to calculate the local development factors. These are displayed in Table 5. Notice that in each column the average of the local development factors is quite close to the development factor calculated by the chain-ladder technique. From the analysis of Section 3.1.1 this suggests that the mean reserves calculated by the two methods will be similar. That this is so is shown in Table 6, which shows the mean reserve estimates for each accident year and the total for the various models.

Table 6: Predictions of mean reserve estimates for Dataset 1: Chain ladder prediction, the local chain ladder bootstrap, and the six models of England and Verrall (2002)

Accident year	CL mean	Local CL	ODP1	ODP2	ODP3	G1	G2	G3
1	0	0	0	0	0	0	0	0
2	683	683	683	1,085	622	488	675	450
3	1,792	1,811	1,792	3,101	1,998	2,086	3,296	2,205
4	4,363	4,178	4,363	6,129	4,470	5,240	6,818	5,300
5	5,657	5,460	5,657	7,173	5,940	6,169	7,061	6,313
6	8,209	7,817	8,209	8,689	8,106	9,750	9,305	9,427
7	10,914	10,423	10,914	11,031	11,106	15,080	13,029	15,097
8	15,199	14,536	15,199	14,765	15,112	18,498	15,069	17,671
9	21,135	20,457	21,135	24,002	21,293	20,470	24,400	20,896
10	60,335	60,207	60,335	59,625	60,377	60,043	59,576	58,519
Total	128,286	125,572	128,286	135,600	129,024	137,824	139,229	135,878

In table Table 7 we compare the standard deviation of the local chain ladder bootstrap method to the prediction errors of the other methods. Note that the latter are larger. This can be explained in that they also take into account uncertainty in the estimates of the parameters of the model, a component that is not present (in an obvious manner) in the non-parametric local chain ladder bootstrap bootstrap method.

Now let us examine one of the assumptions of the local chain ladder bootstrap procedure. Consider the sampling for the cell $(n, 2)$. We sample the local development factors from the nine given in column 2 of Table 5, and will use this to multiply by the initial claim amount in cell $(n, 1)$ to get a random draw of the cumulative claim amount for cell $(n, 2)$. Now in sampling with replacement, we are making the tacit assumption that the choice of local development factor to be used does not depend on the size of the initial claim amount in cell $(n, 1)$. A simple way to test this assumption is to draw a scatter-plot of the local development factors in column to against the corresponding initial claim amounts. Such plot is shown in Figure 2. From the plot there does not appear to be much

Table 7: Standard deviations for local stochastic chain ladder, and prediction errors for the other models, expressed as percentages of mean estimates. Dataset 1

Accident year	Local sd	PE ODP1	PE ODP2	PE ODP3	PE G1	PE G2	PE G3
1	-						
2	0%	159%	95%	110%	62%	46%	43%
3	20%	100%	61%	62%	43%	36%	33%
4	29%	63%	46%	43%	36%	32%	29%
5	26%	50%	43%	38%	32%	30%	28%
6	22%	40%	39%	33%	31%	29%	28%
7	18%	34%	34%	29%	31%	29%	29%
8	18%	28%	30%	25%	32%	30%	31%
9	23%	24%	23%	22%	36%	35%	35%
10	13%	17%	17%	16%	52%	48%	48%
Total	8%	15%	15%	12%	25%	23%	24%

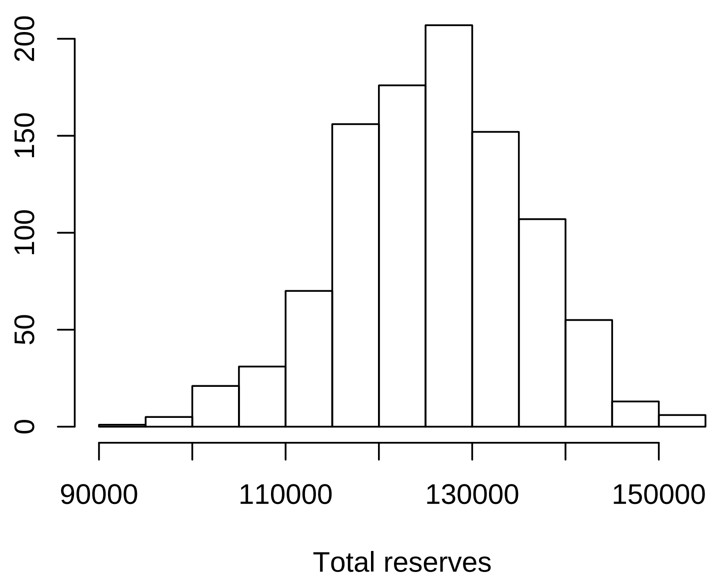


Figure 1: Distribution of total reserves for Dataset1, according to the horizontal local chain ladder bootstrap method, based on a sample of 1000 simulations.

dependence on the two factors, except perhaps for large claim amounts: however as there are only a few data-points this might not be the case. We are going to apply the sampled local development factor to a claim amount of 76013, so it would appear that sampling at random from all of the local development factors would appear to be ok. If the pattern of slightly lower development factor for higher claim amounts as displayed by the scatter-plot is correct, then perhaps this means our estimates are slightly above what they should be. We shall see different behaviours in the next two datasets.

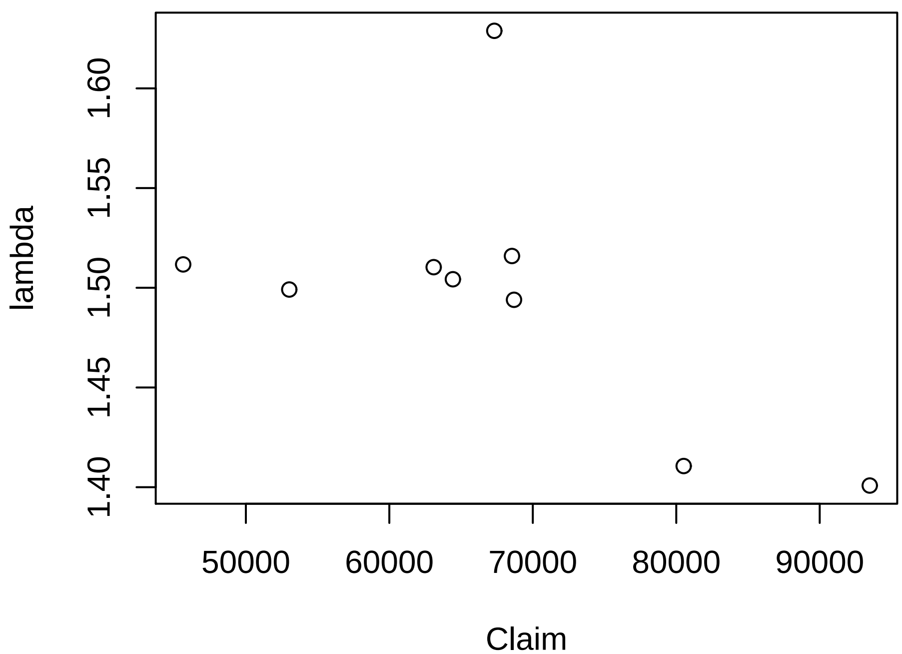


Figure 2: Scatter-plot of local development factors in development year 2 against the claim amounts that they are obtained from. For dataset 1

4.2 Data set 2

The second dataset we analyse is given in Appendix B, was presented by Taylor and Ashe (1983). This dataset has been analysed by a number of authors whose results have been collected together in the paper of England and Verrall (1999) that we take from here for further comparison to the local chain ladder bootstrap method.

As before, the first step in the local chain ladder bootstrap method is to find the local development factors, these are shown in Table 8.

Table 8: Local development factors for Dataset 2. The first column shows the initial claims amounts in the first development year, the remainder of the cells in the triangle show the local development factors. the column means of the local development factors are also show, together with development factor calculated using the chain-ladder technique.

Initial	Development year								
Claim	2	3	4	5	6	7	8	9	10
357848	3.14320	1.54281	1.27830	1.23772	1.20921	1.04408	1.04037	1.06301	1.01772
352118	3.51058	1.75549	1.54529	1.13293	1.08449	1.12811	1.05727	1.08650	
290507	4.44845	1.71672	1.45826	1.23208	1.03686	1.12001	1.06058		
310608	4.56800	1.54705	1.71178	1.07252	1.08736	1.04708			
443160	2.56420	1.87296	1.36154	1.17422	1.13831				
396132	3.36559	1.63568	1.36916	1.23644					
440832	2.92280	1.87810	1.43939						
359480	3.95329	2.01565							
376686	3.61918								
344014									
Mean	3.56614	1.74556	1.45196	1.18098	1.11125	1.08482	1.05274	1.07475	1.01772
CL	3.49061	1.74733	1.45741	1.17385	1.10382	1.08627	1.05387	1.07656	1.01772
Mean Sq	13.11458	3.07200	2.12556	1.39856	1.23830	1.17838	1.10834	1.15523	1.03576

Note again that the average of the local development factors in each column is close to that of the standard chain-ladder development factors, thus again we would expect the predictions of the mean reserves should be similar for the two methods. That this is so is shown in Table 9, which shows the mean reserve estimates for each accident year and the total for the various models. This table shows that the prediction of means reserves for the local chain ladder bootstrap model is in line with the chain ladder model and the various other models.

Next we looked at the standard deviation of the local chain ladder bootstrap model,

Table 9: Predictions of mean reserve estimates for Dataset 2 for the local chain ladder bootstrap, and from various other models as given presented by England and Verrall(1999). (See this paper for the sources of the other models.) Values have been scaled by a factor of 1000.

Accident year	CL mean	local clb	Poisson GLM	Gamma GLM	Mack(1991)	Verrall(1991)	Renshaw/ Christofides	Zehnwirth
2	95	95	95	93	93	96	111	109
3	470	461	470	447	447	439	482	473
4	710	695	710	611	611	608	661	648
5	985	965	985	992	992	1011	1091	1069
6	1,419	1,433	1419	1453	1453	1423	1531	1500
7	2,178	2,227	2178	2186	2186	2150	2311	2265
8	3,920	3,954	3920	3665	3665	3529	3807	3731
9	4,279	4,301	4279	4122	4122	4056	4452	4364
10	4,626	4,753	4626	4516	4516	4340	5066	4965
Total	18,681	18,883	18,682	18,085	18,085	17,652	19,512	19,124

Table 10: Prediction errors as a percentage of reserve estimates according to the various models presented by England and Verrall(1999). (See this paper for the sources of the other models.)

Accident year	Local clb	Mack's dist. free	Poisson GLM analytic	Bootstrap chain ladder	Gamma Chain GLM	Mack (1991)	Verrall (1991)	Renshaw/ Christofides	Zehnwirth
2	0	80	116	117	48	40 (49)	49	54	49
3	13	26	46	46	36	30 (37)	37	39	35
4	10	19	37	36	29	24 (30)	30	32	29
5	19	27	31	31	26	21 (26)	27	28	25
6	23	29	26	26	24	20 (25)	25	26	24
7	22	26	23	23	24	20 (25)	25	26	24
8	21	22	20	20	26	21 (26)	27	28	26
9	20	23	24	24	29	24 (30)	30	31	30
10	25	29	43	43	37	31 (38)	38	40	39
Total	10	13	16	16	15	-	15	16	16

and compared it with the prediction errors of the various other models: these are shown in Table 10. Again that for the local chain ladder bootstrap is smaller than for the other methods. Figure 3 show the distribution of total reserves based on a bootstrap sample of 1000 observations.

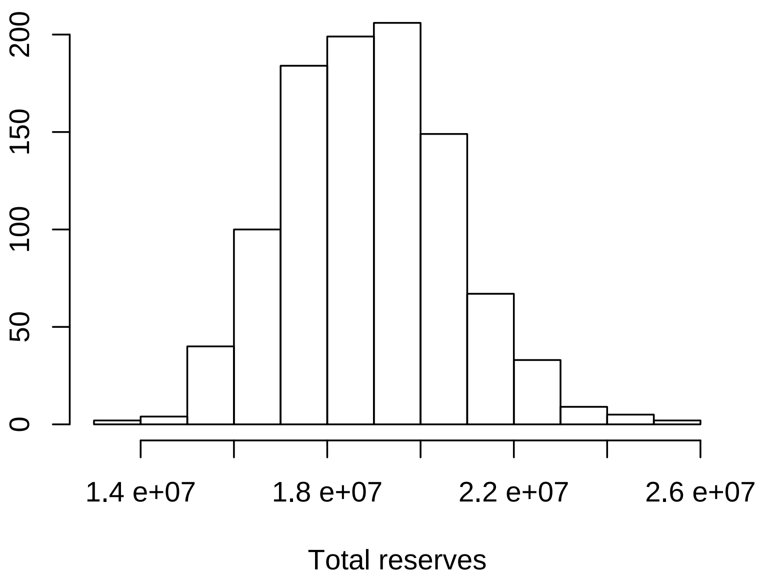


Figure 3: Distribution of total reserves for Dataset 2, according to the horizontal local chain ladder bootstrap method, based on a sample of 1000 simulations

Now let us plot the local development factors in development year 2 against the initial claims amounts: this is done in Figure 4. Here we see a strong dependence of the two variables, with a linear correlation coefficient of -0.91. This suggest that the sampling with replacement is not be the appropriate procedure for this dataset. Now the total claim amount that the sampled development factor is going to be applied to is 344014. Fitting a linear regression line yields a predicted value of 3.8662 for the mean local development factor that should be applied here. This is slightly above the mean of 3.566 of the local development factors in development year 2, which suggest our estimate of reserves for accident year 10 is on the low side, by about 8 percent or so. In addition it would appear that the two large development factors at the top left of the plot together with the smallest at the bottom right are perhaps less likely values and so a simple sampling of all of the factors is perhaps not appropriate for this dataset. By including them equally with the other factors, we have inflated the variability in our bootstrap sample. Two options suggest themselves: (i) do not use these three values in the set of local development factor when drawing a sample; (ii) put weights on the various local development factors so that they are not sampled equally likely. As there could be many ways of doing this, I did not pursue this.

Note that as a corollary, if the claim amount were around 30000, then this simple diagnosis would suggest an underestimation the reserves. Conversely, it if were high, say around 45000, then it would be overestimating the reserves. In both cases we would be overestimating the variability.

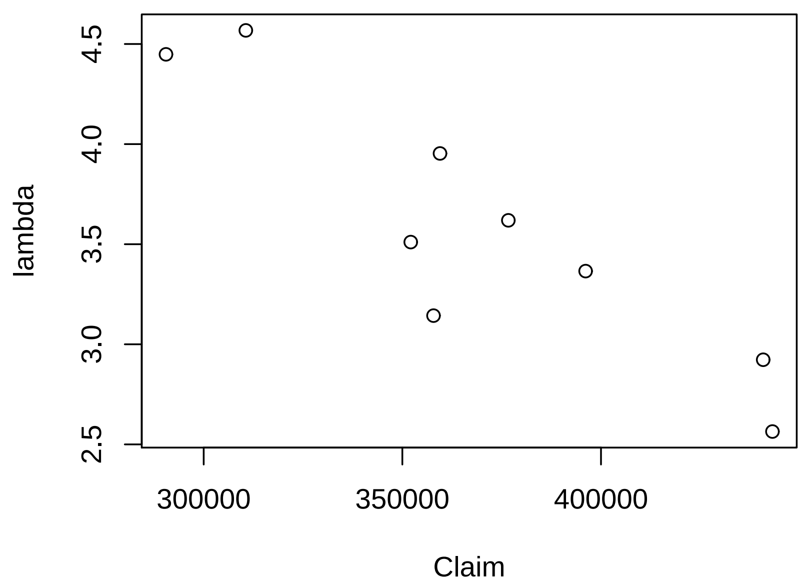


Figure 4: Scatter-plot of local development factors in development year 2 against the claim amounts that they are obtained from. For Dataset 2

4.3 Data set 3

We now come to the third dataset, which is taken from (Mack 1991), and is given in Appendix C. Notice this table has a negative incremental claim at position $(2, 7)$. The set of local development factors is shown in Table 11

Table 11: Local development factors for Dataset 3

Initial	Development year								
claim	2	3	4	5	6	7	8	9	10
5012	1.64984	1.31902	1.08233	1.14689	1.19514	1.11297	1.03326	1.00290	1.00922
106	40.42453	1.25928	1.97665	1.29214	1.13184	0.99340	1.04343	1.03309	
3410	2.63695	1.54282	1.16348	1.16071	1.18570	1.02922	1.02637		
5655	2.04332	1.36443	1.34885	1.10152	1.11347	1.03773			
1092	8.75916	1.65562	1.39991	1.17078	1.00867				
1513	4.25975	1.81567	1.10537	1.22551					
557	7.21724	2.72289	1.12498						
1351	5.14212	1.88743							
3133	1.72199								
2063									
Mean	8.20610	1.69589	1.31451	1.18293	1.12696	1.04333	1.03436	1.01799	1.00922
CL	2.99936	1.62352	1.27089	1.17167	1.11338	1.04193	1.03326	1.01694	1.00922
Mean Sq	202.70628	3.07272	1.81412	1.40303	1.27450	1.09043	1.06994	1.03654	1.01852

One point to note is that again the means of the local development factors are close to the chain ladder development factors, with the marked exception for this in development year 2. Here there a large range of values for the local development factors, with a large value of 40.42 in the second accident year which looks like an outlier. With the mean in this column being 2.73 times the chain ladder development factor, we could anticipate that the estimates of reserves for this accident year will be almost 3 times that of the chain-ladder estimate, and this is indeed the case, as shown in Table 12

Table 12: Mean reserves for Dataset 3, according to chain ladder, local chain ladder bootstrap, and a range of other models presented in England and Verrall (2002). (See this paper for details and references about these other models.)

Accident year	CL mean	Local clb	O.D. Poisson	Negative Binomial	Normal/ negbin	Mack1 Tab13	Mack2 Tab14	LogNormal	Gamma
2	154	154	154	154	154	154	154	357	133
3	617	642	617	617	617	617	617	1,020	588
4	1,636	1,696	1,636	1,636	1,636	1,636	1,636	3,064	1,659
5	2,747	2,846	2,747	2,747	2,747	2,747	2,747	3,753	2,242
6	3,649	3,955	3,649	3,649	3,649	3,649	3,649	6,010	3,330
7	5,435	5,887	5,435	5,435	5,435	5,435	5,435	7,742	4,654
8	10,907	12,363	10,907	10,907	10,907	10,907	10,907	18,806	9,863
9	10,650	12,381	10,650	10,650	10,650	10,650	10,650	25,367	13,630
10	16,339	53,718	16,339	16,339	16,339	16,339	16,339	56,475	17,842
Total	52,135	93,643	52,134	52,135	52,135	52,135	52,135	122,595	53,940

Table 13: Prediction errors as a percentage of reserve estimates according to the various models presented by England and Verrall(2002). (See this paper for the sources of the other models.)

Localsl sd	O.D. Poisson	Negative Binomial	Normal/ negbin	Mack1 Tab13	Mack2 Tab14	Lognormal	Gamma
0%	361%	367%	241%	134%	324%	210%	139%
56%	181%	185%	104%	101%	140%	139%	98%
28%	109%	110%	47%	46%	62%	107%	83%
46%	81%	83%	50%	53%	59%	94%	74%
37%	67%	68%	50%	55%	57%	87%	76%
28%	57%	59%	37%	41%	42%	86%	74%
50%	46%	47%	46%	49%	49%	87%	76%
52%	57%	58%	56%	59%	60%	98%	84%
161%	79%	81%	143%	150%	150%	137%	107%
93%	35%	36%	49%	53%	52%	70%	48%

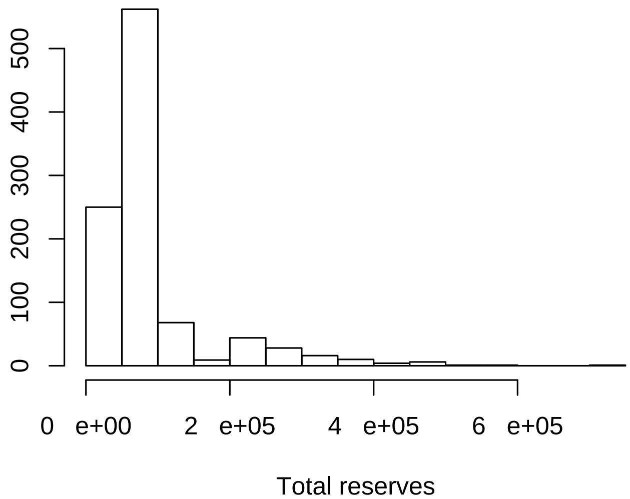


Figure 5: Distribution of total reserves for Dataset 3, according to the local chain ladder bootstrap method, based on a sample of 1000 simulations

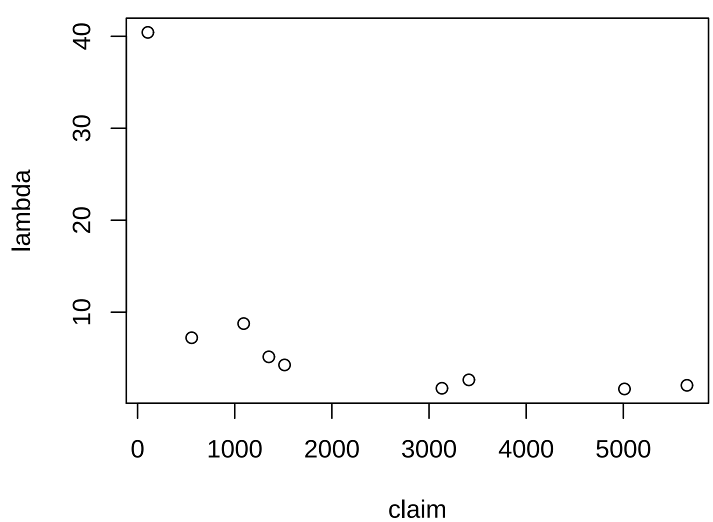


Figure 6: Scatter-plot of local development factors in development year 2 against the claim amounts that they are obtained from. For dataset 3

Table 14: Modified mean reserve estimates after removing the large local development factor from analysis

Origin year	E[R]	sd(u)
2	154	0
3	642	357
4	1,696	469
5	2,846	1,300
6	3,955	1,472
7	5,887	1,648
8	12,363	6,161
9	12,381	6,431
10	26,342	20,869
total	66,267	22,842

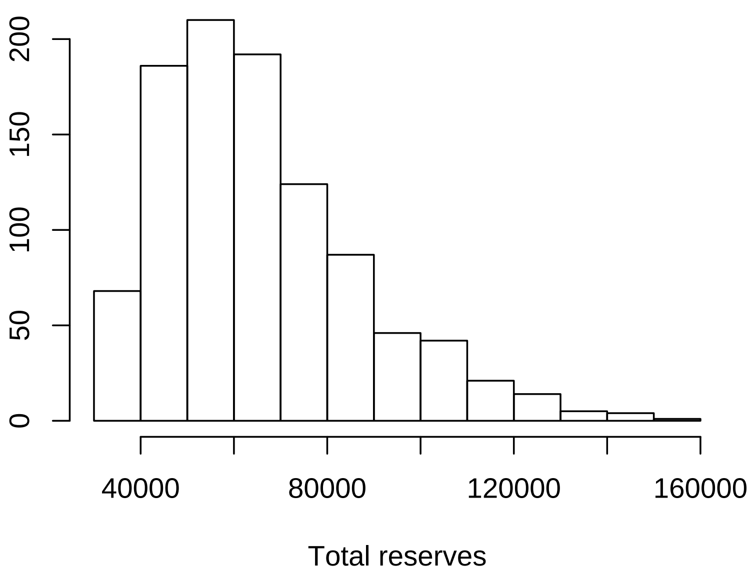


Figure 7: Distribution of total reserves for Dataset 3, according to the local chain ladder bootstrap method, based on a sample of 1000 simulations, after removing the large local development factor.

5 A simulation study

The comparisons carried out in Section 4 indicate that the non-parametric (horizontal) bootstrap method gives comparable results for the estimate of the mean reserves when compared to other methods, but also generally estimates of the variance for the reserves compared to other methods. But such comparisons are quite limiting, in that it is not possible to tell which method is performing best in terms of matching the true reserve distribution. For this reason, a simulation study was carried out, in which claims triangles were simulated and the theoretical distribution of reserves was known for each given claims triangle. This it was possible to compare the true distribution of reserves against the predicted distribution. In this simulation study estimated reserves based on the non-parametric bootstrap methods (all three variants), the over-dispersed Poisson model of England and Verrall (1999), and the model of Mack (1993) were compared.

For the simulation study two methods for simulating claims triangles were used. One is a method proposed by Schiegl (2004) as described by Verdonck *et al.* (2007), the other by Professor Kaishev of the Cass Business School (personal communication). For both models the following approach was used.

1. Use the model to simulate an upper claims triangle.
2. Use the model to simulate 1000 completions of the lower triangle.
   - This gives a sample from the true distribution of the reserves.
3. Use the local chain ladder bootstrap method (all three variants), the over-dispersed Poisson method, and Mack's method, to estimate the distributions of reserves, using as input the upper triangle of Step 1. For each of these six methods a bootstrap procedure was employed to generate 1000 samples.
4. Compare the true and estimated reserve distribution using quantile-quantile plots and boxplots.

The above procedure was repeated eight times: four times using claims triangles simulated with Schiegl's method, and four times using claims triangles simulated with Kaishev's method. We now present the details of the methods for simulating the claims triangles, and the results of the simulation study.

5.1 Schiegl's triangle simulation method

In the modified approach of Schiegl (Verdonck *et al.* 2007), the number of claims in the cell (i, j) of a claims triangle has a Poisson distribution with rate

$$\lambda_{ij} = \lambda_0(1 + (i-1)\eta_2)e^{2(j-1)\log(\eta_1)/n}$$

where (Verdonck *et al.* 2007) "λ_0 *is the expectation value of claim number at first occurrence and first development year,* η_2 *is the rate of increase for number of claims over occurrence years,* η_1 *measures the decrease of claim numbers for passing development years and* n *is the number of accident years*".

Following (Verdonck *et al.* 2007), we took $\lambda_0 = 100$ $\eta_1 = 0.3$, $\eta_2 = 0.05$, and considered claims triangles over $n = 10$ development years.

Claims are taken to be independent, hence it is possible to simulate from the distribution of the lower triangle given the upper triangle, and thus estimate the true distribution of reserves, for each claim year and also the total.

The simulation algorithm is as follows:

- Initialize $C_{i,j} = 0$ for all i and $j \le n$.
- Set $\lambda_{ij} = \lambda_0(1 + (i-1)\eta_2)e^{\frac{2\log\eta_1}{n}(j-1)}$
- Simulate $N_{ij} \sim \text{Poisson}(\lambda_{ij})$
- Set $C_{ij} = \sum_{k=1}^{N_{ij}} Z_k$ where $Z_k \sim \text{Gamma}(r, 1)$.

In fact, for the last step it was quicker and simpler to simulate C_{ij} from $\text{Gamma}(rn_{ij}, 1)$ once using the simulated value n_{ij} of N_{ij}.

One important thing to note which makes this simulation method a little unsatisfactory for the present purpose is that, given the same set of parameter values for λ_0, η_1, η_2 and n, the distribution of reserves is the same regardless of the sample drawn for the upper triangle. This is because each C_{ij} value is sampled independently of the others.

The results of the simulations are shown as plots in the following four figures. The format of the plots is the same in all figures, and is as follows.

The top set of plots shows quantile-quantile plots. In the top left, the sample of 1000 draws from the true distribution has been split into two set of 500, and a quantile-quantile plot for each half has been formed: theoretically the points should lie close to the diagonal line, and this is observed in all simulations.

The remaining quantile-quantile plots on the top row are the bootstrap samples from Mack's method plotted against the true distribution sample, and the over dispersed Poisson sample plotted against the true distribution sample. On the bottom row are the quantile-quantile plots for the horizontal local chain ladder bootstrap (NPNa), the vertical local chain ladder bootstrap (NPBd), and the combined or mixed local chain ladder bootstrap method (NPBb) respectively.

The lower plot shows comparative boxplots of the true distribution samples against the reserve samples for the five methods of fitting the claims triangle.

Figure 8: Simulation 1: In these plots, the local chain ladder bootstrap appears to have variability closer to the true distribution, with the horizontal and combined variants performing better. All methods underestimate the mean.

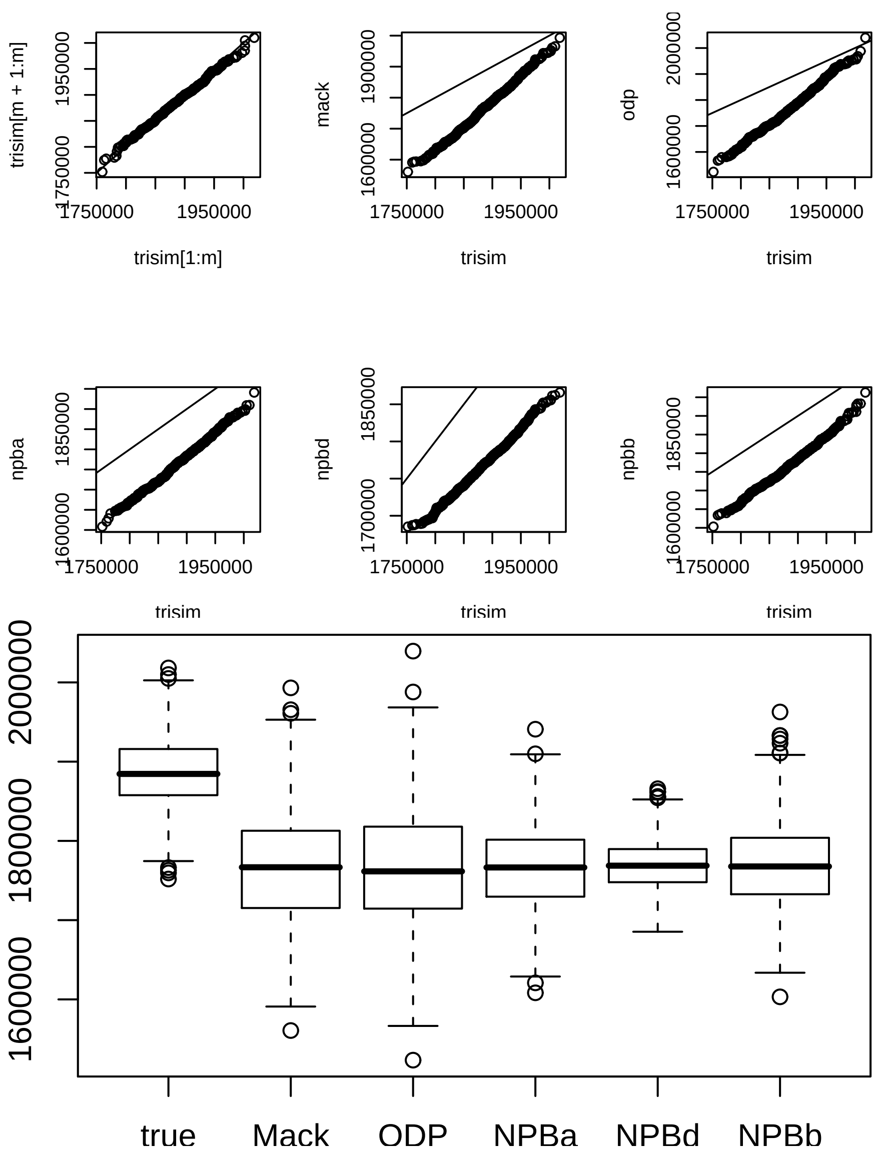


Figure 9: Simulation 2: In these plots, the vertical local chain ladder bootstrap appears to have variability closer to the true distribution. The other methods are noticeably out. All methods overestimate the mean.

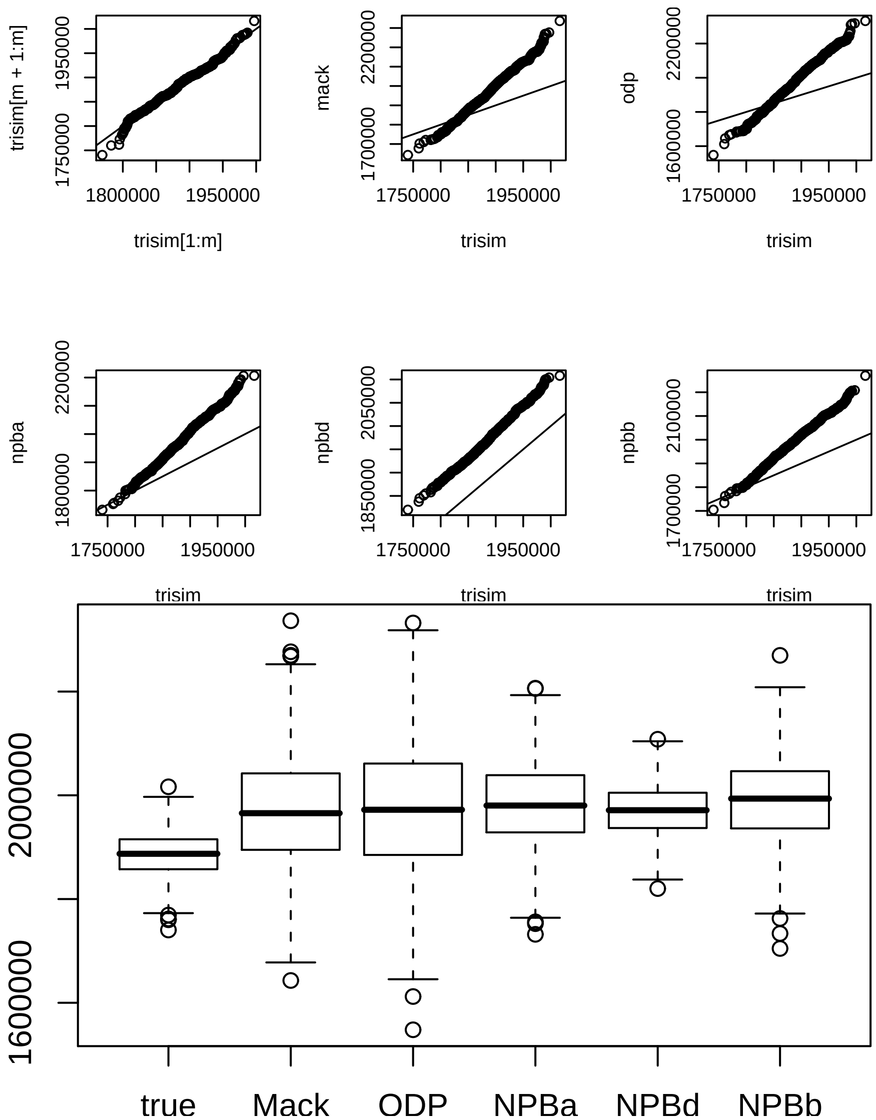


Figure 10: Simulation 3: In these plots, the vertical local chain ladder bootstrap appears to be very close to the true distribution, the qqplot indicating the distributions match but with a slight bias in the localtion of the distirbution

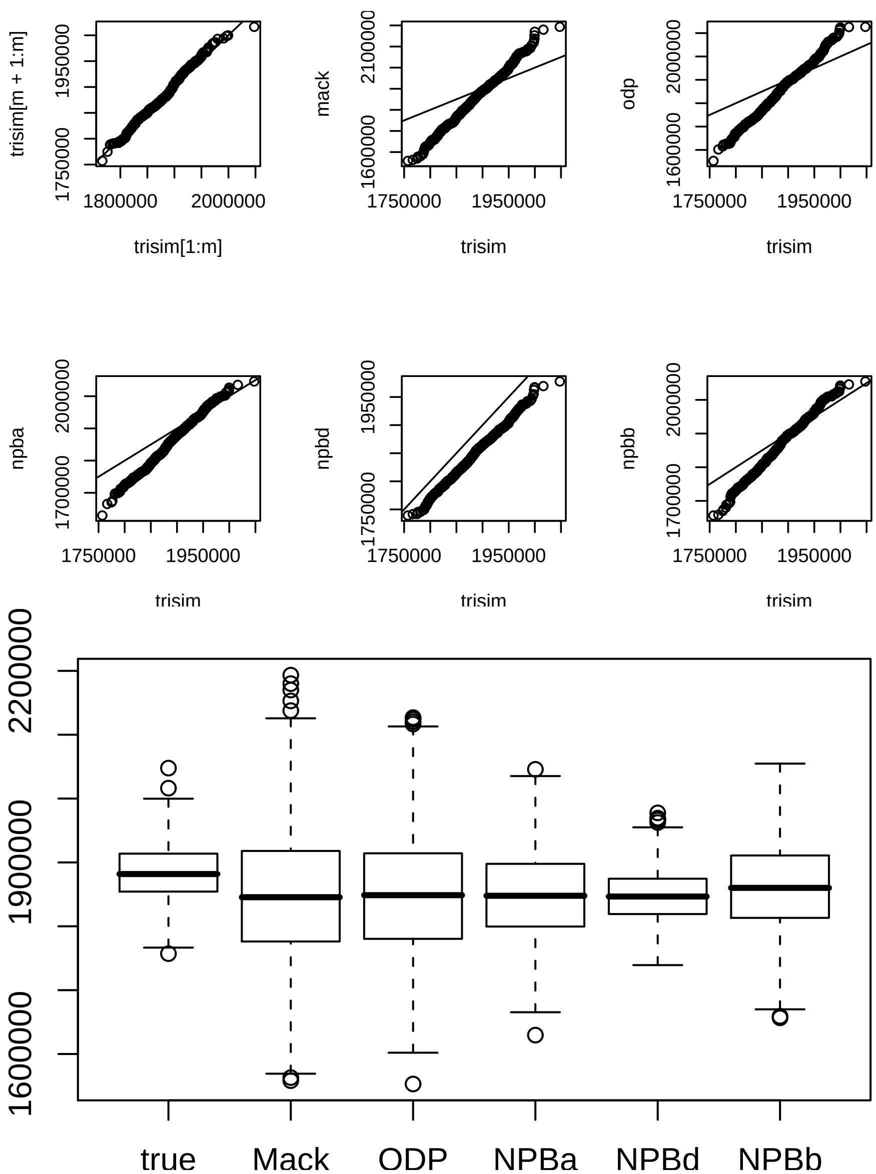
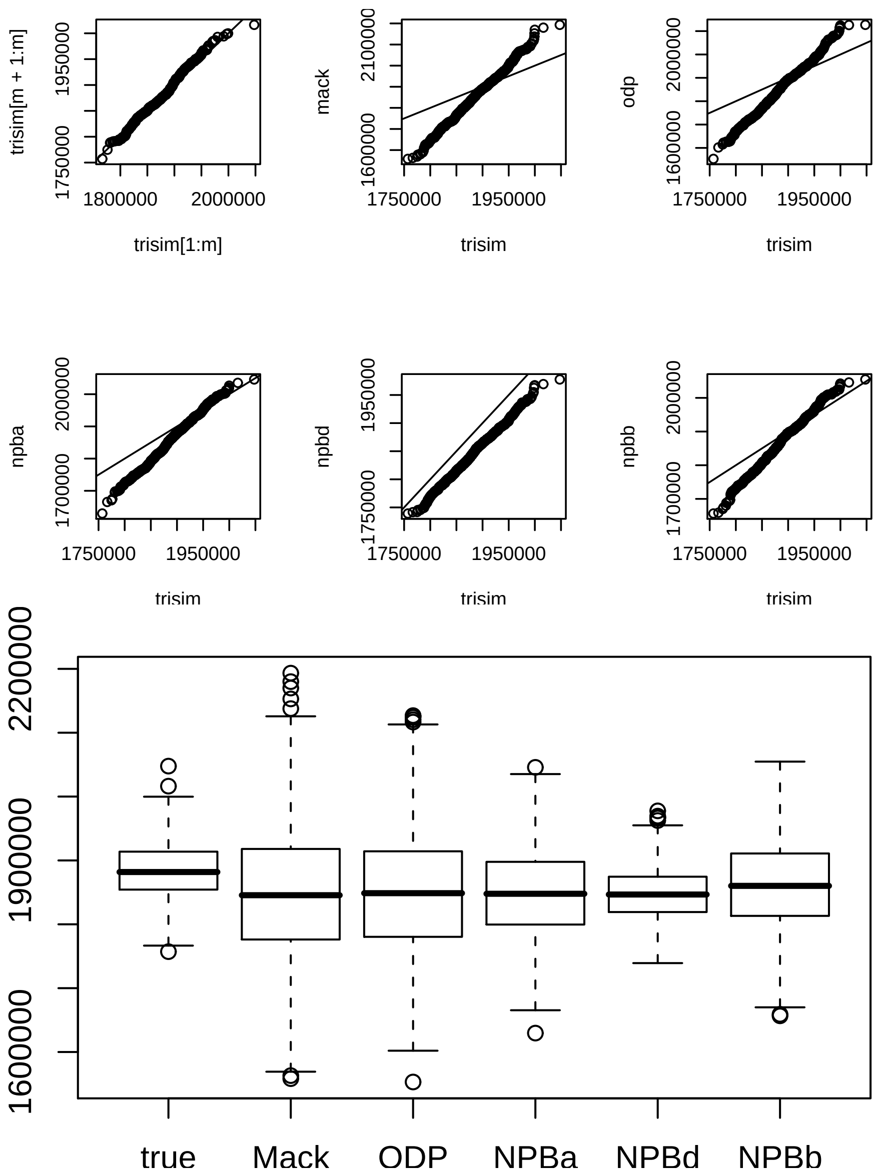


Figure 11: Simulation 4: Similar results to Simulation 3, again the vertical local chain ladder bootstrap is very closre to the true distribution than the other methods, which have much larger variability, and the wrong shape.



5.2 Kaishev's triangle simulation method

Kaishev's method uses a heavier tailed lognormal distribution instead of a gamma distribution for the size of individual claims. The model is a bit more involved than Schiegl's model.

- Initialize $C_{i,j} = 0$ for all i and $j \leq n$.
- For each row i
  - Simulate $N_i \sim \text{Poisson}(\lambda)$
  - for (i = 1 to N_i) do
    - Simulate: $k = \lfloor n \times (Beta)(\alpha, \beta) \rfloor$
    - Simulate: $X \sim \text{LogNorm}(\mu, \sigma)$
    - Increment: $C_{i,k}$ by X

Essentially, N_i claims are simulated for the i-th origin year. Each claim has a lognormal distribution, and is allocated randomly to a development year (the values of k) by sampling from a Beta distribution.

In the simulations we used use $n = 10$, $\lambda = 1000$, $\alpha = 1$, $\beta = 1.6$, $\mu = 1$, $\sigma = 1.8$; these were values suggested to me as generating realistic looking claims triangles by Professor Kaishev.

As an additional important point, if the N_i values are stored for each row, and the number of k values in the i-th row up to the main diagonal entry of the claims triangle are found, then this can be used to conditionally sample incremental claim amounts in the rest of the row. Hence, conditional on a given sample in the upper claims triaingle, it is possible to simulate from the distribution of reserves in the lower triangle. Crucially, even if the parameters $n =$, λ, α, β, μ and σ are ketp fixed, different simulations will lead to different distirbutions of reserves, in marked contrast to the Schiegl method of simulating claims triangles.

The results for four simulations are now presented, in the form of plots similar in format to those in the previous section.

Figure 12: Simulation 5: All methods are underestimating the median, the mixed vertical method appears to be the best, though all perform badly in the high tail of the true distribution.

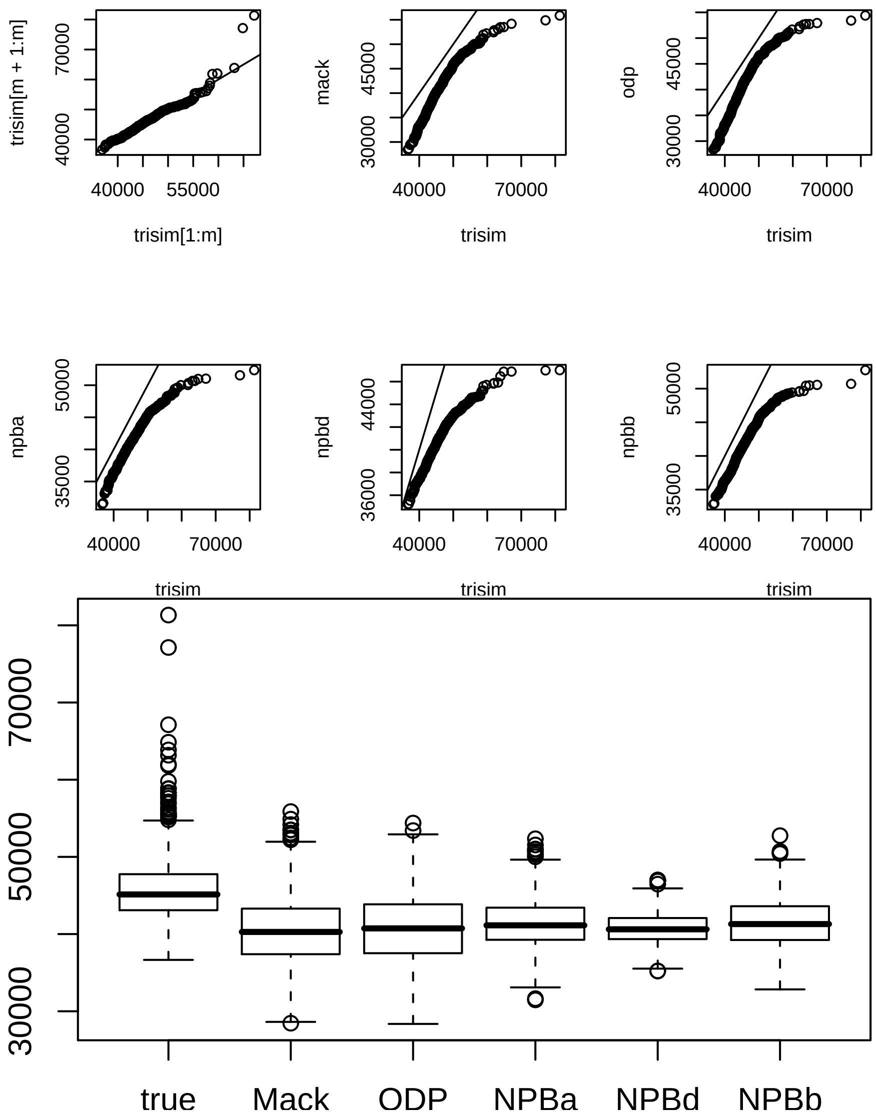


Figure 13: Simulation 6: The vertical local chain ladder bootstrap method appears to follow the true distribution very closely, except in the high tail of the true distribution.

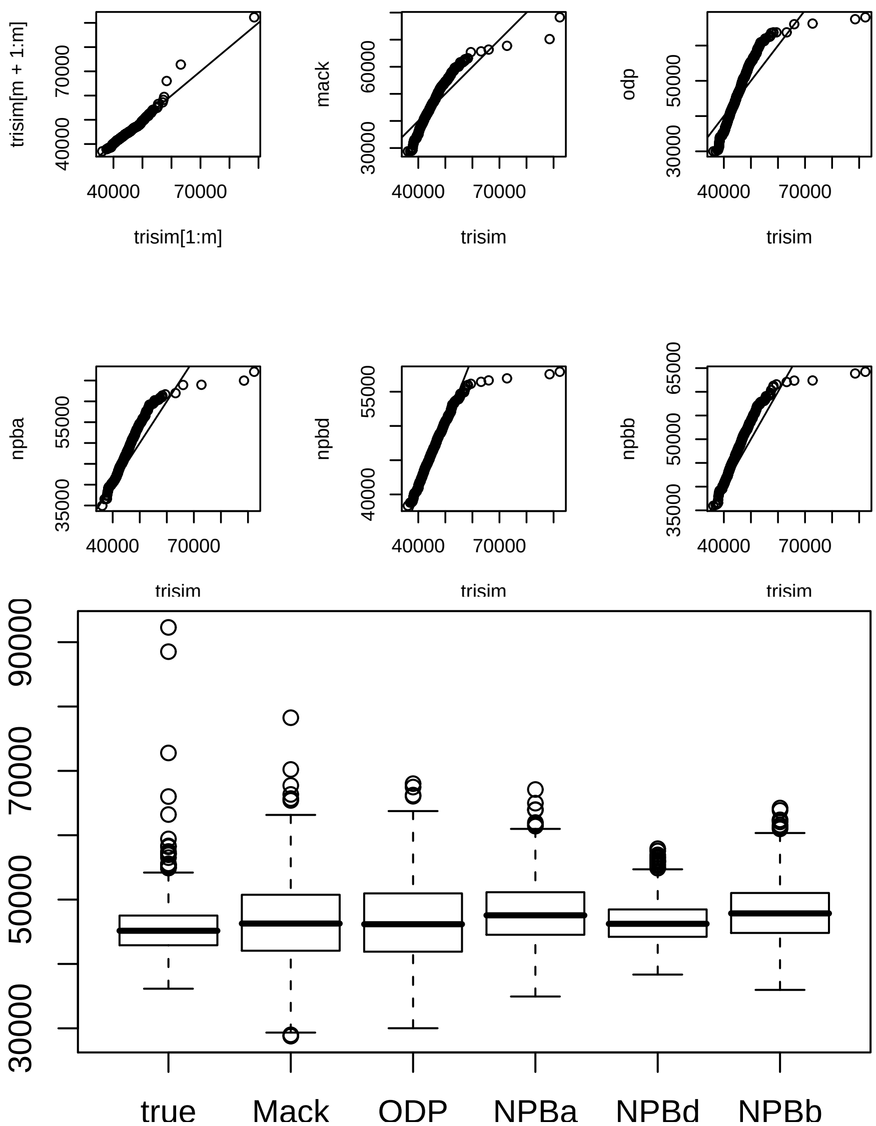


Figure 14: Simulation 7: The vertical local chain ladder bootstrap method fares best in terms of shape, the location of the median is too high. All other methods have much larger variability.

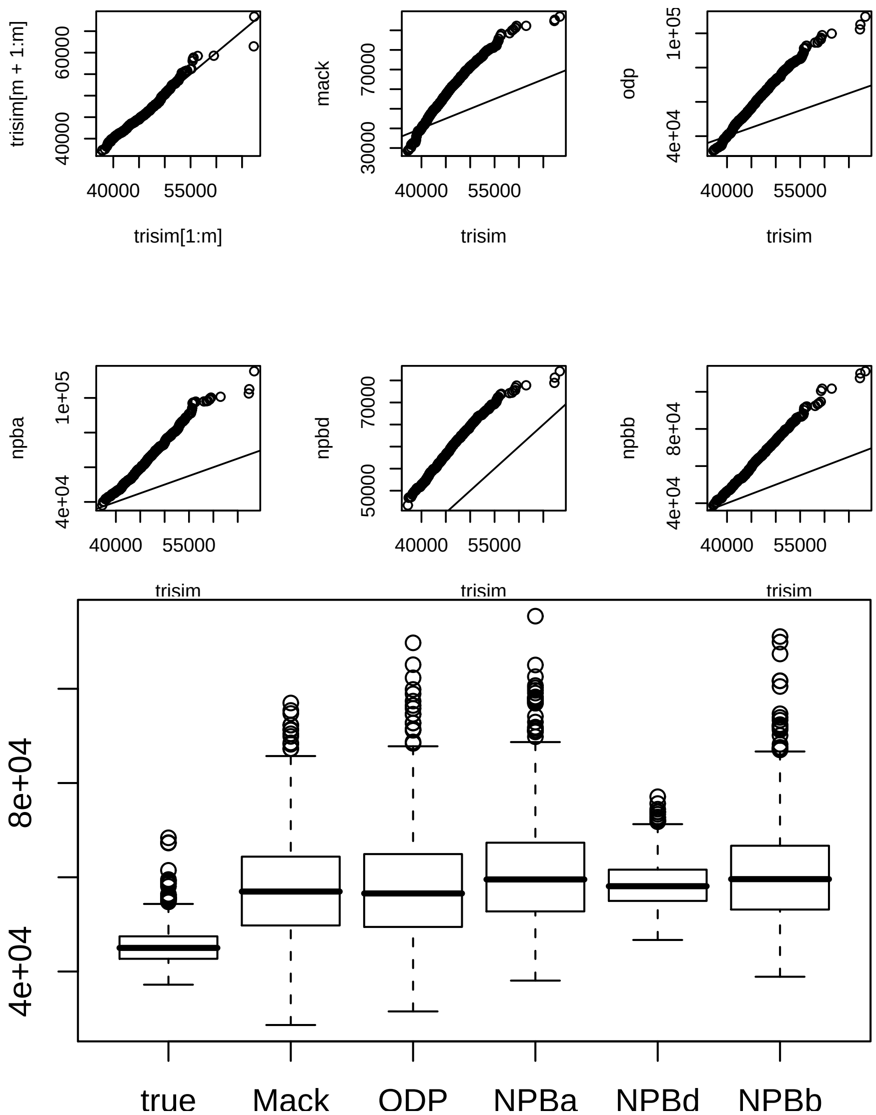
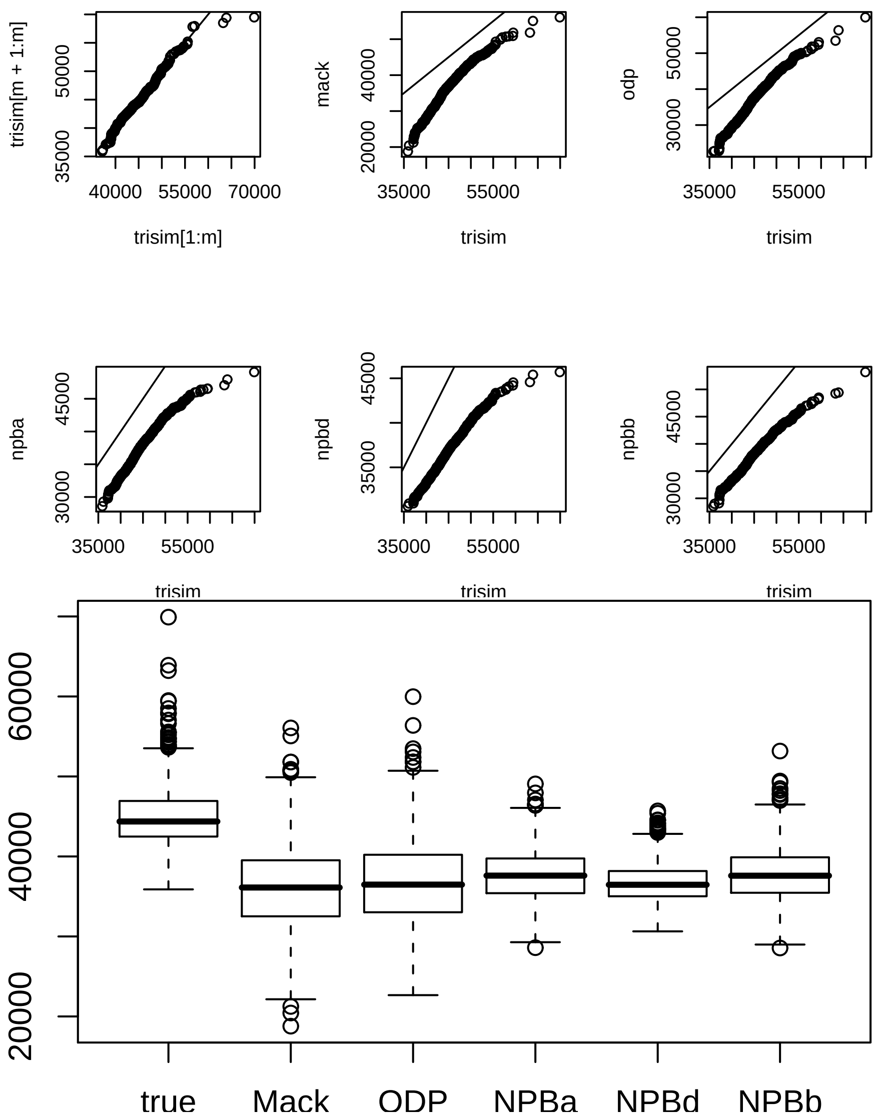


Figure 15: Simulation 8: The horizontal and combined local chain ladder bootstrap methods fare best in terms of shape, all methods underestimating the median.



6 Summary, and suggestions for future work

This report has summarized the results of an exploratory study of a simple and novel non-parametric bootstrap method for estimating the distribution of reserves given data in the form of a claims triangle. A simple mathematical proof of the equivalence of standard and non-standard chain ladder calculations was presented, that is believed to be new. This inspired three variants of the local chain ladder bootstrap method, which was then examined using both historical data and in a simulation study.

The analysis of historical data showed that the horizontal local chain ladder bootstrap method produced values for the expected values of reserves similar to a number of other stochastic claims reserving methods. However the standard deviation was generally much lower than returned by the other stochastic methods. One explanation for this is that the local chain ladder bootstrap method does not have a component corresponding to *parameter estimation error*, that is present in the other stochastic claims reserving methods examined. This would suggest that the method would always inherently underestimate uncertainty. However the simulation study showed that the local chain ladder bootstrap methods, particularly the vertical variant, produced reserve distributions that usually match the shape of the true distribution quite well, with other methods producing much larger variation in all simulations.

The simulation study showed that the local chain ladder bootstrap method tend to be biased in location of the distribution (median), but in an unpredicatble way. Curiuosly, in each simulation, the local chain ladder bootstrap method was biased in the same direction and about the same amount as the over-dispersed Poisson method and Mack's (1991) method.

This suggests that if it could be understood when the bias is happening, and could be corrected for by a simple estimable translation, the local chain ladder bootstrap method might provide a reasonable simple and robust method for estimating the distribution of claims reserves. The analysis of the historical data suggested a simple way via scatterplots in which the assumptions of the local chain ladder bootstrap method could be tested, and outliers detected. I suggest that such diagnostics should always be carried out.

Acknowledgements

The financial support of The Actuarial Profession for this project is gratefully acknowledged, without it this work may never have been done.

I would like to thank Professor Kaishev for discussions, and for providing me with his method for simulating realistic looking claims triangles.

I would also like to thank the people at the Cass Business School who dealt with the administrative side of this grant, in particular John Montgomery.

References

England, P. D. and Verrall, R. J. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: Mathematics and Economics*, **25**, 281–93.

England, P. D. and Verrall, R. J. (2001). A flexible framework for stochastic claims reserving. *Proceedings of the Casualty Actuarial Society (US)*, **88**, 1–38.

England, P. D. and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, **8**, (37), 443–544.

Historical Loss Development Study (1991). Reinsurance Association of America. Washington D.C.

Mack, T. (1991). A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *Astin Bulletin*, **21**, (1), 93–109.

Mack, T. (1993). Distribution-free calculation of the standard error of chain-ladder reserve estimates. *ASTIN Bulletin*, **23**, 213–25.

Renshaw, A. E. and Verrall, R. J. (1998). A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, **4**, 903–23.

Schiegl, M. (2004). Simulation. *Handbuch der Schadenreservierung.*

Taylor, G. (2000). *Loss reserving: An actuarial perspective.* Kluwer Academic Publishers.

Taylor, G. C. and Ashe, F. R. (1983). Second moments of estimates of outstanding claims. *J. Econometrics*, **23**, 37–61.

Verdonck, T., Brys, G., and Wouwe, M. V. (2007). A robustification of the chain-ladder method. In *Proceedings of the APRIA Conference, Taipei (Taiwan), 21-25 July 2007.*

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

Wüthrich, M. V. and Merz, M. (2008). *Stochastic claims reserving methods in insurance.* Wiley. ISBN 978-0-470-72346.

A Data set 1

Data taken from England and Verrall (2001): "incremental paid losses from an aggregation of classes of business". Note the negative increment at $(3, 3)$.

Table 15: Incremental claims triangle

Origin	Development year									
year	1	2	3	4	5	6	7	8	9	10
1	45630	23350	2924	1798	2007	1204	1298	563	777	621
2	53025	26466	2829	1748	732	1424	399	537	340	
3	67318	42333	-1854	3178	3045	3281	2909	2613		
4	93489	37473	7431	6648	4207	5762	1890			
5	80517	33061	6863	4328	4003	2350				
6	68690	33931	5645	6178	3479					
7	63091	32198	8938	6879						
8	64430	32491	8414							
9	68548	35366								
10	76013									

Table 16: Corresponding cumulative claims triangle

Origin	Development year									
year	1	2	3	4	5	6	7	8	9	10
1	45630	68980	71904	73702	75709	76913	78211	78774	79551	80172
2	53025	79491	82320	84068	84800	86224	86623	87160	87500	
3	67318	109651	107797	110975	114020	117301	120210	122823		
4	93489	130962	138393	145041	149248	155010	156900			
5	80517	113578	120441	124769	128772	131122				
6	68690	102621	108266	114444	117923					
7	63091	95289	104227	111106						
8	64430	96921	105335							
9	68548	103914								
10	76013									

B Data set 2

This data is taken from England and Verrall (1999), who cite Taylor and Ashe (1983) as the original source.

Table 17: Incremental claims triangle

Origin	Development year									
year	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	
3	290507	1001799	926219	1016654	750816	146923	495992	280405		
4	310608	1108250	776189	1562400	272482	352053	206286			
5	443160	693190	991983	769488	504851	470639				
6	396132	937085	847498	805037	705960					
7	440832	847631	1131398	1063269						
8	359480	1061648	1443370							
9	376686	986608								
10	344014									

Table 18: Corresponding cumulative claims triangle

Origin	Development year									
year	1	2	3	4	5	6	7	8	9	10
1	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
2	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	
3	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315		
4	310608	1418858	2195047	3757447	4029929	4381982	4588268			
5	443160	1136350	2128333	2897821	3402672	3873311				
6	396132	1333217	2180715	2985752	3691712					
7	440832	1288463	2419861	3483130						
8	359480	1421128	2864498							
9	376686	1363294								
10	344014									

C Data set 3

Historical Loss Development Study (1991). Automatic Facultative General Liability data (excluding asbestos and environmental). Note the negative increment at $(2, 7)$.

Table 19: Incremental claims triangle

Origin	Development year									
year	1	2	3	4	5	6	7	8	9	10
1	5012	3257	2638	898	1734	2642	1828	599	54	172
2	102	4179	1111	5270	3116	1817	-103	673	535	
3	3410	5582	4881	2268	2594	3479	649	603		
4	5655	5900	4211	5500	2159	2658	984			
5	1092	8473	6271	6333	3786	225				
6	1513	4932	5257	1233	2917					
7	557	3463	6926	1368						
8	1351	5596	6165							
9	3133	2262								
10	2063									

Table 20: Corresponding cumulative claims triangle

Origin	Development year									
year	1	2	3	4	5	6	7	8	9	10
1	5012	8269	10907	11805	13539	16181	18009	18608	18662	18834
2	102	4281	5392	10662	13778	15595	15492	16165	16700	
3	3410	8992	13873	16141	18735	22214	22863	23466		
4	5655	11555	15766	21266	23425	26083	27067			
5	1092	9565	15836	22169	25955	26180				
6	1513	6445	11702	12935	15852					
7	557	4020	10946	12314						
8	1351	6947	13112							
9	3133	5395								
10	2063									

Cass Business School
CITY UNIVERSITY LONDON

FACULTY OF ACTUARIAL SCIENCE AND INSURANCE

Actuarial Research Papers since 2001

Report Number	Date	Publication Title	Author
160.	January 2005	Mortality Reduction Factors Incorporating Cohort Effects. ISBN 1 90161584 7	Arthur E. Renshaw Steven Haberman
161.	February 2005	The Management of De-Cumulation Risks in a Defined Contribution Environment. ISBN 1 901615 85 5.	Russell J. Gerrard Steven Haberman Elena Vigna
162.	May 2005	The IASB Insurance Project for Life Insurance Contracts: Impart on Reserving Methods and Solvency Requirements. ISBN 1-901615 86 3.	Laura Ballotta Giorgia Esposito Steven Haberman
163.	September 2005	Asymptotic and Numerical Analysis of the Optimal Investment Strategy for an Insurer. ISBN 1-901615-88-X	Paul Emms Steven Haberman
164.	October 2005.	Modelling the Joint Distribution of Competing Risks Survival Times using Copula Functions. I SBN 1-901615-89-8	Vladimir Kaishev Dimitrina S, Dimitrova Steven Haberman
165.	November 2005.	Excess of Loss Reinsurance Under Joint Survival Optimality. ISBN1-901615-90-1	Vladimir K. Kaishev Dimitrina S. Dimitrova
166.	November 2005.	Lee-Carter Goes Risk-Neutral. An Application to the Italian Annuity Market. ISBN 1-901615-91-X	Enrico Biffis Michel Denuit
167.	November 2005	Lee-Carter Mortality Forecasting: Application to the Italian Population. ISBN 1-901615-93-6	Steven Haberman Maria Russolillo
168.	February 2006	The Probationary Period as a Screening Device: Competitive Markets. ISBN 1-901615-95-2	Jaap Spreeuw Martin Karlsson
169.	February 2006	Types of Dependence and Time-dependent Association between Two Lifetimes in Single Parameter Copula Models. ISBN 1-901615-96-0	Jaap Spreeuw
170.	April 2006	Modelling Stochastic Bivariate Mortality ISBN 1-901615-97-9	Elisa Luciano Jaap Spreeuw Elena Vigna.
171.	February 2006	Optimal Strategies for Pricing General Insurance. ISBN 1901615-98-7	Paul Emms Steve Haberman Irene Savoulli
172.	February 2006	Dynamic Pricing of General Insurance in a Competitive Market. ISBN1-901615-99-5	Paul Emms
173.	February 2006	Pricing General Insurance with Constraints. ISBN 1-905752-00-8	Paul Emms
174.	May 2006	Investigating the Market Potential for Customised Long Term Care Insurance Products. ISBN 1-905752-01-6	Martin Karlsson Les Mayhew Ben Rickayzen

Report Number	Date	Publication Title	Author
175.	December 2006	Pricing and Capital Requirements for With Profit Contracts: Modelling Considerations. ISBN 1-905752-04-0	Laura Ballotta
176.	December 2006	Modelling the Fair Value of Annuities Contracts: The Impact of Interest Rate Risk and Mortality Risk. ISBN 1-905752-05-9	Laura Ballotta Giorgia Esposito Steven Haberman
177.	December 2006	Using Queuing Theory to Analyse Completion Times in Accident and Emergency Departments in the Light of the Government 4-hour Target. ISBN 978-1-905752-06-5	Les Mayhew David Smith
178.	April 2007	In Sickness and in Health? Dynamics of Health and Cohabitation in the United Kingdom. ISBN 978-1-905752-07-2	Martin Karlsson Les Mayhew Ben Rickayzen
179.	April 2007	GeD Spline Estimation of Multivariate Archimedean Copulas. ISBN 978-1-905752-08-9	Dimitrina Dimitrova Vladimir Kaishev Spiridon Penev
180.	May 2007	An Analysis of Disability-linked Annuities. ISBN 978-1-905752-09-6	Ben Rickayzen
181.	May 2007	On Simulation-based Approaches to Risk Measurement in Mortality with Specific Reference to Poisson lee-Carter Modelling. ISBN 978-1-905752-10-2	Arthur Renshaw Steven Haberman
182.	July 2007	High Dimensional Modelling and Simulation with Asymmetric Normal Mixtures. ISBN 978-1-905752-11-9	Andreas Tsanakas Andrew Smith
183.	August 2007	Intertemporal Dynamic Asset Allocation for Defined Contribution Pension Schemes. ISBN 978-1-905752-12-6	David Blake Douglas Wright Yumeng Zhang
184.	October 2007	To split or not to split: Capital allocation with convex risk measures. ISBN 978-1-905752-13-3	Andreas Tsanakas
185.	April 2008	On Some Mixture Distribution and Their Extreme Behaviour. ISBN 978-1-905752-14-0	Vladimir Kaishev Jae Hoon Jho
186.	October 2008	Optimal Funding and Investment Strategies in Defined Contribution Pension Plans under Epstein-Zin Utility. ISBN 978-1-905752-15-7	David Blake Douglas Wright Yumeng Zhang
187.	May 2008	Mortality Risk and the Valuation of Annuities with Guaranteed Minimum Death Benefit Options: Application to the Italian Population. ISBN 978-1-905752-16-4	Steven Haberman Gabriella Piscopo
188.	January 2009	The Market Potential for Privately Financed Long Term Care Products in the UK. ISBN 978-1-905752-19-5	Leslie Mayhew
189.	June 2009	Whither Human Survival and Longevity or the Shape of things to Come. ISBN 978-1-905752-21-8	Leslie Mayhew David Smith
190	September 2009	ilc: A Collection of R Functions for Fitting a Class of Lee Carter Mortality Models using Iterative fitting Algorithms* ISBN 978-1-905752-22-5	Zoltan Butt Steven Haberman
191.	October 2009	Decomposition of Disease and Disability Life Expectancies in England, 1992-2004. ISBN 978-1-905752-23-2	Domenica Rasulo Leslie Mayhew Ben Rickayzen

Report Number	Date	Publication Title	Author
192.	October 2009	Exploration of a Novel Bootstrap Technique for Estimating the Distribution of Outstanding Claims Reserves in General Insurance. ISBN 978-1-905752-24-9	Robert Cowell

Statistical Research Papers

1.	December 1995.	Some Results on the Derivatives of Matrix Functions. ISBN 1 874 770 83 2	P. Sebastiani
2.	March 1996	Coherent Criteria for Optimal Experimental Design. ISBN 1 874 770 86 7	A.P. Dawid P. Sebastiani
3.	March 1996	Maximum Entropy Sampling and Optimal Bayesian Experimental Design. ISBN 1 874 770 87 5	P. Sebastiani H.P. Wynn
4.	May 1996	A Note on D-optimal Designs for a Logistic Regression Model. ISBN 1 874 770 92 1	P. Sebastiani R. Settimi
5.	August 1996	First-order Optimal Designs for Non Linear Models. ISBN 1 874 770 95 6	P. Sebastiani R. Settimi
6.	September 1996	A Business Process Approach to Maintenance: Measurement, Decision and Control. ISBN 1 874 770 96 4	Martin J. Newby
7.	September 1996.	Moments and Generating Functions for the Absorption Distribution and its Negative Binomial Analogue. ISBN 1 874 770 97 2	Martin J. Newby
8.	November 1996.	Mixture Reduction via Predictive Scores. ISBN 1 874 770 98 0	Robert G. Cowell.
9.	March 1997.	Robust Parameter Learning in Bayesian Networks with Missing Data. ISBN 1 901615 00 6	P.Sebastiani M. Ramoni
10.	March 1997.	Guidelines for Corrective Replacement Based on Low Stochastic Structure Assumptions. ISBN 1 901615 01 4.	M.J. Newby F.P.A. Coolen
11.	March 1997	Approximations for the Absorption Distribution and its Negative Binomial Analogue. ISBN 1 901615 02 2	Martin J. Newby
12.	June 1997	The Use of Exogenous Knowledge to Learn Bayesian Networks from Incomplete Databases. ISBN 1 901615 10 3	M. Ramoni P. Sebastiani
13.	June 1997	Learning Bayesian Networks from Incomplete Databases. ISBN 1 901615 11 1	M. Ramoni P.Sebastiani
14.	June 1997	Risk Based Optimal Designs. ISBN 1 901615 13 8	P.Sebastiani H.P. Wynn
15.	June 1997.	Sampling without Replacement in Junction Trees. ISBN 1 901615 14 6	Robert G. Cowell
16.	July 1997	Optimal Overhaul Intervals with Imperfect Inspection and Repair. ISBN 1 901615 15 4	Richard A. Dagg Martin J. Newby
17.	October 1997	Bayesian Experimental Design and Shannon Information. ISBN 1 901615 17 0	P. Sebastiani. H.P. Wynn
18.	November 1997.	A Characterisation of Phase Type Distributions. ISBN 1 901615 18 9	Linda C. Wolstenholme

19.	December 1997	A Comparison of Models for Probability of Detection (POD) Curves. ISBN 1 901615 21 9	Wolstenholme L.C
20.	February 1999.	Parameter Learning from Incomplete Data Using Maximum Entropy I: Principles. ISBN 1 901615 37 5	Robert G. Cowell
21.	November 1999	Parameter Learning from Incomplete Data Using Maximum Entropy II: Application to Bayesian Networks. ISBN 1 901615 40 5	Robert G. Cowell
22.	March 2001	FINEX : Forensic Identification by Network Expert Systems. ISBN 1 901615 60X	Robert G.Cowell
23.	March 2001.	Wren Learning Bayesian Networks from Data, using Conditional Independence Tests is Equivalant to a Scoring Metric ISBN 1 901615 61 8	Robert G Cowell
24.	August 2004	Automatic, Computer Aided Geometric Design of Free-Knot, Regression Splines. ISBN 1-901615-81-2	Vladimir K Kaishev, Dimitrina S.Dimitrova, Steven Haberman Richard J. Verrall
25.	December 2004	Identification and Separation of DNA Mixtures Using Peak Area Information. ISBN 1-901615-82-0	R.G.Cowell S.L.Lauritzen J Mortera,
26.	November 2005.	The Quest for a Donor : Probability Based Methods Offer Help. ISBN 1-90161592-8	P.F.Mostad T. Egeland., R.G. Cowell V. Bosnes Ø. Braaten
27.	February 2006	Identification and Separation of DNA Mixtures Using Peak Area Information. (Updated Version of Research Report Number 25). ISBN 1-901615-94-4	R.G.Cowell S.L.Lauritzen J Mortera,
28.	October 2006	Geometrically Designed, Variable Knot Regression Splines : Asymptotics and Inference. ISBN 1-905752-02-4	Vladimir K Kaishev Dimitrina S.Dimitrova Steven Haberman Richard J. Verrall
29.	October 2006	Geometrically Designed, Variable Knot Regression Splines : Variation Diminishing Optimality of Knots. ISBN 1-905752-03-2	Vladimir K Kaishev Dimitrina S.Dimitrova Steven Haberman Richard J. Verrall
30.	November 2008	Scheduling Reentrant Jobs on Parallel Machines with a Remote Server. ISBN 978-1-905752-18-8	Konstantin Chakhlevitch Celia Glass

Papers can be downloaded from

http://www.cass.city.ac.uk/facact/research/publications.html

Faculty of Actuarial Science and Insurance

Cass Business School

The financial support of The Actuarial Profession is gratefully acknowledged.

ISBN 978-1-905752-24-9