KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

A multi-state approach and flexible payment distributions for micro-level reserving in general insurance

Katrien Antonio, Els Godecharle, and Robin Van Oirbeek

AFI_16106

A multi-state approach and flexible payment distributions for micro-level reserving in general insurance

Katrien Antonio[1], Els Godecharle [*2], and Robin Van Oirbeek[3]

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

April 9, 2016

Abstract

Insurance companies hold reserves to be able to fulfill future liabilities with respect to the policies they write. Micro-level reserving methods focus on the development of individual claims over time, providing an alternative to the classical techniques that aggregate the development of claims into run-off triangles. This paper presents a discrete-time multi-state framework that reconstructs the claim development process as a series of transitions between a given set of states. The states in our setting represent the events that may happen over the lifetime of a claim, i.e. reporting, intermediate payments and closure. For each intermediate payment we model the payment distribution separately. To this end, we use a body-tail approach where the body of the distribution is modeled separately from the tail. Generalized Additive Models for Location, Scale and Shape introduced by Stasinopoulos and Rigby (2007) allow for flexible modeling of the body distribution while incorporating covariate information. We use the toolbox from Extreme Value Theory to determine the threshold separating the body from the tail and to model the tail of the payment distributions. We do not correct payments for inflation beforehand, but include relevant covariate information in the model. Using these building blocks, we outline a simulation procedure to evaluate the RBNS reserve. The method is applied to a real life data set, and we benchmark our results by means of a back test.

Keywords: micro-level reserving, extreme value theory, splicing, multi-state model

1 Introduction

An important feature in the insurance market is the precedence of premium income to the claim costs of an insurance policy. This characteristic is commonly referred to as an *inverted production*

[*]Corresponding author: els.godecharle@gmail.com

cycle. Due to this feature, it is important insurance companies hold sufficient reserves in order to be able to fulfill future liabilities with respect to claims that occur within the insurance coverage period. These reserves are a key factor on the liability side of the balance sheet of the insurance company. Accurate, reliable and stable reserving methods for a wide range of products and lines of business are crucial to safeguard solvency, stability and profitability. For example, according to Swiss Re (2008) deficient loss reserves were the main cause of financial insolvency in the US property and casualty (also called: general or non-life) market during the period 1969–2002. With the introduction of new regulatory guidelines for the European insurance business in the form of Solvency II, the insurance industry has regained interest in using more elaborate methodology to model future cash flows and meet regulators' increasing requirements. Insurance companies are strongly encouraged to replace their ad hoc, deterministic methods with fully stochastic approaches, aiming at accurately reflecting the riskiness in the portfolio under consideration. Current techniques for loss reserving will have to be improved, adjusted or extended to meet the requirements of the new regulations. This paper fits within this research avenue and contributes to the literature on statistical models for reserving in general insurance.

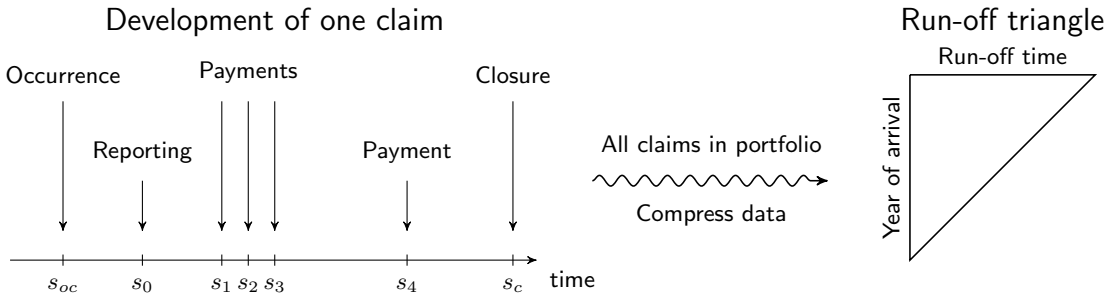

Figure 1: Time line representing the development of a non-life claim and summarized in a run-off triangle.

Figure 1 illustrates the run-off (or development) process of a non-life insurance claim (see Taylor (2000), England and Verrall (2002) and Wüthrich and Merz (2008)). A non-life insurance claim starts its lifetime or development at a certain point in time with the occurrence of an accident or claim event (e.g. a car accident). The occurrence date (s_{oc} in Figure 1) of a claim refers to the date at which the claim event occurs. After the occurrence, the insured reports the claim to the insurance company. The reporting date (s_0 in Figure 1) refers to the date at which this happens. Once the insurer is aware of the claim and accepts the claim for reimbursement, some payments follow (at times s_1, s_2, s_3, s_4 in Figure 1) to compensate the insured for his loss. Eventually, when the loss covered by the policy is completely compensated for, the claim closes. The closing date (s_c in Figure 1) refers to the date at which the claim closes. This finalizes the development of a claim. At the moment of evaluation (typically: end of a quarter, mid year or end of book year), say $s^\star$, the insurer has to set reserves aside to fulfill his future liabilities and safeguard the solvency of the company. Loosely speaking, the insurer must predict, with maximum accuracy, the outstanding loss amount with respect to claims not yet closed at the moment of evaluation. Our interest goes out to the reserve necessary to cover outstanding liabilities from incurred claims that are not yet finalized.

Existing methods for claims reserving (see England and Verrall (2002) and Wüthrich and Merz (2008)) are designed for aggregated data, conveniently summarized in a so-called *run-off triangle*. A run-off triangle summarizes the information registered on individual claims by aggregating

payments into two-dimensional cells, representing the year of occurrence of the claim and the period of development during which the payment took place. See Figure 1 for a visualization of this process. Through this data compression many useful information of the claim is lost. In particular, policy(holder) characteristics (e.g. own risk or deductible, policy limit), characteristics of the accident, the claim, and expert information are ignored, or heavily compressed, in the run-off design. Recent literature challenges the appropriateness of reserving methods based on run-off triangles. For example, both Halliwel (2007) and Schiegl (2015) discuss and demonstrate conditions under which the traditional reserving methods for aggregated data, in particular the chain ladder method, are biased.

A recent focus within the literature on loss reserving is the possible added value of using more extensive data when calculating reserves. Such data are available within insurance companies, as the left part of Figure 1 illustrates, and it fits within the avenue of *big data* and *insurance analytics* to explore their use, see Frees (2015).

On the one hand, initiated by the work of Verrall et al. (2010), Martínez Miranda et al. (2012) and Martínez Miranda et al. (2013) extend the traditional chain ladder framework to Double Chain Ladder (DCL) and even continuous chain ladder setting. The DCL method combines the information of a classical run-off triangle with reported count data, whereas the continuous chain ladder improves the classical actuarial technique, which can be formulated as a histogram type of approach, by replacing this histogram by a kernel smoother. Hiabu et al. (2016a) and Hiabu et al. (2016b) show different ways to extend the traditional chain ladder framework through the inclusion of extra data resources in an aggregated format such as reporting delay and expert knowledge in the form of incurred payments.

On the other hand, micro-level loss reserving models focus on the development of individual claims over time, as launched originally by Norberg (1993) and Norberg (1999). Haastrup and Arjas (1996) and Larsen (2007) model the development of a set of claims over time as a marked Poisson process. Antonio and Plat (2014) apply the continuous time framework of Norberg (1993) and Norberg (1999) to a real life data set, using maximum likelihood estimation, and add a discussion of distributional choices made in their analysis. In their approach, hazard rates drive the time to 'events' in the development of a claim (e.g. a payment, or settlement) and a lognormal regression is used to model intermediate payments (e.g. those at time s_1, s_2, s_3 and s_4 in Figure 1) corrected for inflation using the Consumer Price Index (CPI). Zhao et al. (2009) and Zhao and Zhou (2010) also work in continuous time and propose a semi-parametric model to develop individual claims. Another strand within the micro-level reserving literature works in discrete time and aggregates payments per development period (e.g. a development year) while keeping focus on the development of an individual claim. Pigeon et al. (2014) extend the work of Pigeon et al. (2013), who introduce a claim specific run-off viewpoint with chain ladder like development factors, by including information on incurred losses. Drieskens et al. (2012) follow up on the work of Murphy and McLennan (2006) and propose to develop individual large claims through a non-parametric method based on historical simulation. Rosenlund (2012) presents a deterministic reserving method that allows to condition on specific characteristics of an individual claim. Godecharle and Antonio (2015) integrate this conditioning approach with the historical simulation used in Drieskens et al. (2012). Recent methodological work shows how the use of micro-level claims reserving methods significantly improves accuracy compared to the aggregate methods. Jin and Frees (2013) evaluate the performance of traditional run-off triangle techniques compared to a micro-level model, highlighting scenarios in which the latter outperform the former. Jin (2013) contribute to this topic in the form of a case study

on a workers compensation insurance portfolio. Huang et al. (2015a), Huang et al. (2015b) and Huang et al. (2016) demonstrate, theoretically and numerically, the advantage one can gain from using a micro-level approach compared to the traditional run-off triangle techniques.

We extend the micro-level loss reserving model of Antonio and Plat (2014) in multiple ways and connect it to recent contributions in the statistical literature on multiple state models, the framework of Generalized Additive Models for Location, Scale and Shape (GAMLSS) introduced by Stasinopoulos and Rigby (2007) and tools from Extreme Value Theory (EVT). As such, we avoid some of the rigid choices made in previous work on this topic. First, we adjust the loss reserving model of Antonio and Plat (2014) to a discrete-time framework. To this end, we propose a multi-state framework such that the claim development process can be reconstructed as a series of transitions between a given set of states. The use of a multi-state framework for pricing and reserving is common in a health insurance context (see Haberman and Pitacco (1999), Olivieri and Pitacco (2009), Chapter 6 in Pitacco (2014) and Czado and Rudolph (2002)) and in life insurance (see Chapter 8 in Dickson et al. (2013) and Chapter 20 in Frees et al. (2014)). Hesselager (1994) was one of the first to apply this popular actuarial model to the non-life insurance reserving context. A second extension lies in the modeling of the distribution of each subsequent payment in the development process. Contrary to Antonio and Plat (2014), we model the claim size distribution for each subsequent payment of the claim development process separately. Because the insurer is often confronted with heavy-tailed data and should safeguard the company against extreme losses, an accurate description of the upper tail of the payment distribution is of utmost importance. Therefore, as Larsen (2007) suggests, we model small and large payments separately and have specific attention for the tail of each of the payment distributions. Splicing (see Klugman et al. (2012) and Panjer (2006)) allows the use of a body-tail approach where the body of the distribution is modeled separately from the tail. GAMLSS introduced by Stasinopoulos and Rigby (2007) allow for flexible modeling of the the body distribution while incorporating covariate information in the location, scale and shape parameters for a wide collection of distributions. We use the toolbox from EVT (see McNeil et al. (2005) and Beirlant et al. (2006)) to determine the threshold separating the body from the tail and to model the tail of the payment distributions. Whereas our previous contributions, Antonio and Plat (2014), Pigeon et al. (2014) and Godecharle and Antonio (2015), corrected the observed payments beforehand for inflation based on the Consumer Price Index (CPI), we now do not apply any discounting and capture inflation effects directly by including appropriate covariates in our model.

The paper is organized as follows. In Section 2 we describe the multi-state model for the development of a non-life insurance claim. Section 3 introduces the payment distribution model. We demonstrate our methodology in a case study on a data set from a European insurance company in Section 4 and end with a conclusion in Section 5.

2 A multi-state model for the development of a non–life insurance claim

2.1 Claim dynamics

Consider the non-life claim for which the development is illustrated in Figure 1. Insurance companies distinguish three types of claims depending on how far a claim is in the development

process at the moment of evaluation $s^\star$. For an 'IBNR' or 'Incurred But Not Reported' claim a claim event has happened, but the insurer is not aware of it yet at the moment of evaluation (i.e. $s^\star \geq s_{oc}$ and $s^\star < s_0$). We call a claim 'RBNS' or 'Reported But Not Settled' when the insurer is aware of the claim, but the claim is not closed yet (i.e. $s^\star > s_0$ and $s^\star \leq s_c$). Lastly, a claim is closed when we have observed its complete development at the moment of evaluation (i.e. $s^\star > s_c$). The insurer has to set a reserve for both the IBNR and RBNS claims. In this paper, we present a framework to evaluate the RBNS reserve. We refer to Pigeon et al. (2013), Pigeon et al. (2014) and Antonio and Plat (2014) for a method to evaluate the IBNR reserve.

2.2 The multi-state approach

We model the development of a non-life insurance claim as a sequence of events using the multi-state model $(\mathscr{S}, \mathscr{T})$ in Figure 2 with state space $\mathscr{S}$ and set of direct transitions $\mathscr{T}$ (Haberman and Pitacco (1999) and Denuit and Robert (2007)).

At occurrence the claim starts its development in state S_{oc}. Afterwards it is reported to the insurance company, corresponding to a transition to the reporting state $S_{re} := S_0$. Once reported, a first payment can occur, implying a transition from state S_0 to state S_1. Just like S_{oc} and S_0, the states S_j $(j \in \{1, \ldots, n_{p\max} - 1\})$ are strictly transient, i.e. the claim can leave the state but not re-enter. Index j refers to the number of payments made in the past; thus a transition to state S_j represents the j^{th} payment of a claim. A claim in such a state can move on to another strictly transient state S_{j+1} or to one of the absorbing states, S_{tn} or S_{tp}, which are impossible to leave once entered. Index tn stands for *'Terminal, No payment'*, a transition to S_{tn} means the claim closes without payment. The claim closes with a payment in case of a transition to S_{tp} where tp stands for *'Terminal with Payment'*. The maximal number of payments throughout the development of a claim is denoted by n_{pmax}. Therefore, the only possible transitions from state $S_{n_{pmax}-1}$ are to the absorbing states (see Figure 2). A claim that moves from state S_0 directly to the absorbing state S_{tn} does not receive any payments.

The state space $\mathscr{S}$ is given by:

$$\mathscr{S} = \{S_{oc}, S_{re} \equiv S_0, S_1, S_2, \ldots, S_{n_{p\max}-1}, S_{tn}, S_{tp}\}. \tag{1}$$

An event refers to the transition from one state in $\mathscr{S}$ to another. These events include claim occurrence, reporting, the j^{th} payment $(j \in \{0, 1, \ldots, n_{p\max} - 1\})$ and closure with or without a payment. The set of direct transitions $\mathscr{T}$ defines all possible transitions in the multi-state model which are indicated by the arrows in Figure 2.

We detect the three types of claims discussed in Section 2.1. A claim in state S_{oc} is an IBNR claim because a claim event has happened, but the insurer is not aware of it as the claim did not make the transition to state S_0 yet. A claim in state S_j $(j \in \{1, \ldots, n_{p\max} - 1\})$ is an RBNS claim. The insurer is aware of the claim, but the claim is not settled as it did not reach an absorbing state yet. Hence, the transition from S_{oc} to S_0 terminates the IBNR part, but initializes the RBNS part in the development of a claim. A claim is closed when it reaches one of the absorbing states S_{tp} or S_{tn}. When a claim is in the RBNS part of the multi-state model in Figure 2 it faces competing risks (Klein and Moeschberger (2003), Pintilie (2006), Steele et al. (2006), Beyersmann et al. (2011) and Durrant et al. (2013)). A transition out of state S_j implies no turning back. Moreover, the transition out of this state into S_{j+1}, S_{tn} or S_{tp} are three competing events, as choosing one of the three for the latter transition excludes the possibility of choosing the other two events for the same transition.

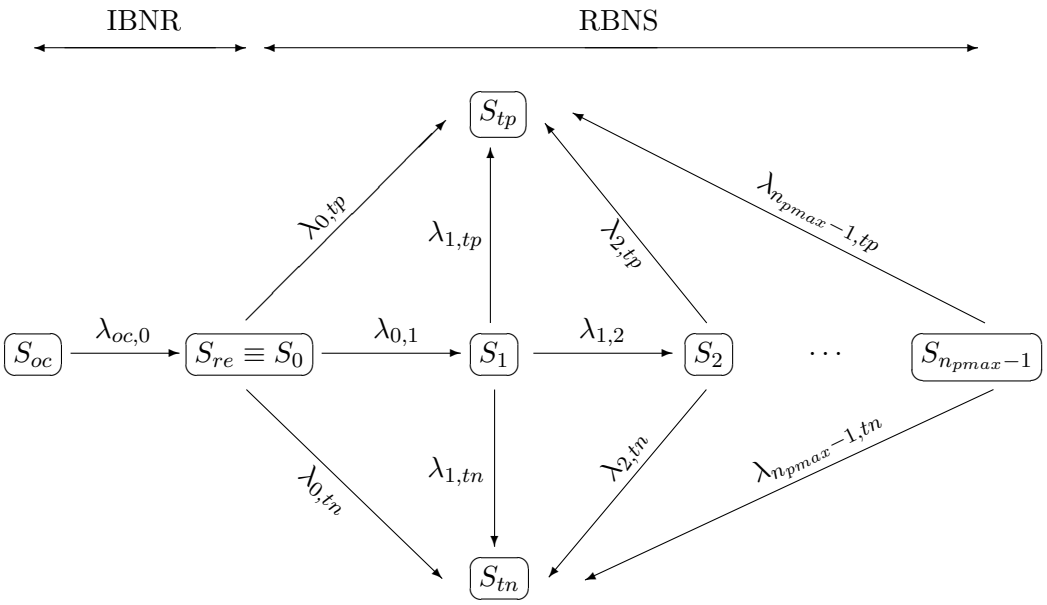

Figure 2: Multi-state model for the development of a non-life insurance claim: states and hazard functions driving transitions between states.

2.3 A multinomial logit model for discrete-time transitions

Time discretization As an insurer does not always record data from Figure 1's timeline in continuous time, we approach the multi-state model in Figure 2 in discrete time. We use calender years in this paper, but the framework is easily extendable to other time discretizations. Denote the occurrence year of a claim κ by $i(\kappa) \in \{0, 1, \ldots, n\}$. Hereby, $n+1$ is the number of occurrence years observed in the data set. We also record information on the development of the claim in discrete time and denote the development year by $t \in \{0, 1, \ldots n\}$ for which 0 corresponds to the occurrence year itself. To obtain a single payment per discrete time period, we aggregate all individual payments within the same time period (here: one calender year) into a single overall payment for that period. Define $S(\kappa, t) \in \mathscr{S}$ as the state occupied by claim κ at the end of development year t ($\in \{0, \ldots, n\}$). Then, $\mathcal{S} = \{S(\kappa, t) | t = 0, 1, \ldots, n\}$ is a discrete-time process describing the evolution of claim κ. As it is common for a claim to be reported and to receive its first payment within the same year, we allow the claim to transition from S_{oc} to S_0 and from S_0 to S_1 in the same development year. For later transitions we consider a single transition per time period.

We illustrate the time discretization and corresponding notation in Table 1 (see Section 4 for empirical illustrations using these data). Let us consider a data set where the observation period starts at 01/01/1997, making 1997 the first observed occurrence year. Therefore, claim κ originating in 1998 corresponds to occurrence year $i(\kappa) = 1$. Note that a double transition occurs in 1999: first the claim transitions from S_{oc} to S_0 as the claim is reported, followed by a first payment in the same year implying a transition from S_0 to S_1. A transition from S_1 to S_2 happens in 2001 and corresponds to the second payment in the development of this claim for an amount of €400 (in total). Finally, a third payment follows in 2002 and the claim closes, which translates to a transition from S_2 to S_{tp} in the multi-state framework. The transition into state S_{tp} corresponds to the third payment. The other notation used in Table 1 is clarified throughout the remainder of the text.

Event		**Date** *Continuous*	**Development period** t *Discrete*	**Payment number**	**Payment**	$S(\kappa, t)$	**Time to event** T
Occurrence		08/23/1998	0			S_{oc}	
Reporting		01/13/1999	1			S_0	$T_{oc} = 1$
Cash flow	€200	09/31/1999		1	$Y_1 = 200$	S_1	$T_0 = 0$
	€150	03/15/2001	3	2	$Y_2 = 400$	S_2	$T_1 = 2$
	€250	10/20/2001					
	€100	12/03/2002	4	3	$Y_3 = 100$	S_{tp}	$T_2 = 1$
Closing		12/13/2002					

Table 1: Illustration of the development of a claim κ as registered in continuous time (first three columns). Column 4 demonstrates the discrete time notation, columns 5 & 6 the payment number and corresponding payment amount, column 7 the stochastic process $S(\kappa, t)$ and column 8 the development year T_a at which the claim transitions out of a state S_a given that time is reset upon arrival in state S_a (Section 2.3 paragraph 'Multinomial logit model'). This claim is fictional.

Multinomial logit model We use the hazard function $\lambda_{a,b}$ to model the transition from state S_a to state S_b in the multi-state model in Figure 2. Let T_a denote the development year at which the claim moves out of state S_a. We reset time at each transition. T_a equal to zero implies the claim transitions into and out of state S_a in the same year, which is the case for state S_0 in the illustration in Table 1. As $T_{oc} = T_2 = 1$ in this example, the claim transitions out of respectively state S_{oc} and S_2 in development year 1 since entering the latter state. The claim transitions out of state S_1 in development year 2 since entry in this state, so $T_1 = 2$. The label δ_a tells us which of the competing risks the claim transitions to, where $\delta_a = b$ means a transition to state S_b. The hazard function from state S_a to S_b evaluated in τ,

$$\lambda_{a,b}(\tau) = P\left(T_a = \tau,\ \delta_a = b | T_a \geq \tau\right), \qquad \tau \in \{0, 1, \ldots, n\} \tag{2}$$

is the probability the claim transitions from S_a into S_b in year τ, given that the transition out of state S_a did not happen in the years before τ.

We allow the hazard functions to depend on covariate information and use a multinomial logit framework to model the discrete-time hazard functions (Allison (1982) and Beyersmann et al. (2011), Chapter 7):

$$\lambda_{a,b}(\tau | \boldsymbol{x}_{a,\kappa}(\tau)) = \frac{\exp\{\alpha_{a,b}(\tau) + \boldsymbol{\beta}'_{a,b}\boldsymbol{x}_{a,\kappa}(\tau)\}}{1 + \sum_l \exp\{\alpha_{a,l}(\tau) + \boldsymbol{\beta}'_{a,l}\boldsymbol{x}_{a,\kappa}(\tau)\}} \tag{3}$$

where $\boldsymbol{x}_{a,\kappa}(\tau)$ is the (possibly) time-dependent covariate information for claim κ in year τ since transition into state S_a, $\alpha_{a,b}(\tau)$ and $\boldsymbol{\beta}_{a,b}$ are the regression parameters and the sum over l in the denominator runs over all possible states claim κ can directly transition into from state S_a. For each transient state S_a, the parameters used in the discrete-time hazard functions $\lambda_{a,b}(\tau)$ are estimated simultaneously for all possible subsequent states S_b using maximum likelihood[1].

[1] We fit this model using the function 'multinom' from the `nnet` `R` package introduced by Venables and Ripley (2002).

We hereby treat all observed time periods of the individual claims as separate, independent observations.

Without additional covariate information $\boldsymbol{x}_{a,\kappa}(\tau)$, the specification in (3) simplifies to the Nelson-Aalen estimator (Nelson (1969), Aalen (1976)):

$$\hat{\lambda}_{a,b}(\tau) = \frac{d_{a,b}(\tau)}{n_a(\tau)} \tag{4}$$

with $n_a(\tau)$ the number of claims in state S_a at the start of the τ^{th} year since transition into state S_a and $d_{a,b}(\tau)$ the number of these claims transitioning into state S_b during year τ.

3 A flexible body-tail model for the payment distribution

In the second part of our internal model, we model the payments in the development of a claim as visualized in Figure 1. We aggregate a claim's intermediate payments (at s_1, ..., s_4 in Figure 1) within the same calendar year and obtain a single payment for that period. We label the payments consequently with a payment number, i.e. the first, second, ... payment in the development of a claim. A separate model is built for each payment number. Let Y_j ($j \in \{1, \ldots, n_{p_{\max}}\}$) denote the j^{th} payment of a random claim. For the example discussed in Section 2.3, Table 1 illustrates these payments in the fifth column. We need a model for Y_j that is flexible, allows for the inclusion of claim-specific covariate information and correctly captures the skewness of the right tail. In the context of micro-level reserving Antonio and Plat (2014) include covariate information in the location and scale parameter of a lognormal distribution for intermediate payments which are not stratified by payment number. Distributional models for payments are also highly relevant in insurance pricing. Frees and Valdez (2008) model payments with a Generalized Beta of the second kind (GB2) distribution, with four parameters, including covariate information. Klein et al. (2014) use Generalized Additive Models for Location, Scale and Shape (GAMLSS) in a Bayesian framework for their observed claim severities, and specifically investigate the use of zero-adjusted versions of the gamma, inverse-Gaussian and lognormal distributions where covariate information is included in three parameters (location and shape or scale parameter, as well as probability of a claim). Verbelen et al. (2015) propose mixtures of Erlangs as a flexible, yet tractable tool for loss modeling, while accounting for truncation and/or censoring. A model that fits the attritional losses (i.e. the 'small' payments, also called the body of the loss distribution) does not necessarily capture large payments well. Pigeon and Denuit (2011) consider a composite lognormal-Pareto model to capture both attritional and large losses. EVT (McNeil (1997), McNeil et al. (2005) and Beirlant et al. (2006)) suggests the use of the Generalized Pareto Distribution (GPD) to model the tail of the loss distribution, i.e. the losses above a certain, high threshold. As we want a flexible model for both attritional and large payments, we naturally opt for a global loss model, obtained as a spliced distribution with a body – below the threshold – and a tail – above the threshold – component. This way, we allow the density function of Y_j to be a spliced distribution with two components (see Klugman et al. (2012), Panjer (2006) and Peters and Shevchenko (2015), Nadarajah and Bakar (2014), Aue and Kalkbrener (2006) for examples in modeling operational risk data):

$$f_{Y_j}(y) = \begin{cases} p_{j,1} \cdot f_{j,1}(y), & \text{if } 0 < y \leq u_j \\ p_{j,2} \cdot f_{j,2}(y), & \text{if } u_j < y. \end{cases} \tag{5}$$

$f_{j,1}(y)$ is a well-defined density function on the interval $(0, u_j]$. The tail of the distribution is treated separately by a well-defined density function $f_{j,2}(y)$ on the interval (u_j, ∞). We call u_j the threshold which separates the body of the data from the tail, or the attritional losses from the large losses. The support of the body of the distribution is therefore $(0, u_j]$ whereas the support of the tail is given by (u_j, ∞). $p_{j,1}$ is the probability (or weight) that Y_j pertains to the body of the distribution whereas $p_{j,2}$ expresses the probability of Y_j belonging to the tail of the distribution, hence $p_{j,1} + p_{j,2} = 1$. We describe the choice of the splicing thresholds u_j, components $f_{j,1}$ and $f_{j,2}$, probabilities $p_{j,1}$ and $p_{j,2}$ and the inclusion of claim specific covariate information in the remainder of this section.

3.1 Modeling the tail: threshold selection and GPD fit

Let F_j denote the cdf of Y_j, the j^{th} payment in the discretized development of a claim. We use the excess distribution over a threshold u to model the tail of F_j:

$$F_{ju}(y) := P(Y_j - u \leq y | Y_j > u) = \frac{F_j(y+u) - F_j(u)}{1 - F_j(u)}. \tag{6}$$

We determine the optimal threshold u_j from which the tail of Y_j starts by state of the art techniques from EVT (McNeil (1997), McNeil et al. (2005) and Beirlant et al. (2006)). The selected threshold, u_j, is included in (5) as the threshold separating the body from the tail of the distribution.

Balkema and de Haan (1974) and Pickands III (1975) prove that the excess distribution of common continuous distribution functions used in loss modeling converges to a Generalized Pareto Distribution (GPD) if the threshold u is high enough. Following their theorem, we assume $F_{ju}(y) = G(y)$ for some high threshold u where G is the GPD, defined as

$$G(y) = \begin{cases} 1 - \left(1 + \frac{\gamma y}{\sigma}\right)^{\frac{-1}{\gamma}}, & y \in (0, \infty) \quad \text{if } \gamma > 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right), & y \in (0, \infty) \quad \text{if } \gamma = 0 \\ 1 - \left(1 + \frac{\gamma y}{\sigma}\right)^{\frac{-1}{\gamma}}, & y \in (0, -\frac{\sigma}{\gamma}) \quad \text{if } \gamma < 0 \end{cases} \tag{7}$$

where γ is the Extreme Value Index (EVI) and σ is a scale parameter. The selection procedure for the optimal threshold u_j depends on the sign of the EVI. Graphical tools to determine this sign include the mean excess plot and the exponential QQ plot (see Section 4.1 in McNeil (1997), Section 7.2 in McNeil et al. (2005) and Section 1.2 in Beirlant et al. (2006)).

For data with a GPD tail, the mean excess plot[2] becomes increasingly linear. The trend in the mean excess plot gives an indication of the sign of the EVI: a downward trend is associated with

[2]The mean excess function of a random variable Y with finite mean, is defined as

$$e(u) = E\left(Y - u | Y > u\right). \tag{8}$$

The mean excess function is empirically estimated as

$$\hat{e}_{n_Y}(u) = \frac{\sum_{i=1}^{n_Y} y_i 1_{(u,\infty)}(y_i)}{\sum_{i=1}^{n_Y} 1_{(u,\infty)}(y_i)} - u \tag{9}$$

where y_i is the i^{th} observation and n_Y the number of observations of Y. The mean excess plot is given by $\{Y_{n_Y-k,n_Y}, \hat{e}_{n_Y}(Y_{n_Y-k,n_Y}), 1 \leq k \leq n_Y - 1\}$ where Y_{n_Y-k,n_Y} denotes the $(n_Y - k)^{\text{th}}$ order statistic or the $(k+1)^{\text{th}}$ largest observation of Y.

$\gamma < 0$, a horizontal trend with $\gamma = 0$ and an upward trend with $\gamma > 0$[3].

We illustrate this approach with the observations on the first payment, i.e. Y_1, from the case study presented in Section 4. Figure 3 shows the empirical mean excess plot (left) and the exponential QQ plot (right). The increasing empirical mean excess function indicates a heavy-tailed distribution (i.e. $\gamma > 0$), which is confirmed by the convex shape of the exponential QQ plot. We therefore conclude that the data used in this illustration are heavy-tailed ($\gamma > 0$). When modeling insurance losses (or: payments in a claim's run-off) we rarely observe the other cases in (7), i.e. $\gamma = 0$ or $\gamma < 0$.

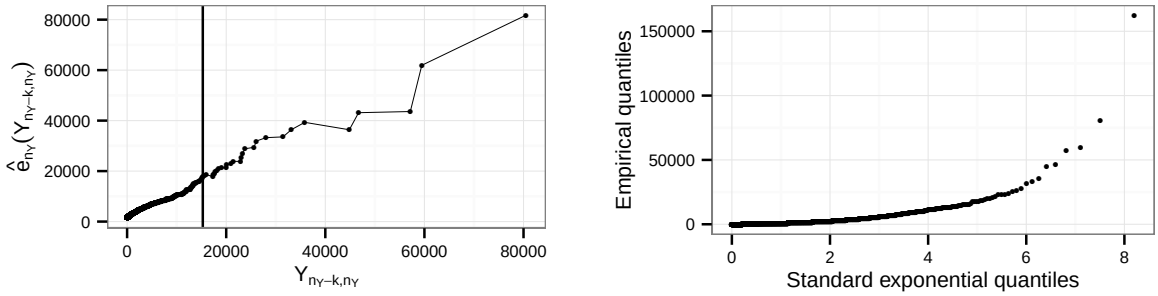

Figure 3: Graphical tools to determine the sign of the EVI γ: sample mean excess plot (*left*) where the vertical line displays the selected threshold u_j, and the exponential QQ plot (*right*). We illustrate the tools on first payment (Y_1) observations for the data set from Section 4

Threshold selection In the $\gamma > 0$ case, we determine the optimal threshold u_j using the Hill (Section 7.2.4 in McNeil et al. (2005) and Section 4.2 in Beirlant et al. (2006)), Zipf (Section 4.3 in Beirlant et al. (2006)) and second order Hill estimator (Section 4.5.1 in Beirlant et al. (2006)) for the EVI γ in (7). We plot these estimators as a function of the number of tail observations taken into account, denoted by k, and look for a stable region in the graphs of these three γ-estimators. Figure 4 (top) visualizes the resulting estimates on the Y_1 data from our case study. We choose the threshold u_1 to be €15,273.90 (the 30^{th} largest observation of Y_1), indicated with the vertical line in the graph, at the point where the Hill and the second order Hill estimator meet. For other threshold selection methods we refer to Beirlant et al. (2006) and Scarrott and MacDonald (2012).

Fitting the GPD tail Given the selected threshold u_j, we use maximum likelihood estimation to determine the parameters in the GPD distribution for the exceedances above this threshold[4]. Figure 5 shows the Peaks Over Threshold (POT) plot for the observations of Y_1 and the selected threshold. The horizontal line in this plot corresponds to the selected threshold u_1, whereas the height of each bar represents an observation on the first payment Y_1. For a payment above the selected threshold u_1, the length of the peak over the threshold represents the exceedance of the payment above u_1. The payments are grouped and color-coded per calendar year. We examine the goodness of fit of the GPD distribution using a QQ and PP plot as in Figure 4 (bottom).

[3]The last observations in the mean excess plot average over a small number of excesses. As these values can distort the graph of the mean excess plot, we often do not consider these.

[4]The `gpd.fit` function from the `ismev` package allows for this estimation in `R`.

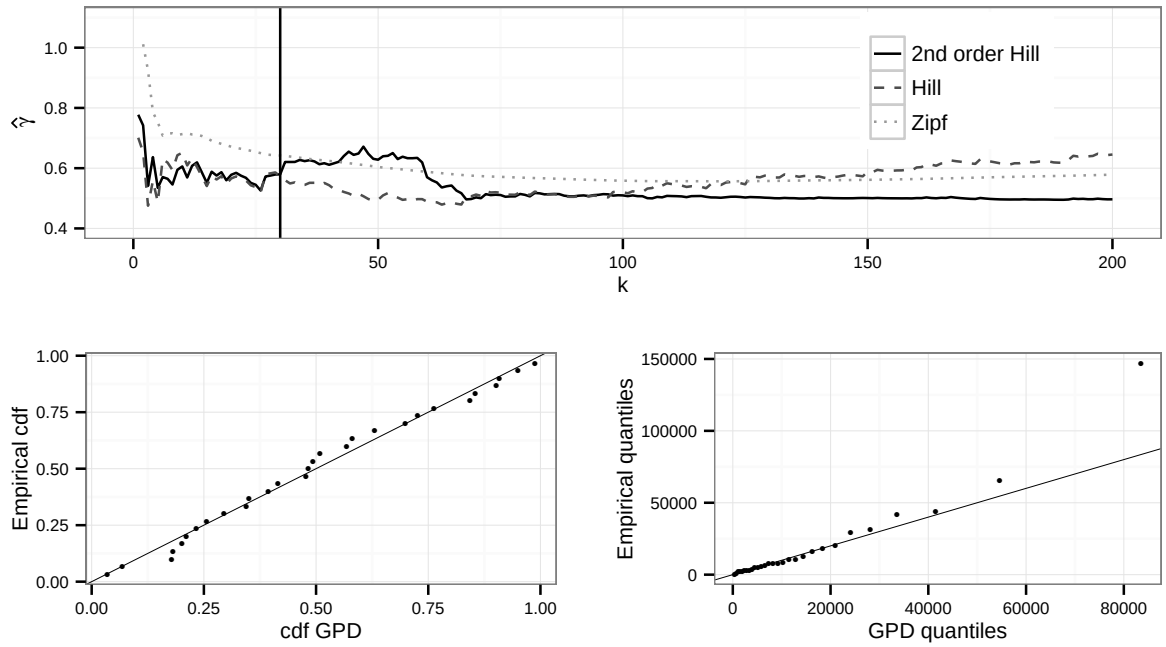

Figure 4: Graphical tools to determine the optimal threshold u_j separating the body of the distribution from its tail: Hill, Zipf and second order Hill plot for the 200 largest observations (*top*) where the vertical line displays the selected threshold u_j. The PP plot (*bottom, left*) and the QQ plot (*bottom, right*) of the GPD fit for exceedances above selected threshold u_j. We illustrate the tools on the first payment (Y_1) observations for the data set from Section 4.

These plots underline the good fit for the tail of the data observed on the first payment number, Y_1.

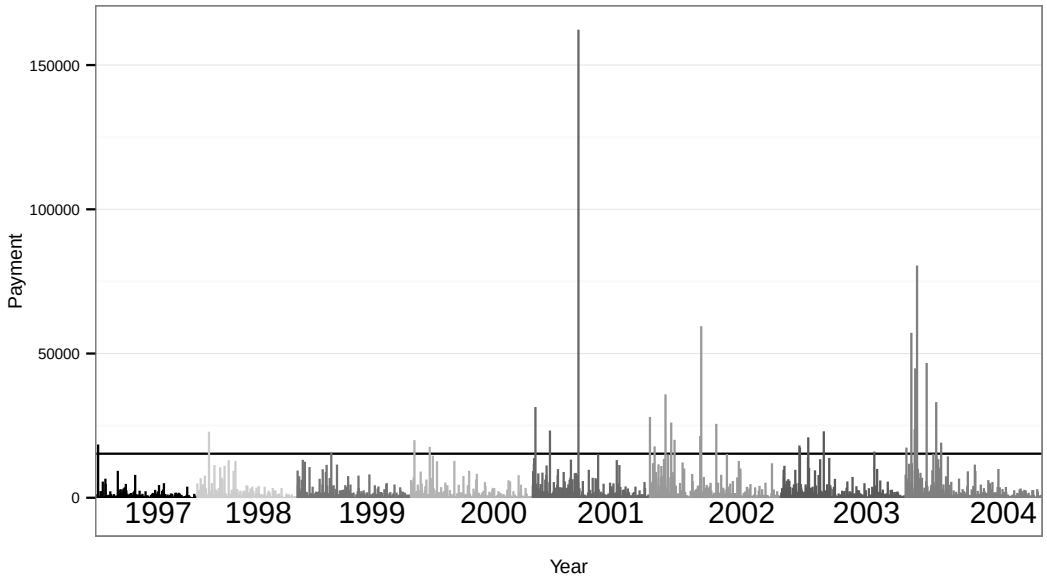

Figure 5: The exceedances over the selected threshold u_j as a Peaks Over Threshold (POT) plot: the length of each bar represents the height of an Y_j observation, grouped and color-coded per calendar year. The horizontal line corresponds to the selected threshold u_j. We illustrate the tool on the first payment (Y_1) observations for the data set from Section 4

3.2 Modeling the body: a GAMLSS approach

We use a parametric GAMLSS fit (Stasinopoulos and Rigby (2007)) in the body distribution, denoted by $f_{j,1}(y)$ in (5). These flexible models extend the framework of Generalized Linear Models (GLM), widely used in actuarial tarification and reserving (De Jong and Heller (2008), Kaas et al. (2008) and Ohlsson and Johansson (2010)), by not restricting the distribution of the response variable to the exponential family. Moreover, they allow the inclusion of covariate information in up to four model parameters. Recent research demonstrates the usefulness of these models in a wide range of applications. For example to model mortgage loan losses as in Tong et al. (2013), to estimate measures of market risk as in Scandroglio et al. (2013), to analyze insurance data as in Klein et al. (2014) and income data as in Klein et al. (2015).

We investigate the GAMLSS model specifications where the density function $f_{j,1}(y|\boldsymbol{x}_\kappa(t))$ is conditional on at most four distribution parameters $\boldsymbol{\theta}_\kappa = (\theta_{1\kappa}, \theta_{2\kappa}, \theta_{3\kappa}, \theta_{4\kappa}) = (\mu_\kappa, \sigma_\kappa, \nu_\kappa, \tau_\kappa)$. The first two parameters μ_κ and σ_κ in the density of claim κ are usually referred to as respectively location and scale parameter, whereas the remaining parameters, if present, are additional shape parameters. Notice that we drop the subscript $j, 1$ (see (5)) for these parameters to ease notation. We incorporate claim specific covariates in a linear way, namely $g_r(\theta_{r\kappa}) = \boldsymbol{\beta}_r'\boldsymbol{x}_\kappa$ where the parametric link function g_r for $r \in \{1, 2, 3, 4\}$ expresses the relation between the systematic component $\boldsymbol{\beta}_r'\boldsymbol{x}_\kappa$ and the distribution parameter $\theta_{r\kappa}$.

Let $\boldsymbol{x}_\kappa(t)$ be the vector with (possibly) time-dependent covariate information of claim κ in development year t. Given the covariate information, we model $f_{j,1}(y|\boldsymbol{x}_\kappa(t))$ with a truncated parametric GAMLSS model since the observation on the j^{th} payment in the development of a claim is left truncated by 0 and right-truncated by u_j. We examine the goodness of fit of a

selection of distributions commonly used in loss modeling that are available within R's GAMLSS package[5]. We use the AIC (see Akaike (1974)) to choose the preferred GAMLSS distribution for $f_{j,1}$.

3.3 Probability of belonging to the body or tail of a payment distribution

The probabilities $p_{j,r}(x_\kappa(t))$ in (5) depend on covariate information $x_\kappa(t)$ of claim κ in development year t. We use a binomial logit framework to model these probabilities:

$$\begin{aligned} p_{j,2}(\boldsymbol{x}_\kappa(t)) &= \frac{\exp\{\alpha + \boldsymbol{\beta}'\boldsymbol{x}_\kappa(t)\}}{1+\exp\{\alpha + \boldsymbol{\beta}'\boldsymbol{x}_\kappa(t)\}} \\ p_{j,1}(\boldsymbol{x}_\kappa(t)) &= 1 - p_{j,2}(\boldsymbol{x}_\kappa(t)). \end{aligned} \tag{10}$$

where α and $\boldsymbol{\beta}$ are the model parameters which we estimate by maximum likelihood[6]. $p_{j,1}(\boldsymbol{x}_\kappa(t))$ and $p_{j,2}$ represent the probability of belonging respectively to the body or tail of the distribution.

4 Case study

We demonstrate the methodology on a data set from a European insurance company. The data consist of the bodily injury (BI) claims arising from general liability insurance contracts. Within the actuarial literature, Antonio and Plat (2014), Pigeon et al. (2013), Pigeon et al. (2014) and Godecharle and Antonio (2015) work with the same data set. These papers first discount payments using the relevant Consumer Price Index (CPI) and then model discounted payments. We choose not to correct the observed payments for inflation beforehand but instead want to capture inflation effects from the data at hand, using appropriate covariate information in our regression models. We split the data set in a training (January 1997 – December 2004) and a validation data set (January 2005 – August 2009) and we discretize time using annual periods starting from 01/01/1997 and running until 31/12/2004. As such we have eight full years of observations in the training data set. At the moment of evaluation, i.e. at the end of the day at 31/12/2004, the training data set contains 4,483 claims of which 3,452 are closed. The 1,031 remaining claims are RBNS claims for which we will simulate the reserve. One full simulation of the RBNS reserve consists of two steps. Starting from the multi-state model as depicted in Figure 2, we first use the estimated hazard functions from Section 2 to complete the stochastic process $S(\kappa, t)$ for each RBNS claim κ by simulation. Define the maximum number of development years observed in the data set by n. In this application, when a claim reaches development year $t = n = 8$ and the claim is still open, we force a transition to closure without payment in the next development year such that $S(\kappa, 9) = S_{tn}$, limiting the development of a claim in time. As a consequence $n_{p_{\max}} = 9$ in the multi-state model represented by Figure 2. Second, we use the calibrated distributions from Section 3 to simulate a payment for each simulated transition involving a payment. We incorporate the policy limit of €2.5 million from policy conditions underneath our data set. We construct the distribution for the RBNS reserve by repeating this two-step simulation process.

[5] The `gamlss.tr` package in R allows for maximum likelihood parameter estimation of a GAMLSS distribution left-truncated by 0 and right-truncated by u_j.

[6] The `nnet` R package introduced by Venables and Ripley (2002) allows to compute the maximum likelihood estimators of these parameters.

4.1 The multi-state model: estimation and simulation

Summary statistics Table 2 shows summary statistics on the number of claims transitioning out of state S_0 (left) and state S_1 (right). Similar tables can be constructed for states S_j ($j \in \{2, \ldots, 8\}$). We recall the notation introduced in Section 2.3: $n_a(\tau)$ is the number of claims in state S_a at the start of the τ^{th} year since transition into state S_a and $d_{a,b}(\tau)$ is the number of these claims transitioning into state S_b during year τ. The top left cells in each table show the number $n_0(0)$ and $n_1(0)$ of claims we observe entering respectively state S_0 and S_1. Note that $n_1(0)$ consists of the claims that entered state S_0 (i.e. $n_0(0)$) and did not transition into an absorbing state nor were censored before transitioning out of state S_0. Our model does not allow a claim to transition out of state S_j in the year of entrance into this state when $j \geq 1$. As a consequence, $d_{1,2}(0) = d_{1,tp}(0) = d_{1,tn}(0) = 0$. We do allow a claim to transition out of state S_0 in the same year this state was entered, explaining the non-zero values for $d_{0,1}(0), d_{0,tp}(0), d_{0,tn}(0)$ in the left table. From the claims that entered state S_0, 1,705 received a payment which did not close the claim (i.e. transition to S_1), 1,028 received a payment which closed the claim (i.e. transition to S_{tp}) and 309 claims closed without a payment (i.e. transition to S_{tn}) within the same year, $\tau = 0$, of entrance into state S_0. The columns corresponding to $\tau = 1, 2$ and ≥ 3 (left) and $\tau = 1, 2, 3$ and ≥ 4 display these same numbers for claims that have been one, two and at least three years in state S_0, respectively state S_1. To make sure we have enough observations for each value of τ, we do not distinguish values of τ beyond 3 for S_0 and beyond 4 for S_0, but consider these as a single group.

τ	0	1	2	≥ 3
$n_0(\tau)$	4,483	1,185	151	83
$d_{0,1}(\tau)$	1,705	388	36	11
$d_{0,tp}(\tau)$	1,028	439	36	12
$d_{0,tn}(\tau)$	309	177	15	12

τ	0	1	2	3	≥ 4
$n_1(\tau)$	2,140	1,834	242	84	82
$d_{1,2}(\tau)$	0	635	41	8	10
$d_{1,tp}(\tau)$	0	540	57	12	3
$d_{1,tn}(\tau)$	0	381	37	14	13

Table 2: Summary statistics on the transition out of S_0 (left) and out of S_1 (right).

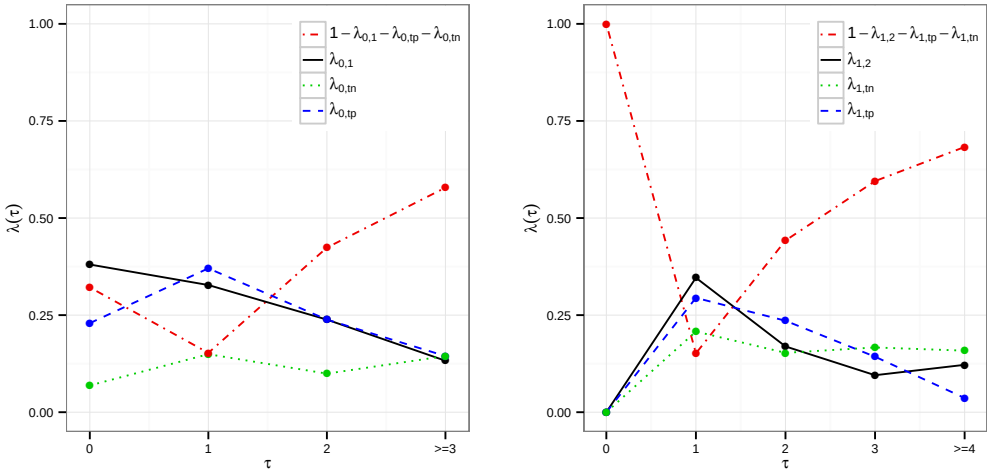

Figure 6: *Left*, with τ the time since entry in state S_0: $\hat{\lambda}_{0,1}(\tau)$ (*full black line*), $\hat{\lambda}_{0,tn}(\tau)$ (*dotted green line*), $\hat{\lambda}_{0,tp}(\tau)$ (*dashed blue line*) and the probability of not making a transition $1-\hat{\lambda}_{0,1}(\tau)-\hat{\lambda}_{0,tn}(\tau)-\hat{\lambda}_{0,tp}(\tau)$ (*dotted-dashed red line*). *Right*, as a function of time since entry τ in state S_1: $\hat{\lambda}_{1,2}(\tau)$ (*full black line*), $\hat{\lambda}_{1,tn}(\tau)$ (*dotted green line*), $\hat{\lambda}_{1,tp}(\tau)$ (*dashed blue line*) and the probability of not making a transition $1-\hat{\lambda}_{1,2}(\tau)-\hat{\lambda}_{1,tn}(\tau)-\hat{\lambda}_{1,tp}(\tau)$ (*dotted-dashed red line*).

Estimation results We use the multinomial logit model from (3) without additional covariate information $\boldsymbol{x}_{a,\kappa}(\tau)$, resulting in the Nelson-Aalen estimator (4), to estimate the hazard functions $\lambda_{a,b}(\tau)$. We leave the inclusion of covariate information as a subject for future research. To illustrate our results, Figure 6 visualizes the estimated hazard functions for transitions out of state S_0 and S_1.

Simulation of future paths for RBNS claims The stochastic process $S(\kappa,t)$ describes the movement of a claim κ through the multi-state model as time evolves. For an RBNS claim, this process did not yet reach an absorbing state at the moment of evaluation. We simulate the further run-off of such an RBNS claim by simulating its future path – driven by estimated hazard functions $\hat{\lambda}_{a,b}$ – until $S(\kappa,t)$ reaches an absorbing state or until t reaches the maximum number n of observed development years. The simulation procedure for claim κ starts with determining the last state this claim occupies in its development process $S(\kappa,t)$, say S_a, and the first unobserved development period τ since entrance into this state. We then select the relevant hazard function values $\hat{\lambda}_{a,b}(\tau)$ for all possible transitions to some state b out of state S_a in development period τ. Given these probabilities we simulate what happens in period τ:

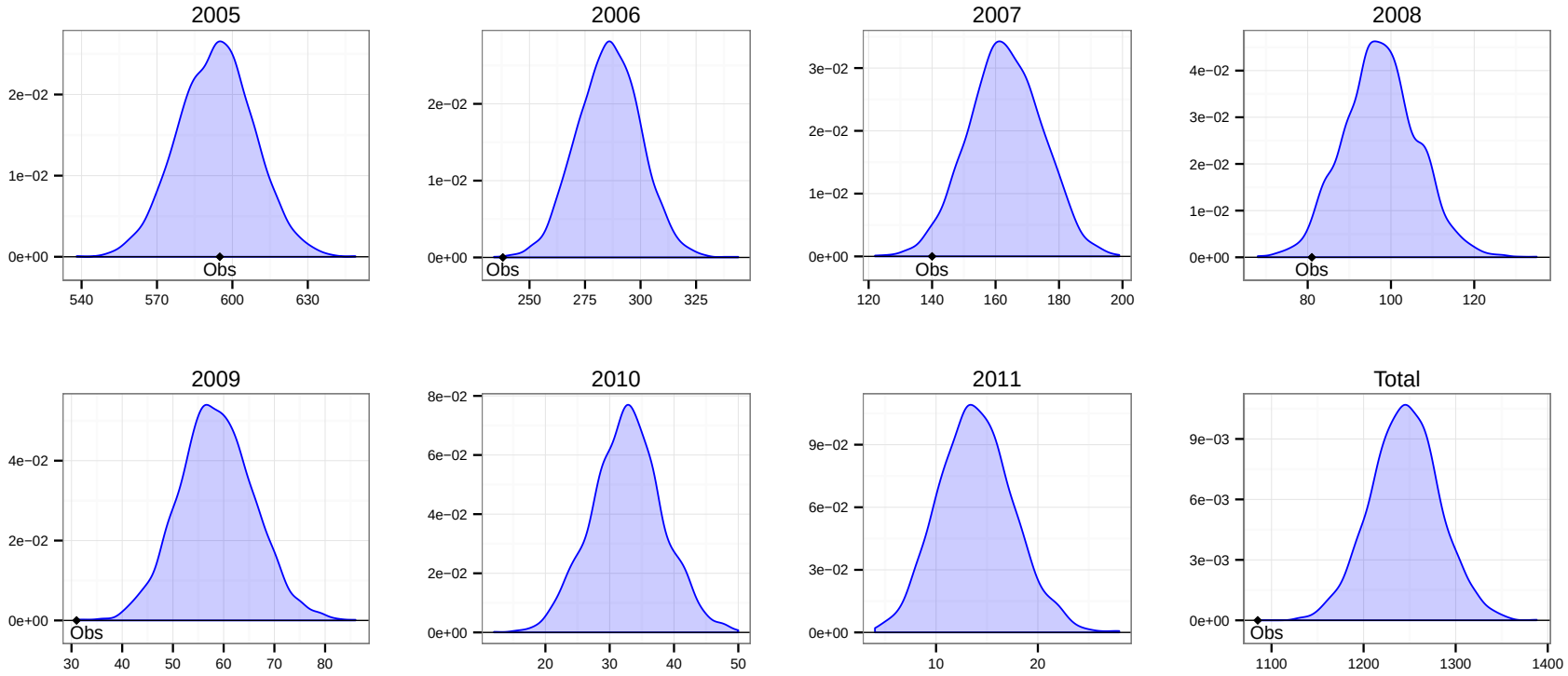

Figure 7: Distribution for the number of payments from the RBNS claims per calendar year obtained by the simulation procedure (2,000 simulations) based on the estimated hazard rates. The point labeled 'Obs' is used to indicate number of payments observed in the validation data set of RBNS claims. Note that the validation data set contains observations until August 2009. As a consequence, the number of payments observed in 2009 is incomplete and we do not observe payments in 2010 and 2011. The total number of payments is therefore also incomplete.

a non-terminal payment transition (to state S_{a+1}), a terminal payment transition (S_{tp}), closure without payment (S_{tn}) or no transition at all (S_a). In case the simulated state is an absorbing one, the stochastic process $S(\kappa, t)$ ends here as the claim closes. If the development period t reaches the maximum number $n = 8$ of observed development years without the claim attaining an absorbing state, we close the claim by a transition to state S_{tn} in the following period. When the claim did not reach an absorbing state and $t < n$, we repeat the procedure and simulate the next state in the stochastic process of the open claim.

For each RBNS claim observed in the training data set, we simulate its future path through the multi-state model. By repeating this procedure, we can construct empirical distributions of quantities of interest using the simulated paths. For example, Figure 7 shows the empirical distribution of the number of payments per calendar year (i.e. 2005, 2006, and so on). Table 3 shows summary statistics of these empirical distributions per calendar year. We simulate 2,000 paths and compare the simulated number of payments to the actual observed number of payments in the validation data set. Note that the year 2009 is only observed until August in the validation data set, making the observed number for this calendar year incomplete.

Calendar year	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**Total**
Minimum	538	234	122	68	31	12	4	1,129
25% quantile	583	276	155	92	54	29	11	1,220
Median	594	286	163	97	58	33	14	1,245
Mean	594	286	163	97.5	58.6	32.7	13.9	1,245
75% quantile	603	295	171	103	63	36	16	1,269
95% quantile	618	309	182	112	71	42	20	1,307
Maximum	649	344	199	135.0	86	50	28	1,388
Observed	595	238	140	81	≥ 31	NA	NA	$\geq 1,085$

Table 3: Summary statistics for the number of payments from the RBNS claims per calendar year obtained by the simulation procedure (2,000 simulations) based on the estimated hazards. The row labeled 'Observed' shows the number of payments observed in the validation data set of RBNS claims for the corresponding calendar year.

Figure 7 and Table 3 show the empirical distribution captures the number of payments observed in the validation data set in calendar year 2005 very well. For calendar year 2006, 2007 and 2008 the observed value lies more in the left tail of the distribution. Results may probably be improved by introducing covariate information in the hazard functions, as suggested in Section 5. The observed number of payments in 2009 and in total are also in the left tail of the empirical distribution. Note, however, that we only observe 2009 until August in the validation data set, making these observed numbers incomplete. As a consequence, we do not observe any payments in 2010 and 2011 and the total observed number is incomplete as well.

4.2 Flexible payment distributions: estimation and simulation

Descriptive statistics Table 4 summarizes the empirical distribution per payment number as observed in the training data set. Recall from Section 3 that we stratify the distribution of non-zero payments in a claim's run-off based on payment number and use notation Y_j for the j^{th} payment during the development of a claim.

We construct the spliced density function of Y_j (see (5)) per payment number j, following the procedure outlined in Section 3. Because we have very few payment observations for high

payment numbers, in particular: payment numbers 5–8 in this data set, we fit one body-tail model on the collection of these observations. We denote the payment distribution of these payment numbers by f_{Y_5} and the corresponding random variable by Y_5.

Payment	Y_1	Y_2	Y_3	Y_4	Y_5
♯ observed	3,655	1,306	399	154	98
Minimum	6.58	8.09	11.60	26.32	30.00
Mean	1,419.24	3,562.92	6,446.61	12,075.07	12,615.79
Median	442.44	1,130.04	1,956.31	3,624.73	5,000.00
95% quantile	5,530.02	14,225.36	25,705.87	39,165.62	41,707.68
Maximum	162,274.50	135,583.00	230,137.40	394,788.80	234,509.99

Table 4: Training data set, with occurrence years 1997–2004: summary statistics of the empirical distribution per payment number.

Tail of the distribution We apply the strategy from Section 3.1 to all payment number distributions. Graphics illustrating this strategy were shown in Section 3.1. We apply the same strategy to Y_2, Y_3, Y_4 and Y_5. In all cases we observe an increasing mean excess plot and convexity of the exponential QQ plots (cfr. our illustration in Figure 3 with data observed on Y_1). Thus, we conclude that Y_1, Y_2, Y_3, Y_4 and Y_5 are heavy-tailed. Table 5 shows the selected thresholds. For example, the threshold splicing the distribution of Y_2 is $u_2 = €12{,}320.11$ which corresponds to the 83^{rd} largest observation. The GPD fit on the tail of $Y_1, \ldots, Y_5$ is verified by means of a QQ and a PP plot. Table 6 shows the corresponding parameter estimates and their standard error (in brackets). Note that the standard errors for the highest two payment numbers (Y_4 and Y_5) becomes large as a result of the low number of observations included in the fit.

	u_1	u_2	u_3	u_4	u_5
Threshold	€15,273.90	€12,320.11	€19,129.6	€28,815.35	€29,226.73
Order statistic	30	83	32	12	9

Table 5: Optimal splicing thresholds u_j for $j \in \{1, 2, 3, 4, 5\}$ together with their corresponding order statistic.

	$f_{1,2}$	$f_{2,2}$	$f_{3,2}$	$f_{4,2}$	$f_{5,2}$
$\hat{\gamma}$	0.46 (0.27)	0.51 (0.17)	0.51 (0.26)	0.54 (0.36)	0.49 (0.74)
$\hat{\sigma}$	10,068.95 (3,258.57)	7,416.94 (1,420.20)	12,633.43 (3,229.18)	24,151.34 (7,844.73)	25,760.67 (28,345.51)

Table 6: Parameter estimates and their standard error (*in brackets*) for $f_{\cdot,2}$ in equation (5) with the GPD distribution given by (7).

			$f_{1,1}$		$f_{2,1}$		$f_{3,1}$		$f_{4,1}$		$f_{5,1}$	
			μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
		$\hat{\alpha}$	5.54 (0.07)*	0.38 (0.03)*	6.76 (0.14)*	0.62 (0.05)*	7.94 (0.26)*	0.66 (0.10)*	11.33 (0.90)*	1.30 (0.16)*	14.96 (0.95)*	1.50 (0.18)*
t	1	$\hat{\beta}_1$	0.73 (0.06)*	0.16 (0.03)*								
	2	$\hat{\beta}_2$	1.68 (0.16)*	0.30 (0.06)*	0.33 (0.14)*	0.22 (0.05)*						
	3	$\hat{\beta}_3$	2.01 (0.21)*	0.11 (0.09)	0.74 (0.25)*	0.11 (0.09)	-0.09 (0.22)	-0.07 (0.09)				
	4	$\hat{\beta}_4$			0.16 (0.31)	0.01 (0.12)	-0.34 (0.47)	0.46 (0.12)*	2.76 (0.57)*	0.36 (0.12)*		
	5	$\hat{\beta}_5$									-0.75 (0.58)	-0.02 (0.15)
i	1	$\hat{\delta}_1$	0.09 (0.09)	-0.09 (0.05)$^\circ$	0.02 (0.19)	-0.11 (0.07)	-0.17 (0.35)	-0.05 (0.12)	0.19 (1.15)	-0.01 (0.19)	-6.15 (1.00)*	-0.55 (0.20)*
	2	$\hat{\delta}_2$	0.33 (0.10)*	-0.07 (0.05)	0.18 (0.19)	-0.08 (0.07)	-0.84 (0.33)*	-0.13 (0.13)	-3.94 (0.95)*	-0.71 (0.19)*	-5.62 (0.96)*	-0.66 (0.19)*
	3	$\hat{\delta}_3$	0.37 (0.09)*	-0.07 (0.04)	0.16 (0.19)	-0.04 (0.07)	-0.17 (0.32)	-0.11 (0.12)	-1.51 (0.99)	-0.28 (0.18)		
	4	$\hat{\delta}_4$	0.48 (0.09)*	-0.10 (0.05)*	0.82 (0.20)*	-0.02 (0.07)	0.26 (0.38)	0.23 (0.13)$^\circ$				
	5	$\hat{\delta}_5$	0.42 (0.09)*	-0.11 (0.05)*	0.23 (0.19)	-0.17 (0.08)*	-0.12 (0.33)	-0.65 (0.17)*				
	6	$\hat{\delta}_6$	0.47 (0.09)*	-0.10 (0.05)*	0.23 (0.20)	-0.21 (0.08)*						
	7	$\hat{\delta}_7$	0.44 (0.10)*	-0.10 (0.05)*								
Distribution			log-normal		log-normal		log-normal		log-normal		log-normal	
♯ obs.			3626		1224		368		143		90	
♯ parameters			22		20		16		10		8	
AIC			56698		20855		6631		2737		1746	

Table 7: Body – Tail split: Parameter estimates and their standard error (*in brackets*) for $f_{\cdot,1}$ in equation (5). Covariate information consists of development year t (as factor information) and occurrence year i (as factor information). Significance of the parameter at 5% level is denoted by $^\circ$, at 1% level is denoted by *.

Body of the distribution We examine the fit of five truncated GAMLSS error distributions[7] to the body distribution of Y_j. We include the same set of (possibly) time-dependent covariate information in the model for the location parameter μ and scale parameter σ of each of the five distributions under investigation. The vector of covariate information, say $\boldsymbol{x}_\kappa(t)$, is observed for each individual claim κ at the start of development period t, where t is the development period in which the claim's intermediate payment Y_j is registered. Our case study examines the use of the occurrence year i and development year t since occurrence, both expressed as factor information. Inclusion of these two variables allows us to capture inflation in the model through the combined $i+t$ effect. Based on the AIC we determine the preferred model for the body of Y_j. The selected GAMLSS distribution is the log-normal distribution for all $j \in \{1,2,3,4,5\}$.

Table 7 shows the parameter estimates and their standard errors for the selected distributions. We explain the coding of the covariates for the case of $f_{4,1}(y)$, which is the body distribution for the fourth payment in a claim's development process. This model uses five regression parameters ($\alpha, \beta_4, \delta_1, \delta_2$ and δ_3). Remember that occurrence year $i=0$ corresponds to the first occurrence year observed in the data set, i.e. 1997. Development year t since occurrence has value 0 at the year of occurrence, 1 the year afterwards, and so on. As the fourth payment can be made earliest in development year $t=3$, the reference category captured by the intercept α is $t=3, i=0$. To have enough observations in each level of the factor variables we combine levels representing longer tailed run-off lengths. For example, in the density $f_{4,1}(y)$, β_4 captures the effect of development periods $t \geq 4$. More generally the parameter β_l with the highest l in Table 7 corresponds to the effect of development periods $t \geq l$. The parameter δ_l with the highest l corresponds to the effect of occurrence period $i \geq l$. For example, for $f_{Y_4}(y)$, δ_3 captures the effect of $i \geq 3$, whereas the remaining parameters δ_1 and δ_2 correspond to occurrence year $i=1$ and $i=2$ respectively.

We discuss some striking features in Table 7. As development period t increases, so does the location parameter μ in the body distribution of the first payment Y_1. Up until $i=4$, the occurrence year i has the same effect on the location parameter. The effect stabilizes more or less from $i=5$ onwards. For the second payment Y_2, we see the location parameter μ increases with development year up until $t=3$. We detect significant effects of i and t in the systematic component of the location and scale parameter. We do not simplify the model in this case study

[7] The table below displays the GAMLSS error distributions, including their pdf and the link functions used in the GAMLSS framework, we investigate as possible models for the body of the various payment distributions.

Distribution	pdf		link function μ	link function σ
log-normal	$f(y\|\mu,\sigma) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\log(y)-\mu)^2}{2\sigma^2}\right\}$	for $y>0$ where $\mu>0$ and $\sigma>0$	identity	log
gamma	$f(y\|\mu,\sigma) = \frac{1}{(\sigma^2\mu)^{1/\sigma^2}} \frac{y^{\frac{1}{\sigma^2}-1}\exp\{-y/(\sigma^2\mu)\}}{\Gamma(1/\sigma^2)}$	for $y>0$ where $\mu>0$ and $\sigma>0$	log	log
exponential	$f(y\|\mu) = \frac{1}{\mu}\exp\left\{-\frac{y}{\mu}\right\}$	for $y>0$ where $\mu>0$	log	
Gumbel	$f(y\|\mu,\sigma) = \frac{1}{\sigma}\exp\left\{\left(\frac{y-\mu}{\sigma}\right) - \exp\left(\frac{y-\mu}{\sigma}\right)\right\}$	for $-\infty<y<\infty$ where $-\infty<\mu<\infty$ and $\sigma>0$	identity	log
Weibull	$f(y\|\mu,\sigma) = \frac{\sigma y^{\sigma-1}}{\mu^\sigma}\exp\left\{-\left(\frac{y}{\mu}\right)^\sigma\right\}$	for $y>0$ where $\mu>0$ and $\sigma>0$	log	log

by regrouping factor levels. This is the subject of future research, together with the search of other relevant covariates to include in the model and the use of non-linear covariate effects.

Probability of belonging to the body or tail of a payment distribution Table 8 shows the parameter estimates (and their standard errors) for the probabilities of belonging to the tail of the distribution $p_{j,2}(\boldsymbol{x}_\kappa(t))$ given in equation (10). Interpretation of the notation used in this table is analogous to the interpretation of Table 7. Given that the selected threshold u_j is constant over time, the inclusion of the development year t and occurrence year i as covariates allows the probability of a payment Y_j to belong to the tail of the distribution to increase over time as a consequence of inflation. The results show that the probability of a payment being in the tail of the distribution increases with the development year t for this specific data set. The effect of the occurrence year i is less outspoken.

Simulation of payments We explain at the end of Section 4.1 how to simulate a path for an RBNS claim through the multi-state model from Figure 2. If a claim makes a transition that corresponds to a payment (e.g. a transition to state S_1 implies a first payment is made), we simulate a payment from the relevant payment distribution constructed in Section 3. Consider an RBNS claim κ for which we simulated a transition involving a payment in development period t. We start by determining the payment number j corresponding to this payment and the vector of covariate information $\boldsymbol{x}_\kappa(t)$ at the start of development period t. Next, we simulate a payment from the distribution $f_{Y_j}(y|\boldsymbol{x}_\kappa(t))$. When all future payments of RBNS claim κ are simulated, we cap the total paid amount on claim κ at the policy limit of the policy under consideration, namely €2,500,000.

			$p_{1,2}$	$p_{2,2}$	$p_{3,2}$	$p_{4,2}$	$p_{5,2}$
		$\hat{\alpha}_2$	-6.81 (0.76)	-2.85 (0.29)	-2.95 (0.52)	-2.08 (0.70)	-3.85 (1.25)
Development	1	$\hat{\beta}_{2,1}$	2.28 (0.56)				
year t	2	$\hat{\beta}_{2,2}$	2.97 (0.73)	0.63 (0.27)			
	3	$\hat{\beta}_{2,3}$	3.77 (0.81)	0.93 (0.44)	0.68 (0.47)		
	4	$\hat{\beta}_{2,4}$		1.29 (0.66)	1.51 (0.57)	0.09 (0.67)	
	5	$\hat{\beta}_{2,5}$		1.84 (0.56)			1.56 (1.11)
Occurrence	1	$\hat{\beta}_{2,6}$	-0.24 (0.93)	-0.62 (0.45)	-0.57 (0.68)	-0.39 (0.86)	0.55 (0.97)
year i	2	$\hat{\delta}_{2,1}$	0.21 (0.83)	-0.00 (0.38)	0.17 (0.62)	-1.57 (1.19)	0.12 (1.05)
	3	$\hat{\delta}_{2,2}$	-0.12 (0.84)	-0.37 (0.40)	-0.71 (0.68)	-0.35 (0.86)	1.91 (1.65)
	4	$\hat{\delta}_{2,3}$	0.57 (0.76)	-0.52 (0.44)	0.29 (0.62)	-0.69 (1.25)	
	5	$\hat{\delta}_{2,4}$	1.01 (0.75)	0.15 (0.40)	1.04 (0.75)		
	6	$\hat{\delta}_{2,5}$	1.28 (0.75)	0.03 (0.48)			
	7	$\hat{\delta}_{2,6}$	-8.87 (0.00)				

Table 8: Parameter estimates and their standard error (*in brackets*) for the probability of belonging to the tail of the distribution, $p_{\cdot,2}$ in equation (10).

4.3 RBNS reserve

One full simulation of the total RBNS reserve for this portfolio consists of two steps: (1) we simulate the future path in the multi-state model for each RBNS claim as explained in the last paragraph of Section 4.1, (2) for each of these simulated transitions involving a payment, we

simulate a corresponding payment as discussed in the last paragraph of Section 4.2. By repeating this procedure, we generate a large number of possible values for the total RBNS reserve and the reserve per calendar year which can be summarized and represented by empirical distribution functions or densities. Figures 8, 9, 10 and 11 show the distribution of the RBNS reserve per calendar year (i.e. 2005 until 2011) as well as the distribution of the total RBNS reserve over the complete run-off.

To check robustness of our model, we also fit a density function of Y_j consisting of only one GAMLSS component that is left-truncated at 0. In this case, there is no distinction between the body and the tail of the distribution. Table 9 shows the fitting results for $j \in \{1, 2, 3, 4, 5\}$. We select the best fit based on the AIC, resulting in the log-normal distribution for $f_{Y_1}(y), f_{Y_2}(y)$ and $f_{Y_3}(y)$ and the Weibull distribution for $f_{Y_4}(y)$ and $f_{Y_5}(y)$. Interpretation of the results is analogous to Table 7.

We compare our results to the distribution for the RBNS reserve per calendar year as obtained from a bootstrap implementation of the Double Chain Ladder (DCL) method with and without including parameter uncertainty (see Martínez Miranda et al. (2012)[8]). This method works on aggregate data and combines the classical payments run-off triangle with a reported counts triangle. We use DCL as a benchmark because it explicitly allows to distinguish the IBNR from the RBNS reserve and most classical methods, such as the chain ladder method, do not allow this explicit separation. Figures 8 and 9 compare the micro-level results to the results of the bootstrap DCL method where parameter uncertainty is not taken into account, whereas Figures 10 and 11 compare the micro-level results to the results of the bootstrap DCL method where parameter uncertainty is taken into account. Table 10 and Table 11 show summary statistics on the empirical distribution of the RBNS reserve shown in blue in Figures 8 and 9 or 10 and 11. On the other hand, Table 12 and Table 13 show summary statistics on the empirical distribution of the RBNS reserve shown in red in Figures 10 and 11 respectively in Figures 8 and 9. A striking feature is that the observed value for the calendar year 2009 contains an extreme observation of €1,022,376.36 corresponding to a deadly accident[9].

Discussion We first compare the RBNS distributions with a single GAMLSS component to the body-tail split (as in equation (5)). The former approach leads to a slightly heavier tail when aggregated per calendar year. For each calendar year the distributions in blue in Figures 8, 9, 10 and 11 capture the observed value in the validation data set, even the extreme observation in 2009. However, the observed value for this year is very far in the right tail. Recall that the observed number of payments in Figure 7 for calender year 2006, 2007 and 2008 were in the left tail of the empirical distribution. Therefore, a reasonable consequence is that the RBNS reserve as observed in the validation data set for these calendar years is also in the left tail of the micro-level distribution.

[8] Martínez Miranda et al. (2012) implement the DCL method in the `R` package `DCL`.

[9] In the multi-state model represented by Figure 2 the claim corresponding to this extreme event arrives in state S_{oc} in 2002 and makes a transition to state S_0 in the same year. The claim transitions from state S_0 to state S_1 by a first payment of €5,083.11 in the year 2004. In 2005 (i.e. the first year in the validation data set) the claim transitions from S_1 to S_2 with a payment of €17,325.15. The year after, a transition to state S_3 takes place with a payment of €8,406.67. The claim then transitions from state S_3 to S_4 and from S_4 to S_5 with payments of respectively €20,750.44 and €159,564.04. Finally, the claim closes by a transition from state S_5 to state S_{tn} with a sixth payment of €1,022,376.36 bringing the total paid amount to €1,233,505.77.

			$f_{1,1}$		$f_{2,1}$		$f_{3,1}$		$f_{4,1}$		$f_{5,1}$	
			μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
		$\hat{\alpha}$	5.53 (0.07)*	0.38 (0.03)*	6.67 (0.13)*	0.56 (0.05)*	7.56 (0.22)*	0.50 (0.09)*	9.09 (0.44)*	-0.73 (0.16)*	9.01 (0.38)*	-0.34 (0.18)$^\circ$
t	1	$\hat{\beta}_1$	0.69 (0.06)*	0.13 (0.03)*								
	2	$\hat{\beta}_2$	1.36 (0.14)*	0.16 (0.06)*	0.10 (0.12)	0.15 (0.05)*						
	3	$\hat{\beta}_3$	1.81 (0.18)*	-0.01 (0.09)	0.51 (0.21)*	0.08 (0.09)	0.21 (0.19)	0.05 (0.08)				
	4	$\hat{\beta}_4$			0.33 (0.42)	0.14 (0.16)	-0.33 (0.38)	0.49 (0.12)*	0.06 (0.29)	-0.08 (0.12)		
	5	$\hat{\beta}_5$			0.94 (0.45)*	0.20 (0.16)					0.21 (0.34)	-0.29 (0.16)$^\circ$
i	1	$\hat{\delta}_1$	0.10 (0.09)	-0.09 (0.05)*	0.01 (0.18)	-0.11 (0.07)	-0.18 (0.29)	-0.06 (0.12)	-0.37 (0.52)	0.20 (0.19)	-0.29 (0.45)	-0.03 (0.20)
	2	$\hat{\delta}_2$	0.32 (0.09)*	-0.08 (0.05)$^\circ$	0.20 (0.18)	-0.07 (0.07)	-0.43 (0.30)	0.02 (0.12)	-0.56 (0.50)	0.32 (0.19)$^\circ$	-0.14 (0.41)	0.05 (0.19)
	3	$\hat{\delta}_3$	0.36 (0.09)*	-0.08 (0.04)$^\circ$	0.09 (0.17)	-0.09 (0.07)	-0.21 (0.27)	-0.15 (0.11)	-0.16 (0.48)	0.32 (0.18)$^\circ$		
	4	$\hat{\delta}_4$	0.48 (0.09)*	-0.09 (0.05)*	0.45 (0.17)*	-0.15 (0.07)*	-0.16 (0.31)	0.14 (0.12)				
	5	$\hat{\delta}_5$	0.43 (0.09)*	-0.10 (0.05)*	0.35 (0.18)*	-0.11 (0.07)	0.57 (0.32)	-0.25 (0.16)				
	6	$\hat{\delta}_6$	0.50 (0.09)*	-0.08 (0.05)$^\circ$	0.31 (0.18)$^\circ$	-0.18 (0.08)*						
	7	$\hat{\delta}_7$	0.44 (0.10)*	-0.11 (0.05)*								
Distribution			log-normal		log-normal		log-normal		Weibull		Weibull	
♯ obs.			3655		1306		399		154		98	
♯ parameters			22		22		16		10		8	
AIC			57627		23188		7533		3085		1994	

Table 9: No body-tail split, but a global fit with a GAMLSS model per payment number: parameter estimates and their standard error (*in brackets*) for f_{Y_j} for $j \in \{1, 2, 3, 4, 5\}$. Covariate information consists of development year t (as factor information) and occurrence year i (as factor information). Significance of the parameter at 5% level is denoted by $^\circ$, at 1% level is denoted by *.

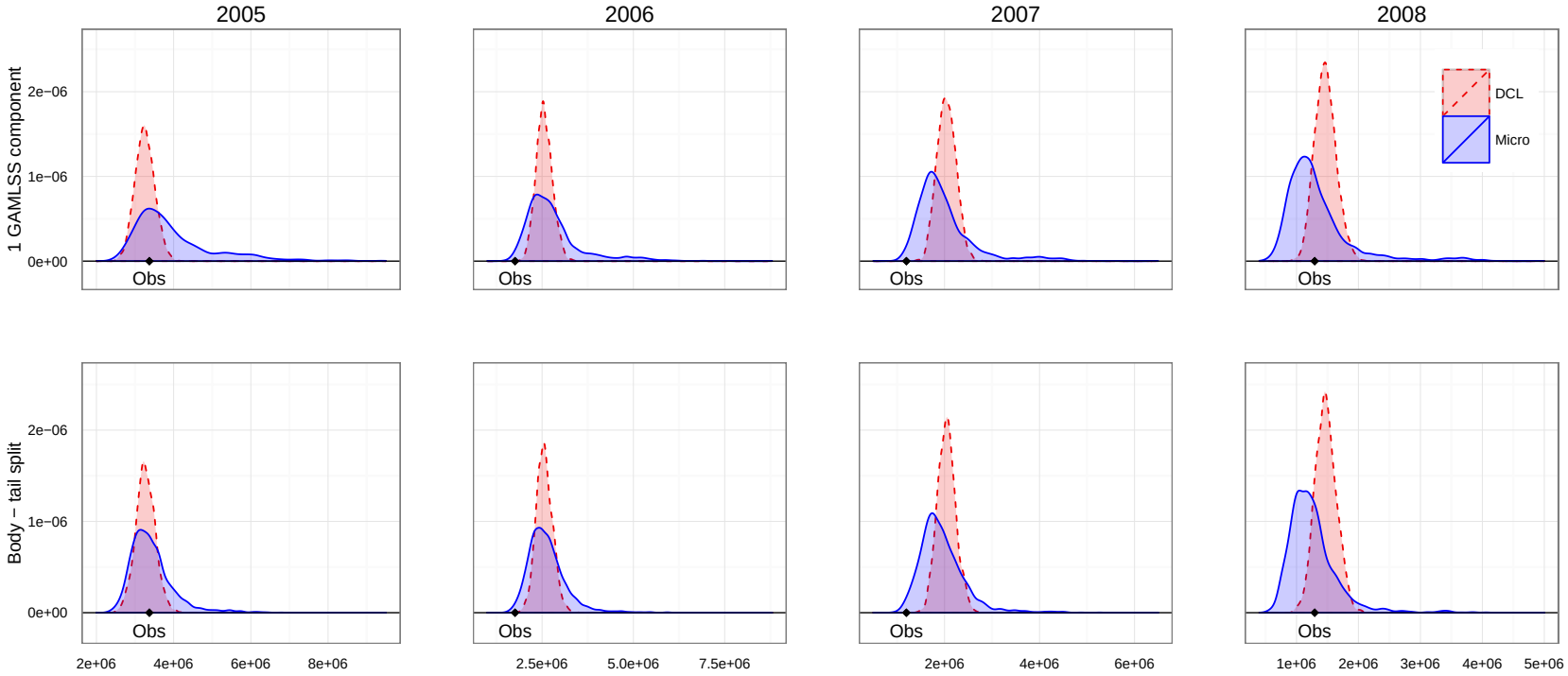

Figure 8: RBNS reserve per calendar year (*columns*) obtained by 2,000 full simulations. We compare the results of our micro-level approach (*full blue line*) to the results obtained by the bootstrap DCL method where parameter uncertainty is not taken into account (*dashed red line*). The first row of figures shows the results of our micro-level approach where the payment distributions f_{Y_j} are modeled by a single truncated GAMLSS component. The second row of figures shows the results of our micro-level approach where the payment distributions f_{Y_j} are spliced in a body and tail component. The label 'Obs' indicates the corresponding observed RBNS reserve in the validation data set.

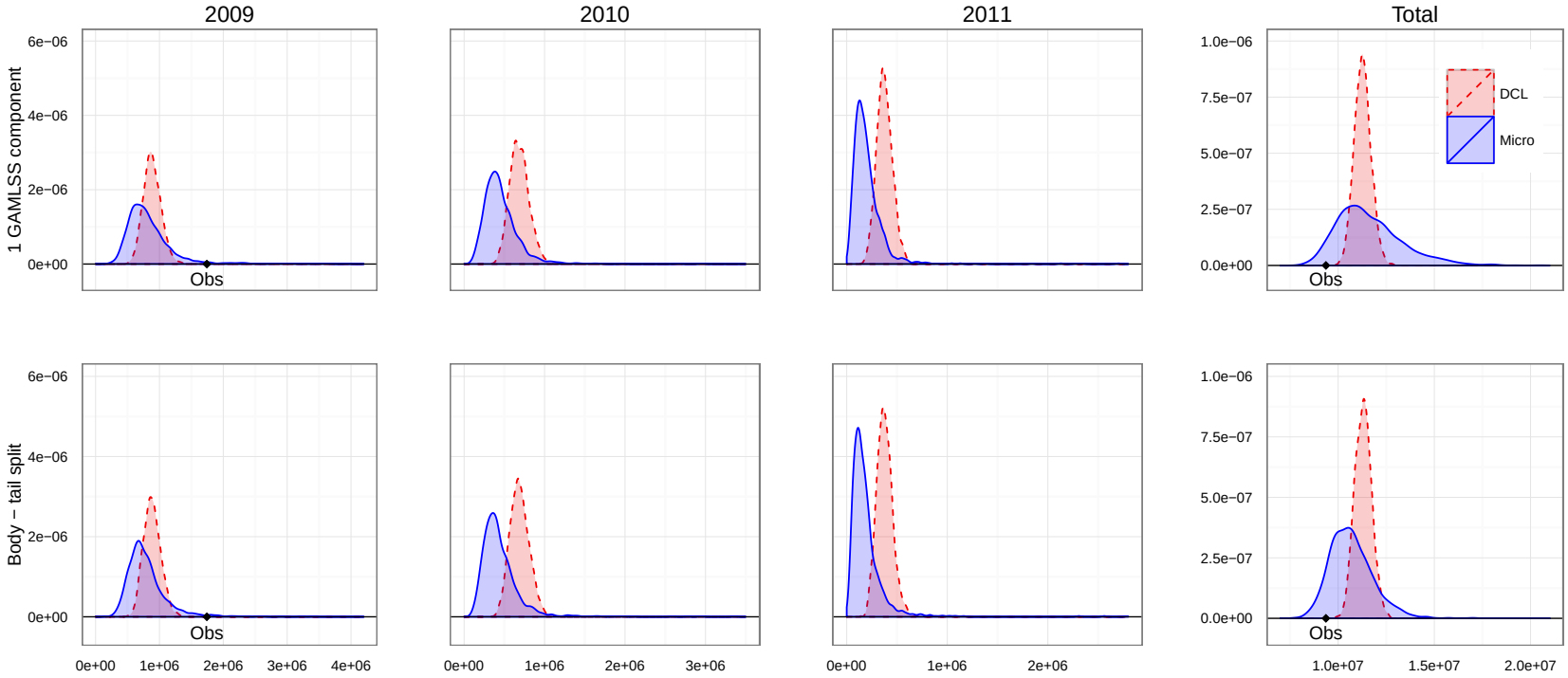

Figure 9: RBNS reserve per calendar year (*first three columns*) and in total (*last column*) obtained by 2,000 full simulations. We compare the results of our micro-level approach (*full blue line*) to the results obtained by the bootstrap DCL method where parameter uncertainty is not taken into account (*dashed red line*). The first row of figures shows the results of our micro-level approach where the payment distributions f_{Y_j} are modeled by a single truncated GAMLSS component. The second row of figures shows the results of our micro-level approach where the payment distributions f_{Y_j} are spliced in a body and tail component. The label 'Obs' indicates the corresponding observed RBNS reserve in the validation data set.

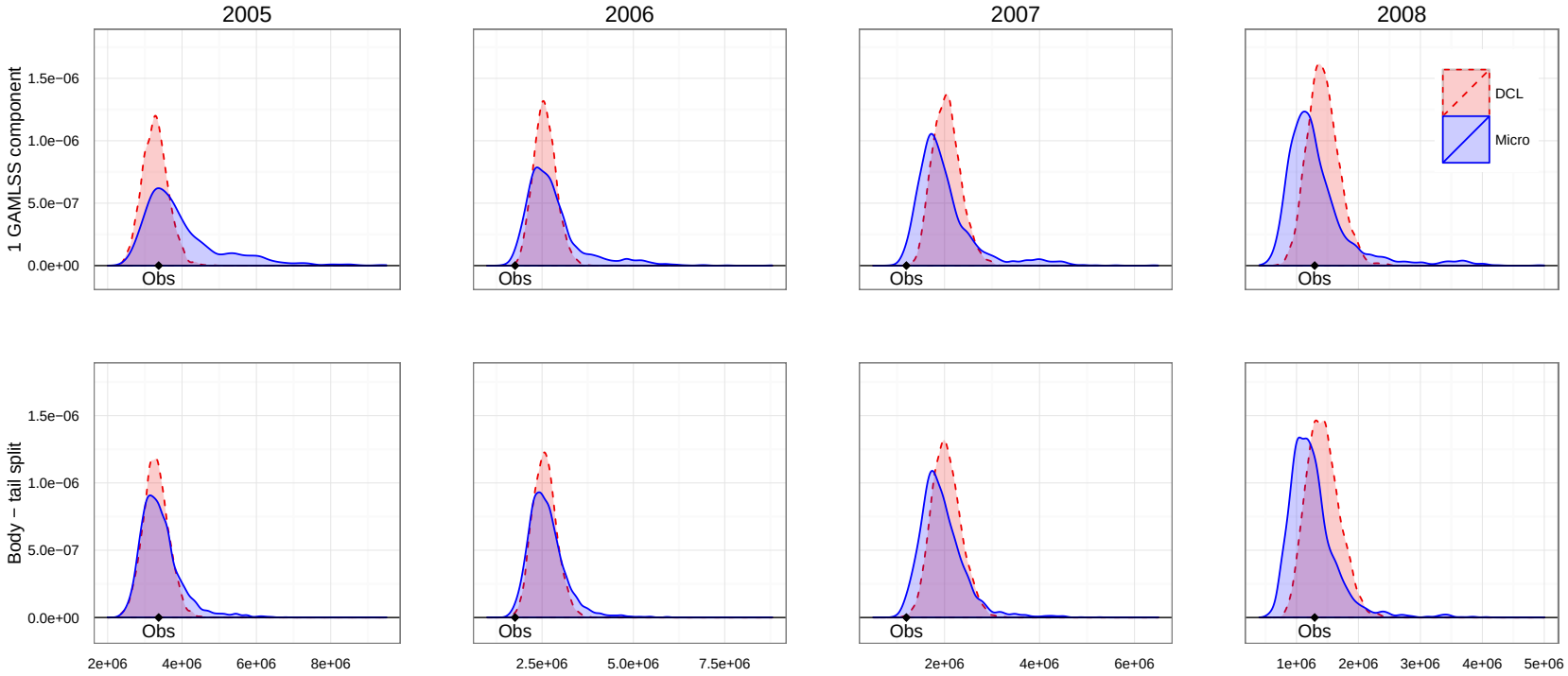

Figure 10: RBNS reserve per calendar year (*columns*) obtained by 2,000 full simulations. We compare the results of our micro-level approach (*full blue line*) to the results obtained by the bootstrap DCL method where parameter uncertainty is taken into account (*dashed red line*). The first row of figures shows the results of our micro-level approach where the payment distributions f_{Y_j} are modeled by a single truncated GAMLSS component. The second row of figures shows the results of our micro-level approach where the payment distributions f_{Y_j} are spliced in a body and tail component. The label 'Obs' indicates the corresponding observed RBNS reserve in the validation data set.

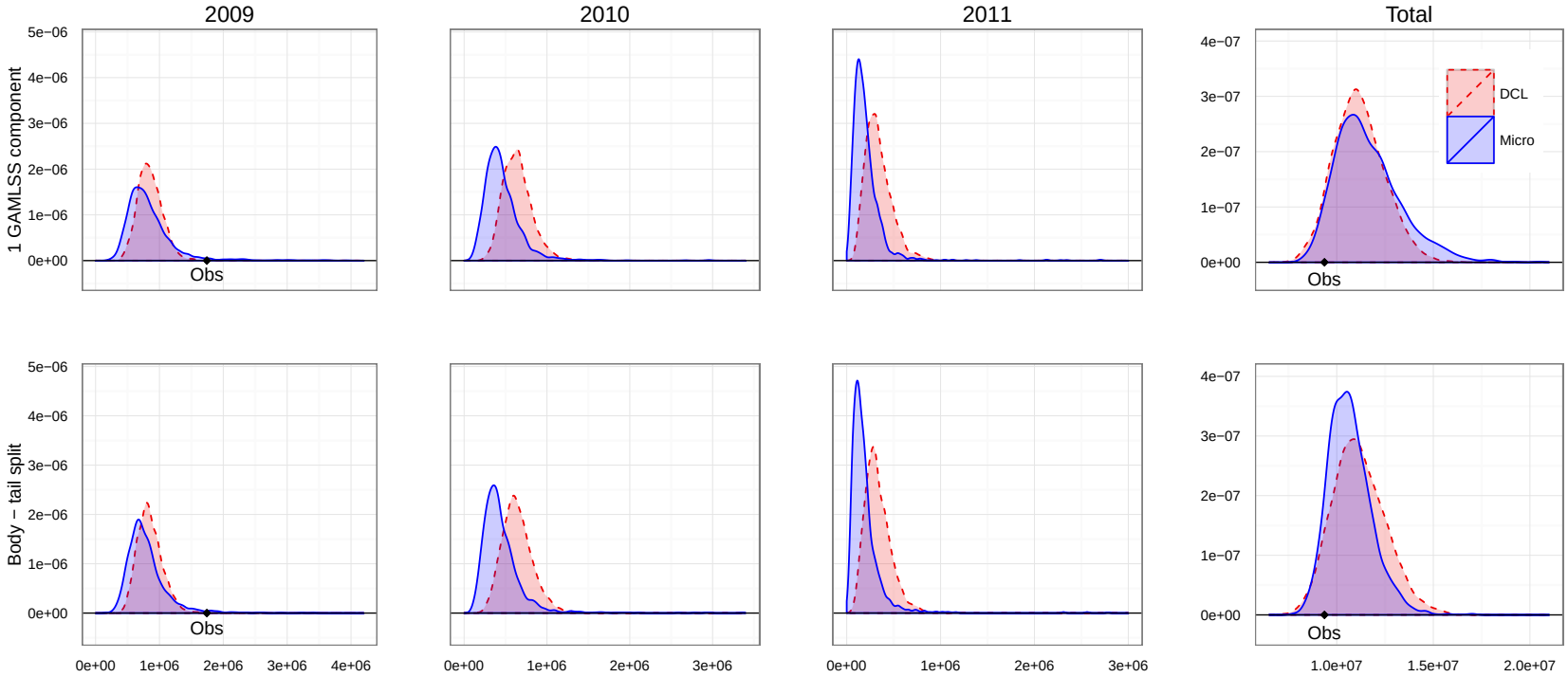

Figure 11: RBNS reserve per calendar year (*first three columns*) and in total (*last column*) obtained by 2,000 full simulations. We compare the results of our micro-level approach (*full blue line*) to the results obtained by the bootstrap DCL method where parameter uncertainty is taken into account (*dashed red line*). The first row of figures shows the results of our micro-level approach where the payment distributions f_{Y_j} are modeled by a single truncated GAMLSS component. The second row of figures shows the results of our micro-level approach where the payment distributions f_{Y_j} are spliced in a body and tail component. The label 'Obs' indicates the corresponding observed RBNS reserve in the validation data set.

Calendar year	2005	2006	2007	2008	2009	2010	2011	Total
Minimum	2,306	1,545	938	510	174	89	12	7,730
25% quantile	3,052	2,284	1,639	1,007	603	299	99	9894
Median	3327	2548	1866	1189	734	396	152	10,574
Mean	3,423	2,630	1,948	1,259	795	445	192	10,691
75% quantile	3,646	2,875	2,173	1,404	907	527	227	11,344
95% quantile	4,412	3,518	2,771	1,914	1,352	829	443	12,717
Maximum	6,789	5,934	5,268	4,060	3,411	3,360	2,784	17,067
Observed	3,372	1,762	1,196	1,293	1,742			9,366

Table 10: Summary statistics (in thousands) on the simulation results for the RBNS reserve per calendar year using the micro-level reserving method introduced in this paper. The number of simulations used: 2,000. The row labeled 'observed' corresponds to the total paid amount observed in the validation data set.

Calendar year	2005	2006	2007	2008	2009	2010	2011	Total
Minimum	2,365	1,447	906	473	225	72	3	8,219
25% quantile	3,262	2,289	1,616	1,005	602	317	112	10,392
Median	3,675	2,605	1,848	1,209	757	417	169	11,325
Mean	3,978	2,792	2,001	1,335	839	458	204	11,609
75% quantile	4,348	3,014	2,174	1,476	963	550	246	12,523
95% quantile	6,136	4,637	3,410	2,393	1,456	839	422	14,841
Maximum	9,474	8,793	6,407	4,921	3,934	2,970	2,715	20,498
Observed	3,372	1,762	1,196	1,293	1,742			9,366

Table 11: Summary statistics (in thousands) on the simulation results for the RBNS reserve per calendar year using the micro-level reserving method where the payment distributions f_{Y_j} are modeled by a single truncated GAMLSS component. The number of simulations used: 2000. The row labeled 'observed' corresponds to the total paid amount observed in the validation data set.

Calendar year	2005	2006	2007	2008	2009	2010	2011	Total
Minimum	2,278	1,640	1,241	718	359	238	57	7,487
25% quantile	3,036	2,363	1,834	1,257	716	519	237	10,255
Median	3,269	2,563	2,027	1,414	835	630	314	11,071
Mean	3,280	2,576	2,043	1,429	851	643	337	11,161
75% quantile	3,494	2,774	2,224	1,588	974	746	413	12,008
95% quantile	3,872	3,105	2,553	1,872	1,179	955	596	13,449
Maximum	4,833	3,742	3,032	2,476	1,642	1,366	934	16,302
Observed	3,372	1,762	1,196	1,293	1,742			9,366

Table 12: Summary statistics (in thousands) on the simulation results for the RBNS reserve per calendar year using the bootstrap DCL method including parameter uncertainty. The number of simulations used: 2000. The row labeled 'observed' corresponds to the total paid amount observed in the validation data set.

Calendar year	2005	2006	2007	2008	2009	2010	2011	Total
Minimum	2,516	1,887	1,408	946	525	358	142	9,639
25% quantile	3,082	2,404	1,905	1,347	794	603	324	10,973
Median	3,249	2,546	2,038	1,463	880	679	372	11,260
Mean	3,254	2,559	2,042	1,467	886	686	377	11,270
75% quantile	3,421	2,704	2,175	1,578	974	761	426	11,545
95% quantile	3,680	2,944	2,370	1,766	1,120	892	509	11,987
Maximum	4,142	3,380	2,777	2,093	1,407	1,058	625	12,847
Observed	3,372	1,762	1,196	1,293	1,742			9,366

Table 13: Summary statistics (in thousands) on the simulation results for the RBNS reserve per calendar year using the bootstrap DCL method without taking into account parameter uncertainty. The number of simulations used: 2000. The row labeled 'observed' corresponds to the total paid amount observed in the validation data set.

resulting from our simulation procedure. For calendar years 2006 and 2007 the observed value is in the left tail of the distribution. However, this is not the case for calendar year 2008. Note that year 2009 was only observed until August in the validation data set, implying the observed RBNS reserve for this calendar year is incomplete.

Next, we compare the micro-level results with the results from the bootstrap DCL method. Firstly, the distributions obtained with the bootstrap DCL method are generally more symmetrical in this particular case study compared to the resulting distributions from the micro-level methods. Apart from being more symmetrical, they also seem to be more shifted to the right which is in line with earlier results reported in Antonio and Plat (2014), Pigeon et al. (2014). The distribution obtained with the bootstrap DCL method without parameter uncertainty is more narrow compared to the resulting distribution from the bootstrap DCL method that does take parameter uncertainty into account. Caution should be taken with the DCL method that does not include parameter uncertainty: in this particular case study, the observed RBNS reserve is lower than the minimum value simulated for calendar years 2006 and 2007, whereas the incomplete observed RBNS reserve for 2009 is above the maximum of the simulated values.

5 Conclusion

We propose a multi-state framework for RBNS claims reserving for time-discrete data. Herein, we represent the claim development process by a series of transitions between a given set of states. In this framework, we model the payments per payment number conditional on covariate information and with attention for the tail of the distributions. For each payment number, we combine a GAMLSS fit for the body of the payment distribution with a GPD fit for the tail. Moreover, the parameters in the GAMLSS body of the distribution as well as the probability of belonging to the tail of the distribution depend on covariate information. Covariate information included in the model is meant to capture inflation effects and consists of the occurrence year of the claim and the development period the payment is made in. Other covariate information can be incorporated easily.

We investigate the performance of our micro-level RBNS reserving method in a case study with a portfolio of general liability insurance policies for private individuals. We compare the results obtained by our proposed micro-level method to a simplified method with no specific attention to the tail of the payment distributions, as well as to results obtained by the DCL method with

and without taking into account parameter uncertainty as introduced by Martínez Miranda et al. (2012). In this specific case study, we find that the micro-level reserving method performs well. However, more case studies are needed to confirm this performance as well as simulation studies to investigate the performance of the method in diverse circumstances.

To finish, we indicate some possible directions for future research. A first and obvious direction is to extend our approach to the estimation of the Incurred But Not Reported (IBNR) reserve. This would require modeling the claim arrival process and the distribution of reporting delays. Second, we can investigate the inclusion of other covariate information in the hazard rates included in the multi-state model as well as in the payment models. Besides these more or other covariates, we could also investigate flexible effects of covariates through additive modelling within the GAMLSS class. Third, we could examine strategies to project our reserve calculations beyond triangle boundary.

Acknowledgement:

We thank María Dolores Martínez Miranda for her kind help with the R application of the bootstrap DCL method and Liivika Tee for comments on an earlier draft of this paper.

References

Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scandinavian Journal of Statistics*, 3(1):15–27.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological methodology*, 13(1):61–98.

Antonio, K. and Plat, R. (2014). Micro–level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, 2014(7):649–669.

Aue, F. and Kalkbrener, M. (2006). LDA at work: Deutsche Bank's approach to quantifying operational risk. *Journal of Operational Risk*, 1(4):49–93.

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2(5):792–804.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Beyersmann, J., Allignol, A., and Schumacher, M. (2011). *Competing risks and multistate models with R*. Springer Science & Business Media.

Czado, C. and Rudolph, F. (2002). Application of survival analysis methods to long-term care insurance. *Insurance: Mathematics and Economics*, 31(3):395 – 413.

De Jong, P. and Heller, G. Z. (2008). *Generalized Linear Models for insurance data*. Cambridge University Press.

Denuit, M. and Robert, C. (2007). *Actuariat des Assurances de Personnes: Modélisation, Tarification et Provisionnement*. Economica.

Dickson, D. C., Hardy, M. R., and Waters, H. R. (2013). *Actuarial mathematics for life contingent risks*. Cambridge University Press.

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

Durrant, G. B., D'Arrigo, J., and Steele, F. (2013). Analysing interviewer call record data by using a multilevel discrete time event history modelling approach. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1):251–269.

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

Frees, E. W. (2015). Analytics of insurance markets. *Annual Review of Financial Economics*, 7(1):253–277.

Frees, E. W., Derrig, R. A., and Meyers, G. (2014). Predictive modeling in actuarial science. *Predictive Modeling Applications in Actuarial Science: Volume 1, Predictive Modeling Techniques*, page 1.

Frees, E. W. and Valdez, E. A. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, 103(484):1457–1469.

Godecharle, E. and Antonio, K. (2015). Reserving by conditioning on markers of individual claims: a case study using historical simulation. *North American Actuarial Journal*, 19(4):273–288.

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

Haberman, S. and Pitacco, E. (1999). *Actuarial models for disability insurance*. CRC Press.

Halliwel, L. J. (2007). Chain-ladder bias: its reason and meaning. *Variance*, 1:214–247.

Hesselager, O. (1994). A Markov model for loss reserving. *ASTIN Bulletin*, 24:183–193.

Hiabu, M., Margraf, C., Martínez Miranda, M. D., and Nielsen, J. P. (2016a). Cash flow generalisations of non-life insurance expert systems estimating outstanding liabilities. *Expert Systems with Applications*, 45:400–409.

Hiabu, M., Margraf, C., Martínez Miranda, M. D., and Nielsen, J. P. (2016b). The link between classical reserving and granular reserving through Double Chain Ladder and its extensions. *British Actuarial Journal*, 21:97–116.

Huang, J., Qiu, C., and Wu, X. (2015a). Stochastic loss reserving in discrete time: Individual vs. aggregate data models. *Communications in Statistics - Theory and Methods*, 44(10):2180–2206.

Huang, J., Qiu, C., Wu, X., and Zhou, X. (2015b). An individual loss reserving model with independent reporting and settlement. *Insurance: Mathematics and Economics*, 64:232 – 245.

Huang, J., Wu, X., and Zhou, X. (2016). Asymptotic behaviors of stochastic reserving: Aggregate versus individual models. *European Journal of Operational Research*, 249(2):657–666.

Jin, X. (2013, working paper). Micro–level loss reserving models with applications in workers compensation insurance. https://sites.google.com/site/xiaolijin2013/research/working-paper2.

Jin, X. and Frees, E. W. J. (2013, working paper). Comparing micro– and macro–level loss reserving. https://sites.google.com/site/xiaolijin2013/research/working-paper1.

Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008). *Modern Actuarial Risk Theory: Using R*. Springer.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data*. Springer.

Klein, N., Denuit, M., Lang, S., and Kneib, T. (2014). Nonlife ratemaking and risk management with Bayesian Generalized Additive Models for Location, Scale, and Shape. *Insurance: Mathematics and Economics*, 55:225–249.

Klein, N., Kneib, T., Lang, S., and Sohn, A. (2015). Bayesian structured additive distributional regression with an application to regional income inequality in Germany. *The Annals of*

Applied Statistics, 9(2):1024–1052.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss models: from data to decisions*. John Wiley & Sons.

Larsen, C. R. (2007). An individual claims reserving model. *ASTIN Bulletin*, 37:113–132.

Martínez Miranda, M. D., Nielsen, J. P., Sperlich, S., and Verrall, R. (2013). Continuous chain ladder: Reformulating and generalizing a classical insurance problem. *Expert Systems with Applications*, 40(14):5588 – 5603.

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

McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 27:117–137.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton university press.

Murphy, K. and McLennan, A. (2006). A method for projecting individual large claims. *Casualty Actuarial Society Forum*, pages 205–236.

Nadarajah, S. and Bakar, S. (2014). New composite models for the Danish fire insurance data. *Scandinavian Actuarial Journal*, 2014(2):180–187.

Nelson, W. (1969). Hazard plotting for incomplete failure data. *Journal of Quality Technology*, 1(1).

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*, 29:5–27.

Ohlsson, E. and Johansson, B. (2010). *Non–life insurance pricing with Generalized Linear Models*. Springer-Verlag Berlin Heidelberg.

Olivieri, A. and Pitacco, E. (2009). Stochastic models for disability. approximations and applications to sickness and personal accident insurance. *The ICFAI University Journal on Risk and Insurance*, 6:19–43.

Panjer, H. H. (2006). *Operational risk: modeling analytics*. John Wiley & Sons.

Peters, G. W. and Shevchenko, P. V. (2015). *Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk*. John Wiley & Sons.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119 – 131.

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

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

Pigeon, M. and Denuit, M. (2011). Composite Lognormal-Pareto model with random threshold. *Scandinavian Actuarial Journal*, 2011(3):177–192.

Pintilie, M. (2006). *Competing risks: a practical perspective*. John Wiley & Sons, Ltd.

Pitacco, E. (2014). *Health insurance: Basic actuarial models*. Springer International Publishing.

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

Scandroglio, G., Gori, A., Vaccaro, E., and Voudouris, V. (2013). Estimating VaR and ES of the spot price of oil using futures-varying centiles. *International Journal of Financial Engineering and Risk Management*, 1(1):6–19.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT – Statistical Journal*, 10(1):33–60.

Schiegl, M. (2015). A model study about the applicability of the chain ladder method. *Scandinavian Actuarial Journal*, 2015(6):482–499.

Stasinopoulos, D. M. and Rigby, R. A. (2007). Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software*, 23(7):1–46.

Steele, F., Kallis, C., and Joshi, H. (2006). The formation and outcomes of cohabiting and marital partnerships in early adulthood: the role of previous partnership experience. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(4):757–779.

Swiss Re (2008). Non-life claims reserving: improving on a strategic challenge. *Sigma*, 2:1–39.

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

Tong, E. N., Mues, C., and Thomas, L. (2013). A zero-adjusted gamma model for mortgage loan loss given default. *International Journal of Forecasting*, 29(4):548–562.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.

Verbelen, R., Gong, L., Antonio, K., Badescu, A., and Lin, S. (2015). Fitting mixtures of Erlangs to censored and truncated data using the EM algorithm. *ASTIN Bulletin*, 45:729–758.

Verrall, R., Nielsen, J. P., and Jessen, A. H. (2010). Prediction of RBNS and IBNR claims using claim amounts and claim counts. *ASTIN Bulletin*, 40:871–887.

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

Zhao, X. B. and Zhou, X. (2010). Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics*, 46(2):290–299.

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

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