DET
RAN
OTE

DETRA NOTE 2020-1

SEMI-MARKOV MULTISTATE INDIVIDUAL LOSS RESERVING MODEL IN GENERAL INSURANCE

By Carole Bettonville, Robin Van Oirbeek,
Michel Denuit & Julien Trufin

Detralytics

DISCLAIMER

The content of the Detra Notes for a pedagogical use only. Each business case is so specific that a careful analysis of the situation is needed before implementing a possible solution. Therefore, Detralytics does not accept any liability for any commercial use of the present document. Of course, the entire team remain available if the techniques presented in this Detra Note required your attention.

Detralytics
Rue Belliard/Belliardstraat 2
1040 Brussels
www.detralytics.com
info@detralytics.eu

ABSTRACT

This paper proposes a multistate model with a Semi-Markov dependence structure describing the different stages in the settlement process of individual claims in general insurance. Every trajectory, from reporting to closure is combined with a modeling of individual link ratios to obtain the ultimate cost of each claim. Quantiles and other risk measures can then easily be obtained by simulation. A case study based on individual claim developments in a motor third party liability insurance portfolio illustrates the relevance of the proposed approach.
Keywords: IBNR, RBNP, RBNS, loss development, technical provisions, solvency calculation, finan-cial reporting.

Contents

Chapter 1

Introduction

Within the Solvency II prudential framework as well as in the new IFRS environment, the evaluation of future cash flows and technical provisions have become increasingly important to assess the financial strength of insurance portfolios. The standard collective techniques based on aggregated data, conveniently summarized in a run–off triangle with occurrence and development years indexing rows and columns, respectively, must be supplemented with more granular models. It is also important to reconcile risk and reserving models so that all actuarial evaluations remain consistent. The individual claim development model proposed in this paper effectively responds to these challenges.

For the sake of completeness, let us briefly recall the main aspects of loss reserving models in general insurance. After the occurrence of the insured event, it may take some time before the policyholder reports a claim to the insurer. Also, the claim sometimes cannot be finalized by the end of the period but requires more time to be settled, for instance because of the long legal procedures in liability insurance. Meanwhile, the insurer must constitute a reserve representing the future costs to be paid in relation to this claim. Figure 1.1 illustrates the time scale for a claim in general insurance. The period during which the insured event occurs is referred to as the accident period. The time from occurrence to final settlement, or closure is divided into the reporting delay and the settlement delay, which is itself divided into a RBNP phase followed with a RBNS one. More specifically,

- Between occurrence of the insured event and notification to the insurance company, the insurer is liable for the claim amount but is unaware of the claim's existence. The claim is said to be Incurred But Not Reported (IBNR).
- Once notified, the company is aware of the claim but it may take some time before the first payment (if any) is made. The claim is said to be Reported But Not Paid (RBNP), meaning a reported claim for which no payments have been made yet.
- Then, either the claim closes without payment or the insurer makes an initial payment and several partial payments or refunds may follow. The claim finally closes at the closure or settlement date. From the first payment until closure of the claim, the final amount remains unknown and the claim is said to be Reported But Not Settled (RBNS).

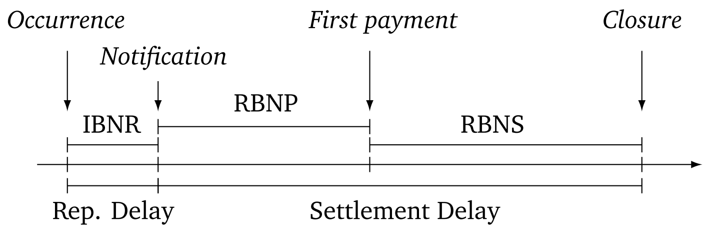

Figure 1.1: Decomposition of the time scale from occurrence to final settlement for a claim in general insurance.

At any time after reporting, the claim may terminate with or without a final payment made by the insurer to the beneficiaries.

For the sake of reporting and risk assessment, the total amount needed to pay for all the claims falling under the coverage must be evaluated by the actuary. At the evaluation date, one part of the total amount comes from the completion of RBNS claims. Predictions for costs related to RBNP and IBNR claims form the second part of the total amount. Individual loss reserving aims to model the settlement dynamics for each claim reported to the insurer. The goal is to design a stochastic model for trajectories until final settlement so that the associated cash flows can be used to determine the required reserve for outstanding claims.

The literature devoted to this topic can be traced back to the 1990s with the development of a mathematical framework in continuous time by Arjas (1989) and Norberg (1993, 1999). Since then, many models have been proposed and we refer the interested reader to Boumezoued and Devineau (2017) for a survey. In this paper, we work in the multistate approach to loss reserving proposed by Hachemeister (1980) and further considered e.g. by Hesselager (1994), Hurlimann (2015), and Antonio et al. (2016). Precisely, we adopt a discrete-time approach and model the reporting delay, the RBNP stage, and the times between consecutive payments by discrete random variables representing occupation times in a multistate process. The state space consists in an IBNR state, a RBNP state, a cascade of RBNS states, and two final states corresponding to closure with or without terminal payment. The trajectory across these states correspond to the decomposition visible in Figure 1.1. The occupation times in each state are studied with tools from discrete survival analysis that have initially been developed for two states, only, but easily extend to the multistate setting. The state-specific discrete hazard rate function corresponds to the probability that a transition takes place given that the claim has stayed in the current state for some time.

As in Hesselager (1994), we assume that payments occur at the times of transition between states. Individual development factors, or link ratios together with an initial payment structure the cash-flows along the development pattern of each claim. Several candidate distributions are available to model the resulting development factors at an individual claim level. Following Antonio et al. (2016), we adopt a discrete-time setting and we allow for general severity distributions, using the GAMLSS models that were successfully applied in an insurance context by Klein et al. (2014). In this paper, we extend the approach proposed by Antonio et al. (2016) in several directions:

- firstly, we allow for duration dependence by switching from a Markov modeling to a Semi-Markov setting. The duration effect appears to be significant in the case study worked out in the present paper so that this extension improves the modeling of claim dynamics.

- secondly, we explicitly account for the IBNR part in loss reserving calculations by estimating the occurrence intensity and the distribution of reporting delay.

The multistate approach proposed in the present paper turns out to be particularly well-suited for internal modeling under the prudential Solvency II framework imposed to insurance companies operating in the European Union. Individual link ratios can be used to simulate the run-off of open claims and obtain the sequence of the associated future cash-flows as well as their ultimate cost. The number of IBNR claims can also be simulated as well as the development of these claims until final settlement. If necessary, each trajectory from reporting to settlement can be impacted by reinsurance treaties or other risk mitigation techniques deployed to protect the portfolio under consideration, in order to assess the insurer's solvency or provide accurate financial reporting.

As an illustration, we perform a detailed case study based on data from a motor insurance portfolio considered in Denuit and Trufin (2017, 2018). The multistate loss reserving model is embedded in the hybrid approach proposed in Denuit and Trufin (2018). Specifically, the individual multistate modeling is applied to those claims with longer path to settlement. The majority of claims that are rapidly reported and closed are still modeled collectively (fitting the actuarial model to individual observations) and the detailed multistate modeling is used for claims with a longer settlement process, only. This is particularly effective from a computational point of view since it avoids to simulate numerous short trajectories corresponding to claims that are rapidly reported and settled (during the accident year and the year after, in the numerical study), concentrating the effort on the more complicated cases that generally also turn out to be more expensive. This hybrid approach appears to be a good compromise between accuracy and computing effort.

The remainder of this paper is organized as follows. We introduce the multistate loss reserving model in Section 2. Section 3 is devoted to a detailed case study to demonstrate the advantage of the proposed approach compared to traditional loss reserving procedures. The final Section 4 briefly concludes the paper.

Chapter 2

Model

2.1 Multistate approach

The different stages in the settlement process are modeled using a discrete-time multistate process. Time is measured in discrete steps, corresponding to periods $(t-1,t)$, $t=1,2,\ldots$, that can represent weeks, months, quarters, semesters or years. At the end of each time step $t=1,2,\ldots$, the claim can move from one state to another according to its specific settlement dynamics. These transitions are accompanied by cash-flows, payments or refunds. Because we work in discrete time, all the cash-flows of a period are aggregated into a single value. Henceforth, as in the majority of applications considered in the literature, we assume without loss of generality that time steps correspond to accounting, or calendar years.

Consider a claim related to an insured event occurring during a given accident year, taken as time origin 0. The stochastic process $\mathscr{S}=\{S_t,\ \ t=0,1,2,\ldots\}$ describes the claim settlement procedure, with $S_0=$IBNR. The meaning is that the claim moves to state S_t during period t, that is, the transition takes place between time $t-1$ and time t, $t=1,2,\ldots$. Hence, we actually have two time scales: calendar time and time elapsed since accident year. Depending on the calculation, one time scale or the other may be more convenient.

The state space and possible transitions are represented in Figure 2.1. Broadly speaking, all claims of the accident year under consideration are assumed to occupy the IBNR state at time 0. Then, claims start their path to settlement by moving to RBNP and/or a sequence of RBNS states until closure. Here, n denotes the maximal number of RBNS states that can be visited before final settlement. Notice that the time to settlement can exceed n because each claim can generally spend more than one year in every state. The different elements of the multistate model are precisely defined as follows.

IBNR state

The IBNR state is the initial one, where the claim stays until it is reported to the insurer. The numbers of insured events taking place in each accident year are assumed to be mutually independent, Poisson distributed. The means are allowed to vary between accident years. This specification is in line with

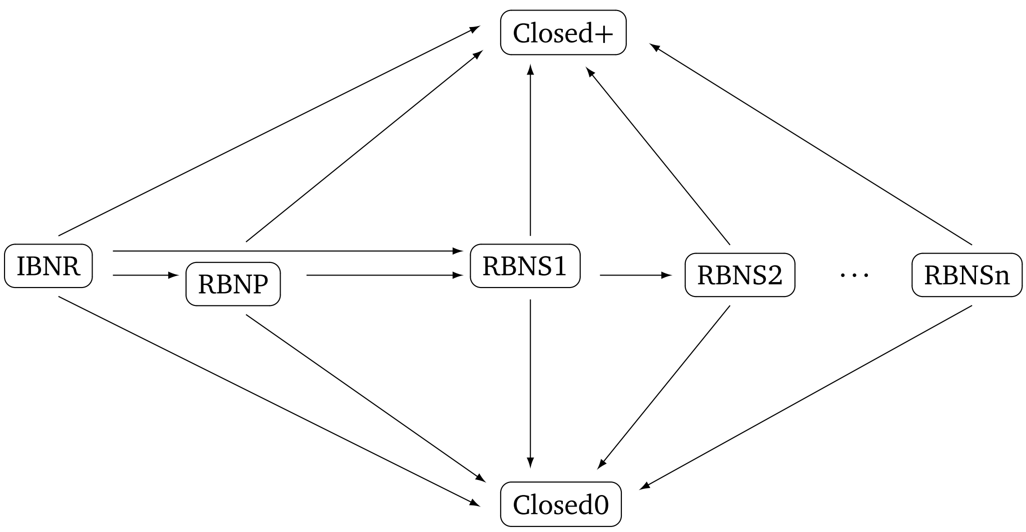

Figure 2.1: Multistate model describing the claim's settlement process, from occurrence to closure.

claim occurrence times obeying a non-homogeneous Poisson process (after Hesselager, 1994). The time-varying Poisson means account for

- time trends: in most industrialized countries, the accident frequency is known to decrease over time at market level in motor insurance, for instance;
- volume: the size of the portfolio may change over time and this obviously impacts on the number of claims, too.

If the time step is less than one year, time-varying Poisson means can also capture seasonality effects (more traffic accidents generally occur during winter periods in motor insurance, for instance).

Since claims may be reported at a later moment than when they actually occurred, the sojourn time in the IBNR state is a random variable denoted as D_{IBNR} and defined as

$$D_{\text{IBNR}} = \min\{t = 1, 2, \ldots | S_t \neq \text{IBNR}\}.$$

Hence, $D_{\text{IBNR}} = 1$ for those claims that are reported during the first development year (such that $S_1 \neq \text{IBNR}$). Then, $D_{\text{IBNR}} = 2$ holds for those claims that are reported one year after the accident year, that is, during the second development year, and so on.

Notice that the IBNR state is hidden: the insurer only observes claims when they are reported, that is, when they leave the IBNR state. At that time, the insurer gets information about the insured event, including its occurrence date. Thus, the reporting lag D_{IBNR} can be determined and the claim can be allocated to the right accident period.

Once reported to the insurer, the claim leaves the IBNR state and there are several possibilities:

- either the claim is reported but the insurer makes no payment at this stage such that it moves to the RBNP state defined below;

- or it is reported and the insurer makes an initial payment such that the claim moves directly to the first payment state RBNS1 defined below;
- or the claim is terminated during the year it is reported and directly moves from the IBNR state
    - either to the terminal state with payment Closed+
    - or to the terminal state without payment Closed0,

  both corresponding to closure. Here, we do not allow for possible re-opening of a closed claim.

The states corresponding to these different situations are defined in more details hereafter.

RBNP state

If the reported claim is not terminated during the reporting period but the insurer has not made any payment yet then the claim moves from the IBNR state to the RBNP one. It remains there as long as it stays open but the insurer does not pay anything to beneficiaries. When leaving the RBNP state, the claim may terminate with or without payments, or payments can start so that it enters the first RBNS state defined below.

RBNS states

If the insurer makes at least one payment in addition to the terminal one then the claim enters the cascade of RBNS states shown in Figure 1.1. More specifically, the claim moves along the series RBNS1, RBNS2, ... of RBNS states until closure, with a transition occurring each time an additional payment is made. Claims moving from IBNR to RBNS1 are those for which the insurer makes an initial payment but which are not closed during the same period that they are reported. Claims moving from state RBNP to RBNS1 are those for which a first payment is made by the insurer but which remain open after that.

At transition from IBNR or RBNP to the first RBNS state RBNS1, the insurer makes a first, non-terminal payment. Remember that we work in discrete time, such that several payments can actually be made during the development year and the total amount paid during this period is recorded as the first payment.

Then, each time a payment is made by the insurer, the claim moves to the next RBNS state, until final settlement. We assume that all claims are settled after a maximum number of payments n, such that we consider RBNS states indexed from 1 to n. Given that a claim can stay for several periods in each RBNS state, the total time to settlement may be longer than n. The value of n does not need to be specified in the application as letting n tend to infinity can be considered as a tail factor capturing unusually long developments.

Closure states Closed+ and Closed0

From each IBNR, RBNP or RBNS states (i.e., RBNS1, RBNS2, ...), the claim can terminate with or without terminal payment. This is formalized by a transition from one of these states to the final state Closed+ or Closed0. The difference between the two types of closure depends on the presence of a final payment upon termination. When there is no amount paid by the insurer at closure then the claim moves to the final state Closed0 whereas a transition to Closed+ indicates that the insurer makes a final payment at termination.

A direct transition from the IBNR state or RBNP state to Closed0 or Closed+ corresponds to the case where the claim is settled during the reporting period and is closed without any payment or with a (set of) payment(s) occurring during the first period only, respectively.

As mentioned before, Closed0 and Closed+ are the final states since we do not allow for possible re-openings of closed claims. By convention, a claim which re-opens is considered to have remained in a transient state RBNP or RBNS.

2.2 Claim occurrence rate and reporting lag

Once a claim is reported, the insurer gets information about its reporting lag D_{IBNR}: based on the date of occurrence, the length of the stay in the IBNR state can be determined. This allows the actuary to estimate the distribution function of the reporting delay D_{IBNR} and to infer the number of IBNR claims. In practice, this can also be done by studying an aggregated run-off triangle filled with numbers of claims cross-classified according to accident, or occurrence period and reporting lag, as explained below. This offers a consistency check of the results derived from the multistate model.

Formally, let us denote as N_{tj} the number of claims relating to insured events that occurred in accident year t, that is, from calendar time $t-1$ to t and were reported to the insurer at development j, i.e. during calendar year $t+j-1$. Development 1 thus corresponds here to the accident year.

In line with the classical Chain-Ladder model, we use the multiplicative specification

$$\mathrm{E}[N_{tj}] = \alpha_t \beta_j \tag{2.1}$$

subject to the usual identifiability constraint

$$\sum_{j\geq 1} \beta_j = 1.$$

In accordance with actuarial practice, we assume that the random variables N_{tj} are independent and Poisson distributed. This ensures that the total number

$$N_t = \sum_{j\geq 1} N_{tj}$$

of claims for accident year t also obeys the Poisson distribution with mean $\mathrm{E}[N_t] = \alpha_t$, and that β_j is the probability that a claim is reported at lag j, that is $j-1$ periods after its occurrence. The parameters α_t and β_j can be estimated from the triangle with observed N_{tj} using Poisson regression.

The expected claim number α_t for each accident year t allows the actuary to track possible trends (portfolio volume or market trends). Notice that a risk exposure (number of years of risk coverage, for instance) is useful to perform inference as α_t is the expected number of claims meaning the product between the claim rate and the corresponding risk exposure. For the evaluation of the technical provisions, only the expected numbers of claims α_t for past occurrence periods matter. If available, a volume measure can be included as an offset in the Poisson regression analysis allowing to estimate the coefficients α_t and β_j.

Notice that here, we have assumed some stationarity in the reporting pattern, in that the same proportions β_j apply to all accident years. In case the analyst suspects that some change may have occurred, which speed up or slow down the reporting of claims to the insurer, then different β_j can be specified depending on the accident year.

The distribution of the reporting lag D_{IBNR}, that is, of the sojourn time in state IBNR, can then be deduced from the resulting β_j. Precisely, the reporting lag is 1 if the claim is reported during the accident year, that is, at development $j = 1$. More generally, the reporting lag is j if the claim is reported at development j. Hence,

$$\mathrm{P}[D_{\text{IBNR}} = j] = \beta_j \text{ for } j = 1, 2, \ldots$$

The probability that the claim is reported at, or before development j is then given by

$$\begin{aligned} \mathrm{P}[S_j \neq \text{IBNR} | S_0 = \text{IBNR}] &= \mathrm{P}[D_{\text{IBNR}} \leq j] \\ &= \beta_1 + \ldots + \beta_j. \end{aligned}$$

This is the probability that a claim corresponding to an insured event that occurred in calendar year t is reported to the insurer in calendar year $t + j - 1$ at the latest.

2.3 Transition probabilities

The probability distribution of the stochastic process $\mathscr{S} = \{S_t, \quad t = 0, 1, 2, \ldots\}$ corresponding to path towards final settlement, starting from IBNR, is described by the transition probabilities. Working in a Semi-Markov setting, these transition probabilities depend on the current state and the time elapsed since the claim entered the current state. If available, other features can also be introduced into these probabilities, to refine risk evaluation.

Transition probabilities may depend on observable claim features. Information available for a claim in state S_t includes time t, i.e. time elapsed since occurrence, reporting lag D_{IBNR}, current state S_t so that the number of payments made so far may also influence the trajectory to final settlement, and time spent in the current state S_t (to recover duration effects in a Semi-Markov setting). In addition, some other features like the number of development years without payment can also be included in the analysis. If available, claim-specific information such that the presence of bodily injuries or policy-specific information such as the sales channel, for instance, can also be included as explanatory features for transition probabilities. Notice that the amount of information may depend on the occupied state S_t because the modeling proceeds separately for each transition. This information is included in both the transition probabilities and the cash-flows (as explained below).

2.4 Cash-flows

Cash-flows correspond to transition between some pairs of states (remember that there is no cash-flow when the claims enters the RBNP state, or moves to the Closed0 state). The corresponding amounts are modeled using an initial payment P_1 possibly multiplied with a sequence of link ratios describing the evolution of cumulative payments to the final cost. More specifically, when the claim enters the RBNS1 state, the insurer makes a first payment P_1 that will be subsequently followed by payments $P_2, P_3, \ldots$ along the RBNS cascade. The cumulative amount paid so far by the insurer for a claim in state RBNSj is given by

$$C_j = \sum_{k=1}^{j} P_k,$$

assuming that payments are made when the claim enters the RBNS state.

Notice that some of the payments P_k may be negative. This happens when the insurer has paid some benefits but the involved third party is recognized liable for the accident such that it must refund the insurer. The cumulative amount C_j must nevertheless remain non-negative (even positive because some costs are never reimbursed, such as internal settlement expenses). To account for this particularity, actuaries generally model payments using link ratios Λ_j defined as

$$\Lambda_j = \frac{C_{j+1}}{C_j}, \ \ j = 1, 2, \ldots$$

The advantage of working with such link ratios is that they can easily deal with negative payments, the random variable Λ_j being just less than 1 in such a case. The LogNormal distribution is certainly the natural choice for the distribution of the link ratios since ultimate amounts corresponding to products of link ratios are therefore also LogNormal. Depending on the line of business, a mixture model combining a light-tailed distribution and a heavier-tailed one may also be a good candidate to model the link ratios. For this purpose, the `gamlss` and `gamlss.mx` packages provide actuaries with a variety of distributions including the classical LogNormal, Pareto, and Gamma distributions but also their combinations with the help of finite mixtures.

Each claim is accompanied with its own sequence of cash-flows described by the random variables $P_1, \Lambda_1, \Lambda_2, \ldots$ such that

$$C_1 = P_1 \text{ and } C_j = P_1 \prod_{k=1}^{j-1} \Lambda_k \text{ for } j \geq 2.$$

The link ratios Λ_j have a distribution specific to each RBNS state RBNSj. They are independent from one claim to another (i.e. we assume that the claim-specific developments are mutually independent, except for inflation, for instance, which can be accounted for by using a mixed model). For a given claim, we assume that these random variables are independent for different values of j, given the features included in the analysis. Unconditionally, there is thus some dependence across the trajectory of each claim to settlement. We will see in the application that some dependence between P_1 and Λ_1 is nevertheless needed to prevent final costs that are unrealistically large, without economic relevance. This is because a large initial payment P_1 is generally accompanied with smaller link ratios compared to smaller initial payments.

Notice that if the claim does not enter the RBNS states, being terminated with a single payment, then link ratios are not needed. In case the claim moves directly from IBNR to Closed+ there is only a single payment P_1 made by the insurer, which has a specific distribution (that may differ from the one associated to a claim that enters state RBNS1), if suggested by the data. For convenience, link ratios can be set to 1 for states with no cash flow associated (like RBNP and Closed0).

The multistate model described in this paper is a payment-to-payment model, in the sense that each transition corresponds to an actual payment made by the insurer. For this reason, there is no need to include a positive probability mass at the origin for the initial payment, or at 1 for the link ratios.

Chapter 3

Case study

3.1 Data set

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 European Union. The observation period consists in calendar years $2,004$ till $2,014$. The database records all claims that originate from insured event occurred in accident years $2,004$ to $2,013$ so that we have observed developments up to 11 calendar years. We refer to Denuit and Trufin (2017, 2018) for a detailed description of the database and we only provide the reader here with relevant information for the analysis performed in the present paper.

The length of the settlement period appears to be a natural criterion to distinguish two types of claims:

- claims with short development to settlement, i.e. those claims that are reported and fully settled during the year of occurrence, or the year after;
- and claims with longer development to settlement, i.e. that are not yet reported or that are still open two years after occurrence.

Such a separation based on the time to settlement provides the actuary with a simple relevant criterion to isolate cheaper losses from the other, more expensive ones.

The multistate model is not needed for the claims that are rapidly settled because trajectories with

$$\begin{aligned} S_0 &= \text{IBNR} \\ S_1 &\in \{\text{IBNR, RBNP, RBNS}_1,\ \text{Closed0, Closed+}\} \\ S_2 &\in \{\text{Closed0, Closed+}\} \end{aligned}$$

are so short (with at most two states visited) that developing these claims on an individual basis does not bring any additional insight. This is why we isolate claims that are not yet reported or settled two years after occurrence and model their individual development using the proposed multistate model. Restricting accident years to 2,004-2,013 allows us to classify each claim into one of the two categories.

The database comprises the following information for each reported claim. A claim ID allows us to follow the whole trajectory, from reporting to closure. The database records the yearly amount of payments (equal to 0 if no payment has been made) together with the accident year, the current development year (the first development corresponding to accident year), the reporting year and the settlement year. The development increases by one unit whereas the other years do not change among the records for the same claim. Notice that the first development period and the development period in which the claim terminates generally do not correspond to an entire calendar year. Here is an example of individual claim development recorded in the database (for the claim with ID 6):

Claim ID	Payment	Accident year	Development year	Reporting year	Settlement year
6	0	2,004	1	2,004	2,007
6	0	2,004	2	2,004	2,007
6	0	2,004	3	2,004	2,007
6	559	2,004	4	2,004	2,007

This claim relates to an insured event that occurred in calendar year 2,004 that has been reported during the same year. This claim stayed in the RBNP state until 2,006 before termination in 2,007 (transition from RBNP to Closed+, with a unique payment of amount 559).

Because we restrict the data to claims with longer developments, the settlement year varies from $2,006$ to $2,014$. It is recorded as 0 if the claim is still open by the end of $2,014$. All payments are expressed in euro currency and have been corrected for inflation (using consumer price index).

3.2 Augmented data file

Based on the original data set, we build an augmented database comprising the following new variables:

- a variable indicating the RBNP or RBNS state at the end of the considered time period, here calendar year.
- a variable representing the time spent in the state, indicated by the payment state variable.
- a variable denoted as Trans which is equal to 1 if there is a transition and 0 otherwise.

Here is an example of records for claim with ID 611, augmented with the new features described above:

Claim ID	Payment	Acc. year	Dev. year	Rep. year	Set. year	Payment state	Time in state	Trans.
611	6,641	2,004	1	2,004	2,014	RBNS1	0	1
611	61,138	2,004	2	2,004	2,014	RBNS2	0	1
611	6,403	2,004	3	2,004	2,014	RBNS3	0	1
611	0	2,004	4	2,004	2,014	RBNS3	1	0
611	0	2,004	5	2,004	2,014	RBNS3	2	0
611	0	2,004	6	2,004	2,014	RBNS3	3	0
611	0	2,004	7	2,004	2,014	RBNS3	4	0
611	0	2,004	8	2,004	2,014	RBNS3	5	0
611	4,560	2,004	9	2,004	2,014	RBNS4	0	1
611	0	2,004	10	2,004	2,014	RBNS4	1	0
611	0	2,004	11	2,004	2,014	RBNS4	2	1

Claim 611 has been filed during the accident year 2,004 and developed until final settlement in 2,014. Following Allison (1982) and due to the definition of the multistate model, the augmented database has a single line for each development period. In 2,004, it moves directly from the IBNR state to the RBNS1 state (with an initial payment of 6,641) and then to RBNS2 and RBNS3 states (with successive payments of 61,138 and 6,404). For the three first records relating to Claim 611, Trans is equal to 1 because the claim makes a transition each year to the next RBNS state, and Time in state is 0 meaning that the transition occurred during the year. Then the claim remains in state RBNS3 for 5 years before moving to state RBNS4 (with an additional payment of 4,560). Two years after, it leaves state RBNS4 to enter state Closed0 when it terminates without final payment. The cash-flows for this claim are represented by the initial payment $P_1 = 6,641$ made at entrance into RBNS1 and the link ratios

$$\begin{aligned}
\Lambda_1 &= \frac{6,641 + 61,138}{6,641} = 10.2061 \\
\Lambda_2 &= \frac{6,641 + 61,138 + 6,403}{6,641 + 61,138} = 1.0945 \\
\Lambda_3 &= \frac{6,641 + 61,138 + 6,403 + 4,560}{6,641 + 61,138 + 6,403} = 1.0615
\end{aligned}$$

need to be applied to P_1 to obtain the cumulative payments when entering the next RBNS state. The features shown in this augmented database are then used to explain the transition probabilities, the mean initial payment and the expected link ratios.

3.3 Estimation of transition probabilities

3.3.1 Transition probabilities from IBNR

A Binomial GLM is used to model exit probabilities from the IBNR state. The response of interest is the time D_{IBNR} spent in the IBNR state. Recall that $D_{\text{IBNR}} = 1$ if the claim is reported during the accident year (at development 1, thus) and $S_1 \neq \text{IBNR}$. Exit probabilities correspond to the discrete hazard rate of D_{IBNR}, that is, to $\text{P}[D_{\text{IBNR}} = j | D_{\text{IBNR}} \geq j-1]$ for $j = 1, 2, \ldots$ In order to fit these probabilities, we use the binary variable Trans which is equal to 1 if the transition takes place. Its mean therefore corresponds to the exit probability from the IBNR state, depending on the time $j-1$ spent in that state. Claims are reported after at most 4 development years in the database under consideration. Since the accident year has been taken as the origin of time and corresponds to development 0, a claim spends at most 3 years in the IBNR state.

The analysis is performed with a Binomial GLM using the canonical logit link function. The score does not include an intercept and we start with a model allowing for a specific regression coefficient for each of the 4 categories $j \in \{0,1,2,3\}$. Several regression coefficients appeared to be sufficiently close (given the standard errors) to merge the corresponding levels. After grouping as much as possible, we obtain estimated coefficients 2.2318 (with standard error 0.0449) for $j=0$ and 1.8489 (with standard error 0.1167) for $j \geq 1$.

The estimated exit probabilities from IBNR state are as follows:

$$\mathrm{P}\big[D_{\text{IBNR}} = j \big| D_{\text{IBNR}} \geq j-1\big] = \begin{cases} 0.903 \text{ for } j=1 \\ 0.864 \text{ for } j \in \{2,3\} \\ 1.000 \text{ for } j=4. \end{cases} \tag{3.1}$$

In case some later reporting is suspected, the actuary can specify a Geometric tail factor corresponding to $\mathrm{P}[D_{\text{IBNR}} = j | D_{\text{IBNR}} \geq j-1] = q$ for $j \geq 4$ and some appropriate value for q. In the current setting, only 0.179% of the claims are not reported by the end of development period 3 so that the overall effect of imposing a tail factor is negligible. Equivalently,

$$\mathrm{P}\big[S_t \neq \text{IBNR} \big| S_{t-1} = \text{IBNR}\big] = \begin{cases} 0.903 \text{ for } t=1 \\ 0.864 \text{ for } t \in \{2,3\} \\ 1.000 \text{ for } t=4. \end{cases}$$

The exit probabilities (3.1) give the distribution of the sojourn time D_{IBNR} in the IBNR state. The next step is to determine the destination state when a claim leaves the IBNR state. There are 4 possibilities : either the claims terminates and moves to the final state Closed0 or Closed+ or it further develops and then moves to the RBNP state or to the RBNS1 state. In the database under consideration, no transition from IBNR to Closed0 has been observed such that we set the corresponding probability equal to 0, that is,

$$\mathrm{P}\big[S_t = \text{Closed0} \big| S_{t-1} = \text{IBNR}\big] = 0 \text{ for all } t.$$

The transition probabilities are then decomposed into two parts. First, we model the probability of moving to the terminal state Closed+. No significant duration effect is detected such that we end up with

$$\mathrm{P}\big[S_t = \text{Closed+} \big| S_{t-1} = \text{IBNR}\big] = \begin{cases} 0.000 \text{ for } t \in \{1,2\} \\ 0.507 \text{ for } t \geq 3. \end{cases}$$

Recall that we restricted our analysis to claims with longer developments such that no closure is possible during the two first development years.

For claims which do not terminate, we then model the probability to move either to the RBNP state or to the RBNS1 state. The corresponding probabilities are fitted using a Binomial GLM with the canonical link function and binary response which is equal to 1 if the claim moves to the RBNS1 state and to 0 otherwise. Reporting lag is treated as a categorical feature, as before. Starting from a score without intercept including a regression coefficient specific to each sojourn time in the state IBNR and grouping as much as possible, we finally obtain estimated coefficients equal to -0.8428 (with standard error 0.0305) for $j = 0$, to -0.4495 (with standard error 0.0941) for $j = 1$ and to -2.1102 (with standard error 0.5294) for $j \geq 2$. The corresponding conditional transition probabilities are as follows:

$$\mathrm{P}\big[S_t = \mathrm{RBNS1}\big|S_{t-1} = \mathrm{IBNR}, S_t \in \{\mathrm{RBNP}, \mathrm{RBNS1}\}\big] = \begin{cases} 0.301 \text{ for } t = 1 \\ 0.390 \text{ for } t = 1 \\ 0.108 \text{ for } t \geq 2. \end{cases}$$

3.3.2 Exit probabilities from RBNP or RBNS

The exit probabilities from RBNP or RBNS states are modeled using a Binomial GLM with the canonical logit link function. The response is the binary variable Trans whose mean is the exit probability from the current state, RBNP or RBNS, depending on the time spent in the current state. Because data become scarce at later development periods, data relating to RBNS states RBNS4, RBNS5, etc. are pooled together for estimation purposes. This means that sojourn times are identically distributed in each RBNSj state with $j \geq 4$. The time spent in state has to be categorized as well to have a sufficient volume of data in each class. We consider the categories "1 year", "2 years" and "$\geq$ 3 years" for the time spent in the current RBNP or RBNS state.

The score does not include an intercept and we start with a model allowing for a specific regression coefficient for each of the 15 cases (RBNP and 4 types of RBNS crossed with 3 categories for the time spent in the current state). Several regression coefficients appeared not to differ significantly with one another such that the corresponding levels were merged together. A deviance analysis is performed to compare the initial model (without grouping, counting 15 parameters) to the final one. The obtained p-value of 98.13% does not reveal any significant difference between the two models. Denoting as D_t the time spent in state occupied at time t, the estimated exit probabilities from RBNP and RBNSj states are as follows:

$$\mathrm{P}\big[S_t \neq \mathrm{RBNP}\big|D_t = j\big] = \begin{cases} 0.566 \text{ for } j = 1 \\ 0.646 \text{ for } j \geq 2 \end{cases}$$

$$\mathrm{P}\big[S_t \neq \mathrm{RBNS1}\big|D_t = j\big] = 0.684 \text{ for all } j \geq 1$$

$$\mathrm{P}\big[S_t \neq \mathrm{RBNS2}\big|D_t = j\big] = \begin{cases} 0.798 \text{ for } j = 1 \\ 0.684 \text{ for } j = 2 \\ 0.579 \text{ for } j \geq 3 \end{cases}$$

and, for any $l \geq 3$,

$$\mathrm{P}\big[S_t \neq \mathrm{RBNS}l\big|D_t = j\big] = \begin{cases} 0.844 \text{ for } j = 1 \\ 0.684 \text{ for } j = 2 \\ 0.579 \text{ for } j \geq 3. \end{cases}$$

3.3.3 Transition probabilities from RBNP or RBNS states

The exit probabilities obtained in the previous section express the distribution of the sojourn time in each state RBNP and RBNS. The next step is to determine the destination state when a claim leaves the current RBNP or RBNS state. The response is now a binary variable equal to 1 if the claim moves to the terminal state (Closed0 or Closed+) and 0 otherwise. If the response is 0 then the claim moves either from RBNP to RBNS1, or from RBNSj to RBNS$j+1$, depending on its current state. This part of the analysis thus gives the termination probabilities given that the claim leaves its current RBNP or RBNS state.

The estimated termination probabilities from each state RBNP and RBNS are displayed below:

$$\mathrm{P}\Big[S_t \in \{\text{Closed0}, \text{Closed+}\} \Big| S_{t-1} = \text{RBNP}, D_{t-1} = j\Big] = \begin{cases} 0.087 \text{ for } j = 1 \\ 0.794 \text{ for } j \geq 2, \end{cases}$$

$$\mathrm{P}\Big[S_t \in \{\text{Closed0}, \text{Closed+}\} \Big| S_{t-1} = \text{RBNS1}, D_{t-1} = j\Big] = \begin{cases} 0.556 \text{ for } j = 1 \\ 0.844 \text{ for } j = 2 \\ 0.893 \text{ for } j \geq 3, \end{cases}$$

$$\mathrm{P}\Big[S_t \in \{\text{Closed0}, \text{Closed+}\} \Big| S_{t-1} = \text{RBNS2}, D_{t-1} = j\Big] = \begin{cases} 0.556 \text{ for } j = 1 \\ 0.660 \text{ for } j = 2 \\ 0.792 \text{ for } j \geq 3, \end{cases}$$

$$\mathrm{P}\Big[S_t \in \{\text{Closed0}, \text{Closed+}\} \Big| S_{t-1} = \text{RBNS3}, D_{t-1} = j\Big] = \begin{cases} 0.462 \text{ for } j \in \{1, 2\} \\ 0.362 \text{ for } j \geq 3, \end{cases}$$

and for every $l \geq 4$

$$\mathrm{P}\Big[S_t \in \{\text{Closed0}, \text{Closed+}\} \Big| S_{t-1} = \text{RBNS}l, D_{t-1} = j\Big] = \begin{cases} 0.362 \text{ for } j = 1 \\ 0.462 \text{ for } j \geq 2. \end{cases}$$

The probability that the claim terminates with a final payment when leaving the current RBNP or RBNS state is finally modeled with a Binomial GLM resulting in

$$\mathrm{P}\Big[S_t = \text{Closed+} \Big| S_{t-1} = \text{RBNP}, D_{t-1} = j, S_t \in \{\text{Closed0}, \text{Closed+}\}\Big] = 1 \text{ for all } j$$

$$\mathrm{P}\Big[S_t = \text{Closed+} \Big| S_{t-1} = \text{RBNS1}, D_{t-1} = j, S_t \in \{\text{Closed0}, \text{Closed+}\}\Big] = \begin{cases} 0.437 \text{ for } j = 1 \\ 0.403 \text{ for } j = 2 \\ 0.478 \text{ for } j \geq 3 \end{cases}$$

and for all $l \geq 2$,

$$\mathrm{P}\Big[S_t = \text{Closed+} \Big| S_{t-1} = \text{RBNS}l, D_{t-1} = j, S_t \in \{\text{Closed0}, \text{Closed+}\}\Big] = 0.711 \text{ for all } j.$$

3.3.4 Model adequacy

To check model adequacy, we compare the observed average time spent in each state and the corresponding expected time calculated according to the fitted model. The numerical results are displayed next:

Average time	RBNP	RBNS1	RBNS2	RBNS3	RBNSj with $j \geq 4$
Observed	1.4197	1.3126	1.1845	1.1709	1.1219
Model	1.4250	1.3211	1.1855	1.1433	1.1433
Difference	0.0053	0.0085	0.001	−0.0276	0.0220

We can see that model predictions are very close to the observed averages. This shows that the model adequately captures the claim settlement dynamics.

Weights	$\widehat{\pi}_1 = 0.1154$		$\widehat{\pi}_2 = 0.8846$	
	$\widehat{\mu}_1$	Std. Error	$\widehat{\mu}_2$	Std. Error
Time in state $= 0$	5.7932	0.1690	7.4481	0.0285
Time in state≥ 1	4.1257	0.1828	6.9604	0.0328
	$\ln \widehat{\sigma}_1$	Std. Error	$\ln \widehat{\sigma}_2$	Std. Error
	0.9163	0.0351	0.1831	0.0127

Table 3.1: Estimated regression coefficients for the first payment from RBNP to RBNS1 to Closed+ in the two-component LogNormal mixture.

3.4 Estimation of the payment process

3.4.1 Discrete mixture setting

To model the payment process, we need appropriate distributions for both the first payment P_1 and the subsequent link ratios Λ_j. The first payment and link ratios are assumed to obey a finite mixture with two components having respective probability density function $f_k(\cdot)$, $k = 1, 2$. The corresponding density function $f(\cdot)$ is then given by

$$f(y) = \pi_1 f_1(y) + \pi_2 f_2(y),$$

where π_k is the probability that the observed realization comes from component k, with $0 \leq \pi_k \leq 1$ and $\pi_1 + \pi_2 = 1$. Here, $f_1(\cdot)$ and $f_2(\cdot)$ are taken from the LogNormal and Pareto families. The function `gamlssMXfits` comprised into the `R` package `gamlss.mx` is used to fit this model to the available data. The optimal model corresponds to a mixture of two LogNormal distributions for the first payment whereas link ratios are modeled using a combination of a LogNormal distribution and a Pareto distribution.

3.4.2 First payment

Here, we fit the distribution of the first payment, which is associated to the transition from IBNR to RBNS1 or Closed+ or from RBNP to RBNS1 or Closed+. Several models have been tried out and the best-fitting one corresponds to a mixture of two LogNormal distributions with respective parameters (μ_1, σ_1^2) and (μ_2, σ_2^2). Parameter estimates are displayed in Table 3.1 for transition from RBNP to RBNS1 or Closed+ and in Table 3.2 for transition from IBNR to RBNS1 or Closed+ together with the corresponding standard errors. The explanatory variable Time in state corresponds to the one introduced in Subsection 3.2.

In order to assess the goodness of the resulting fit for first payments, we have transformed the observation using the distribution function of the fitted mixture and check whether the resulting values were uniformly distributed over the unit interval (as they should be if the model was correct). Figures 3.1 and 3.2 display the empirical distribution function of these values compared to the 45-degree line corresponding to the unit uniform distribution function. The graph clearly suggests a close agreement. In addition to this visual inspection, a formal Kolmogorov-Smirnov test for uniformity has been performed using the `kolmogorov.unif.test` funtion included in the `R` package `uniftest`. This results in a p-value equal to 0.0747 above the usual 5% level.

Weights	$\widehat{\pi}_1 = 0.0416$		$\widehat{\pi}_2 = 0.9584$	
	$\widehat{\mu}_1$	Std. Error	$\widehat{\mu}_2$	Std. Error
Time in state≥ 0	5.8822	0.3168	7.4254	0.0296
	$\ln \widehat{\sigma}_1$	Std. Error	$\ln \widehat{\sigma}_2$	Std. Error
	1.0003	0.0824	0.1976	0.0172

Table 3.2: Estimated regression coefficients for the first payment from IBNR to RBNS1 or Closed+ in the two-component LogNormal mixture.

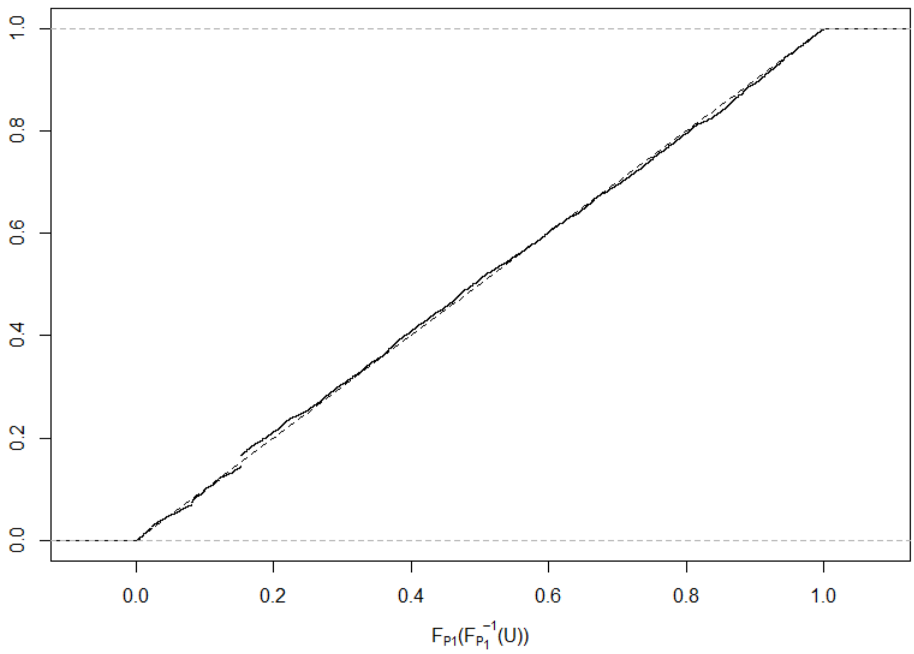

Figure 3.1: Empirical distribution function of the ranks in the two-component LogNormal mixture model for the first payment P_1 from RBNP to RBNS1.

3.4.3 Impact of P_1 on Λ_1 and of C_j on Λ_j

To prevent unrealistically large costs for some claims, there is a need to introduce some dependence between the first payment P_1 and the first link ratio Λ_1. Figure 3.3 displays the observed pairs (P_1, Λ_1) recorded in the database. We clearly see there that a large P_1 cannot be followed by a large Λ_1. Otherwise, the ultimate cost for that claim becomes so large that it has no economic relevance. This is because some claims have large payments at early developments which are followed by moderate ones (such that the corresponding link ratios remain small) whereas other claims exhibit an opposite behavior, with small payments at early developments followed by larger ones at later stages, such that the corresponding link ratios are pretty large.

This is why we introduce a threshold for P_1, with a distribution for Λ_1 that differs according to whether P_1 exceeds this threshold or not. To account for this phenomenon, the state RBNS1 is split into a state RBNS1a where P_1 is smaller than $5,000$ and a state RBNS1b where P_1 exceeds this threshold. The threshold is determined graphically on the basis of Figure 3.3 which shows

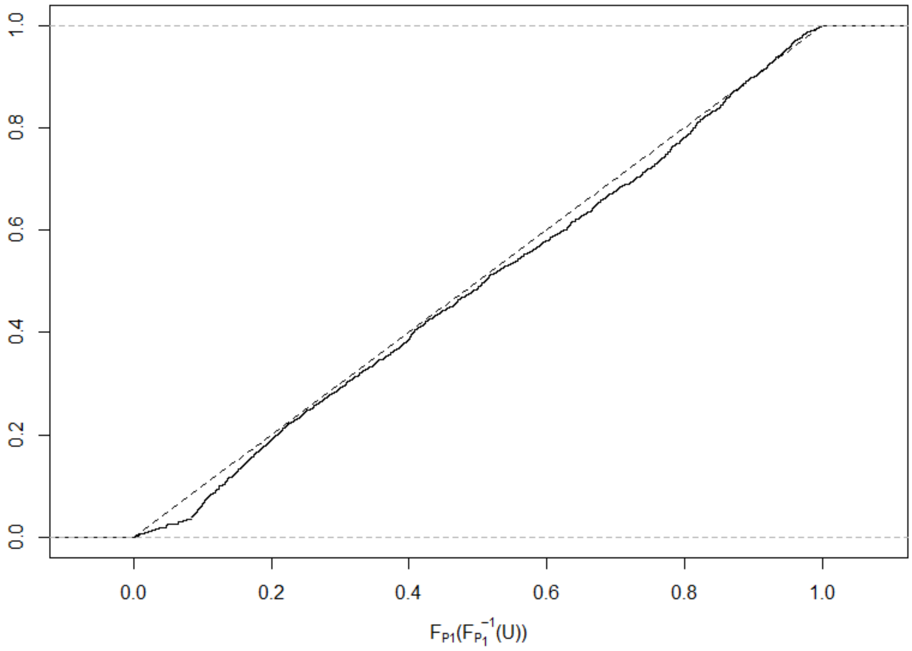

Figure 3.2: Empirical distribution function of the ranks in the two-component LogNormal mixture model for the first payment P_1 from IBNR to RBNS1.

that the majority of the very large link ratios are related to a first payment lower than $5,000$. The latter observation also holds for the next link ratios where large cumulative amounts C_j are typically accompanied with smaller link ratios Λ_j.

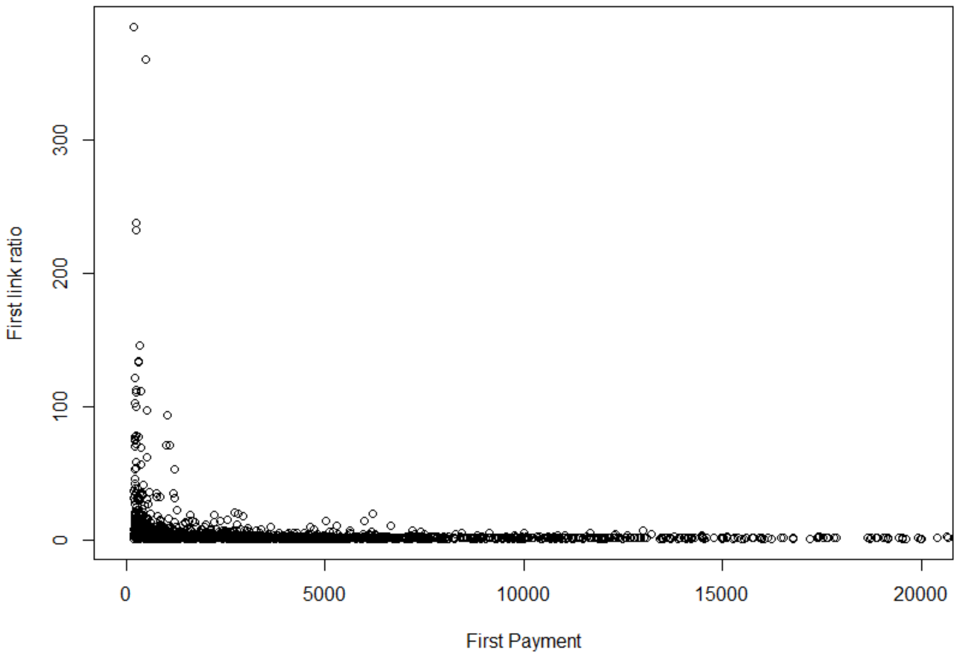

Figure 3.3: Observed pairs (P_1, Λ_1) recorded in the database.

	LogNormal component		Pareto component	
Weights	$\widehat{\pi}_1 = 0.0889$		$\widehat{\pi}_2 = 0.9111$	
	$\widehat{\mu}$	Std. Error	$\widehat{\lambda}$	Std. Error
Time in state 0 First payment $\leq 5,000$	2.6907	0.0758	0.1515	0.0316
Time in state 1 First payment $\leq 5,000$	2.6907	0.0758	0.4448	0.0615
Time in state ≥ 2 First payment $\leq 5,000$	2.6907	0.0758	0.7519	0.088
Time in state 0 First payment $> 5,000$	1.2612	0.1294	1.1328	0.0469
Time in state 0 First payment $> 5,000$	1.2612	0.1294	1.8506	0.1226
Time in state ≥ 2 First payment $> 5,000$	1.2612	0.1294	2.2911	0.1996
	$\ln\widehat{\sigma}$	Std. Error		
	−0.1047	0.0514		

Table 3.3: Estimated regression coefficients for the first link ratio Λ_1.

3.4.4 Link ratios

The first link ratio Λ_1 is best modeled with a mixture of a LogNormal distribution with parameters (μ_1, σ_1^2) and a one-parameter Pareto distribution with parameter λ. The resulting fit is described in Table 3.3 where point estimates are given as well as standard errors.

The next link ratios are also modeled using a finite mixture with LogNormal and Pareto components. As mentioned before, the cumulative payment C_j is split into two categories based on a threshold of 5,000.

Tables 3.4, 3.5 and 3.6 present the fitted models for link ratios Λ_2, Λ_3 and Λ_j, $j \geq 4$. Notice that we do not explicitly associate a link ratio to the transition to Closed+. Instead, if l payments are needed to settle the claim and there is a final payment, the last link ratio corresponds to the transition to Closed+.

In order to assess the goodness of the resulting fit for link ratios, we proceed as for the first payment and transform the observation using the distribution function of the fitted LogNormal-Pareto mixture to check whether the resulting values were uniformly distributed over the unit interval. Figure 3.4 displays the empirical distribution function of these values compared to the 45-degree line corresponding to the unit uniform distribution function. The graph suggests close agreement, except for some moderate departure in the left tail. In addition to visual inspection, A Kolmogorov-Smirnov test for uniformity has been performed with the help of the function `kolmogorov.unif.test` available from the `R` package `uniftest`. Uniformity cannot rejected at usual probability levels, as shown by the p-values equal to 22.57% for Λ_1, 45.02% for Λ_2, 33.95% for Λ_3, and 10.75% for Λ_j, $j \geq 4$.

3.5 Reserve calculations

Now that transition probabilities and payments have been modeled, we can use the multistate model to perform actuarial calculation.

	LogNormal component		Pareto component	
Weights	$\widehat{\pi}_1 = 0.093$		$\widehat{\pi}_2 = 0.907$	
	$\widehat{\mu}$	Std. Error	$\widehat{\lambda}$	Std. Error
Time in state 1 payment $\leq 5,000$	1.5823	0.0619	0.833	0.0501
Time in state 2 payment $\leq 5,000$	2.5291	0.2258	1.0495	0.0895
Time in state 3 payment $\leq 5,000$	1.5823	0.0619	1.0495	0.0895
Time in state 1 payment $> 5,000$	0.942	0.0787	1.5647	0.0693
Time in state 2 payment $> 5,000$	0.5207	0.1246	3.1686	0.1855
Time in state 3 payment $> 5,000$	0.5207	0.1246	3.1686	0.1855
	$\ln\widehat{\sigma}$	Std. Error		
	−0.9347	0.0801		

Table 3.4: Estimated regression coefficients for the second link ratio Λ_2.

	LogNormal component		Pareto component	
Weights	$\widehat{\pi}_1 = 0.104$		$\widehat{\pi}_2 = 0.896$	
	$\widehat{\mu}$	Std. Error	$\widehat{\lambda}$	Std. Error
Time in state 1 payment $\leq 5,000$	0.0328	0.0037	0.7093	0.0784
Time in state 2 payment $\leq 5,000$	0.3785	0.0138	0.7093	0.0786
Time in state 3 payment $\leq 5,000$	0.4539	0.0162	1.7085	0.2577
Time in state 1 payment $> 5,000$	0.0231	0.0049	1.3572	0.1029
Time in state 2 payment $> 5,000$	0.5531	0.0105	1.3572	0.1029
Time in state 3 payment $> 5,000$	0.3008	0.0146	1.7085	0.2577
	$\ln\widehat{\sigma}$	Std. Error		
	−4.1941	0.1259		

Table 3.5: Estimated regression coefficients for the third link ratio Λ_3.

	LogNormal component		Pareto component	
Weights	$\widehat{\pi}_1 = 0.1163$		$\widehat{\pi}_2 = 0.8837$	
	$\widehat{\mu}$	Std. Error	$\widehat{\lambda}$	Std. Error
Time in state 1 payment $\leq 5,000$	0.2204	0.0018	0.754	0.0959
Time in state 2 payment $\leq 5,000$	0.0161	0.0024	0.754	0.096
Time in state 3 payment $\leq 5,000$	0.0962	0.0041	0.754	0.096
Time in state 1 payment $> 5,000$	0.0119	0.0015	1.143	0.0983
Time in state 2 payment $> 5,000$	0.0161	0.0024	1.143	0.0983
Time in state 3 payment $> 5,000$	0.9804	0.0052	1.143	0.0983
	$\ln\widehat{\sigma}$	Std. Error		
	−5.2699	0.1338		

Table 3.6: Estimated regression coefficients for link ratios Λ_j, $j \geq 4$.

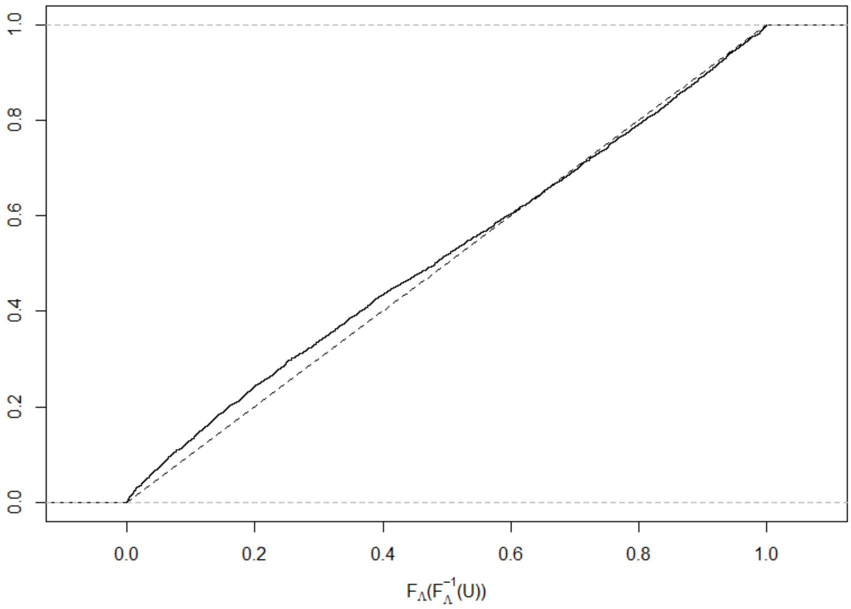

Figure 3.4: Empirical distribution function of the ranks of link factors Λ_j in the two-component mixture with a LogNormal distribution and a Pareto one.

3.5.1 Simulation

Assume that we are at time $t^\star$ and that we want to simulate the total cost of the portfolio. More specifically, we put the portfolio in run-off and we are interested in the remaining costs for all insured events that occurred in accident years up to $t^\star$. We proceed separately for each type of claim, RBNS, RBNP or IBNR claims.

RBNS claims For all claims occupying a RBNS state, we have to simulate the whole trajectory until final settlement. First, we simulate the remaining sojourn time in the current RBNS state. More specifically, if the claim is in state RBNSj at time $t^\star$ and entered that state at time $t^\star - w$, we determine the remaining duration d in state RBNSj as

$$d = \min\left\{k \in \mathbb{N} \middle| \mathrm{P}[S_{t^\star+k} = \mathrm{RBNS}j | S_t = \mathrm{RBNS}j \text{ for } t = t^\star - w, \ldots, t^\star] \leq u\right\}$$

where u is a realization of a random variable U obeying the unit uniform distribution and

$$\mathrm{P}[S_{t^\star+k} = \mathrm{RBNS}j | S_t = \mathrm{RBNS}j \text{ for } t = t^\star - w, \ldots, t^\star]$$

$$= \prod_{l=1}^{k} \mathrm{P}[S_{t^\star+l} = \mathrm{RBNS}j | S_t = \mathrm{RBNS}j \text{ for } t = t^\star - w, \ldots, t^\star + l - 1].$$

Each factor entering the product has been estimated from the data.

At time $t^\star + d$, the claim leaves the state RBNSj and we simulate the transition to RBNS$j+1$ or Closed0/Closed+. If the claim moves to state RBNS$j+1$ then the procedure described above is repeated (with $w = 0$ since the claim just entered the next RBNS state) until the claim finally moves to Closed0/Closed+.

The corresponding link ratios $\Lambda_{j+1}, \Lambda_{j+2}, \ldots$ are also simulated and applied to the current observed total payment C_j for the claim occupying state RBNSj, to obtain its ultimate cost.

RBNP claims For all claims occupying the RBNP state, we first simulate the remaining sojourn time in this state. More specifically, if the claim entered the state RBNP at time $t^\star - w$, we determine the remaining duration d of the sojourn in state RBNP as

$$d = \min\left\{k \in \mathbb{N} \middle| \mathrm{P}[S_{t^\star+k} = \mathrm{RBNP}|S_t = \mathrm{RBNP} \text{ for } t = t^\star - w, \ldots, t^\star] \leq v\right\}$$

where v is a realization of a random variable V obeying the unit uniform distribution and

$$\mathrm{P}[S_{t^\star+k} = \mathrm{RBNP}|S_t = \mathrm{RBNP} \text{ for } t = t^\star - w, \ldots, t^\star]$$

$$= \prod_{l=1}^{k} \mathrm{P}[S_{t^\star+l} = \mathrm{RBNP}|S_t = \mathrm{RBNP} \text{ for } t = t^\star - w, \ldots, t^\star + l - 1].$$

Each factor entering the product has been estimated from the data.

At time $t^\star + d$, the claim leaves the state RBNP and we simulate the transition to RBNS1 or Closed0/Closed+. If the claim moves to state RBNS1 then we proceed as described above for RBNS claims (with $j = 1$ and $w = 0$ since the claim just entered the state RBNS1). The amount of the initial payment P_1 is simulated, as well as the subsequent link ratios $\Lambda_1, \Lambda_2, \ldots$ and the product gives the ultimate cost of the claim.

IBNR claims For all accident years $t \leq t^\star$ for which some claims may still be reported, i.e. such that

$$\sum_{j=1}^{t^\star - t} \beta_j < 1$$

we first simulate the numbers $N_{t,j}$, $j \geq t^\star - t + 1$, of claims originating from insured events which happened in year t but were reported at development j. This is done by generating independent realizations from Poisson random variables with means $\alpha_t \beta_j$. The trajectory of each of these $N_{t,j}$ claims is then simulated as explained before.

For illustration purposes, knowing that the observed maximum number of payments in the database is equal to 11, the maximum number of payments ω is set arbitrarily at 15. Also, to avoid unrealistic trajectory, a claim can not exceed two times of the biggest cumulative payment observed in the database. This amount rises to $1,130,039$ in this case.

Figures 3.5, 3.6 and 3.7 show empirical distributions for the total cost of IBNR claims, RBNP claims and RBNS claims, respectively, obtained from $10,000$ simulations.

3.5.2 Best estimates and risk measures

On each trajectory, the actuary can apply the risk mitigation technique in force within the portfolio under consideration. For instance, if the portfolio is protected by an excess-of-loss reinsurance treaty, the part of the cost exceeding the priority will be paid by the reinsurer. Also, financial discounting can be introduced, if necessary.

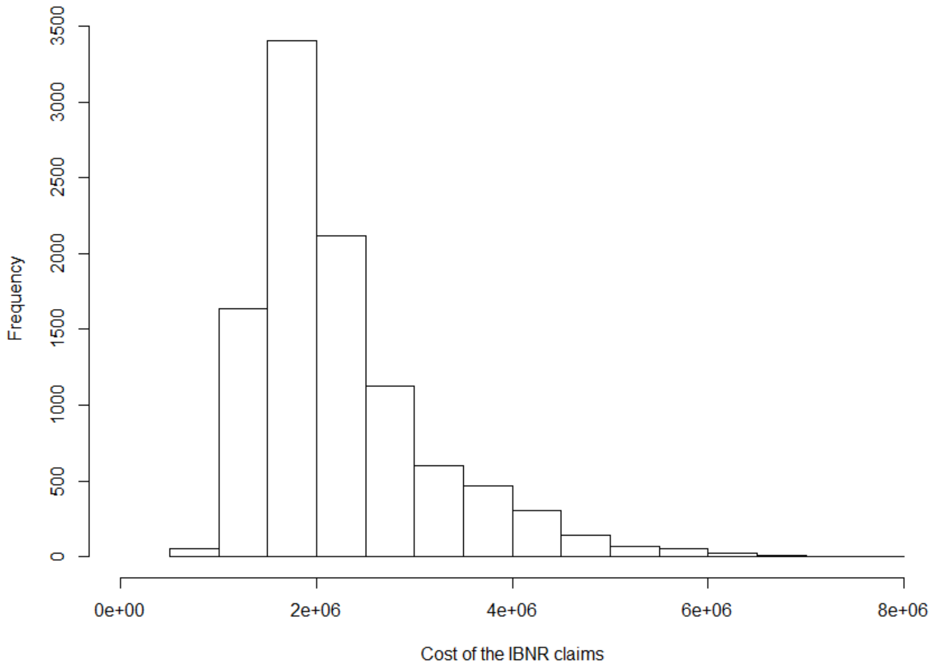

Figure 3.5: Empirical distribution for the total cost of IBNR claims.

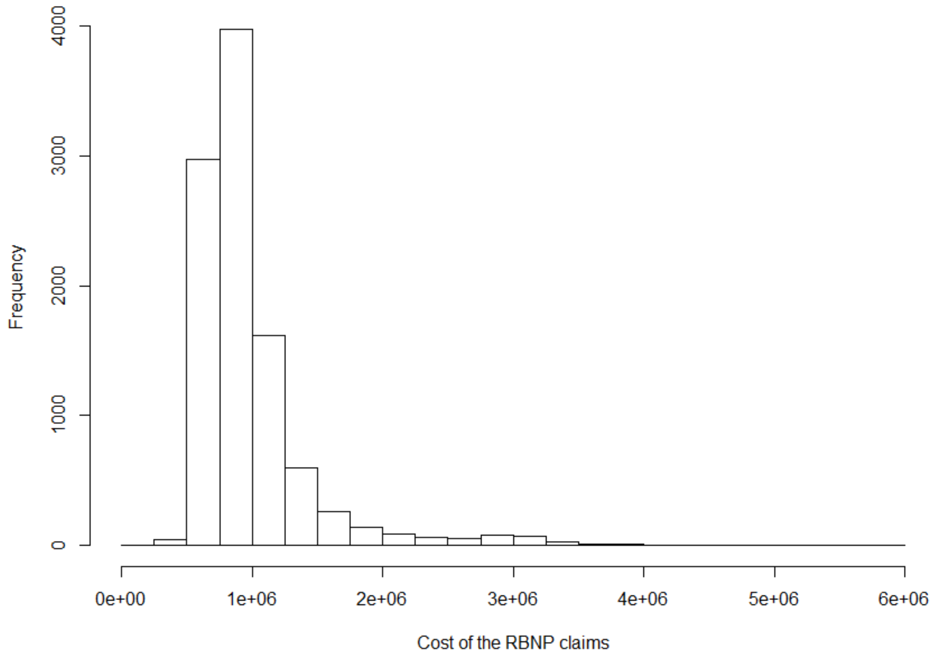

Figure 3.6: Empirical distribution for the total cost of RBNP claims.

After completion of the whole process, we have cash-flows for each claim and thus also for the entire portfolio, in calendar years $t^\star, t^\star + 1, t^\star + 2, \ldots$ until all claims relating to accident year before $t^\star$ are finally settled. Repeating the procedure a large number of times, here $10,000$ times, allows the actuary to estimate best-estimates by averaging the outcomes and to derive risk measures, such as Value-at-Risk (VaR) ant Tail-VaR.

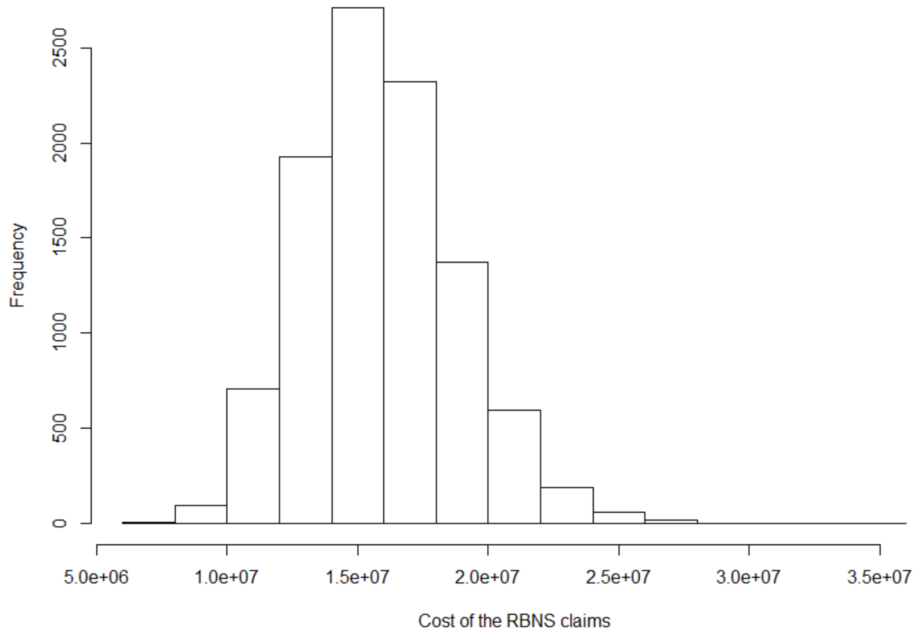

Figure 3.7: Empirical distribution for the total cost of RBNS claims.

	IBNR	RBNP	RBNS	All
Mean	2, 246, 225	975, 122	15, 876, 253	19, 097, 600
VaR 99.5%	5, 820, 638	3, 229, 190	24, 561, 222	28, 166, 869

Table 3.7: Estimated means and VaRs for IBNR, RBNP and RBNS claims as well as for the entire portfolio.

Table 3.7 displays means and VaRs at 99.5% estimated from the 10, 000 simulations for the three types of claims as well as for the entire portfolio.

We notice that the difference between the sum of the VaR at 99.5% for the three types of claims and the VaR at 99.5% for the entire portfolio is given by 5, 444, 181, which can be interpreted as the diversification benefit resulting from the aggregation of the three types of claims.

Chapter 4

Conclusion

In this paper, a Semi-Markov model has been proposed to model the claim settlement process in general insurance. Cash-flows are associated to transitions between states and the total outstanding losses can easily be simulated while analytical calculation is possible for the moments. A case study performed on motor TPL insurance data demonstrates the usefulness of this approach.

Acknowledgements

The authors wish to thank our colleague Cindy Courtois for interesting exchanges about individual loss reserving methods.

References

- Allison, P. D. (1982). Discrete-time methods for the analysis of event histories Sociological Methodology 13, 61-98.

- Antonio, K., Godecharle, E., Van Oirbeek, R. (2016). A multi-state approach and flexible payment distributions for micro-level reserving in general insurance. Available at SSRN: https://ssrn.com/abstract=2777467

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

- Boumezoued, A., Devineau, L. (2017). Individual claims reserving: A survey. Available at `https://hal.archives-ouvertes.fr/hal-01643929/`.

- Denuit, M., Trufin, J. (2017). Beyond the Tweedie reserving model: The collective approach to loss development. North American Actuarial Journal 21, 611-619.

- Denuit, M., Trufin, J. (2018). Collective loss reserving with two types of claims in motor third party liability insurance. Journal of Computational and Applied Mathematics 335, 168-184.

- Hachemeister, C. A. (1980). A stochastic model for loss reserving. Proceedings of the International Congress of Actuaries, 185-194.

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

- Hurlimann, W. (2015). A simple multi-state Gamma claims reserving model. International Journal of Contemporary Mathematical Sciences 10, 65-77.

- Klein, N., Denuit, M., Lang, S., Kneib, Th. (2014). Nonlife ratemaking and risk management with Bayesian additive models for location, scale and shape. Insurance: Mathematics and Economics 55, 225-249.

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

- Norberg, R. (1999). Prediction of outstanding liabilities II. Model variations and extensions. ASTIN Bulletin 29, 5-25.

Chapter 5

About the serie and the authors...

5.1 The Detra Notes

The Detra Notes are a series of educational papers dedicated to the insurance sector. Those notes are published by members of the Detralytics team and written in a clear and accessible language. The team combines academic expertise and business knowledge. Detralytics was founded to support companies in the advancement of actuarial science and the solving of the profession?s future challenges. It is within the scope of this mission that we make our work available through our Detra Notes and FAQctuary?s series.

5.2 Authors' biographies

Carole Bettonville

Carole is part of Detralytics' Talent Consolidation Program (TCP). The TCP offers an opportunity for actuaries with at least 2 years of experience to acquire greater coaching, business and technical skills in order to bring their career to the next level.

During her recent missions, Carole worked in various fields such as calibration of lapse rate and mortality models and a project as part of the IFRS 17 framework, aiming at identifying a Best Estimate allocation key based on some claims characteristics. She worked also on a credibility model for truncation in third party liability. Currently she works on the implementation of the tool Risk Integrity IFRS 17.

Robin Van Oirbeek

Robin holds a Master and PhD degree in Statistics, and soon a Master degree in Artificial Intelligence as well. He has worked for over 5 years in the insurance industry now, with a focus on pricing and reserving, and has mainly published in the field of survival analysis, measures of the predictive quality of a model and reserving. In the academic year 2020-2021, he will be an Invited Lecturer at the University of Antwerp.

Michel Denuit

Michel is Scientific Director at Detralytics, as well as a Professor in Actuarial Science at the Université Catholique de Louvain. Michel has established an international career for some two decades and has promoted many technical projects in collaboration with different actuarial market participants. He has written and co-written various books and publications. A full list of his publications is available at : `https://uclouvain.be/en/directories/michel.denuit`.

Julien Trufin

Julien is Scientific Director at Detralytics, as well as a Professor in Actuarial Science at the department of mathematics of the Université Libre de Bruxelles. Julien is a qualified actuary ofthe Instute of Actuaries in Belgium (IA|BE) and has experience as a consultant, as well as a compelling academic background developed in prominent universities such as Université Laval (Canada), UCL and ULB (Belgium).He has written and co-written various books and publications. A full list of his publications is available at : `http://homepages.ulb.ac.be/~jtrufin/`.

Expertise and innovation at the service of your future

www.detralytics.com - info@detralytics.eu