INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

DISCUSSION PAPER

2016/29

COLLECTIVE LOSS RESERVING WITH TWO TYPES OF CLAIMS IN MOTOR THIRD PARTY LIABILITY INSURANCE

Denuit, M. and J. Trufin

COLLECTIVE LOSS RESERVING WITH TWO TYPES OF CLAIMS IN MOTOR THIRD PARTY LIABILITY INSURANCE

Michel Denuit
Institute of Statistics, Biostatistics and Actuarial Science
Université Catholique de Louvain (UCL)
Louvain-la-Neuve, Belgium

Julien Trufin
Department of Mathematics
Université Libre de Bruxelles (ULB)
Bruxelles, Belgium

July 15, 2016

Abstract

This paper adopts the new loss reserving approach proposed by Denuit and Trufin (2016), inspired from the collective model of risk theory. But instead of considering the whole set of claims as a collective, two types of claims are distinguished, those claims with relatively short development patterns and claims requiring longer developments. In each case, the total payment per cell is modelled by means of a Compound Poisson distribution with appropriate assumptions about the severities. A case study based on a motor third party liability insurance portfolio observed over 2004-2014 is used to illustrate the approach proposed in this paper. Comparisons with Chain-Ladder are performed and reveal significant differences in best estimates as well as in Value-at-Risk at high probability levels.

Key words and phrases: general insurance, reserving, collective model, mixture models, GAMLSS, solvency evaluation.

1 Introduction

In Property and Casualty (P&C) insurance, claims often need several years to be settled. Meanwhile, insurers have to build reserves representing their estimate of outstanding liabilities for claims that occurred on or before the valuation date. Reserving calculation has traditionally been performed on the basis of aggregated data summarized in run-off triangles with rows corresponding to accident years and columns corresponding to development years. Such data exhibit three dimensions: for each accident (or occurrence, or underwriting) year i and development period $j = 1, 2, \ldots$, we read in cell (i, j) inside the triangle the total amount paid by the insurer in calendar year $i + j - 1$ for claims originating in year i. Techniques dealing with such aggregated triangular arrays of data go back to the pre-computer era, at a time where the available computing resources, data storage facilities and statistical methodologies were extremely limited. The Chain-Ladder (CL, in short) approach is certainly the most popular technique falling in this category. See for instance Kaas et al. (2008) for an introduction and Wuthrich and Merz (2008) for a detailed account of the topic.

Departing from these aggregated run-off triangles, Arjas (1989) and Norberg (1993, 1999) developed a mathematical framework for the development of individual claims in continuous time. Individual reserving models describe how each claim evolves over time, from the occurrence of the accident until settlement of the claim. In addition to the pioneering works by Arjas (1989) and Norberg (1993, 1999), let us also mention the contributions by Larsen (2007), Zhao et al. (2009), Drieskens et al. (2012), Rosenlund (2012), Antonio and Plat (2014), Pigeon et al. (2013, 2014), and Huang et al. (2015, 2016).

To bridge the gap between aggregated CL techniques and fully individual reserving models, Denuit and Trufin (2016) tried to find the best compromise, inspired from the individual and collective models in actuarial risk theory (see, e.g., Chapters 2-3 in Kaas et al., 2008, for an introduction). Specifically, the number of payments and the amounts of each of them in cell (i, j) are used to fit a Compound Poisson model. The severities are described by a mixture model, combining light-tailed and heavy-tailed distributions to capture the mix of large and attritional claims. Compared to individual reserving methods, payments related to the same claim are not modelled jointly.

In the present paper, we adopt the same approach but we separate the numerous small (or attritional in the Solvency 2 parlance) claims that are reported to the insurer soon after occurrence and are rapidly settled, and the larger claims that develop more slowly over time.

Large losses are generally defined as those exceeding a large loss threshold at least once during their development. Several authors proposed reserving methods which separate large and attritional losses. See Riegel (2014, 2016) and the references therein. In this constext, the actuary is faced with a major technical problem: for older accident years, most large losses have already been identified whereas for more recent accident years, less developments are available and a higher percentage of large losses have not yet exceeded the threshold. This is why appropriate bifurcation techniques have been proposed by Riegel (2014, 2016) to treat large losses on a consistent basis. To avoid these difficulties, we separate here claims with short and long developments. Typically, claims with short developments are those reported during the year of occurrence or the year after, and finalized at most one year after the accident year. In motor third party liability insurance, this simple classification rule creates two sets of claims with very different characteristics, both in terms of settlement dynamics and amounts paid by the insurer. Even if the time to settlement does not reveal the ultimate cost of the claim, those claims that are rapidly settled are generally cheaper compared to the ones requiring longer developments. Hence, we approximately recover in a simple and efficient way the dichotomy between large and attritional losses.

In this approach, the vast majority of claims can be classified into one of these two categories, except only for the very last accident year. This difficulty can nevertheless easily be circumvented using simple moment-based approximations or mixture models. Even if other criteria can be used, such as the separation between claims with bodily injuries and claims with material damages, only, the classification rule based on the length of the settlement period proposed in this paper appears to be particularly efficient in motor third party liability insurance.

The collective model exploits several triangles of counts. The numbers of payments are built from the numbers of reported and closed claims to get the number of open claims and finally allow for zero payments to isolate the number of effective payments made by the insurer. Notice that we aggregate all the payments related to a single claim and work with total yearly payments, in line with the individual and collective models of actuarial risk theory. To favor tools that are familiar to non-life actuaries (such as Poisson and GLM regression analysis, their GAMLSS extensions, compound sums and Panjer algorithm), we opt here for the incremental payments, for which independence is usually considered as a reasonable assumption in regression-based reserving techniques applied to aggregated triangular data.

Yearly payments are modelled by means of discrete mixture models. A probability mass at the origin accounts for the possibility of zero yearly payments, which means that nothing has been paid by the insurer during the calendar year for that particular claim. A continuous component with a density over the half positive real line is then added to model positive payments. For claims with longer developments, this continuous component is further decomposed into a mixture of light-tailed and heavy-tailed distributions to capture large amounts paid by the insurer. Model parameters are explained by the combined effect of accident year, development lag, and calendar time in a regression setting using GAMLSS techniques.

The remainder of this paper is organized as follows. The motor third party liability insurance data basis used to illustrate the model proposed in this paper is presented in Section 2. Section 3 explains how claims are divided in two categories. Sections 4 and 5 are devoted to the modelling of claims with a relatively short development and claims developing over longer periods, respectively. Section 6 is devoted to the calculation of reserves and to comparisons with the classical CL approach. The final Section 7 discusses the results and concludes.

2 Notation and data

2.1 Accident and development indices

We assume that we have n years of observations. Accident years range from $i = 1$ to n and developments from $j = 1$ to n. These data fill a triangle: Accident year i is followed from development $j = 1$ (corresponding to the accident year itself) to the last observed development $n - i + 1$ (corresponding to the last calendar year n for which observations are available, located along the last diagonal of the triangle).

Henceforth, we denote as ω the time needed to settle all the claims occurred during a given accident year i, i.e. these claims are closed in calendar year $i + \omega - 1$ at the latest. For business lines with long developments, some claims for accident year 1 may still be open in calendar year n so that $\omega > n$. Precisely, $\omega = n$ if all claims of the first accident year are settled at the end of the observation period. If not, $\omega > n$ and we must introduce a tail factor to account for the last developments before final settlement.

2.2 The data

The approach proposed in this paper is applied on a data set extracted from the motor third party liability insurance portfolio of an insurance company operating in the EU. The observation period consists in calendar years 2004 till 2014. The available information concerns accident years 2004 to 2014 so that we have observed developments j up to $n = 11$. In the numerical illustrations, we let i range between $i = 2004$ and 2014, instead of $1, \ldots, n$, in order to make the results easier to interpret.

There are 52,155 claims in the data set. Among them, 4,023 claims are still open at the end of the observation period. Table 1 presents the information available for two claims of the database. Claim #16,384 corresponds to an accident occurred in 2009 that has been reported during the same calendar year. Payments have been made in years 2009 to 2013, but no payment has been recorded for 2014. At the end of the observation period, claim #16,384 is still open. Claim #20,784 corresponds to an accident occurred in 2010 that has been reported during the same calendar year. A payment has been made in 2010, there was no payment in 2011, and the claim has been closed in 2011, one year after its reporting to the insurer. Notice that in our data set, the declaration of a claim corresponds to the first time there is a payment or a positive case estimate for that claim. Hence, late reporting (i.e. at lags 3-4) is due here to the definition adopted for reporting as motor insurance contract typically impose that policyholders rapidly file the claim against the company.

Table 2 displays for each accident year i the total numbers of claims reported at various lags j. We can see from Table 2 that all claims are reported after 4 developments, i.e. during the accident year and the three following calendar years. If the claim appears at lag $j \geq 3$, this usually means that the claim manager initially thought that the insured driver was not liable for the accident but that this opinion has been contradicted later on (recall that reporting means here the first lag j at which there is a payment or a positive case estimate).

Table 3 gives the total numbers of claims closing at various settlement periods j, per accident year i. Even if the majority of claims are settled after two years of development, we also see the emergence of a group of claims with longer path to settlement. Comparing the observed totals 4,196 of reported claims in 2004 and 4,187 of closed claims for the same year reported in Tables 2-3, we can see that some claims are not closed at the end of the observation period, i.e. $n = 11 < \omega$. For the first accident year, there remain 9 claims still open at lag 11. A tail factor will be included in the model to account for the presence of these claims.

Table 4 displays descriptive statistics for the payments per accident year i and lag j. We

Event	No	Year	Amount
Occurrence	16,384	2009	-
	20,784	2010	-
Declaration	16,384	2009	-
	20,784	2010	-
Payments	16,384	2009	5,022
	16,384	2010	67,363
	16,384	2011	903
	16,384	2012	6,295
	16,384	2013	13,850
	16,384	2014	0
	20,784	2010	1,605
	20,784	2011	0
Closure	16,384	Not settled	-
	20,784	2011	-

Table 1: Information available for claims No $16,384$ and No $20,784$ in the data set. Claim No $16,384$ is still open end of year 2014.

	1	2	3	4	5	6	7	8	9	10	11
2004	4,022	165	8	1	0	0	0	0	0	0	0
2005	4,190	174	5	1	0	0	0	0	0	0	
2006	4,331	210	2	2	0	0	0	0	0		
2007	4,743	255	9	3	0	0	0	0			
2008	5,046	222	8	1	0	0	0				
2009	5,168	191	10	0	0	0					
2010	4,612	217	7	1	0						
2011	4,394	200	9	1							
2012	4,299	162	7								
2013	4,557	169									
2014	4,753										

Table 2: Observed numbers of reported claims per accident year $i = 2004, \ldots, 2014$ and lag $j = 1, \ldots, 11$. Whole data basis.

can see there the number of payments, the proportion of claims with no payment, the mean of the payments as well as the standard deviation and skewness per accident year i and lag j. Table 4 shows that there is a break in the average amounts paid by the insurer, which considerably increase after two lags. The standard deviation is often about twice the mean while the large skewness values suggest highly asymmetric distributions.

	1	2	3	4	5	6	7	8	9	10	11
2004	2,266	1,509	203	94	42	38	9	14	5	3	4
2005	2,428	1,582	172	62	46	29	17	12	8	4	
2006	2,433	1,607	228	139	65	29	22	5	7		
2007	2,451	1,853	433	136	59	38	14	8			
2008	2,643	2,079	287	141	58	30	21				
2009	2,607	2,105	385	127	71	29					
2010	1,782	2,442	340	146	66						
2011	1,793	2,223	337	123							
2012	1,852	2,017	373								
2013	1,859	2,197									
2014	1,925										

Table 3: Observed numbers of claims that are definitively settled at development $j = 1, \ldots, 11$ per accident year $i = 2004, \ldots, 2014$. Whole data basis.

3 Separating losses with short and long development to settlement

Claims that are settled relatively rapidly are usually cheaper than those requiring longer settlement periods. This is why we isolate claims that are reported and settled after a few development periods. We see from Table 4 that there is a considerable increase in the average yearly payment after development 2. This suggests to consider that claims are rapidly settled if they are closed during the accident year itself, or during the year after. This choice is also relevant given the absence of reporting delay, all claims being filed during the accident year or the year after (in case of accidents occurring during the last weeks of the calendar year).

Table 5 displays the descriptive statistics for yearly payments at lags $j \in \{1, 2\}$ after exclusion of claims needing more than 2 development periods to be settled. Considering Table 5 and columns $j \geq 3$ in Table 4 clearly shows a break in these averages after lag 2. Also, standard deviations and skewness tend to decrease after the separation between the two types of claims.

4 Losses with short development to settlement

4.1 Collective model

Let ω_1 be the maximum number of developments to qualify as a claim with rapid settlement. Precisely, claims with short development are those claims reported and finalized at most $\omega_1 - 1$ years after occurrence. In our example, we have $\omega_1 = 2$. All claims from accident year i reported and settled during calendar years $\{i, \ldots, i + \omega_1 - 1\}$ are modelled in a collective way. The total payment X_{ij} at development j (i.e. in calendar year $i + j - 1$) for these losses

	1	2	3	4	5	6	7	8	9	10	11
2004											
Num. pay.	2,848	1,459	236	124	68	39	18	18	12	8	7
% no pay.	0.292	0.240	0.438	0.431	0.452	0.524	0.591	0.486	0.429	0.500	0.462
Mean	1,133	1,877	2,713	4,349	4,446	9,894	16,765	4,422	18,072	12,314	21,263
Std. dev.	2,378	5,317	4,861	9,405	7,918	26,576	27,037	9,768	24,203	14,436	50,490
Skewness	12.423	11.878	4.157	4.472	2.948	5.028	1.967	3.476	2.015	1.136	2.039
2005											
Num. pay	3,001	1,492	207	97	53	42	24	21	11	11	
% no pay.	0.284	0.229	0.423	0.484	0.579	0.475	0.529	0.382	0.500	0.214	
Mean	1,112	1,659	3,168	5,455	5,132	14,882	25,781	8,997	4,230	1,347	
Std. dev.	1,847	2,932	6,081	18,278	10,270	41,070	77,046	19,947	2,817	883	
Skewness	4.113	5.509	3.709	8.387	4.089	4.353	3.135	3.464	0.413	0.241	
2006											
Num. pay	3,007	1,659	268	117	61	41	21	10	10		
% no pay.	0.306	0.213	0.467	0.578	0.558	0.438	0.523	0.545	0.412		
Mean	1,164	1,624	5,799	4,494	7,287	6,055	6,141	4,688	12,205		
Std. dev.	2,972	2,932	49,737	7,632	22,190	12,682	9,173	4,594	26,907		
Skewness	23.651	6.020	15.840	3.104	4.485	3.248	1.624	0.506	2.440		
2007											
Num. pay	3,246	1,893	322	170	79	48	24	16			
% no pay.	0.316	0.257	0.542	0.377	0.423	0.385	0.400	0.385			
Mean	1,159	1,905	2,679	3,500	7,401	8,243	12,140	13,148			
Std. dev.	2,258	4,984	6,159	5,831	13,989	14,717	20,625	22,292			
Skewness	9.382	13.781	6.688	3.093	3.029	2.872	2.731	1.921			
2008											
Num. pay	3,574	1,816	304	125	71	37	22				
% no pay.	0.292	0.308	0.451	0.534	0.441	0.464	0.436				
Mean	1,104	1,720	2,189	4,203	4,611	7,775	6,310				
Std. dev.	1,837	3,644	3,524	8,791	9,908	12,249	7,275				
Skewness	5.521	7.533	3.851	4.451	4.774	2.352	0.983				
2009											
Num. pay	3,545	1,877	300	131	90	51					
% no pay.	0.314	0.318	0.543	0.518	0.379	0.311					
Mean	1,142	1,919	3,981	4,379	6,896	9,129					
Std. dev.	1,926	5,710	19,797	11,584	17,446	18,474					
Skewness	4.610	18.270	14.896	7.229	5.379	3.265					
2010											
Num. pay	2,874	2,072	338	161	75						
% no pay.	0.377	0.320	0.448	0.410	0.409						
Mean	1,663	1,984	3,637	5,147	14,935						
Std. dev.	4,012	5,832	11,419	13,420	60,912						
Skewness	18.229	16.910	11.563	5.037	5.849						
2011											
Num. pay	2,777	1,930	327	119							
% no pay.	0.368	0.311	0.443	0.526							
Mean	1,601	1,982	2,441	5,171							
Std. dev.	2,333	4,004	4,119	14,476							
Skewness	5.628	7.586	3.607	4.742							
2012											
Num. pay	2,860	1,749	282								
% no pay.	0.335	0.330	0.529								
Mean	1,716	2,328	4,390								
Std. dev.	4,587	10,085	31,803								
Skewness	36.917	31.363	16.083								
2013											
Num. pay	2,924	1,844									
% no pay.	0.358	0.357									
Mean	1,637	2,230									
Std. dev.	4,120	11,414									
Skewness	31.519	35.894									
2014											
Num. pay	2,723										
% no pay.	0.427										
Mean	1,662										
Std. dev.	2,360										
Skewness	7.018										

Table 4: Descriptive statistics for payments per accident year $i = 2004, \ldots, 2014$ and lag $j = 1, \ldots, 11$, namely the number of payments (Num. pay.), the proportion of claims with no payment (% no pay.), the mean of the payments as well as the standard deviation and skewness. Whole data basis.

is disaggregated into the compound sum

$$X_{ij} = \sum_{k=1}^{N_{ij}^{(o)}} X_{ijk}, \;\; j = 1, 2, \ldots, \omega_1,$$

where

		1	2
2004	$N_{ij}^{(p)}$	2,700	1,220
	% no pay.	0.258	0.192
	Mean	1,001	1,261
	Std. dev.	1,748	2,482
	Skewness	6.333	13.377
2005	$N_{ij}^{(p)}$	2,879	1,299
	% no pay.	0.256	0.179
	Mean	1,027	1,343
	Std. dev.	1,687	1,910
	Skewness	4.432	3.746
2006	$N_{ij}^{(p)}$	2,852	1,370
	% no pay.	0.263	0.147
	Mean	1,011	1,250
	Std. dev.	1,545	1,888
	Skewness	4.412	5.332
2007	$N_{ij}^{(p)}$	2,999	1,498
	% no pay.	0.273	0.192
	Mean	1,002	1,362
	Std. dev.	1,675	2,186
	Skewness	6.416	6.934
2008	$N_{ij}^{(p)}$	3,398	1,524
	% no pay.	0.252	0.267
	Mean	1,010	1,316
	Std. dev.	1,626	2,262
	Skewness	6.085	6.853
2009	$N_{ij}^{(p)}$	3,358	1,527
	% no pay.	0.264	0.275
	Mean	1,070	1,302
	Std. dev.	1,712	1,895
	Skewness	3.998	5.479
2010	$N_{ij}^{(p)}$	2,688	1,753
	% no pay.	0.339	0.282
	Mean	1,499	1,447
	Std. dev.	2,837	1,859
	Skewness	20.963	5.196
2011	$N_{ij}^{(p)}$	2,591	1,615
	% no pay.	0.331	0.274
	Mean	1,521	1,567
	Std. dev.	2,118	2,891
	Skewness	6.012	9.618
2012	$N_{ij}^{(p)}$	2,652	1,420
	% no pay.	0.294	0.296
	Mean	1,515	1,616
	Std. dev.	1,804	2,421
	Skewness	3.524	6.227
2013	$N_{ij}^{(p)}$	2,734	1,535
	% no pay.	0.305	0.301
	Mean	1,490	1,593
	Std. dev.	1,945	2,293
	Skewness	5.075	6.694

Table 5: Observed number $N_{ij}^{(p)}$ of payments and descriptive statistics for payments per accident year $i = 2004, \ldots, 2013$ and development period $j = 1, 2$ restricted to those claims reported and settled during the accident year or the year after. Accident year 2014 is not included as claims with short or long development patterns cannot be identified.

$N_{ij}^{(o)}$ = number of claims with short development originating in accident year i, reported at or before development j, still open at development j;

X_{ijk} = total payment made in calendar year $i+j-1$ for the kth claim with short development originating in accident year i still open at development j, possibly equal to 0.

All these random variables are assumed to be mutually independent. Notice that here, payments related to individual policies are not tracked, only payments for the collective are modelled.

4.2 Severity modelling

Yearly payments X_{ijk} per open claim are modelled by means of zero-augmented regression model based on a light-tailed severity distribution (such as the Gamma or Inverse Gaussian distributions, for instance), with a probability mass at zero

$$\mathrm{P}[X_{ijk} = 0] = \zeta_j \tag{4.1}$$

and a conditional mean of the form

$$\mathrm{E}[X_{ijk}|X_{ijk} > 0] = \gamma_{i+j-1}\xi_j \tag{4.2}$$

where γ_{i+j-1} models inflation (in an hedonic approach) and ξ_j models the development effect. The choice $\gamma_1 = 1$ makes the inflation parameters identifiable and means that the first accident year is taken as the base year for inflation. Notice that working with single payments made by the insurer solves the severe identifiability issues faced in the aggregated triangle approach. The parameter ξ_j then represents the average amount paid at lag j, corrected for inflation.

The estimated probabilities ζ_j of zero payments are $\widehat{\zeta_1} = 28.4\%$ at lag 1 and $\widehat{\zeta_2} = 23.4\%$ at lag 2. Thus, the insurer does not make a positive payment for about one quarter of the open claims.

Table 5 shows a clear break between accident years 2009 and 2010 for the average yearly amount paid per claim. Hence, as an alternative to (4.2), we rather consider a conditional mean of the form

$$\mathrm{E}[X_{ijk}|X_{ijk} > 0] = \nu_i\gamma_{i+j-1}\xi_j, \tag{4.3}$$

where

$$\nu_i = \mathrm{I}[i \leq 2009] + \kappa\mathrm{I}[i > 2009] = \begin{cases} 1 \text{ for } i = 2004, \ldots, 2009 \\ \\ 1 + \kappa \text{ for } i = 2010, 2011, \ldots \end{cases}$$

The parameters ν_i enable us to account for the break observed between accident years 2009 and 2010. Some structure can be imposed to the inflation effects. Here we specify $\gamma_{2004+l} = (1 + \gamma)^l$ where the unique parameter γ is the constant yearly inflation rate. A Gamma regression gives the parameter estimates $\widehat{\kappa} = 1.265$, $\widehat{\gamma} = 1.34\%$, $\widehat{\xi_1} = 1,019$ and $\widehat{\xi_2} = 1,155$. We see that the average sizes of payments, corrected for inflation, increase by 26.5% between accident years 2009 and 2010. Also, we observe that $\widehat{\xi_1} = 1,019 < 1,155 = \widehat{\xi_2}$ suggesting that payments are somewhat more expensive the year after occurrence. Notice that the constant inflation $\widehat{\gamma} = 1.34\%$ leads to an inflation of 14.2% over the whole observation period 2004 − 2014.

4.3 Modelling counts

4.3.1 Reported and closed cases

Let $N_{ij}^{(r)}$ be the number of claims with short development that occurred in accident year i and were reported to the insurer at development j (i.e. during calendar year $i + j - 1$). Also, let $N_{ij}^{(c)}$ be the number of claims with rapid settlement originating in accident year i

that were reported at development j or before and closed during calendar year $i+j-1$. Let us mention that $N_{ij}^{(r)}$ corresponds to N_{ij0} in Schiegl (2015). Schiegl (2015) then splits every $N_{ij}^{(r)}$ into a sequence of numbers N_{ijk} of claims that are still open in calendar year $i+j+k-1$. This allows the actuary to obtain the number $N_{ij}^{(o)}$ of open claims directly from the N_{ijk} at the cost of the extra (or third) dimension k. Here, we obtain $N_{ij}^{(o)}$ by means of $N_{ij}^{(r)}$ and $N_{ij}^{(c)}$.

In line with the classical CL model, we use the multiplicative specification

$$\mathrm{E}[N_{ij}^{(r)}] = \alpha_i \beta_j^{(r)} \text{ and } \mathrm{E}[N_{ij}^{(c)}] = \alpha_i \beta_j^{(c)} \tag{4.4}$$

subject to the usual identifiability constraints

$$\sum_{j=1}^{\omega_1} \beta_j^{(r)} = \sum_{j=1}^{\omega_1} \beta_j^{(c)} = 1.$$

This ensures that the total number

$$N_i = \sum_{j=1}^{\omega_1} N_{ij}^{(r)} = \sum_{j=1}^{\omega_1} N_{ij}^{(c)}$$

of claims with short development has mean $\mathrm{E}[N_i] = \alpha_i$, and that

$$\begin{aligned} \beta_j^{(r)} &= \text{ probability that a claim with rapid settlement is reported at lag } j \\ \beta_j^{(c)} &= \text{ probability that a claim with rapid settlement is closed at lag } j \end{aligned}$$

with $j \in \{1, \ldots, \omega_1\}$.

The parameters α_i, $\beta_j^{(r)}$ and $\beta_j^{(c)}$ are estimated from the two triangles with observed $N_{ij}^{(r)}$ and $N_{ij}^{(c)}$, combined. As the parameter α_i is shared by these two sets of counts, combining both triangles provides the appropriate estimation procedure.

The observed counts $N_{ij}^{(r)}$ are displayed in Table 6 and $N_{ij}^{(c)}$ in Table 7. As $\omega_1 = 2$, restricting the analysis to calendar years with at least ω_1 observed developments means that we exclude only the last accident year. As this is not expected to impact on the estimations, we follow this route here and we estimate the parameters on the basis of occurrence years 2004 to 2013. For these accident years, the observed claim development patterns can be classified into losses with short (i.e. at most $\omega_1 = 2$) or long (i.e. at least $\omega_1 + 1 = 3$) developments to settlement and separate analyses can be conducted for the two types of losses. Tables 5-6-7 stop at accident year 2013 as for the last 2014 we cannot separate the two types of claims. From Table 7, we see that $N_{i1}^{(c)} > N_{i2}^{(c)}$ for $i \leq 2009$ while $N_{i1}^{(c)} < N_{i2}^{(c)}$ for $i \geq 2010$. This can be explained by a change in the claim handling procedure inside the company. In order to account for this deceleration in the speed of settlement for the claims with short development patterns, we decide to calibrate parameters $\beta_j^{(c)}$ only on accident years $i \geq 2010$.

Table 8 displays the estimated parameters α_i, $\beta_j^{(r)}$ and $\beta_j^{(c)}$ that have been obtained by Poisson maximum likelihood based on the two triangles of counts $N_{ij}^{(r)}$ and $N_{ij}^{(c)}$ sharing the

	1	2
2004	3,641	134
2005	3,872	138
2006	3,872	168
2007	4,127	177
2008	4,541	181
2009	4,561	151
2010	4,065	159
2011	3,871	145
2012	3,754	115
2013	3,934	122

Table 6: Observed numbers $N_{ij}^{(r)}$ of reported claims per accident year $i = 2004, \dots, 2013$ and development period $j = 1, 2$ restricted to those claims reported and settled during the occurrence year or the year after.

	1	2
2004	2,266	1,509
2005	2,428	1,582
2006	2,433	1,607
2007	2,451	1,853
2008	2,643	2,079
2009	2,607	2,105
2010	1,782	2,442
2011	1,793	2,223
2012	1,852	2,017
2013	1,859	2,197

Table 7: Observed numbers $N_{ij}^{(c)}$ of claims that are definitively settled at development $j = 1, 2$ per accident year $i = 2004, \dots, 2013$ restricted to those claims reported and settled during the occurrence year or the year after.

common parameter α_i. Therefore, the estimations $\widehat{\alpha}_i$ exactly replicate the observed totals $\sum_{j=1}^{2} N_{ij}^{(r)} = \sum_{j=1}^{2} N_{ij}^{(c)}$. We can see from Table 8 that the estimations $\widehat{\alpha}_i$ remain roughly stable over time, suggesting that the volume of business stays unchanged, or moderately increases being compensated by the progressive reduction in claim frequencies.

The probability of being reported during the accident year is estimated to $\widehat{\beta}_1^{(r)} = 96.4\%$ whereas the probability of being reported the year after is estimated to the remaining

$\widehat{\beta}_2^{(r)} = 3.6\%$. These 3.6% of claims reported the year after roughly correspond to the accidents occurring during the last few weeks of the calendar year, causing delays in reporting and/or handling because of Christmas and New Year breaks. Turning to settlements, we see that about half of those claims developing rapidly are closed during the accident year and the remaining half are closed the year after. Precisely, the corresponding probabilities are estimated to $\widehat{\beta}_1^{(c)} = 45.1\%$ and $\widehat{\beta}_2^{(c)} = 54.9\%$. This is in line with occurrences spread over the calendar year and an average handling period of about half a year.

i	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
$\widehat{\alpha}_i$	3,775	4,010	4,040	4,304	4,722	4,712	4,224	4,016	3,869	4,056
j	1	2	3	4	5	6	7	8	9	10
$\widehat{\beta}_j^{(r)}$	0.964	0.036	-	-	-	-	-	-	-	-
j	1	2	3	4	5	6	7	8	9	10
$\widehat{\beta}_j^{(c)}$	0.451	0.549	-	-	-	-	-	-	-	-

Table 8: Estimated parameters α_i, $\beta_j^{(r)}$ and $\beta_j^{(c)}$.

4.3.2 Number of open cases

The number $N_{ij}^{(o)}$ of claims from accident year i open at development j is then obtained from the identity

$$\sum_{k=1}^{j} N_{ik}^{(r)} = \sum_{k=1}^{j-1} N_{ik}^{(c)} + N_{ij}^{(o)}$$

so that

$$\mathrm{E}[N_{ij}^{(o)}] = \alpha_i \left(\sum_{k=1}^{j} \beta_k^{(r)} - \sum_{k=1}^{j-1} \beta_k^{(c)} \right). \tag{4.5}$$

4.3.3 Number of payments

The number $N_{ij}^{(p)}$ of payments is obtained here from the number of open claims, allowing for the possibility of zero payments. Specifically,

$$N_{ij}^{(p)} = \sum_{k=1}^{N_{ij}^{(o)}} \mathrm{I}[X_{ijk} > 0]$$

where $\mathrm{I}[A]$ is the indicator variable of the event A, equal to 1 if A is realized and to 0 otherwise, so that

$$\mathrm{E}[N_{ij}^{(p)}] = \mathrm{E}[N_{ij}^{(o)}](1 - \zeta_j). \tag{4.6}$$

The observed counts $N_{ij}^{(p)}$ are displayed in Table 5.

The number $N_{ij}^{(o)}$ of open claims is obtained from the joint reporting and closure dynamics, i.e. from the $N_{ij}^{(r)}$ and $N_{ij}^{(c)}$ counts. As we aggregate all the cash-flows related to a single open claim during a given calendar year into a single yearly payment, we believe that this approach is more appropriate in the present setting compared to the direct modelling of the number of payments generated by the $N_{ij}^{(r)}$ reported claims that have payment delay of $0, 1, 2, \ldots$ years.

4.3.4 Compatibility of the count models

In order to check the compatibility of the different count models, we also model $N_{ij}^{(p)}$ directly and compare the corresponding predictions to those obtained from $N_{ij}^{(o)}$. In line with the classical CL model, we use the specification

$$\mathrm{E}[N_{ij}^{(p)}] = \alpha_i^{(p)} \beta_j^{(p)} \tag{4.7}$$

subject to the usual identifiability constraint

$$\sum_{j=1}^{\omega_1} \beta_j^{(p)} = 1.$$

Table 9 displays the estimated $\alpha_i^{(p)}$ and $\beta_j^{(p)}$. The estimated means $\mathrm{E}[N_{ij}^{(p)}]$ obtained by this direct modelling of the number of payments are given in Table 10 and compared to the corresponding values deduced from the number of open claims and the zero-payment probabilities. We see there that the estimated $\mathrm{E}[N_{ij}^{(p)}]$ in (4.7) are reasonably close to the estimated $\mathrm{E}[N_{ij}^{(o)}](1 - \zeta_j)$ obtained previously.

i	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
$\widehat{\alpha}_i^{(p)}$	3,920	4,178	4,222	4,497	4,922	4,885	4,441	4,206	4,072	4,269
j	1	2	3	4	5	6	7	8	9	10
$\widehat{\beta}_j^{(p)}$	0.662	0.338	-	-	-	-	-	-	-	-

Table 9: Estimated parameters $\alpha_i^{(p)}$ and $\beta_j^{(p)}$.

5 Model specification for losses with longer developments

Let us now turn to losses requiring more than ω_1 development periods to be settled or reported to the insurer after lag ω_1.

	$\widehat{\alpha}_i^{(p)}\widehat{\beta}_j^{(p)}$		$\widehat{\mathrm{E}}[N_{ij}^{(o)}](1-\widehat{\zeta}_j)$	
i	$j=1$	$j=2$	$j=1$	$j=2$
2004	2,593	1,327	2,608	1,360
2005	2,764	1,414	2,771	1,444
2006	2,793	1,429	2,791	1,455
2007	2,975	1,522	2,974	1,550
2008	3,256	1,666	3,263	1,701
2009	3,232	1,653	3,256	1,697
2010	2,938	1,503	2,918	1,521
2011	2,782	1,424	2,775	1,446
2012	2,694	1,378	2,673	1,393
2013	2,824	1,445	2,802	1,461

Table 10: Estimated means $\mathrm{E}[N_{ij}^{(p)}]$

5.1 Individual reporting lag and time to settlement

As we work here with fewer losses with longer developments, we build the loss model from individual claim information, before aggregating in yearly totals. Precisely, the kth loss originating in accident year i is represented as follows. First, we account for a random declaration, or reporting lag D_{ik}: The kth claim occurring during calendar year i is reported in year $i+D_{ik}-1$ where the random variable D_{ik} is valued in $\{1,2,\ldots,\omega\}$. Then, we allow for a random length L_{ik} of settlement, with $L_{ik}\in\{0,1,\ldots,\omega\}$: Claim k reported in year $i+D_{ik}-1$ is settled in year $i+D_{ik}+L_{ik}-1\leq i+\omega-1$, where $D_{ik}+L_{ik}\geq\omega_1+1$. We assume that the pairs (D_{ik},L_{ik}) are independent and identically distributed. Notice that the value of D_{ik} constrains the support of L_{ik} as

$$D_{ik}+L_{ik}\geq\omega_1+1$$

so that

$$D_{ik}=d\leq\omega_1\Rightarrow L_{ik}\geq\omega_1-d+1.$$

Moreover, $D_{ik}+L_{ik}\leq\omega$. As an example, let us consider the loss described in Table 11. The corresponding reporting lag and random length of settlement are $D_{ik}=2$ and $L_{ik}=2$.

Let us now describe the history of these claims with longer developments. Between occurrence of the accident and notification to the insurance company, these claims are said to be Incurred But Not Reported (IBNR). From reporting until closure, they are said to be Reported But Not Settled (RBNS). Thus, the claim is classified as IBNR during calendar years $i,\ldots,i+D_{ik}-1$. Then, the claim is classified as RBNS from calendar year $i+D_{ik}-1$ until year $i+D_{ik}+L_{ik}-1$. At the end of each year spent as RBNS, there is a payment of amount $Y_{i,k,D_{ik}+h}$ made by the insurer, $h=0,\ldots,L_{ik}$. We assume that these annual payments are mutually independent (in line with the standard independent increment assumption in regression-based loss reserving). Coming back to our example in Table 11,

Event	Year	Amount
Occurrence	2005	-
Declaration	2006	-
Payments	2006	6,955
	2007	0
	2008	2,089
Closure	2008	-

Table 11: Evolution of a loss with longer development.

the loss occurred in accident year 2005, stayed IBNR during one year (from 2005 to 2006) and became RBNS in 2006 until 2008. The corresponding payments $Y_{i,k,D_{ik}+h}$ are $6,955$ for $h = 0$, 0 for $h = 1$ and $2,089$ for $h = 2$.

As D_{ik} and L_{ik} are correlated, their joint distribution must be inferred from the pairs (D_{ik}, L_{ik}). Three cases must be distinguished for the (D_{ik}, L_{ik}):

- unavalaible data when $D_{ik} > n - i + 1$;
- incomplete data when

$$D_{ik} \leq n - i + 1 \text{ but } D_{ik} + L_{ik} > n - i + 1;$$

- complete data when $D_{ik} + L_{ik} \leq n - i + 1$.

Henceforth, we denote

$$\varphi_{d,l} = \mathrm{P}[D_{ik} = d, L_{ik} = l].$$

5.2 Aggregated counts

Denote as M_i the total number of claims with long development occurred during accident year i. Let us now aggregate the individual developments to form the three related development triangles filled with

$M_{ij}^{(r)}$ = number of claims with long development originating in accident year i, reported at development j, i.e. during calendar year $i + j - 1$,

$$M_{ij}^{(r)} = \sum_{k=1}^{M_i} \mathrm{I}[D_{ik} = j];$$

$M_{ij}^{(c)}$ = number of claims with long development originating in accident year i, reported at development j or before and closed during calendar year $i + j - 1$,

$$M_{ij}^{(c)} = \sum_{k=1}^{M_i} \mathrm{I}[D_{ik} + L_{ik} = j]$$

such that

$$M_{ij}^{(c)} = 0 \text{ for } j \leq \omega_1;$$

$M_{ij}^{(o)} =$ number of claims with long development originating in accident year i, still open at development j,

$$M_{ij}^{(o)} = \sum_{k=1}^{M_i} \mathrm{I}[D_{ik} + L_{ik} \geq j].$$

In line with the CL model, we consider that the counts $M_{ij}^{(r)}$ and $M_{ij}^{(c)}$ have respective means

$$\mathrm{E}[M_{ij}^{(r)}] = \delta_i \theta_j^{(r)} \text{ and } \mathrm{E}[M_{ij}^{(c)}] = \delta_i \theta_j^{(c)} \tag{5.1}$$

subject to the constraints

$$\sum_{j=1}^{\omega} \theta_j^{(r)} = \sum_{j=1}^{\omega} \theta_j^{(c)} = 1.$$

Hence, $\theta_j^{(r)}$ and $\theta_j^{(c)}$ can be interpreted as probabilities of being reported and of being closed at lag j, respectively. We set

$$\theta_j^{(c)} = 0 \text{ for } j \leq \omega_1$$

in accordance with our condition to qualify as a claim with long development. The marginal distribution of the reporting lag D_{ik} is deduced from the triangle of the reported claims $M_{ij}^{(r)}$, i.e.

$$\mathrm{P}[D_{ik} = d] = \theta_d^{(r)}.$$

In case data do not cover the entire settlement period ω, i.e. $n < \omega$, a tail factor can be included in (5.1) to account for the development between n and ω. This may be needed to ensure that the same δ_i is involved in the expectations of $M_{ij}^{(r)}$ and $M_{ij}^{(c)}$. Indeed, claims are typically reported sooner whereas some complicated cases may require development until ω. For instance, the series of estimated $\theta_j^{(c)}$ can be extrapolated from $j = n+1$ to ω when they exhibit a clear trend.

Then, the number M_i of claims with long development occurred during accident year i is decomposed into

$$M_i = \sum_{j=1}^{\omega} M_{ij}^{(r)} = \sum_{j=1}^{\omega} M_{ij}^{(c)},$$

with mean

$$\mathrm{E}[M_i] = \delta_i. \tag{5.2}$$

Furthermore, as

$$\sum_{k=1}^{j} M_{ik}^{(r)} = \sum_{k=1}^{j-1} M_{ik}^{(c)} + M_{ij}^{(o)}$$

we then get

$$\mathrm{E}[M_{ij}^{(o)}] = \delta_i \left(\sum_{k=1}^{j} \theta_k^{(r)} - \sum_{k=1}^{j-1} \theta_k^{(c)} \right).$$

Table 12 displays the observed numbers $M_{ij}^{(r)}$ for the claims included in the database. Table 12 is thus the analog to Table 2 after having excluded the claims reported and settled in at most 2 years. Similarly, Table 13 displays the observed numbers $M_{ij}^{(c)}$ of closed claims.

	1	2	3	4	5	6	7	8	9	10	11
2004	381	31	8	1	0	0	0	0	0	0	0
2005	318	36	5	1	0	0	0	0	0	0	
2006	459	42	2	2	0	0	0	0	0		
2007	616	78	9	3	0	0	0	0			
2008	505	41	8	1	0	0	0				
2009	607	40	10	0	0	0					
2010	547	58	7	1	0						
2011	523	55	9	1							
2012	545	47	7								
2013	623	47									

Table 12: Observed numbers $M_{ij}^{(r)}$ of reported claims requiring more than two development periods, per accident year $i = 2004, \ldots, 2013$ and development $j = 1, \ldots, 11$, together with totals $\sum_{j=1}^{n-i+1} M_{ij}^{(r)}$. Accident year 2014 is not included as claims with short or long development patterns cannot be identified.

	1	2	3	4	5	6	7	8	9	10	11
2004	0	0	203	94	42	38	9	14	5	3	4
2005	0	0	172	62	46	29	17	12	8	4	
2006	0	0	228	139	65	29	22	5	7		
2007	0	0	433	136	59	38	14	8			
2008	0	0	287	141	58	30	21				
2009	0	0	385	127	71	29					
2010	0	0	340	146	66						
2011	0	0	337	123							
2012	0	0	373								

Table 13: Observed numbers $M_{ij}^{(c)}$ of closed claims requiring more than two development periods, per accident year $i = 2004, \ldots, 2013$ and development $j = 1, \ldots, 11$, together with totals $\sum_{j=1}^{n-i+1} M_{ij}^{(c)}$. Accident years 2013-2014 are not included as claims require at least two developments before settlement.

Table 14 displays the estimated δ_i and $\theta_j^{(r)}$ obtained from the observed counts $M_{ij}^{(r)}$ by assuming that these random variables are independent and obey Poisson distributions with means (5.1). We see that the resulting $\widehat{\delta}_i$ replicate the observed totals $\sum_{j=1}^{n-i+1} M_{ij}^{(r)}$ for earlier accident years (up to 2011), whereas 1 and 10 claims are added to these totals for accident years 2012 and 2013, respectively. This is in line with the observed reporting patterns at lags $j \in \{3, 4\}$ for accident years 2012 and 2013. The estimated probabilities $\widehat{\theta}_j^{(r)}$ of being reported at lag j rapidly decrease with j, starting from 90.1% and ending at 0.2%.

i	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
$\widehat{\delta}_i$	421	360	505	706	555	657	613	588	600	680
j	1	2	3	4						
$\widehat{\theta}_j^{(r)}$	0.901	0.084	0.013	0.002						

Table 14: Estimated parameters δ_i and $\theta_j^{(r)}$.

Table 15 displays the corresponding estimates obtained from the observed counts $M_{ij}^{(c)}$ by Poisson regression. The tail factors $\widehat{\theta}_{12}^{(c)}$ and $\widehat{\theta}_{13}^{(c)}$ have been set to

$$\widehat{\theta}_{11}^{(c)} = \widehat{\theta}_{12}^{(c)} = \widehat{\theta}_{13}^{(c)} = 1\%$$

because the last estimations $\widehat{\theta}_{10}^{(c)}$ and $\widehat{\theta}_{11}^{(c)}$ tend to stabilize around that value. For the first accident year, these two tail factors increase $\sum_{j=1}^{n} M_{ij}^{(c)} = 412$ to 420 so that $\widehat{\delta}_i$ obtained from the counts $M_{ij}^{(c)}$ and $M_{ij}^{(r)}$ by Poisson regression almost coincide. For accident year 2005, $\widehat{\delta}_i$ obtained from $M_{ij}^{(c)}$ and $M_{ij}^{(r)}$ are equal. This motivates the choice $\omega = 13$. However, $\widehat{\delta}_i$ obtained from the counts $M_{ij}^{(c)}$ then exceeds the observed values $\sum_{j=1}^{n-i+1} M_{ij}^{(r)}$ for accident years 2006 and beyond.

i	2004	2005	2006	2007	2008	2009	2010	2011	2012				
$\widehat{\delta}_i$	420	360	514	726	579	683	659	627	715				
j	1	2	3	4	5	6	7	8	9	10	11	12	13
$\widehat{\theta}_j^{(c)}$	0	0	0.522	0.212	0.103	0.059	0.032	0.019	0.015	0.009	0.010	0.010	0.010

Table 15: Estimated parameters δ_i and $\theta_j^{(c)}$.

The compatibility with the triangles of the counts $M_{ij}^{(r)}$, $M_{ij}^{(c)}$ and $M_{ij}^{(o)}$ is then obtained from

$$\mathrm{E}[M_{ij}^{(c)}] = \sum_{k=1}^{j} \mathrm{E}[M_{ik}^{(r)}] \frac{\varphi_{k,j-k}}{\theta_k^{(r)}}$$

and

$$\mathrm{E}[M_{ij}^{(o)}] = \sum_{k=1}^{j} \mathrm{E}[M_{ik}^{(r)}] \left(1 - \frac{\varphi_{k,0} + \ldots + \varphi_{k,j-k-1}}{\theta_k^{(r)}}\right).$$

5.3 Claim severities

In order to model the sequence of yearly payments $Y_{i,k,D_{ik}+h}$, $h = 0, \ldots, L_{ik}$, we need to account for zero values (i.e. no payment for that claim during that particular year) as well as possibly large values. Therefore, we resort to a discrete mixture with three components:

- a lighter-tailed component with probability τ_h such as Gamma or Inverse Gaussian distributions;

- a heavier-tailed component with probability ρ_h with Pareto type 2 distribution;

- as well as a probability mass at zero

$$\mathrm{P}[Y_{i,k,D_{ik}+h} = 0] = 1 - \tau_h - \rho_h.$$

The parameters (probabilities assigned to each component as well as distributional parameters) are explained by means of several explanatory variables using appropriate regression models. Specifically, yearly payments are assumed to be mutually independent and explained by means of

- a large-claim inflation effect ϑ_t related to an appropriate time scale t; here, we consider the time of payment ($t = i + D_{ik} + h - 1$);

- a reporting lag effect D_{ik};

- a claim-specific development effect h (i.e. h measures the development from reporting, not from the occurrence year).

If needed, the inflation effect may be structured (by specifying a constant inflation rate, for instance). Of course, other effects may also be included.

Considering the amounts of payments $Y_{i,k,D_{ik}+h}$, we use a mixture model with a Gamma and a Pareto type 2 component augmented with a probability mass at zero. We denote by $F^{(1)}_{h,i+D_{ik}+h-2}$ the cumulative distribution function of the Gamma component and by $F^{(2)}_{h,i+D_{ik}+h-2}$ the cumulative distribution function of the Pareto type 2 component. Precisely, the average payment is of the form

$$\chi_{1,h}(1 + g_1)^{i+D_{ik}+h-2}$$

for the Gamma component and of the form

$$\chi_{2,h}(1 + g_2)^{i+D_{ik}+h-2}$$

for the Pareto type 2 component. In these averages, g_1 (resp. g_2) can be interpreted as a constant inflation rate for the Gamma (resp. Pareto type 2) component and $\chi_{1,h}$ (resp. $\chi_{2,h}$) models the claim-specific development effect h for the Gamma (resp. Pareto type 2) component.

Considering the estimations displayed in Table 16, the weights for the Gamma and Pareto type 2 components are comparable at developments $h \in \{0, 1\}$ whereas the Gamma component decreases between $h = 2$ and $h = 3$ to become negligible when $h \geq 4$.

The estimated inflation rates g_1 and g_2 are $\widehat{g}_1 = 0.03\%$ and $\widehat{g}_2 = 1.96\%$ and Table 17 shows the estimated $\chi_{1,h}$ and $\chi_{2,h}$ obtained by using cubic spline smoothers. The way these estimations vary with the development h conforms with intuition. The mathematical expectations of the Gamma components decrease with lag h whereas those of the Pareto component increase with h, capturing the largest losses with expensive payments and longer development to settlement.

h	0	1	2	3	4	5	6	7	8	9	10
$\widehat{\tau}_h$	0.170	0.262	0.153	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\widehat{\rho}_h$	0.193	0.293	0.355	0.450	0.549	0.564	0.502	0.545	0.576	0.607	0.500
$1-\widehat{\tau}_h-\widehat{\rho}_h$	0.637	0.445	0.492	0.487	0.451	0.436	0.498	0.455	0.424	0.393	0.500

Table 16: Estimated parameters τ_h and ρ_h, as well as the probability mass at zero $1-\tau_h-\rho_h$, by development h for a lighter-tailed component Gamma and a heavier-tailed component Pareto type 2.

h	0	1	2	3	4	5	6	7	8	9	10
$\widehat{\chi}_{1,h}$	3,432	3,765	1,903	652	-	-	-	-	-	-	-
$\widehat{\chi}_{2,h}$	3,061	4,019	4,154	5,318	6,011	8,253	9,709	10,167	10,145	10,231	10,178

Table 17: Estimated parameters $\chi_{1,h}$ and $\chi_{2,h}$ by development h for a lighter-tailed component Gamma and a heavier-tailed component Pareto type 2.

6 Reserve calculations

6.1 Best estimates

6.1.1 Claims with short development

For calendar years $i \in \{n-\omega_1+2, \ldots, n\}$, the expected value of the outstanding claims with short development is

$$\begin{aligned} \mathrm{E}\left[\sum_{j=n-i+2}^{\omega_1} X_{ij}\right] &= \sum_{j=n-i+2}^{\omega_1} \mathrm{E}\left[N_{ij}^{(o)}\right] \mathrm{E}\left[X_{ij1}\right] \\ &= \sum_{j=n-i+2}^{\omega_1} \mathrm{E}\left[N_{ij}^{(p)}\right] \mathrm{E}\left[X_{ij1}^{+}\right] \end{aligned}$$

where X_{ij1}^{+} is distributed as X_{ij1} given $X_{ij1} > 0$. Of course, discounting at some appropriate interest rate can be included, if needed. By equations (4.5), (4.2) and (4.6), it comes

$$\mathrm{E}\left[\sum_{j=n-i+2}^{\omega_1} X_{ij}\right] = \sum_{j=n-i+2}^{\omega_1} \alpha_i \left(\sum_{k=1}^{j} \beta_k^{(r)} - \sum_{k=1}^{j-1} \beta_k^{(c)}\right)(1-\zeta_j)\gamma_{i+j-1}\xi_j.$$

In our case study, $\omega_1 = 2$ so that only the last accident year 2014 is concerned with reserve calculations for claims with short development. Also, in the setting of our case study, instead of relying on (4.2) for $\mathrm{E}\left[X_{ij1}^{+}\right]$, we rather use (4.3). Hence, since $\nu_i = \kappa$ and $\gamma_{i+1} = (1+\gamma)^{11}$ for the last accident year $i = 2014$, we get

$$\mathrm{E}\left[X_{i2}\right] = \alpha_i \left(\sum_{k=1}^{2} \beta_k^{(r)} - \beta_1^{(c)}\right)(1-\zeta_2)\nu_i\gamma_{i+1}\xi_2 = \alpha_i \left(1 - \beta_1^{(c)}\right)(1-\zeta_2)\kappa(1+\gamma)^{11}\xi_2.$$

The parameter α_i still needs to be estimated (see Table 8). Since $\mathrm{E}\left[N_{i1}^{(c)}\right] = \alpha_i \beta_1^{(c)}$, we can estimate α_i by $N_{i1}^{(c)}/\hat{\beta}_1^{(c)} = 1,925/0.451 = 4,271$. Hence, we get $\hat{\mathrm{E}}\left[X_{i2}\right] = 3,040,309$.

6.1.2 Claims with longer development

For all accident years (except the first one if $n = \omega$), we must add the corresponding amount for claims with longer development. To develop the total payments related to the whole portfolio, we can simply use the following aggregate representation

$$Z_{ij} = \sum_{k=1}^{M_{ij}^{(o)}} Z_{ijk}$$

where Z_{ijk} is the payment in calendar year $i+j-1$ for a claim originating in accident year i, still open at development j. The distribution of Z_{ijk} can be obtained as a mixture by conditioning with respect to D_{ik}. Precisely,

$$\mathrm{P}[Z_{ijk} \leq z] = \sum_{d=1}^{j} \mathrm{P}[Z_{ijk} \leq z | D_{ik} = d] \mathrm{P}[D_{ik} = d | D_{ik} \leq j]$$

where

$$\begin{aligned} \mathrm{P}[Z_{ijk} \leq z | D_{ik} = d] &= \mathrm{P}[Y_{i,k,d+(j-d)} \leq z] \\ &= 1 - \tau_{j-d} - \rho_{j-d} + \tau_{j-d} F_{j-d,i+j-2}^{(1)}(z) + \rho_{j-d} F_{j-d,i+j-2}^{(2)}(z) \end{aligned}$$

and

$$\mathrm{P}[D_{ik} = d | D_{ik} \leq j] = \frac{\theta_d^{(r)}}{\theta_1^{(r)} + \ldots + \theta_j^{(r)}}.$$

Then,

$$\mathrm{E}\left[\sum_{j=n-i+2}^{\omega} Z_{ij}\right] = \sum_{j=n-i+2}^{\omega} \mathrm{E}\left[M_{ij}^{(o)}\right] \mathrm{E}\left[Z_{ij1}\right]$$

with

$$\mathrm{E}\left[Z_{ij1}\right] = \sum_{d=1}^{j} \mathrm{E}[Y_{i,1,d+(j-d)}] \frac{\theta_d^{(r)}}{\theta_1^{(r)} + \ldots + \theta_j^{(r)}}$$

and

$$\mathrm{E}[Y_{i,1,d+(j-d)}] = \tau_{j-d} \chi_{1,j-d} (1+g_1)^{i+j-2} + \rho_{j-d} \chi_{2,j-d} (1+g_2)^{i+j-2}.$$

6.1.3 Last accident year

In the context of our case study, we still need to estimate δ_i for the last accident year 2014 (see Table 14). As we know that $\mathrm{E}\left[N_{i1}^{(r)} + M_{i1}^{(r)}\right] = \alpha_i \beta_1^{(r)} + \delta_i \theta_1^{(r)}$, we estimate δ_{2014} by

$$\widehat{\delta}_{2014} = \frac{N_{2014,1}^{(r)} + M_{2014,1}^{(r)} - \widehat{\alpha}_{2014} \widehat{\beta}_1^{(r)}}{\widehat{\theta}_1^{(r)}} = 704.$$

Also, we need parameters estimates for τ_h, ρ_h, and $\chi_{2,h}$ for lags $h = 11, 12$. Considering the values reported in Tables 16 and 17 for h up to 10, it seems reasonable to set $\widehat{\tau}_h = 0$, $\widehat{\rho}_h = 0.5$ and $\widehat{\chi}_{2,h} = 10,200$ for $h = 11, 12$. The reserve estimate corresponding to claims with longer development is $24,384,172$.

6.2 Outstanding loss distribution

In a compound Poisson setting, X_{ij} is distributed as

$$\sum_{k=1}^{N_{ij}^{(p)}} X_{ijk}^{+}$$

where the number of payments $N_{ij}^{(p)}$ is Poisson distributed with mean (4.6) and X_{ijk}^{+} is distributed as X_{ijk} given $X_{ijk} > 0$. So, all X_{ij} and Z_{ij} are independent and compound Poisson distributed. Hence their sum also obeys a compound Poisson distribution and Panjer algorithm can be used to derive the distribution of the outstanding claim amount

$$\sum_{j=n-i+2}^{\omega_1} X_{ij} + \sum_{j=n-i+2}^{\omega} Z_{ij}. \tag{6.1}$$

However, a computational problem may happen at initialization of the Panjer recursion to compute $\exp(-\lambda)$ for large values of λ. We refer the reader for instance to Embrechts and Frei (2009) for more details. In our case study, the Poisson parameter of the outstanding claim amount (6.1) is $7,578$ so that $\exp(-7,578)$ is outside the range of representable numbers in `R` and is thus considered as 0. The same problem occurs with other software, such as `Mathematica`. To prevent this issue, we could decompose the compound Poisson random variable into the sum of m compound Poisson sums for a suitably large m such that $\exp(-\lambda/m)$ can be evaluated in `R` and then carry out the $m-$fold convolution. Alternatively, we may initiate Panjer recursion with unit value and then stop before probabilities explode, dividing them by $\exp(\lambda)$. Here we rather choose to perform Monte-Carlo simulations.

Table 18 summarizes the outstanding claim distribution. The Value-at-Risk (VaR) at probability levels 95% and 99.5% are based on $100,000$ simulations while the reserve estimate is calculated analytically as shown previously. We see that about 1,500,000 has to be added to the best estimate of the reserve to reach the 95th percentile. Approximately another 1,500,000 is needed to obtain the 99.5% quantile.

6.3 Comparison with CL and single collective approach

To enable benchmarking, we include the estimation results as obtained with standard reserving techniques designed for run-off triangles. We consider the results obtained with the help of an Overdispersed Poisson (ODP) model with CL structure (as obtained with the `chainladder` package available in `R`) where we apply a tail factor and we perform $10,000$ simulations. We also fit the collective model proposed by Denuit and Trufin (2016), which does not separate claims with short and long developments. Table 18 also shows the reserve estimates and the VaR at probability levels 95% and 99.5% obtained with these two methods.

	Reserve estimate	$\mathrm{VaR}_{0.95}$	$\mathrm{VaR}_{0.995}$
Our approach	27,424,481	29,262,230	30,751,029
Single collective	21,693,829	23,457,061	25,017,697
CL	22,259,690	24,142,573	25,239,963

Table 18: Reserve estimates and VaR at probability levels 95% and 99.5%.

Let us briefly comment on the values listed in Table 18. We see that both the best estimates and the VaRs of the reserve are higher with the collective reserving model separating the two types of claims proposed in this paper, compared to the aggregate CL predictions. Also, working with a single collective leads to a significant underestimation of the insurer's liabilities. Let us stress that the VaR at 99.5% obtained with the approach developed in this paper appears to be closer to the sum of the insurer's case estimates at the end of 2014 compared to the lower CL value.

7 Discussion

The model that has been proposed in this paper proceeds in two steps. For data corresponding to short developments on the one hand, the observed numbers of claims are studied in a Poisson regression setting. Moving to more elaborate models, including zero-inflated or other mixed Poisson specifications, is possible if more appropriate. A zero-augmented Gamma regression model is calibrated to paid amounts, with a specific inflation effect. For claims with longer developments on the other hand, the reporting and settlement lags are modelled at individual level and a 3-component mixture model describes the yearly payments per reported loss at various developments. Such finite mixture models can be fitted to observed loss developments using the `GAMLSS` package of the statistical software `R`.

In our example, we selected $\omega_1 = 2$ on the basis of the observed developments, so that we only excluded the last accident year from the statistical estimation procedure. For larger values of ω_1, this may no longer seem reasonable as excluding a significant volume of data near the end of the observation period may impact on the results. As observed loss development patterns cannot be classified into short and long ones for accident years where less than ω_1 developments are available (i.e. for accident years $i > n - \omega_1 + 1$), we need to resort to a mixture model to account for the co-existence of losses with short and long development patterns. Specifically, denoting as π_1 and $\pi_2 = 1 - \pi_1$ the probability that a given loss develops in less than ω_1 years and more than ω_1 years, respectively, each (possibly zero) payment in these cells obeys the 4-component mixture model consisting in a probability mass at zero, the common distribution of the X_{ijk}^{+}, the lighter-tailed Gamma component, and the heavier-tailed Pareto type 2 component. Such a discrete mixture model can be fitted to the observations in the lower, left cells of the reserving triangle.

Of course, the same approach can be adopted with other dichotomies suitable for motor insurance, such as claims with bodily injuries and claims with material damages, only, or claims with initial case estimate above or below a given threshold.

Acknowledgements

The authors wish to thank the Addactis team, especially Patrick Hennion, for fruitful discussions and for kind assistance in obtaining the data set used to illustrate the present paper.

References

- Antonio, K., Plat, R. (2014). Micro–level stochastic loss reserving for general insurance. Scandinavian Actuarial Journal 2014, 649-669.

- Arjas, E. (1989). The claims reserving problem in non–life insurance: Some structural ideas. ASTIN Bulletin 19, 140-152

- Denuit, M., Trufin, J. (2016). Beyond the Tweedie reserving model: The collective approach to loss development. Working Paper, Institute of Statistics, Biostatistics and Actuarial Science, UCL, Louvain-la-Neuve, Belgium.

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

- Embrechts, P., Frei, M. (2009). Panjer recursion versus FFT for compound distributions. Mathematical Methods of Operations Research 69, 497-508.

- Huang, J., Qiu, C., Wu, X. (2015). Stochastic loss reserving in discrete time: Individual versus aggregate data models. Communications in Statistics: Theory and Methods 44, 2180-2206.

- Huang, J., Wu, X., Zhou, X. (2016). Asymptotic behaviors of stochastic reserving: Aggregate versus individual models. European Journal of Operations Research 249, 657-666.

- Kaas, R., Goovaerts, M.J., Dhaene, J., Denuit, M. (2008). Modern Actuarial Risk Theory Using R. Springer, New York.

- Larsen, C. (2007). An individual claims reserving model. ASTIN Bulletin 37, 113-132

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

- Norberg, R. (1999). Prediction of outstanding liabilities II: Model variations and extensions. ASTIN Bulletin 239, 5-25.

- Pigeon, M., Antonio, K., Denuit, M. (2013). Individual loss reserving with the multivariate Skew Normal framework. ASTIN Bulletin 43, 399-428.

- Pigeon, M., Antonio, K., Denuit, M. (2014). Individual loss reserving using paid–incurred data, Insurance: Mathematics and Economics 58, 121-131.

- Riegel, U. (2014). A bifurcation approach for attritional and large losses in chain ladder calculations. ASTIN Bulletin 44, 127-172.

- Riegel, U. (2016). Bifurcation of attritional and large losses in an additive IBNR environment. Scandinavian Actuarial Journal 2016, 604-623.

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

- Schiegl, M. (2015). A model study about the applicability of the Chain Ladder method. Scandinavian Actuarial Journal 2015, 482-499.

- Wuthrich, M. V., Merz, M. (2008). Stochastic Claims Reserving Methods in Insurance. Wiley, New York.

- Zhao, X., Zhou, X., Wang, J. (2009). Semiparametric model for prediction of individual claim loss reserving. Insurance: Mathematics and Economics 45, 1-8.