FACULTY OF ECONOMICS AND BUSINESS

Reserving by conditioning on markers of individual claims: a case-study using historical simulation

Els Godecharle and Katrien Antonio

AFI_1492

Reserving by conditioning on markers of individual claims: a case–study using historical simulation

Els Godecharle *[1] and Katrien Antonio[2]

[1]Faculty of Economics and Business, KU Leuven, Belgium.
[2]Faculty of Economics and Business, KU Leuven, Belgium, and Faculty of Economics and Business, University of Amsterdam, The Netherlands.

April 23, 2014

Abstract

This paper explores the use of claim specific characteristics, so–called *claim markers*, for loss reserving with individual claims. Starting from the approach of Rosenlund (2012) we develop a stochastic Reserve by Detailed Conditioning ('RDC') method which is applicable to a micro–level data set with detailed information on individual claims. We use historical simulation to construct the predictive distribution of the outstanding loss reserve by simulating payments of a claim, given its claim markers. We explore how to incorporate different types of claim specific information when simulating outstanding loss reserves, and evaluate the impact of the set of markers and their specification on the predictive distribution of the outstanding reserve. We demonstrate the performance of the method on a portfolio of general liability insurance policies for private individuals from a European insurance company.

Keywords: claims reserving, micro-level loss reserving, claim characteristics, historical simulation.

1 Introduction

To be able to fulfill future liabilities, insurance companies set reserves for incurred claims which are not finalized at the moment of evaluation. Traditional reserving methods compress large data sets with information on the development of individual claims into small sample designs, the so–called run–off triangles (see England and Verrall (2002) and Wüthrich and Merz (2008)). As such, triangular reserving methods ignore detailed information on the policyholder, the claim and its development so far. Recently, the necessity and appropriateness of the use of run–off triangles has been challenged. Building upon the fundamental work by Norberg (1993), Haastrup and Arjas (1996) and Norberg (1999), Antonio and Plat (2014) model the development of individual claims in continuous time. Drieskens et al. (2012), Rosenlund (2012) and Pigeon et al. (2013) work in discrete time and aggregate payments per development time period (e.g. a development

*Corresponding author. E-mail: els.godecharle@kuleuven.be

year), while keeping the claim specific run–off viewpoint. Verrall et al. (2010), Martínez Miranda et al. (2011) and Martínez Miranda et al. (2012) extend the traditional chain–ladder framework towards the use of extra data sources.

Starting point of this paper is the *'Reserve by Detailed Conditioning'* (RDC) method, as introduced by Rosenlund in 2012, see Rosenlund (2012). RDC - in its original specification - is a deterministic reserving method, designed for individual claims in discrete time. A remarkable and innovative aspect of the method is its ability to condition on claim characteristics (here called *'claim markers'*), which are used for identification or clustering of similar claims. Conditional on a specific set of claim markers, a best estimate for the reserve attached to an open claim is obtained from the observed, historical development of claims from the same cluster, hence with similar characteristics.

We develop our own stochastic version of the RDC approach, starting from the Rosenlund (2012) paper. Moreover, while Rosenlund (2012) only considers simulated data, we evaluate the performance of the stochastic RDC through an extensive case–study. The stochastic RDC simulates the future development of outstanding liabilities, conditional upon a selection of claim markers. Examples of such markers are the observed reporting delay and the censored observations of cumulative payment and settlement delay. It offers a flexible, generic framework, which can easily be adjusted or extended to - for example - other types of claim markers, like case estimates. We use the technique of *historical simulation* to obtain a predictive distribution of the reserve. Historical simulation is used in finance e.g. to estimate the Value-at-Risk of a portfolio (see Hull (2012), Chapter 14, pages 303–322). In the context of claims reserving, Drieskens et al. (2012) apply historical simulation in reserving with individual, large claims. Our paper integrates the approach of Drieskens et al. (2012) with the conditioning idea from Rosenlund (2012).

We demonstrate the performance of the original, deterministic RDC as well as our extensions on a portfolio of general liability insurance policies for private individuals from a European insurance company. Antonio and Plat (2014) and Pigeon et al. (2013) use the same data set. A back test illustrates the predictive power of the technique.

The paper is organized as follows. In section 2 we describe the structure of our data and the claim markers used in the paper. Section 3 describes the historical simulation approach. Section 4 presents the results of applying the stochastic RDC method to a general liability data set, including a sensitivity analysis. We end in section 5 with a conclusion and an outline of future work.

2 A micro-level data set with information on claim markers

2.1 Structure of the data

2.1.1 An individual claim in discrete time

Starting point is a data set with the development of individual claims in continuous time. We discretize the available information and work with discrete time periods. The occurrence period of claim k is $i(k) \in \{1, \cdots, n\}$. $j \in \{1, \cdots, n\}$ denotes the development period, where the first development period coincides with the period of occurrence. We denote the reporting delay for claim k with $W(k)$, which is the development period during which reporting is registered (or, the *waiting-for-reporting* period). Analogously, we denote the finalization period with $F(k)$. The sum of all payments made in development period j is $Y(k,j)$, where $j = 1, \ldots, F(k)$. Using this notation we summarize claim k as (Rosenlund, 2012)

$$\{i(k), W(k), F(k), Y(k,1), \cdots, Y(k, F(k))\}. \tag{1}$$

A claim reported in its period of occurrence has $W(k) = 1$. If the claim also closes in the same period, its finalization period, $F(k)$, is also 1. The information in (1) is fully observed for a settled (or: closed) claim, but is censored for open claims.

We illustrate this notation in Table 1. Starting point of the first occurrence period is 01/01/1997 and one year periods are considered. The claim in Table 1 occurs in 1997, thus $i(k) = 1$. Reporting takes place in 1998, thus $W(k) = 2$, and the claim settles in 2002, thus $F(k) = 2002 - 1997 + 1 = 6$.

Event		**Date**	**Our Notation**
Accident		05/17/1997	$i = 1$
Reporting		02/02/1998	$W(k) = 2$
Cash flow	€200	11/24/1998	$Y(k,2) = 200$
	€150	02/08/1999	$Y(k,3) = 250$
	€100	05/11/1999	
			$Y(k,4) = 0$
	€50	02/23/2001	$Y(k,5) = 50$
			$Y(k,6) = 0$
Closing		03/13/2002	$F(k) = 6$

Table 1: *Illustration of occurrence, reporting, claim payments and settlement as registered in continuous time for individual claims (first three columns). In column 4 we demonstrate the discrete time notation used in this paper. This is a fictional claim.*

2.1.2 A micro-level data set in discrete time

At portfolio level the data structure from Section 2.1.1 creates the design in Table 2. This table is a loss *'triangle'* at micro–level and replaces the traditional run-off triangles. Each row in the table corresponds to a claim reported in the data set.

We denote the number of fully observed occurrence periods in the data set by n. The moment of evaluation of the portfolio is at the end of the n^{th} occurrence period and the start of occurrence period $n+1$. For claim k the maximum number of observed development periods is $n - i(k) + 1$.

2.2 Claim markers

Following Rosenlund (2012) we determine reserves by conditioning on claim specific information, i.e. so-called claim markers. These markers summarize information registered during the

i	**Claim ID**	**Development period j**					
		1	2	3	$\cdots$	$n-1$	n
1	$c_{1,1}$	$Y(c_{1,1},1)$	$Y(c_{1,1},2)$	$Y(c_{1,1},3)$	$\cdots$	$Y(c_{1,1},n-1)$	$Y(c_{1,1},n)$
	$c_{1,2}$	$Y(c_{1,2},1)$	$Y(c_{1,2},2)$	$Y(c_{1,2},3)$	$\cdots$	$Y(c_{1,2},n-1)$	$Y(c_{1,2},n)$
	$\vdots$	$\vdots$					
2	$c_{2,1}$	$Y(c_{2,1},1)$	$Y(c_{2,1},2)$	$Y(c_{2,1},3)$	$\cdots$	$Y(c_{2,1},n-1)$	
	$c_{2,2}$	$Y(c_{2,2},1)$	$Y(c_{2,2},2)$	$Y(c_{2,2},3)$	$\cdots$	$Y(c_{2,2},n-1)$	
	$\vdots$	$\vdots$					
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
n	$c_{n,1}$	$Y(c_{n,1},1)$					
	$c_{n,2}$	$Y(c_{n,1},1)$					
	$\vdots$	$\vdots$					

Table 2: *Micro–level data set.*

development of a claim. They allow for the identification of similar claims. In this work we consider the (bounded) claim length, the last observed (binned) cumulative payment and the (bounded) reporting delay as claim markers.

2.2.1 Claim length

As in Rosenlund (2012), we denote the length of claim k by $L(k)$ and define

$$L(k) = F(k) - W(k) + 1. \tag{2}$$

We measure the length of a claim as the number of periods starting from and *including* the period of reporting until the closure of the claim. For the example in Table 1, $L = 2002 - 1998 + 1 = 5$. Three types of claims can be identified using the combined information on reporting delay (i.e. $W(k)$) and claim length (i.e. $L(k)$):

- for a reported and closed claim (closed)

$$W(k) \leqslant n - i(k) + 1 \text{ and } L(k) \leqslant n - i(k) - W(k) + 2,$$

- for a reported, but not settled claim (RBNS)

$$W(k) \leqslant n - i(k) + 1 \text{ and } L(k) > n - i(k) - W(k) + 2,$$

- for an incurred, but not reported claim (IBNR)

$$W(k) > n - i(k) + 1.$$

The length of an RBNS claim is not observed, but censored at the moment of evaluation. An open claim from occurrence period $i(k)$, reported in period $W(k)$, has a claim length that is right censored at $n - i(k) - W(k) + 2$. We identify

$$L_{\min}(k) = n - i(k) - W(k) + 2 \tag{3}$$

as the length observed at the time of censoring.

In this paper we do not consider the development of a claim beyond the triangle boundary. It is however straightforward to incorporate a tail factor in our reserving method.

2.2.2 Cumulative payments

Following Rosenlund (2012) we define a time index t with values in $\{0, 1, \cdots, n\}$. t expresses the *number of periods evolved, starting from and including the period of reporting.* $t = 1$ corresponds to the period of reporting itself and $t = 0$ indicates that no history is available, that is the claim is not reported yet. $H(k,t)$ is the cumulative paid amount for claim k, up to and including period t since reporting. That is

$$H(k,t) = \sum_{h=1}^{t} Y(k, h + W(k) - 1). \tag{4}$$

When $t = L(k)$, $H(k, L(k))$ is the ultimate loss for claim k, which is only observed for closed claims. At the moment of evaluation $t = n - i(k) - W(k) + 2$ for an observed claim k. For an RBNS claim $H(k, n - i(k) - W(k) + 2)$ is the last observed cumulative payment at the moment of evaluation. For an IBNR claim $t = 0$ at the moment of evaluation and $H(k, 0) = 0$. For the example in Table 1

$$H(k,1) = 200, \quad H(k,2) = 450, \quad H(k,3) = 450, \quad H(k,4) = 500, \quad H(k,5) = 500. \tag{5}$$

Quantile binning. As in Rosenlund (2012) we use quantile binning to discretize a claim's information flow on cumulative payments. The number of quantiles q_0 in the binning is fixed beforehand and independent of the period since reporting t. At t we consider the set of cumulative payments $H(\cdot, t)$ for all claims still *open at* t, thus with $L > t$. This set of cumulative payments is binned in q_0 intervals with the empirical $\left\{\frac{1}{q_0}, \frac{2}{q_0}, \cdots, \frac{q_0}{q_0}\right\} \cdot 100\%$ quantiles of the distribution of $H(\cdot, t)$ as borders. We denote these boundaries as $\{h_{t,1}, h_{t,2}, \cdots, h_{t,q_0}\}$, and create the following bins: $[0, h_{t,1}]$, $\cdots$, $(h_{t,q_0-1}, h_{t,q_0}]$. The claim marker expressing the cumulative paid amount for claim k at t is a label, say $Q_t(k)$, indicating the interval to which $H(k,t)$ belongs. That is

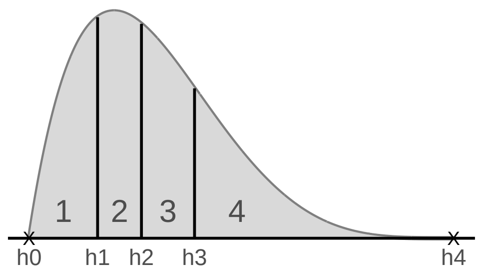


Figure 1: Quantile binning

$$\begin{aligned} Q_t(k) &= \text{quantile interval number of } H(k,t) \text{ with } L(k) > t, \\ Q_t(k) &\in \{1, \cdots, q_0\}. \end{aligned} \tag{6}$$

for $t > 0$. When $t = 0$ the claim has not been reported yet, thus there is no information on the cumulative payments. Consequently, $Q_0(k) = 1$. Quantile binning is illustrated in Figure 1, where the empirical pdf of $H(.,t)$ (for given t) is split into intervals $[0, h_{t,1}]$, $(h_{t,1}, h_{t,2}]$, $(h_{t,2}, h_{t,3}]$ and $(h_{t,3}, h_{t,4}]$. The claim marker associated with an observed $h(k,t)$ is then $q_t(k)$ such that $h_{t,q_t(k)-1} < h(k,t) \leqslant h_{t,q_t(k)}$.

Zero payments. In the presence of many observed zero cumulative payments some of the boundaries in $\{h_{t,1}, h_{t,2}, \cdots, h_{t,q_0}\}$ can coincide at 0. When this happens, our approach creates a separate label for all cumulative payments equal to zero. For example, say $q_0 = 4$ and boundaries are $\{h_{2,1}, h_{2,2}, h_{2,3}, h_{2,4}\}$ where $h_{2,1} = h_{2,2} = 0$. We then create the bins: $[0, 0]$, $(0, h_{2,3}]$ and $(h_{2,3}, h_{2,4}]$, with labels respectively '1', '2' and '3'.

2.2.3 Late payments, long reporting delay and long development

To avoid the construction of very small, almost empty, clusters we identify upper bounds ω_0, τ_0 and λ_0 and define

- as in Rosenlund (2012): claims with *'long reporting delay'*, as claims with $W(k) \geqslant \omega_0$;
- *'long developing'* claims as claims for which $L(k) \geqslant \lambda_0$;
- *'late payments'* as payments taking place in period t since reporting where $t \geqslant \tau_0$.

This changes our claim markers in the following way:

- the bounded reporting delay $W_b(k) = \min\{W(k), \omega_0\}$;
- the bounded claim length $L_b(k) = \min\{L(k), \lambda_0\}$;
- in the historical simulation method, for claims with the same markers, we consider the payments at time t since reporting $Y(\cdot, t + W - 1)$ for $t \geqslant \tau_0$ as a single group.

3 Stochastic Reserving by Detailed Conditioning using historical simulation

We use claim specific information in the reserving process, through the creation of claim markers as discussed in Section 2.2. However, in contrast to the deterministic approach in Rosenlund (2012), our approach is stochastic and uses historical simulation. We do not repeat the deterministic RDC approach here, but interested readers should consult Rosenlund (2012) (Appendix, pages 22–29).

3.1 Simulating claim length

In this subsection we suppress index k for ease of notation. At period t since reporting the distribution of claim length $L|L > t$, conditional on markers Q_t and W_b, is given by

$$p(\lambda; q, w, t) := P\left(L = \lambda \middle| L > t, Q_t = q, W_b = w\right) \tag{7}$$

for

$$0 \leqslant t \leqslant n - w \qquad\qquad t + 1 \leqslant \lambda \leqslant n - w + 1.$$

Therefore, with probability $p(\lambda; q, w, t)$ the claim length is λ, given that the claim is open since t periods (starting from and including reporting), the last observed cumulative payment is in the q^{th} interval and the bounded reporting delay equals w. We define the one-step variant of $p(\lambda; q, w, t)$ as

$$r(\lambda; q, w, t) := P\left(L = \lambda \middle| L \geqslant \lambda, Q_t = q, W_b = w \right), \tag{8}$$

such that (see Rosenlund, 2012)

$$p(\lambda; q, w, t) = \left(\prod_{m=t+1}^{\lambda-1} \left[1 - r(m; q, w, t) \right] \right) \cdot r(\lambda; q, w, t). \tag{9}$$

We estimate the probabilities defined by (8) using maximum likelihood. The corresponding likelihood function is

$$r(\lambda; q, w, t)^{I^F(\lambda; q, w, t)} \cdot (1 - r(\lambda; q, w, t))^{J(\lambda; q, w, t) - I^F(\lambda; q, w, t)}, \tag{10}$$

where

$$\begin{aligned} I^F(\lambda; q, w, t) &= \text{ number of finalized claims with } L = \lambda \text{ given } Q_t = q \text{ and } W_b = w \\ J(\lambda; q, w, t) &= \text{ number of reported claims with } L \geqslant \lambda \text{ given } Q_t = q \text{ and } W_b = w. \end{aligned}$$

Consequently, we set

$$\begin{cases} \hat{r}(n - w + 1; q, w, t) & = 1 \\ \hat{r}(\lambda; q, w, t) & = \frac{I^F(\lambda; q, w, t)}{J(\lambda; q, w, t)} \qquad \text{if } \lambda < n - w + 1. \end{cases} \tag{11}$$

where the claim length is restricted to the triangle boundary.

The probabilities p (see (7)) allow simulation of the length of an open claim, conditional on its observed claim markers. We denote the simulated claim length for claim k by $L^s(k)$. For an RBNS claim k we use $t = n - i(k) - W(k) + 2 = L_{\min}(k)$, i.e. the latest period observed. We (obviously) condition on the claim's most recent information when simulating the length. Thus, we use the probabilities

$$\hat{p}\left(\lambda; Q_{L_{\min}(k)}(k), W_b(k), L_{\min}(k)\right).$$

No information is available for an IBNR claim, and therefore $t = 0$. Drieskens et al. (2012) use r instead of p probabilities and do not condition on claim characteristics.

3.2 Simulating claim development

Consider an open claim k with latest period observed $t = n - i(k) - W(k) + 2$ and simulated claim length $L^s(k)$. To simulate the development of this claim (until closing) we define a set of historical claims with the same characteristics. We identify the claims which have

- the same bounded (simulated) claim length $L^s_b(k) = \min\{L^s(k), \lambda_0\}$;

- the same bounded reporting delay $W_b(k)$;
- the same quantile number $Q_{n-i(k)-W(k)+2}(k)$, at the same time period since reporting.

The claim markers are used as static information. Therefore, we do not adjust them throughout the simulation process. The selected group of claims contains open as well as closed claims at the date of evaluation. From this group (or: cluster of claims), we simulate the future, unknown payments of claim k, i.e. the $Y(k, W+t-1)$'s with $t \in \{L_{\min}(k)+1, \cdots, L^s(k)\}$. Within the cluster we assume independent and identically distributed payments $Y(\cdot, W+t-1)$. When $t < \tau_0$ we simulate a future payment for claim k, say $Y(k, W+t-1)$, from all observed payments in the cluster, paid t periods since reporting. If $t \geqslant \tau_0$ we simulate from all observed payments paid τ_0 periods since reporting, or later.

3.3 Simulating the future development of RBNS and IBNR claims

3.3.1 RBNS claims

The outstanding loss for an RBNS claim k is

$$H(k, L(k)) - H(k, n-i(k)-W(k)+2) = \sum_{h=n-i(k)-W(k)+3}^{L(k)} Y(k, h+W(k)-1). \tag{12}$$

Let K_{RBNS} denote the number of RBNS claims in the data set. The corresponding RBNS reserve is

$$R_{\text{RBNS}} = \sum_{k=1}^{K_{\text{RBNS}}} H(k, L(k)) - H(k, n-i(k)-W(k)+2), \tag{13}$$

where the index k runs over all RBNS claims in the data set. Using historical simulation the predictive distribution of this reserve is obtained from the following steps:

1. conditional on the claim markers (i.e. $Q_{n-i(k)-W(k)+2}$, $W_b(k)$), simulate the claim length $L^s(k)$ for an open claim k using the distribution specified in Section 3.1;
2. for every RBNS claim k simulate the remaining payments until triangle boundary

   $$Y_{n-i(k)-W(k)+3}, \cdots, Y_{\min(L^s(k), n-W(k)+1)},$$

   conditional on the claim markers, namely the bounded simulated claim length $L^s_b(k)$, the bounded reporting period $W_b(k)$ and the latest observed quantile interval number $Q_{n-i(k)-W(k)+2}(k)$;
3. repeat steps 1. to 2. to obtain a distribution of the RBNS reserve.

3.3.2 IBNR claims

Number of IBNR claims. We follow the approach from Pigeon et al. (2013) to simulate the number of IBNR claims per occurrence year i and reporting delay W (with values in $\{n-i+2,\cdots,n\}$ for $i \geqslant 2$). The number of claims from occurrence period i is Poisson distributed with parameter $\theta \cdot w(i)$, where $w(i)$ is an exposure measure for occurrence year i. Following Pigeon et al. (2013), we estimate the reporting delay using a geometric distribution with a degenerate component. We thin the Poisson process by taking the reporting delay into account. For further details we refer to Pigeon et al. (2013).

IBNR Reserve. For an IBNR claim k the outstanding loss is

$$H(k, L(k)) - H(k, 0) = H(k, L(k)). \tag{14}$$

Let K_{IBNR} denote the number of IBNR claims. The corresponding IBNR reserve is

$$R_{\text{IBNR}} = \sum_{k=1}^{K_{\text{IBNR}}} H(k, L(k)) - H(k, 0) = \sum_{k=1}^{K_{\text{IBNR}}} H(k, L(k)), \tag{15}$$

where k indicates an IBNR claim. Using historical simulation the predictive distribution of this reserve is obtained from the following steps:

1. simulate the number of IBNR claims per occurrence year i and reporting delay W (with values from $n-i+2$ to n).
2. simulate the claim length and the development for every IBNR claim; use the approach outlined for RBNS claims in Section 3.3.1.

4 Results

The implementation in R of the deterministic RDC method and the stochastic RDC method is available online [1]. We illustrate the stochastic RDC method on a data set from insurance practice. The data have been used in Antonio and Plat (2014) and Pigeon et al. (2013). We perform a back test to evaluate the predictive power of the reserving method. A sensitivity analysis with respect to the chosen claim markers is included. Results obtained with the deterministic RDC method (as introduced by Rosenlund (2012)) and the Over Dispersed Poisson ('ODP') chain–ladder method are reported as benchmark.

4.1 The data

We consider a portfolio of general liability insurance policies for private individuals, with observation period January 1997 until August 2009. Data from January 1997 up to and including December 2004 are used as training set, data observed after 2004 are used as validation set. The data set contains two types of payments, which we study separately, namely Bodily Injury (BI) and Material Damage (MD) payments. A claim can have both a BI and an MD component.

[1] `http://www.econ.kuleuven.be/els.godecharle`

In this case, both components are analyzed separately. The training data set contains 4,483 BI claims and 224,836 MD claims of which 3,452 and 220,730 claims, respectively, are closed at the moment of evaluation. Payments in the data set are adjusted for inflation to 01/01/1997 using the appropriate consumer price index. We assume that all payments, conditional on the claim markers, are independent and identically distributed.

We discretize the development time line, using one year periods starting from 01/01/1997 and running until 12/31/2004.

4.2 Descriptive statistics

In Table 3 and 4 we display the incremental run-off triangles for respectively the BI and MD claims.

Arrival	Development year							
year	1	2	3	4	5	6	7	8
1997	261	614	359	526	546	137	130	339
1998	202	473	307	336	269	56	179	**78**
1999	238	569	393	270	249	286	**132**	**97**
2000	237	557	429	496	406	**365**	**247**	**275**
2001	389	628	529	559	**446**	**375**	**147**	**239**
2002	260	570	533	**444**	**132**	**122**	**332**	**1,082**
2003	236	743	**558**	**237**	**217**	**205**	**171**	
2004	248	**794**	**401**	**236**	**254**	**98**		

Table 3: Incremental run-off triangle for BI (in thousands). The observed aggregate payments from the validation data set are in bold.

Arrival	Development year							
year	1	2	3	4	5	6	7	8
1997	4,427	992	89	13	39	27	37	11
1998	4,389	984	60	35	76	24	0.6	**16**
1999	5,280	1,239	76	110	113	12	**0.4**	**0**
2000	5,445	1,164	172	16	6	**10**	**0**	**10**
2001	5,612	1,838	156	127	**13**	**3**	**0.4**	**3**
2002	6,593	1,592	74	**71**	**17**	**15**	**9**	**9**
2003	6,603	1,660	**150**	**52**	**37**	**18**	**3**	
2004	7,195	**1,417**	**109**	**86**	**39**	**15**		

Table 4: Incremental run-off triangle for MD (in thousands). The observed aggregate payments from the validation data set are in bold.

Time dynamics Table 5 shows the empirical distribution of the reporting delay W. The majority of the claims is reported in the occurrence year or the year after.

Table 6 summarizes the empirical distribution of claim length L. As expected, BI claims need more time to settle than MD claims. Figure 2 represents this information graphically in a

W	1	2	3	4	5	6	7	8
BI	4,077	332	46	13	9	6	0	0
MD	216,906	7,764	127	23	9	7	0	0

Table 5: Frequencies of observed W's for BI and MD claims.

histogram. In total, 23.00% of the reported BI claims are censored at the date of analysis, whereas 1.83% of the reported MD claims are censored at the moment of evaluation.

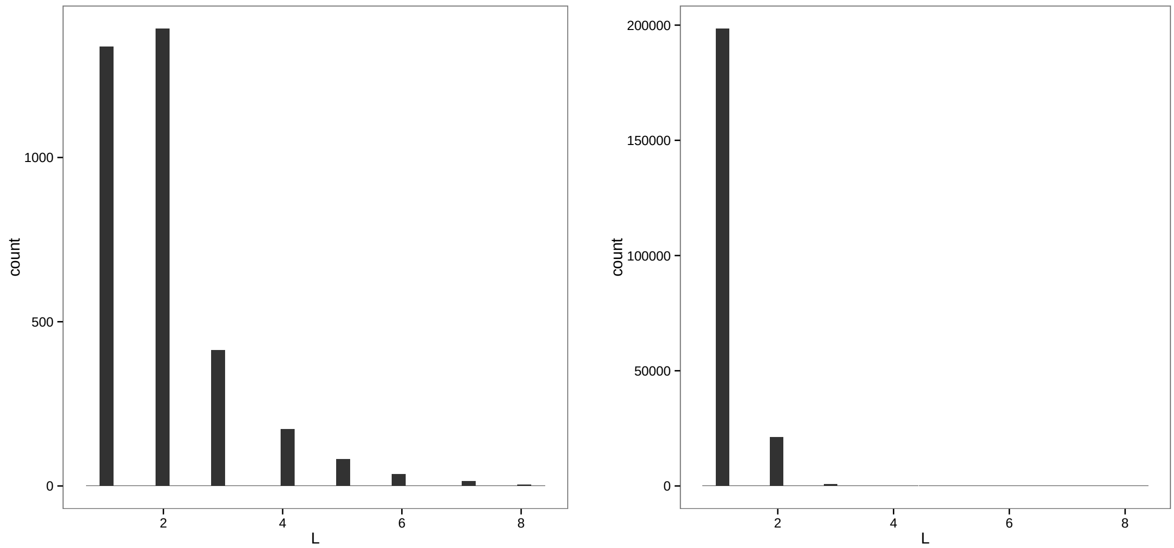


Figure 2: Histogram of the observed claim lengths for BI (left) and MD (right)

L	1	2	3	4	5	6	7	8	♯ observed lengths
BI	1,337	1,392	414	173	81	36	15	4	3,452
	29.82%	31.05%	9.23%	3.86%	1.81%	0.80%	0.33%	0.09%	77.00%
MD	198,426	21,091	757	199	77	123	20	37	220,730
	88.25%	9.38%	0.34%	0.09%	0.03%	0.05%	0.01%	0.02%	98.17%

Table 6: Number of observed L for BI and MD claims. For example, 1,337 of 4,483 reported BI claims (i.e. 29.82%)are observed to close with $L = 1$.

Payments We now consider the empirical distribution of payments at period t since reporting, $Y(\cdot, W + t - 1)$. Table 7 displays the percentage of zero payments for $t = 1, 2, \geqslant 3$ conditional on the observed value of reporting delay W.

Table 8 displays a selection of summary statistics on the observed payments per period t since reporting.

t	$W=1$	$W=2$	$W \geqslant 3$
1	38.78%	39.76%	50%
2	31.98%	37.11%	53.45%
$\geqslant 3$	48.87%	43.95%	57.14%

(a) BI

t	$W=1$	$W=2$	$W \geqslant 3$
1	25.41%	17.62%	48.80%
2	29.23%	57.76%	75.61%
$\geqslant 3$	85.90%	80.00%	86.11%

(b) MD

Table 7: Percentage of zero payments for different values of reporting delay W.

period t since reporting	1	2	3	$\geqslant 4$
BI payments				
♯ observed	4,483	2,669	1,071	1,046
Minimum	0	0	0	0
Maximum	143,247.9	112,824.8	212,167.4	351,859.6
Mean	575.50	1,787.58	2,346.40	3,533.06
Median	89.06	281.01	99.99	0
95% quantile	2,411.91	8,340.17	10,814.06	18,648.53
% of zero payments	39.04%	32.82%	44.44%	53.06%
MD payments				
♯ observed	224,836	22,872	1,489	1,408
Minimum	0	0	0	0
Maximum	153,816.3	174,387.3	105,113.6	872,32.36
Mean	212.04	328.40	356.17	405.60
Median	81.22	126.33	0	0
95% quantile	801.05	1,115.60	1,197.31	271.86
% of zero payments	25.16%	29.66%	78.71%	93.04%

Table 8: Summary statistics of observed payments.

4.3 Results

4.3.1 Predictive distribution of the reserve

Parameters ω_0, q_0, τ_0 and λ_0. We determine the parameters ω_0, q_0, τ_0 and λ_0 based on the descriptive statistics reported in Section 4.2. For ω_0 we assume, see Table 5, that (for BI and MD) a claim is reported late if $W \geqslant 3$. We set $\omega_0 = 3$. The second parameter is the number of quantile intervals q_0 used in the quantile binning of the cumulative payments $H(\cdot, t)$. We distinguish small, moderate and large cumulative payments. Therefore, we choose $q_0 = 3$. From Table 6 we conclude that BI, respectively MD, claims with length $L \geqslant 4$ and $L \geqslant 3$ have a long claim development. Therefore we set $\lambda_0 = 4$ and 3, respectively. We use $\tau_0 \leqslant \lambda_0$, such that the last payments of claims with a long development are also late payments. Based on the statistics in Table 8 we set $\tau_0 = 3$ for both BI and MD claims.

Simulated distribution We calculate reserves using the stochastic RDC method, the deterministic RDC method by Rosenlund (2012) and the bootstrap Over Dispersed Poisson chain–ladder method.

Figures 3 and 4 show the IBNR, RBNS and total reserve for BI and MD claims, as obtained from the stochastic RDC method with parameters ω_0, q_0, τ_0 and λ_0 chosen upfront. These reserves predict the complete lower triangle in Tables 3 and 4. As a benchmark, Figures 3 and 4 display the outstanding loss reserve calculated from the validation data set (i.e. the numbers in bold in the run-off triangles in Tables 3 and 4). Recall that our data set contains observations until August 2009. Thus, we do not observe the complete run-off triangle. As discussed in Antonio and Plat (2014) and Pigeon et al. (2013) the lower triangle for Bodily Injury (see Table 3) shows an extreme payment (779,383 euro) in occurrence year 2002, development year 8. The policy limit of 2.5M Euro is not taken into account in the RDC simulations, though it is straightforward to incorporate a policy limit in this micro–level approach.

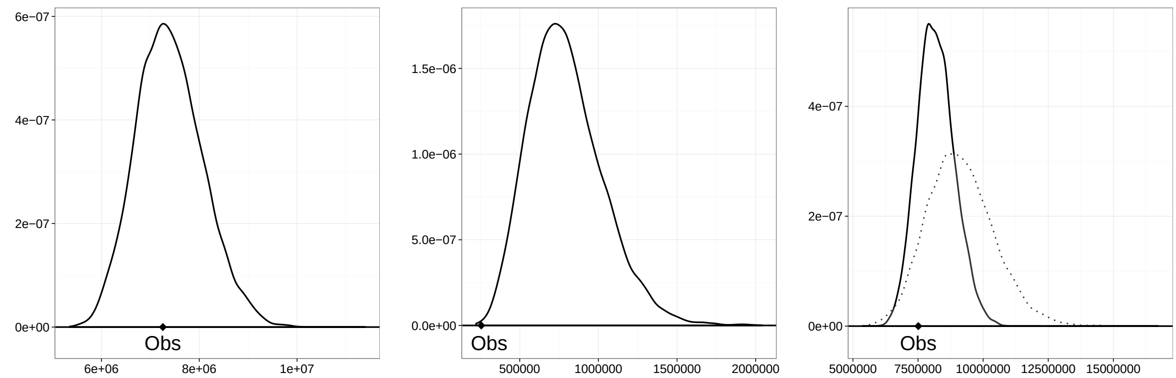


Figure 3: RBNS (left), IBNR (middle) and total (right) kernel density estimate of the BI reserve distribution (10,000 simulations). The dotted line in the right plot represents the bootstrap chain-ladder predictive distribution (10,000 simulations, process distribution: Over Dispersed Poisson).

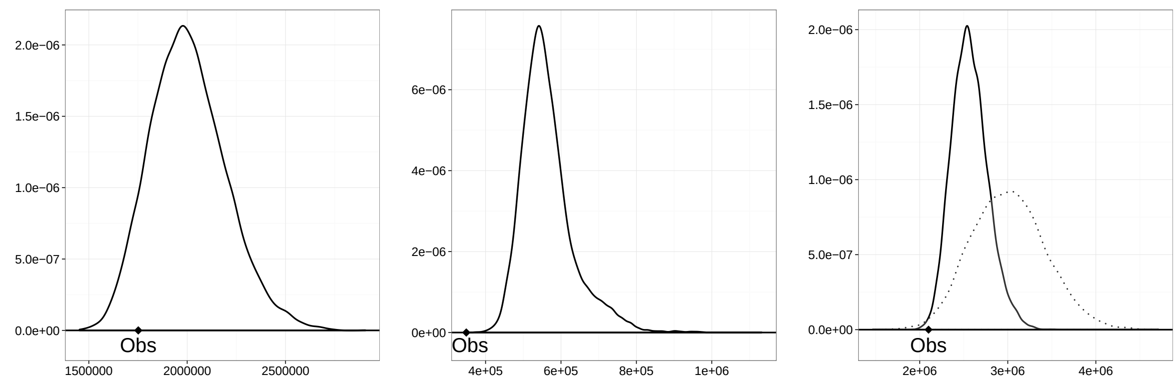


Figure 4: RBNS (left), IBNR (middle) and total (right) kernel density estimate of the MD reserve distribution (10,000 simulations). The dotted line in the right plot represents the bootstrap chain-ladder predictive distribution (10,000 simulations, process distribution: Over Dispersed Poisson).

Table 9 shows corresponding summary statistics for the stochastic RDC method, the bootstrap chain-ladder and the deterministic RDC method by Rosenlund (2012). The deterministic RDC method is using q_0 and ω_0 as identified for stochastic RDC. The back test developed in this case study illustrates the good performance of the stochastic RDC method.

		Min.	25% quantile	Mean	Standard dev.	75% quantile	Max.
BI	IBNR	222	626	794	237	937	2,041
	RBNS	5,352	6,915	7,397	678	7,837	11,389
	Total	5,841	7,693	8,191	715	8,650	11,975
	Bootstrap ODP CL	5,383	8,246	9,133	1,275	9,938	16,723
	Deterministic RDC	9,909 = 1,353 (IBNR) + 8,556 (RBNS)					
	Observed	7,512 = 255 (IBNR) + 7,256 (RBNS)					
MD	IBNR	375	518	564	70	594	1,132
	RBNS	1,453	1,871	2,006	192	2,128	2,907
	Total	1,964	2,424	2,570	205	2,697	3,506
	Bootstrap ODP CL	1,465	2,716	3,011	419	3,287	4,714
	Deterministic RDC	2,485 = 472 (IBNR) + 2,013 (RBNS)					
	Observed	2,100 = 349 (IBNR) + 1,751 (RBNS)					

Table 9: Predictive results (in thousands) stochastic RDC (first three rows of BI and MD), predictive results bootstrap Over Dispersed Poisson chain-ladder ('ODP CL') method (fourth row), deterministic RDC (fifth row) and the observed reserve until August 2009 (sixth row)

For each payment simulated during the routine we store the size of the corresponding cluster. Table 10 displays summary statistics on the sizes of the groups from which payments are simulated. The selected parameters ω_0, q_0, τ_0 and λ_0 lead to reasonable payment group sizes.

	IBNR	RBNS	Total
Minimum	11	3	3
5% quantile	14	37	19
Mean	46.67	382.2	355.5
Median	35	444	391
Maximum	132	612	612

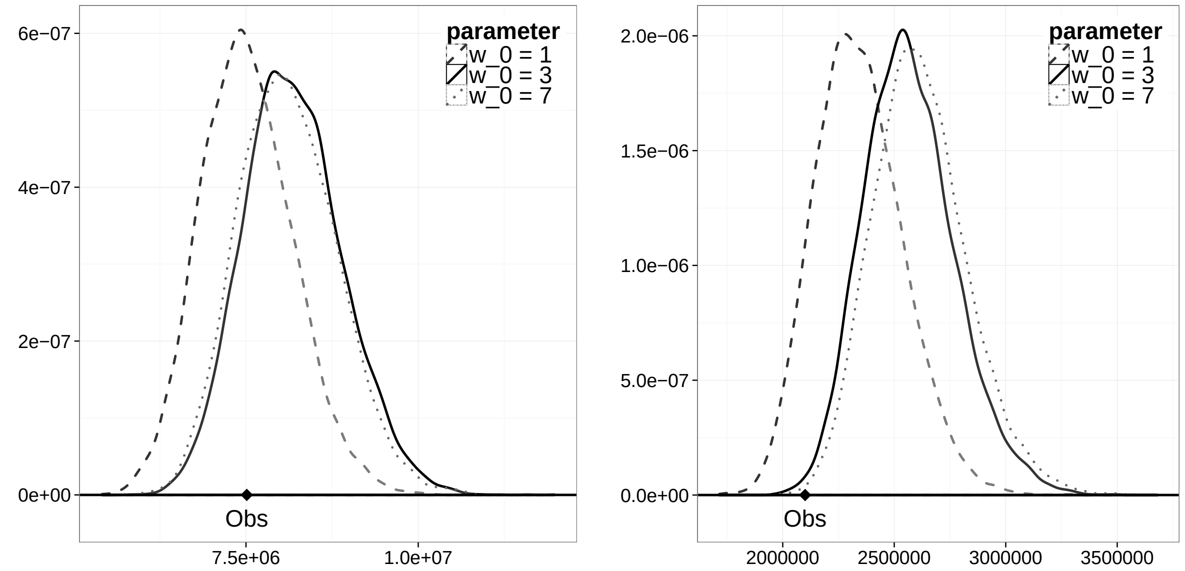
(a) BI

	IBNR	RBNS	Total
Minimum	9	2	2
5% quantile	12	223	24
Mean	3,947	7,700	6,867
Median	7,445	6,038	6,038
Maximum	7,445	14,840	14,840

(b) MD

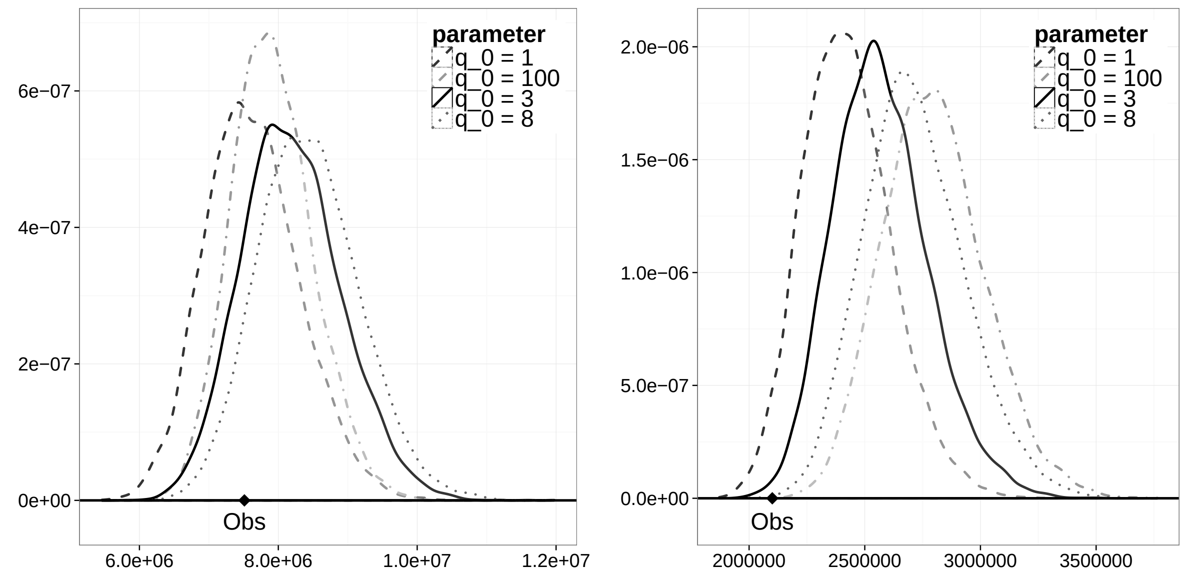
Table 10: Summary statistics of group sizes from which payments are simulated.

4.3.2 Sensitivity analysis



(a) BI, with fixed values $q_0 = 3$, $\tau_0 = 3$, $\lambda_0 = 4$. **(b)** MD, with fixed values $q_0 = 3$, $\tau_0 = 3$, $\lambda_0 = 4$.

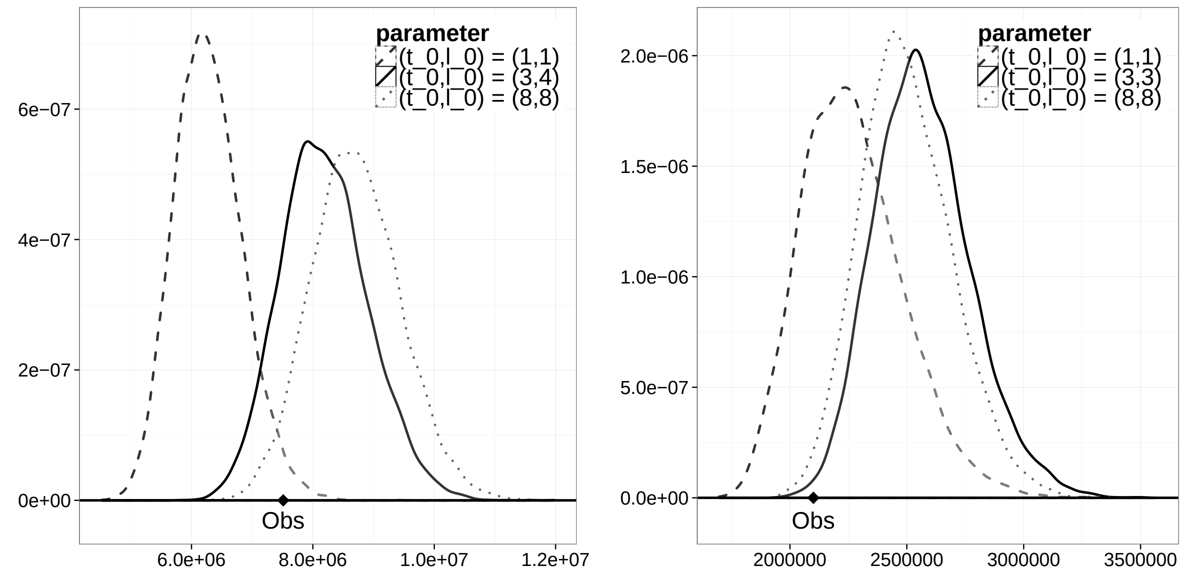
Figure 5: Sensitivity of the total reserve with respect to ω_0: fixed q_0, τ_0 and λ_0, and $\omega_0 = 1, 3, 7$.



(a) BI, with fixed values $\omega_0 = 3$, $\tau_0 = 3$, $\lambda_0 = 4$. **(b)** MD, with fixed values $\omega_0 = 3$, $\tau_0 = 3$, $\lambda_0 = 3$.

Figure 6: Sensitivity of the total reserve with respect to q_0: fixed ω_0, τ_0 and λ_0, and $q_0 = 1, 3, 8, 100$.

We investigate the sensitivity of the results in Section 4.3.1 with respect to the choice of parameters ω_0, q_0, τ_0 and λ_0.



(a) BI, with parameters (τ_0, λ_0) = (1,1), (3,4), (8,8).

(b) MD, with parameters (τ_0, λ_0) = (1,1), (3,3), (8,8).

Figure 7: Sensitivity of the total reserve with respect to τ_0 and λ_0: fixed $\omega_0 = 3$ and $q_0 = 3$.

Effect of ω_0 Figure 5 shows the density of the total reserve as obtained with different values of ω_0. ω_0 equal to 1 implies that reporting delay is not used as claim marker. The influence of the choice of ω_0 on stochastic RDC reserves is limited. Though, the clear difference between $\omega_0 = 1$ and $\omega_0 = 3$ or 7 indicates that reporting delay is a relevant claim marker. Table 11 reports similar conclusions for the deterministic RDC method.

Effect of q_0 While fixing the choice of ω_0, τ_0 and λ_0, we compare the results obtained with q_0 equal to 1 (i.e. no conditioning on the last observed cumulative payment), 3 (our initial choice), 8 and a very high value of 100. The results are in Figure 6. Even though the predicted results of the stochastic RDC method vary more compared to Figure 5, the influence of q_0 (when $q_0 > 1$) is rather limited. Since IBNR claims do not condition on cumulative paid amount (which is zero for an IBNR claim), the variations are only due to differences in the predicted RBNS reserves (apart from simulation differences). The differences between the results with $q_0 = 1$ and $q_1 = 3, 8, 100$ illustrate the relevance of using the cumulative paid amount as a claim marker. The influence of q_0 on the deterministic RDC results is illustrated in Table 11. The impact of changing q_0 on the deterministic RDC reserves is larger than its impact on the stochastic RDC reserves.

Effect of τ_0 and λ_0 We change these parameters simultaneously since our choice of τ_0 depends on the choice of λ_0. Figure 7 compares the reserve distributions obtained with the stochastic RDC method for various choices of (τ_0, λ_0). Setting $\tau_0 = \lambda_0 = 1$ corresponds to no conditioning on the claim length $L^s(k)$ and no conditioning on the timing of the payment. On the other hand, $\tau_0 = \lambda_0 = 8$ corresponds to full conditioning. We conclude that conditioning on the length of the claim and the timing of the payment is relevant. However, from the originally selected parameters onwards, the results are quite stable. The parameters τ_0 and λ_0 are not used in the deterministic RDC specification of Rosenlund (2012).

Conclusion The stochastic RDC method generates stable results, as illustrated in this sensitivity analysis. To avoid overparametrization and - correspondingly - simulation from very small clusters, we prefer values of the parameters ω_0, q_0, τ_0 and λ_0 which are based on careful descriptive analysis of the data (as in Section 4.2). The back test results for our case study are reasonable, and in line with the analysis of Pigeon et al. (2013) of the same data. Moreover, for this particular example the stochastic RDC outperforms the chain-ladder method.

	BI			MD		
	IBNR	RBNS	Total	IBNR	RBNS	Total
$q_0 = 3$, $\omega_0 = 3$	1,352,507	8,556,055	9,908,562	472,022	2,012,754	2,484,776
$q_0 = 3$, $\omega_0 = 1$	478,623	9,552,550	10,031,173	393,196	2,040,165	2,433,361
$q_0 = 3$, $\omega_0 = 7$	1,479,923	8,532,578	10,012,500	489,384	2,011,747	2,501,131
$q_0 = 1$, $\omega_0 = 3$	1,352,507	9,413,110	10,765,617	472,023	1,887,870	2,359,893
$q_0 = 8$, $\omega_0 = 3$	1,352,507	7,464,929	8,817,436	472,023	2,160,033	2,632,055
$q_0 = 100$, $\omega_0 = 3$	1,352,507	5,169,963	6,522,470	472,023	2,066,991	2,539,014

Table 11: Results obtained with the deterministic RDC method introduced in Rosenlund (2012).

5 Conclusions and future work

Starting from the principles of the deterministic Reserving by Detailed Conditioning ('RDC') method in Rosenlund (2012), we develop a stochastic RDC method for micro-level reserving with detailed individual claim characteristics. More specifically, we use the reporting delay, claim length and cumulative paid amount. The method generates a predictive distribution for the outstanding loss reserve using historical simulation. The stochastic RDC method is relatively easy to understand and comes with a generic framework; other claim markers than those used in the paper can be incorporated easily. Except for predicting the number of IBNR claims, the method does not use any parametric distributional assumptions.

We explore the performance of the stochastic RDC method in a case-study with a portfolio of general liability insurance policies for private individuals. We compare the results obtained with stochastic RDC to the reserve calculations with deterministic RDC and the Over Dispersed Poisson bootstrap chain–ladder technique. The stochastic RDC method yields stable results and outperforms the other techniques in the study. Moroever, starting from a careful descriptive analysis of the data, we provide guidelines on how to incorporate claim markers in the reserving process. We illustrate how the reserve distribution reacts to changes in the set of claim markers used, and their specification. The idea of using claim specific markers in the reserving process is promising, and our study should be considered as a starting point of this research agenda.

Several directions for future research can be identified. First, the reserving method is currently limited to the triangle boundary. The introduction of a tail factor should enable the reserving actuary to project his calculations beyond the triangle boundary. Second, we did not present a (statistically) optimal way for binning the claim markers in the study. Analyzing these characteristics with a technique like decision trees is a topic for future work. Third, we may explore parametric distributions as an alternative for the use of empirical distribution functions, including the incorporation of parameter uncertainty.

References

Antonio, K. and Plat, R. (in press, 2014). Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, pages 1–21.

Drieskens, D., Henry, M., Walhin, J.-F., and Wielandts, J. (2012). Stochastic projection for large individual losses. *Scandinavian Actuarial Journal*, 1:1–39.

England, P. D. and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3):443–518.

Haastrup, S. and Arjas, E. (1996). Claims reserving in continuous time: a nonparametric Bayesian approach. *ASTIN Bulletin*, 26(2):139–164.

Hull, J. C. (2012). *Risk Management and Financial Institutions*. Wiley Finance. Wiley, 3rd edition.

Martínez Miranda, M. D., Nielsen, B., Nielsen, J. P., and Verrall, R. J. (2011). Cash flow simulation for a model of outstanding liabilities based on claim amounts and claim numbers. *ASTIN Bulletin*, 41:107–129.

Martínez Miranda, M. D., Nielsen, J. P., and Verrall, R. J. (2012). Double chain ladder. *ASTIN Bulletin*, 42:59–76.

Norberg, R. (1993). Prediction of outstanding liabilities in non-life insurance. *ASTIN Bulletin*, 23(1):95–115.

Norberg, R. (1999). Prediction of outstanding liabilities II. Model variations and extensions. *ASTIN Bulletin*, 29(1):5–27.

Pigeon, M., Antonio, K., and Denuit, M. (2013). Individual loss reserving with the multivariate skew normal distribution. *ASTIN Bulletin*, 43:399–428.

Rosenlund, S. (2012). Bootstrapping individual claim histories. *ASTIN Bulletin*, 42:291–324.

Verrall, R. J., Nielsen, J. P., and Jessen, A. H. (2010). Including count data in claims reserving. *ASTIN Bulletin*, 40(2):871–887.

Wüthrich, M. V. and Merz, M. (2008). *Stochastic claims reserving methods in insurance*, volume 435. John Wiley & Sons.

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be