KU LEUVEN

Methods for claim reserving in non-life insurance: modeling the occurrence, reporting and development of individual claims

Dissertation presented to obtain the
degree of Doctor in Business
Economics

by

Jonas CREVECOEUR

Committee

Advisor:

Prof. Dr. Katrien Antonio   *KU Leuven and University of Amsterdam*

Chair:

Prof. Dr. Gerda Claeskens   *KU Leuven*

Members:

Prof. Dr. Jan Beirlant   *KU Leuven and University of the Free State*

Prof. Dr. Jan Dhaene   *KU Leuven*

Prof. Dr. Ingrid Van Keilegom   *KU Leuven*

Prof. Dr. Tim Verdonck   *University of Antwerp and KU Leuven*

Prof. Dr. Mathias Lindhom   *Stockholm University*

Acknowledgements

This text is the end of a long and at times difficult road towards a PhD. Looking back at the past four years, I realize that the PhD has offered me the chance to meet interesting people, discover new ideas and learn a lot about myself. I'm deeply grateful for the people who shared this journey with me and whose support has contributed tremendously to the completion of this work. It is only right, to dedicate this first chapter to them.

A good supervisor and an interesting topic are key ingredients for any PhD. Thank you Katrien for being the first and providing the later. I'm deeply grateful that you convinced me to start this journey that we completed together. You gave me the freedom to be creative and explore my own research interests, knowing that I could always count on you for guidance when being stuck. I want to thank you especially for allowing me to contribute to a wide range of projects, including collaborations with companies, in-company courses and the COVID-19 project. More than anything, these challenges made me realize where my interests lie and that a master and PhD in actuarial science don't force me to become an actuary.

I strongly enjoyed the collaboration with Argenta Insurance and QBE Re during my PhD. Building pragmatic solutions for actual data has given me a sense of purpose and fulfilment over the years. At Argenta I want to thank Sabine for initiating the project, Maxim for coordinating the research chair and the full Argenta team for their critical, but helpful insights at many meetings. Time after time, these meetings refilled my energy and motivation to complete the project. My history with QBE Re started even before my PhD with an internship. I'm glad that we have kept contact over the years, which has led to an intense collaboration for the last chapter of this thesis. I want to sincerely thank everyone at QBE Re for always making me feel part of the team. Special thanks to Aurélien for repeatedly putting trust in me and our joint project and to Stijn and Alexandre, who took the time to listen to all my chaotic ideas and helped me refine them to a concise story.

While this PhD deals with predicting the future, the COVID-19 pandemic has reminded me that that future remains stochastic with potential outliers beyond my imagination. I'm deeply grateful to Thomas for making living in a pandemic not only the weirdest period, but also the busiest and one of the most rewarding periods of my life so-far. I appreciate the trust you put into me and our almost daily meetings in a period that time was scarce. I'm looking forward to the coming years, collaborating on future challenges that will hopefully be as inspiring.

I would like to thank the members of my supervision committee (Jan Beirlant, Jan Dhaene and Tim Verdonck) for their support over the past years and for sharing their expertise during the doctoral seminars. These suggestions have not only contributed significantly to this thesis, but have also altered the way I will approach outliers, extreme observations or dependency in future projects. Thank you Ingrid Van Keilegom and Mathias Lindhom for joining my doctoral committee and bringing new insights to the preliminary defense.

I'm very grateful to the Fund for Scientific Research-Flanders (FWO) and the Argenta Research chair for financially supporting this doctorate.

I feel extremely lucky to have found a pleasant working environment with nice colleagues at AFI. Before the virus separated us, soup meetings, lunch meetings and hallway chats were an important motivation for me to spend time at the office. Sander and Roel, I'm happy we were able to share an office for such a long time. Sander, as the senior PhD of our office, I want to thank you for always being available for help and advice. I have particularly fond memories of our long morning chats, while waiting for our code to finish. Roel, Tom and Jens, thank you for bringing the Limburg feeling to the office, which immediately made me feel at home. Karim, merci for your patience, while giving me the opportunity to improve my French. Eva and Bavo, thank you for welcoming me in your office on lonely mornings. Eva, I really enjoyed our short morning chats, the perfect start of a working day. Bavo, I appreciate the endless stream of R tips and tricks, a few of which I now apply daily. Roel, sharing my first project with you was the perfect start of the PhD. From you I learned much about what research is and the best practices acquired from our collaboration have helped me tremendously in later years. I will always remember the city trips, i.e. conferences, that I shared with many of you. Some of the best memories involve: a night tour by the Thames, a hunt for the Eiffel tower, a Viennese Waltz and two amazing PARTYs. Thank you for creating and sharing these unforgettable moments.

Staying motivated for the thesis would not have been possible without the support and distractions offered by friends. I'm extremely happy that even during the lockdown and COVID-19 pandemic we have stayed connected.

Contents

Chapter 1

Introduction

Non-life insurance offers individuals and companies the possibility to manage their risk by transferring potential future losses to the insurance company in exchange for a deterministic premium. Such a risk transfer is essential in our modern society as it enables policyholders to take risks, e.g. driving a car, organise events or manufacture products, that could otherwise have significant financial consequences. Managing these risks is the task of the insurance company. For this task, non-life insurers predict the number and size of future claims when setting premiums and set aside funds, the so-called reserve, to cover future losses from policies sold in the past. This work puts focus on quantitative methods and frameworks to estimate the reserve in non-life insurance.

Traditional methods for claim reserving summarize the available claim data in a two dimensional table by aggregating payments by occurrence year and development year, i.e. the number of years elapsed since the occurrence of a claim. In a second step the reserve is estimated with a simple statistical model fitted to the aggregated data. Computational constraints from the past explain the popularity of first aggregating the data, which drastically reduces the number of available data points. However, aggregating the data masks the underlying claim dynamics. This encompasses a risk, when these dynamics vary over time. Detecting and correcting for such changes is a sheer impossible task when only aggregated data is available. Therefore, insurers are nowadays increasingly interested in reserving models that predict the reserve by analysing the development of individual claims.

Figure 1.1 visualizes the development process of a single claim over time. This claim development process starts when an accident happens. After the accident, the insured reports his claim to the insurer at the so-called reporting date. Only

at this point the insurer becomes aware of the existence of the claim. The delay between the occurrence and reporting of the claim is called the reporting delay. These delays are strongly portfolio dependent and can be substantial when the insured does not immediately notice the damage. Once the claim is reported, the insurer reimburses the loss with a single payment or a series of payments.
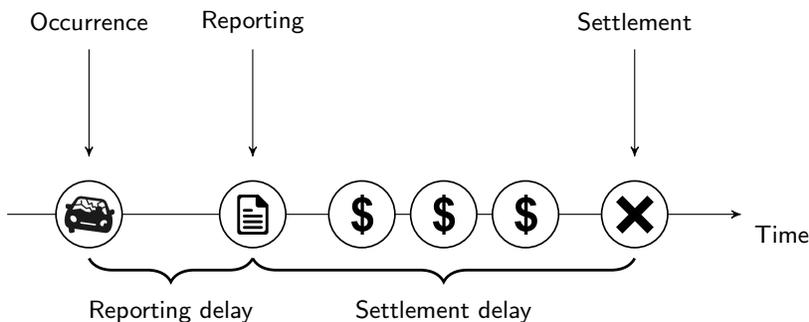
Figure 1.1: Sketch of the development process of a single claim

This thesis presents new tools for estimating the claim reserve by modelling each of the components in the claim development process. At the individual level, reserve computations are often split into a reserve for incurred, but not (yet) reported (IBNR) claims and a reserve for reported, but not (yet) settled (RBNS) claims.

1.1 Modelling the occurrence and reporting of claims

Only reported claims appear in the insurer's claim data set. Since there is no incentive for policyholders to quickly report their insurance claims, a significant fraction of the claims from recent occurrence days is still missing. Correcting the claim data set for these missing claims is an important aspect of IBNR reserving and requires a solid understanding of the claim occurrence and reporting process.

In Chapter 2 we model the occurrence and reporting process using daily data. Modelling the occurrence process with daily data enables our model to flexibly adapt to short term events, such as hail storms, which result in sudden spikes in the occurrence process. Since almost no claims get reported during the weekend and on holidays, a reporting model at daily level should be capable of capturing

these calendar day effects. This chapter develops an intuitive approach, the so-called time change strategy, to model all calendar day effects in the data.

In Chapter 4 we present an alternative approach for IBNR reserving when daily data is not available. This chapter follows an individual reserving approach in which we identify the characteristics of the policyholders, who have not yet reported their claim. Identifying these characteristics enables us to better predict the cost of unreported claims.

1.2   Modelling the development of reported claims

Nowadays insurers collect vast amounts of data over the lifetime of claims. We can divide this information in static claim characteristics (e.g. the cause of the claim) which become available at reporting and dynamic claim characteristics (e.g. the total amount paid, the settlement status of a claim) that can change during the lifetime of a claim. In theory, insurers only need to predict the total amount paid. However, in practice, jointly modelling the amount paid with the other claim characteristics often results in models with a higher predictive accuracy.

In Chapter 3 we present the hierarchical reserving model as an intuitive and flexible approach for jointly modelling claim characteristics. By decomposing the joint model in a set of building blocks, called layers, hierarchical reserving models can be easily adapted to a specific portfolio. An R package accompanies this chapter to assist insurers in implementing this model in practice.

Chapter 4 extends the hierarchical reserving model such that this approach can also be used to predict the cost of IBNR claims. To the best of our knowledge this is the first application of an individual reserving model to reinsurance data. This demonstrates that individual reserving is not necessarily limited to large portfolios with many claims.

1.3   Unifying pricing and reserving methodology

Although both pricing and reserving models rely on data from past claims, insurance companies approach pricing and reserving as separate tasks. Pricing actuaries model the number of claims and (average) claim severity per policyholder, but ignore the fact that exact counts and severities are not available for all policyholders and claims due to delays in the claim development

process (see Figure 1.1). Predicting the future development of claims, and thus completing the claim information, is the task of an individual reserving model.

Chapter 4 presents the occurrence and development model as a unified approach for pricing and reserving. Both of these actuarial tasks benefit from an increased collaboration and knowledge transfer between the pricing and reserving department. In reserving, using policyholder characteristics results in a reserve that is more robust against changes in the portfolio composition. In pricing, better predictions of the future development of claims result in more accurate premium estimates.

The various chapters in this thesis can be found in

  (i) Jonas Crevecoeur, Katrien Antonio, and Roel Verbelen. Modeling the number of hidden events subject to observation delay. *European Journal of Operational Research*, 277(3):930 – 944, 2019. ISSN 0377-2217. URL `https://doi.org/10.1016/j.ejor.2019.02.044`

 (ii) Jonas Crevecoeur and Katrien Antonio. A hierarchical reserving model for non-life insurance claims. 2020a. Available at arXiv: `https://arxiv.org/abs/1910.12692`

(iii) Jonas Crevecoeur and Katrien Antonio. Bridging the gap between pricing and reserving with an occurrence and development model for non-life insurance claims. 2020b. Working paper

The author also contributed to the following original publication

  (i) Roel Verbelen, Katrien Antonio, Gerda Claeskens, and Jonas Crevecoeur. Modeling the occurrence of events subject to a reporting delay via an EM algorithm. 2019. Available at arXiv: `https://arxiv.org/abs/1909.08336`

Chapter 2

Modeling the number of hidden events subject to observation delay

Abstract

We consider the problem of predicting the number of events that have occurred in the past, but which are not yet observed due to a delay. Such delayed events are relevant in predicting the future cost of warranties, pricing maintenance contracts, determining the number of unreported claims in insurance and in modeling the outbreak of diseases. Disregarding these unobserved events results in a systematic underestimation of the event occurrence process. Our approach puts emphasis on modeling the time between the occurrence and observation of the event, the so-called observation delay. We propose a granular model for the heterogeneity in this observation delay based on the occurrence day of the event and on calendar day effects in the observation process, such as weekday and holiday effects. We illustrate this approach on a European general liability insurance data set where the occurrence of an accident is reported to the insurer with delay.

2.1 Introduction

In many domains within operational research analysts are interested in building a stochastic model for the occurrence of events. However, the events of interest are often observed or reported with some delay. Analysts should account for these unobserved events since ignoring them will bias the decisions based on the stochastic model under consideration. Figure 2.1 visualizes this setting. We specify a well defined observation window (on the x-axis) in which we observe the creation of new objects (e.g. products or contracts). Over the course of their lifetimes some objects may experience the event of interest (object 1 and 2 in Figure 2.1) before a given evaluation date, and others will not (object 3 and 4 in Figure 2.1). Upon occurrence the event is initially hidden from the decision maker. The time that elapses between the onset of the object's lifetime and the occurrence of the event is called the event delay. Only after a so-called observation or reporting delay the decision maker becomes aware of the existence of the event. This chapter outlines a data driven strategy to predict the number of events that occurred in the past (before the evaluation date), but which are hidden at the time of evaluation and will only be observed or reported in the future. Subject 2 in Figure 2.1 is an example of such an event.
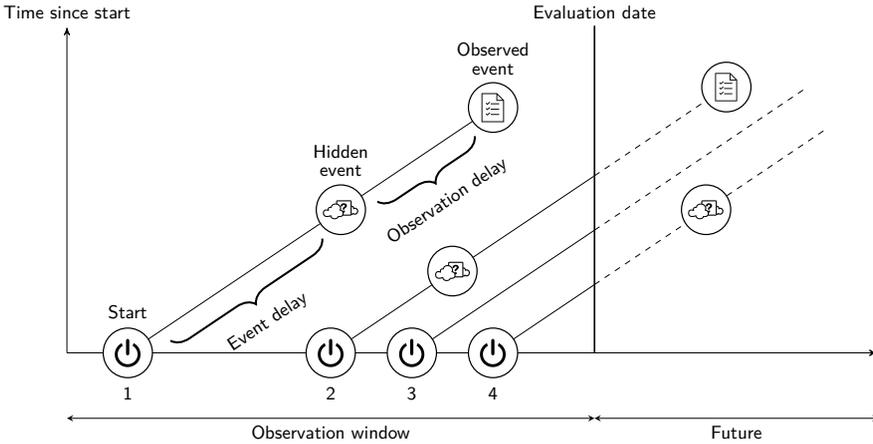

Figure 2.1: Occurrence and observation of events

The modeling of the time to occurrence of an event ('the event delay'), the number of (hidden) events that occurred during a specific time window and the delay between occurrence and observation ('the observation delay') have been active research areas in the literature on operational research, actuarial

science and epidemiology. Typical examples of applications where this predictive problem matters are: a portfolio of maintenance, warranty or insurance contracts, but also an outbreak of a specific disease fits within this framework. We highlight some relevant contributions and explain how this chapter extends the existing literature.

A warranty contract requires the manufacturer to compensate the buyer for all failures occurring within the warranty period. Manufacturers hold capital for future compensations related to goods produced in the past. The amount of capital required depends on the number of defective products that have been sold. Accurate estimation of this number is complicated due to the incompleteness of the data. The diagonal time line in Figure 2.1 begins when a defective product is produced. However, the warranty period only starts when the product is sold to a customer. Manufacturers are typically not aware of these sales and we consider them as a hidden events. Once the defect emerges and the customer calls his warranty contract, the manufacturer is informed of the sale ('the observed event'). Akbarov and Wu [2012] and Ye and Ng [2014] simultaneously model the time to sale and the delay between sale and failure of the product using parametric methods. Since both processes interact in the likelihood, estimation is difficult. Akbarov and Wu [2012] resolve to numerical maximization, whereas Ye and Ng [2014] use a Stochastic Expectation Maximization strategy. While these authors model the time to sale with a simple, parametric distribution without covariates, our framework accounts for the seasonal effects, promotions holidays and weather effects typically present in sales data.

Epidemiologists face a similar statistical problem when modeling the evolution of diseases [Harris, 1990, Salmon et al., 2015]. In this setting, subjects are followed over time and a recent disease infection may remain unobserved due to either delay in disease diagnosis by a medical doctor or incubation time. Modeling these delays allows to take the yet unobserved infections ('the hidden events') into account and thus enables a faster and more accurate identification of disease outbreaks and epidemics [Noufaily et al., 2016].

Maintenance contracts are typically sold together with large industrial appliances. Under these contracts the manufacturer or a third party guarantees the continued use of the equipment. A machine failure ('the observed event') is often the result of previous defects ('the hidden event') which remained unobserved. These defects can be detected by on site inspections and timely repairs will prevent expensive failures or breakdowns of the machine. However, the profitability of these inspections depends largely on the number of hidden defects. Observation delay was first modelled in the context of maintenance contracts by Christer [1973], where it is called delay-time. Since then several papers have focussed on the delay-time concept. Baler and Wang [1993] model delay-time from observed failure data using maximum likelihood estimation. In this approach

both the time to defect as well as the time to observation of the machine failure are tackled with parametric distributions. This literature typically assumes a constant intensity for the occurrences of defects and ignores heterogeneity in the delay-time distribution. Wang [1997] and Apeland and Scarf [2003] rely on expert opinions to formulate a fully subjective delay-time model. Wang [2010] and Berrade et al. [2018] focus on economic decision making when the delay-time distribution is known. In line with the current era of big data analytics (see Mortenson et al. [2015]), our approach goes beyond these assumptions and proposes a data driven strategy to capture heterogeneity in both the occurrence of defects as well as in the delay between a defect and its observation.

The case-study presented in this chapter illustrates our data driven approach with an insurance data set where contracts are sold to policyholders. Some policyholders will be involved in an accident or other type of insured event, while others will not. In insurance parlance the delay between the occurrence ('the hidden event') of an accident and the reporting or filing of the claim to the insurance company ('the observed event') is called the reporting delay. These delays are strongly portfolio dependent and can be substantial when the insured does not immediately notice the damage. In the remainder of the chapter we only consider accidents that will eventually be reported. Accidents that are never reported do not get reimbursed and are not relevant for the balance sheet of the insurer. Once the claim is reported and accepted by the insurer, the insurer reimburses the loss with a single payment or a series of payments. Insurance companies book a reserve to be able to settle the claims that are Incurred But Not yet Reported (IBNR) and refer to this capital as the IBNR reserve. Estimating the number of claims from past exposures that will be reported beyond the evaluation date (the so-called IBNR claim counts) is crucial in setting this reserve. Motivated by computational constraints from the past, many estimation methods in insurance structure the data from Figure 2.1 in a two dimensional table that aggregates the number of accidents by their year of occurrence and year of reporting. We refer the reader to Taylor [2000], Wüthrich and Merz [2008, 2015] for more details on reserving with aggregate methods. Relatively few papers address the problem of specifying a model at granular level for the phenomenon sketched in Figure 2.1. Badescu et al. [2016] and Avanzi et al. [2016] focus on modeling the accident arrival process at a weekly level using Cox processes. These models allow to capture over-dispersion and serial dependence, which is often encountered in such occurrence data. The assumption of independence between the occurrence date and the reporting delay is a disadvantage of the models presented in Badescu et al. [2016] and Avanzi et al. [2016]. Verrall and Wüthrich [2016] were the first to present a model for IBNR counts at a daily level, including the heterogeneity in reporting delays based on the occurrence date of the claim and the strong weekday pattern leading to less claims being reported during the weekend. This weekday pattern

relates to calendar day effects in the reporting process which are difficult to model using classical techniques designed for aggregated data (see Kuang et al. [2008]). Verrall and Wüthrich [2016] provide a method to incorporate this weekday pattern for reporting delays of less than one week. Verbelen et al. [2019] extend this weekday pattern to reporting delays beyond the first week by separately estimating weekly and intra week reporting probabilities. Moreover, Verbelen et al. [2019] present the Expectation Maximization algorithm as a framework for jointly estimating the occurrence and reporting process.

This chapter models the occurrence of hidden events non-parametrically. This allows to capture fluctuations in occurrence counts (for example due to seasonality or weather conditions) without explicitly modeling these events. Moreover, extending the work of Verrall and Wüthrich [2016] and Verbelen et al. [2019] we model the observation delay in the presence of multiple covariates, including calendar day effects. Examples of such calendar day effects are: a reduction in observed events during the weekend, the effect of national holidays and seasonality in observation delay. Our strategy introduces the concept of observation exposure as an intuitive and flexible framework for incorporating (multiple) calendar day effects through regression. This approach elegantly transforms the observation delay distribution by scaling the probability of observing an event on a certain date based on covariates. As such, the transformed observation delay distribution becomes independent of these covariates and is then modelled with a simple, parametric distribution. This makes our approach suitable to a wide range of problems.

This chapter is organized as follows. Section 2.2 describes a statistical framework for modeling the number of hidden events subject to an observation delay. In Section 2.3 we illustrate this approach in a case-study involving an insurance data set. We also investigates the performance of our model in four simulated scenarios. The online appendix provides detailed expressions for implementing the model and links our approach to the non-parametric Kaplan-Meier estimator [Kaplan and Meier, 1958].

2.2 A granular model for the occurrence of events subject to delay

Denote by N_t the number of events occurring on date t, where $t = 1$ is the date of the first event. These events remain hidden until their observation at date s after a delay $s - t$. Let $N_{t,s}$ be the number of events that occurred on date t and are observed on date s. Since all events will be observed at some point in

the future, we find

$$N_t = \sum_{s \geq t} N_{t,s}.$$

Consider an evaluation date τ at which we have to predict the number of hidden events. At τ we split the events from a past occurrence date t into observed ($s \leq \tau$) and hidden events which are not yet observed ($s > \tau$), respectively denoted by

$$N_t^{\text{Obs}}(\tau) = \sum_{s=t}^{\tau} N_{t,s} \quad \text{and} \quad N_t^{\text{Hidden}}(\tau) = \sum_{s=\tau+1}^{\infty} N_{t,s} \quad \text{for } t \leq \tau.$$

We obtain the total number of hidden events by aggregating the unobserved events from all past occurrence dates, i.e.

$$N^{\text{Hidden}}(\tau) = \sum_{t=1}^{\tau} N_t^{\text{Hidden}}(\tau) = \sum_{t=1}^{\tau} \sum_{s=\tau+1}^{\infty} N_{t,s}.$$

This total count is the number that we want to predict. Following Jewell [1990] and Norberg [1993], we formulate two distributional assumptions from which the number of hidden events can be predicted:

(A1) The event occurrence process $(N_t)_{t \geq 1}$ follows an inhomogeneous Poisson distribution with intensity $(\lambda_t)_{t \geq 1}$.

(A2) The observation delay is independent and identically distributed for events occurring on the same date.

Denote by $p_{t,s}$ the probability of observing an event from occurrence date t on date s. We use the notation $p_t^{\text{Obs}}(\tau)$ for the probability that an event from date t is observed by the evaluation date τ. This probability is

$$p_t^{\text{Obs}}(\tau) = \sum_{s=t}^{\tau} p_{t,s}.$$

By assumption (A1) and (A2) the conditions for the Poisson thinning property [Kingman, 1993] are satisfied. The thinning property implies that all $N_{t,s}$ are independent and

$$N_{t,s} \sim \text{Poisson}(\lambda_t \cdot p_{t,s}). \tag{2.1}$$

This allows us to construct the likelihood for the observed data at time τ. Let χ denote the available data, consisting of all events that are observed on the evaluation date τ

$$\chi = \{N_{t,s} \mid t \leq s \leq \tau\}.$$

The loglikelihood of the observed data is

$$\ell(\boldsymbol{\lambda}, \boldsymbol{p}; \boldsymbol{\chi}) = \sum_{t=1}^{\tau} \sum_{s=t}^{\tau} \left[N_{t,s} \cdot \log(\lambda_t) + N_{t,s} \cdot \log(p_{t,s}) - \lambda_t \cdot p_{t,s} - \log(N_{t,s}!) \right] \quad (2.2)$$

where $\boldsymbol{\lambda}$ is a vector with components λ_t for observed occurrence dates t and $\boldsymbol{p} = \{p_{t,s} \mid t \leq s \leq \tau\}$. This chapter puts focus on the observation process without imposing any structure on λ_t. A straightforward computation shows that the loglikelihood in (2.2) is maximal for

$$\lambda_t = \frac{\sum_{s=t}^{\tau} N_{t,s}}{\sum_{s=t}^{\tau} p_{t,s}} = \frac{N_t^{\text{Obs}}(\tau)}{p_t^{\text{Obs}}(\tau)}. \quad (2.3)$$

Replacing λ_t by this expression the loglikelihood in (2.2) becomes

$$\ell(\boldsymbol{p}; \boldsymbol{\chi}) = \sum_{t=1}^{\tau} \sum_{s=t}^{\tau} N_{t,s} \cdot \log(p_{t,s}) - \sum_{t=1}^{\tau} N_t^{\text{Obs}}(\tau) \cdot \log(p_t^{\text{Obs}}(\tau)) + \text{constants}. \quad (2.4)$$

Up to constants this is the loglikelihood for a right truncated observation delay random variable. The truncation point is $\tau - t$, which is the maximal observed delay for an event that occurred on date t.

2.2.1 A time change strategy to model observation delay

We are interested in structuring the observation probabilities $p_{t,s}$ based on covariates corresponding to the occurrence date t and the reporting date s of the event. The probabilistic nature of the data enforces the constraints

$$p_{t,s} \geq 0, \quad \forall t, s \quad \text{and} \quad \sum_{s \geq t} p_{t,s} = 1, \quad \forall t. \quad (2.5)$$

The proposed time change strategy transforms the reporting probabilities such that they can be linked with covariates while preserving these constraints. This transformation is depicted in Figure 2.2, where we consider an event that occurred on a Thursday and for which observation is less likely during the weekend.

First, we view the discrete observation delay as a realization of a continuous random variable U_t under interval censoring. This is graphically illustrated in Figure 2.2a (discrete setting) and 2.2b (continuous setting). Second, we define a time change operator φ_t which assigns a positive length $\alpha_{t,s}$, called the observation exposure, to each combination of an occurrence date t and an observation date s. This time change operator is similar to the concept

of operational time, which is a common technique in continuous financial mathematics, see Swishchuk [2016]. We perceive dates as having variable lengths, whereas prior to this time change an equal length of one time unit was attached to each date. The probability of observing an event on a certain date is scaled by the duration of this date, which motivates calling this length the observation exposure. We define the time-changed delay $\varphi_t(d)$ for an event with occurrence date t and an observation delay of d days as

$$\varphi_t(0) = 0 \quad \text{and} \quad \varphi_t(d) = \sum_{i=1}^{d} \alpha_{t,t+i-1}, \quad d \in \mathbb{N} \setminus \{0\}. \tag{2.6}$$

This is the sum of all observation exposures $\alpha_{t,s}$ assigned to dates in between the occurrence date t and date $t+d-1$. By applying φ_t on the observation delay random variable U_t we obtain a time-changed random variable $\tilde{U} := \varphi_t(U_t)$ which is independent of the occurrence date t of the event. The discrete observation probabilities are easily extracted from this distribution using the relation

$$p_{t,s} = P(U_t \in [s - t, s - t + 1)) \tag{2.7}$$

$$= F_{\tilde{U}}\left(\sum_{i=1}^{s-t+1} \alpha_{t,t+i-1}\right) - F_{\tilde{U}}\left(\sum_{i=1}^{s-t} \alpha_{t,t+i-1}\right).$$

Under the time change transformation the constraints (2.5) become

$$\alpha_{t,s} \geq 0, \quad \forall t, s \quad \text{and} \quad \sum_{s \geq t} \alpha_{t,s} = \infty, \quad \forall t.$$

We specify a regression model for the daily observation exposure as a function of covariates. We set

$$\log(\alpha_{t,s}) = \boldsymbol{x}_{t,s}' \cdot \boldsymbol{\gamma},$$

for a vector $\boldsymbol{x}_{t,s}$ of covariates related to observing on date s an event that occurred on date t and the corresponding parameter vector $\boldsymbol{\gamma}$. In contrast with classical regression methods, the reporting probabilities $p_{t,s}$ not only depend on the characteristics of the observation date, but instead take the full history between the event occurrence and observation date into account through the time change strategy.

Figure 2.2c illustrates this time change. Since less claims get reported during the weekend, we model observation exposure as a function of the reporting day of the week. The time change then assigns lower observation exposures to Saturday and Sunday, hereby transforming the continuous distribution from Figure 2.2b into a time-changed distribution that can be modeled using standard loss distributions.

(a) discrete   (b) continuous   (c) time-changed
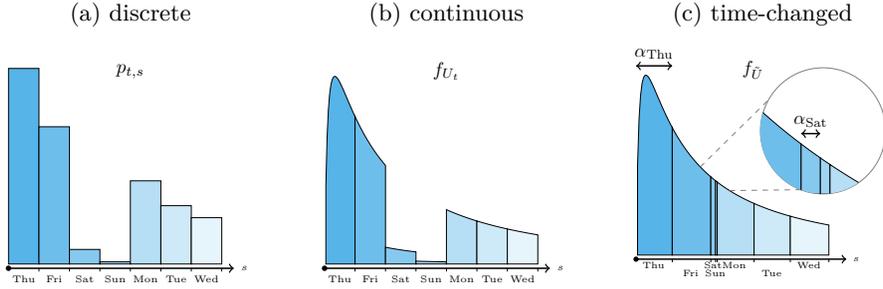
Figure 2.2: Observation delay distribution for an event that occurred on a Thursday. We illustrate **(a)** the discrete observation delay probabilities $p_{t,s}$, **(b)** the density of the continuous observation delay distribution U_t and **(c)** the density of the time-changed observation delay distribution $\tilde{U}$.

2.2.2 Calibration

Our approach divides the observation delay model into two components. The time change transformation φ_t defined in (2.6) captures the heterogeneity in the observation process. This transformation is expressed by the daily observation exposures, which require the calibration of the regression parameters $\boldsymbol{\gamma}$. The time transformed observation delay $\tilde{U}$ is modeled with a simple parametric probability distribution, where the data will assist us in choosing the best candidate. We optimize the loglikelihood in (2.4) with respect to $\boldsymbol{\gamma}$, i.e. we maximize

$$\ell(\boldsymbol{\gamma}; \boldsymbol{\chi}) = \sum_{t=1}^{\tau} \sum_{s=t}^{\tau} N_{t,s} \cdot \log\left[F_{\tilde{U}}\left(\sum_{v=t}^{s} \alpha_{t,v}\right) - F_{\tilde{U}}\left(\sum_{v=t}^{s-1} \alpha_{t,v}\right)\right]$$

$$- \sum_{t=1}^{\tau} N_t^{\mathrm{R}}(\tau) \cdot \log\left[F_{\tilde{U}}\left(\sum_{v=t}^{\tau} \alpha_{t,v}\right)\right],$$

with $\alpha_{t,v} = \exp\left(\boldsymbol{x}'_{t,v} \cdot \boldsymbol{\gamma}\right)$. Online appendix 2.6 describes an optimization strategy for this loglikelihood that is applicable to any sufficiently smooth distribution $F_{\tilde{U}}(\cdot)$. The described strategy is generic and does not immediately take properties from the chosen distribution into account. Significant reductions in computation time can be obtained when $\tilde{U}$ follows a standard exponential

distribution. The loglikelihood then becomes

$$\ell(\boldsymbol{\gamma}; \boldsymbol{\chi}) = -\sum_{t=1}^{\tau} \sum_{s=t}^{\tau} N_{t,s} \cdot \left(\sum_{v=t}^{s-1} \alpha_{t,v} - \log\left(1 - \exp\left(-\alpha_{t,s}\right)\right) \right) \qquad (2.8)$$

$$- \sum_{t=1}^{\tau} N_t^{\mathrm{R}}(\tau) \cdot \log\left(1 - \exp\left(-\sum_{v=t}^{\tau} \alpha_{t,v}\right)\right).$$

The first line in (2.8) is a sum in which each term depends on a single observation exposure, $\alpha_{t,s}$. Since this facilitates computing first and second order derivatives with respect to the reporting exposure, this results in a lower computation time.

2.2.3 Predicting the number of hidden events

At the evaluation date τ we predict the number of events from past occurrence dates t that will be observed on future dates s. Hence our focus is on

$$N_{t,s}, \quad \text{for } t \leq \tau \text{ and } s > \tau.$$

We aggregate these future daily observation counts to find the total number of hidden events

$$N^{\mathrm{Hidden}}(\tau) = \sum_{t=1}^{\tau} N_t^{\mathrm{Hidden}}(\tau) = \sum_{t=1}^{\tau} \sum_{s=\tau+1}^{\infty} N_{t,s}.$$

Following the Poisson assumption in (2.1) each random variable $N_{t,s}$ is independently Poisson distributed with mean

$$E(N_{t,s}) = \lambda_t \cdot p_{t,s}.$$

The observation delay model developed in Section 2.2.1 provides estimates for the observation probabilities $p_{t,s}$, see (2.7)

$$\hat{p}_{t,s} = P(\tilde{U} \in [\varphi_t(s-t), \varphi_t(s-t+1)) \mid \hat{\boldsymbol{\gamma}}).$$

In (2.3) we proposed a pragmatic, non-parametric estimator for the claim occurrence intensity on date t, namely

$$\hat{\lambda}_t = \frac{N_t^{\mathrm{Obs}}(\tau)}{\hat{p}_t^{\mathrm{Obs}}(\tau)}. \qquad (2.9)$$

This estimator depends only on the observed events and the estimated observation delay distribution. This is an advantage when the event generating

process is volatile. For dates with unexpectedly many events the number of observations will be higher and thus we correctly predict more event occurrences. On the downside, (2.9) is less reliable for recent dates when the denominator is close to zero or when the number of daily events is low. When the data set is small, the non-parametric estimator can be replaced by a parametric estimator following the strategy outlined in Bonetti et al. [2016] and Verbelen et al. [2019]. In a parametric framework the estimator for the occurrence intensity may include the daily risk exposure, expressed as the number of policies in effect on a day. Including risk exposure increases the robustness of parametric models to evolutions in the portfolio size and may potentially improve the predictive performance of the model.

2.3 Case-study: reporting delay dynamics in insurance

2.3.1 Data characteristics

We illustrate our approach with the analysis of a liability insurance data set from the Netherlands. The same data is studied in Pigeon et al. [2013], Pigeon et al. [2014] and Godecharle and Antonio [2015] with focus on calculating reserves in discrete time, Antonio and Plat [2014] model reserves in continuous time and Verbelen et al. [2019] who propose a model for the number of hidden claim counts at a daily level. The data registers 506 235 claims related to insured events that occurred and were reported between July, 1996 and August, 2009. From these claims, we remove 75 observations with a reporting date prior to the accident date and 559 claims that are the result of transitions in the reporting system. We focus on the occurrence date of accidents and the corresponding reporting delay in days, i.e. the time (in days) between occurrence of the accident and reporting or filing of the claim to the insurer. To avoid losing valuable insights by aggregation, we study the data at a daily level. This is the most granular timescale at which the data is available.

Occurred accidents Figure 2.3 shows the daily number of accidents that occurred between July, 1996 and August, 2009 and initiated a claim reported to the insurance company before August 31, 2009. Since only claims reported before August 31, 2009 are observed, we see a decrease in observed event counts for the most recent dates which have a substantial number of unreported claims. Two outliers are not shown in this plot, namely 456 accidents on October 27, 2002 and 818 accidents on January 18, 2007. Both outliers correspond to a

storm in the Netherlands causing many insured events.[1] The red line in this figure shows the moving average of the number of occurrences, calculated over the latest 30 days. This trend reveals a seasonal pattern in the occurrence process with more events occurring during the summer months. The trend slightly increases over time due to an increase in portfolio size. Several of the outlying observations in Figure 2.3 correspond to occurrences on the first of January as indicated by the vertical gray bars at the beginning of each year.
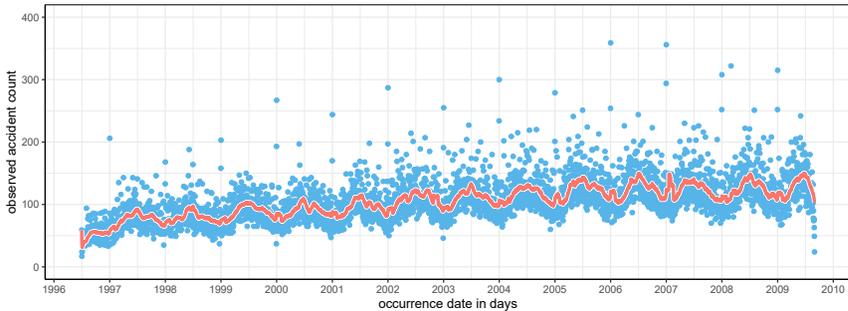
Figure 2.3: Daily number of accidents that occurred between July, 1996 and August, 2009 and were reported before August, 2009. The solid line shows the moving average of occurred accidents, calculated over the latest 30 dates. Two outliers are not shown on the graph: October 27, 2002 (456 accidents) and January 18, 2007 (818 accidents).

Reported claims   Figure 2.4 shows the daily number of claims reported between July 1996 and August 2009. Again the red line shows the moving average of the number of reported claims, calculated over the latest 30 days. The seasonality in event counts observed in Figure 2.3 leads to a similar seasonal pattern in reported claim counts, though with a slight lag due to the delay in reporting a claim. Figure 2.4 reveals two regimes of reporting. On most dates many claims get reported, but there is a substantial number of dates on which few or almost no claims are reported. These dates with few reports correspond to the weekend (Saturday, Sunday) and national holidays.[2] This separation in two regimes is not the case for the occurrence process, since accidents continue to occur during the weekend and on holidays. We further illustrate these calendar day

---

[1] Details (in Dutch) about the storms by the royal national meteorological institute of the Netherlands (KNMI): https://knmi.nl/over-het-knmi/nieuws/storm-van-27-oktober-2002-was-zwaarste-in-twaalf-jaar and https://knmi.nl/over-het-knmi/nieuws/de-zware-storm-kyrill-van-18-januari-2007

[2] List of national holidays in the Netherlands: http://www.officeholidays.com/countries/netherlands/

effects, where reporting is substantially reduced on specific dates, in Figure 2.5. The left hand side lists the average number of reported claims between July, 1996 and August, 2009 on ten national holidays during which all businesses are closed. These averages are compared with the overall daily average of reported claim counts over the observation period. This shows that reporting is strongly reduced on national holidays. We include two non-official holidays, New Year's Eve and Good Friday. These dates show a slight reduction in reporting because many people take a day off from work. The reporting behavior on weekdays is shown in Figure 2.5b. During the weekend and especially on Sunday the number of reports is reduced. These calendar day effects motivate a model for IBNR claim counts at a daily level, capable of incorporating the weekday and holiday effect observed in our empirical analysis.
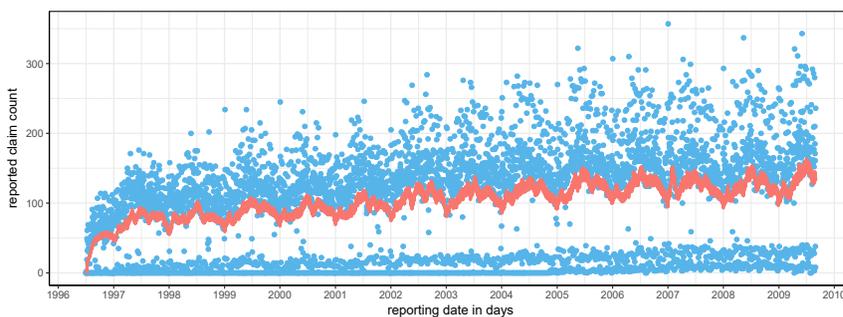
Figure 2.4: Daily number of claims that were reported on each date between July, 1996 and August, 2009. The solid line shows the moving average of reported claims, calculated over the latest 30 dates.

Reporting delay Figure 2.6 illustrates the empirical reporting delay distribution in days over the first three weeks after the occurrence of the insured event. The empirical probability of reporting peaks the day after the claim occurred and strongly decreases afterwards. The increase in reporting after exactly fourteen days is most likely a consequence of data quality issues, where insureds who no longer recall the exact occurrence date report that the accident happened two weeks ago. The same effect to a lesser degree is visible after exactly one week. Figure 2.6b and Figure 2.6c show the empirical reporting delay distribution constructed using only accidents that occurred on Monday and Thursday, respectively. This reveals the effect of the occurrence's day of the week on the reporting delay distribution. An accident that happened on a Monday has a decreased probability of reporting after six or seven days, since these delays correspond to Saturday and Sunday, respectively. Accidents that occurred on a Thursday show the same pattern of reporting delay, but the

(a) national holiday   (b) weekday
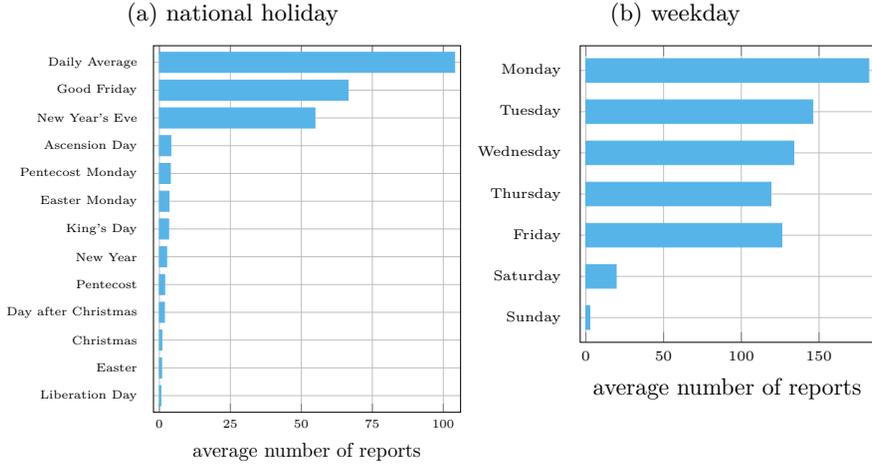
Figure 2.5: Average number of reported claims on **(a)** national holidays and **(b)** weekdays, calculated over all claims that occurred and were reported between July, 1996 and August, 2009.

(a) All claims   (b) Monday   (c) Thursday
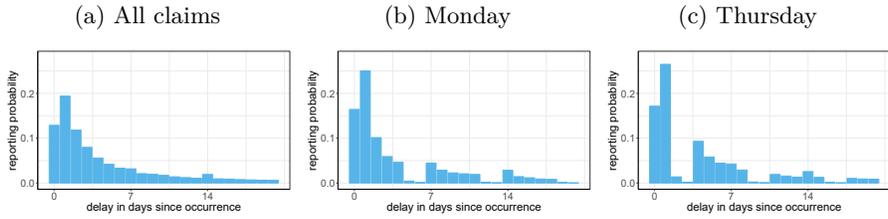
Figure 2.6: Empirical reporting delay distribution in days over the first three weeks after the occurrence of the claim using **(a)** all claims, **(b)** claims that occurred on a Monday and **(c)** claims that occurred on a Thursday.

weekend then corresponds to a different delay. The effect of the weekend is no longer visible in the empirical distribution using all claims (Figure 2.6a), since the weekend then no longer corresponds to a specific reporting delay.

The number of hidden events   The evaluation date refers to the date on which the insurer computes the reserve. In practice this date is often the last day of a quarter or the financial year. Figure 2.7 uses a rolling evaluation date to illustrate the daily number of IBNR claims. For each evaluation date we show the number of claims corresponding to insured events that occurred before this date but were reported afterwards (and before August 31, 2009, the last day of
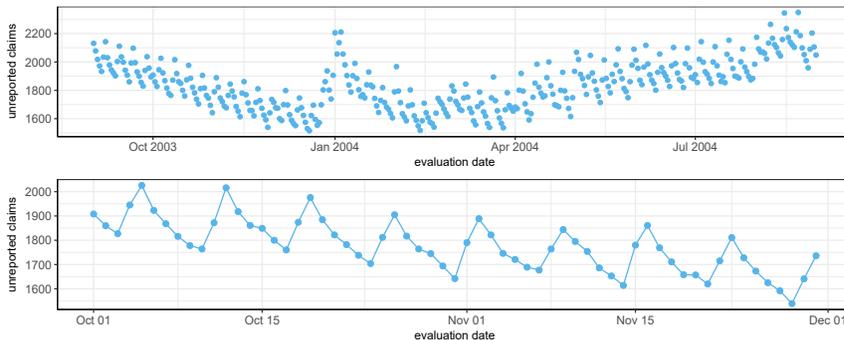
Figure 2.7: Number of unreported claims at each evaluation date between September 2003 and August 2004. These are the number of claims that occurred before this date, but were reported afterwards (but before the end of the observation period, i.e. August 31, 2009). The bottom panel zooms in on evaluation dates in October and November, 2003.

our observation period). The top panel of Figure 2.7 shows the daily number of IBNR claims on each evaluation date between September 1, 2003 and August 31, 2004. The number of unreported claims varies throughout the year with more unreported claims in the summer, when more accidents occur. IBNR counts peak around the start of the new year since many accidents occur on the first of January and reporting is slow due to a clustering of holidays. The bottom panel of Figure 2.7 zooms in on the unreported claims between October 1, 2003 and November 30, 2003. Large fluctuations in unreported claims appear when we evaluate IBNR on a daily basis. These movements follow a seven day pattern where five days of decrease in IBNR are followed by two days of strong upward movement. These upward moves correspond to the weekend when many new insured events occur, but almost no events get reported.

2.3.2 Model specification

We opt for computational efficiency and model the time-changed reporting delay $\tilde{U}$ with an exponential distribution. The reporting exposures include six effects

and are structured as

$$\alpha_{t,s} = \alpha_t^{\texttt{occ.\,dom}} \cdot \alpha_t^{\texttt{occ.\,month}} \cdot \alpha_s^{\texttt{rep.\,holiday}} \cdot \alpha_s^{\texttt{rep.\,month}} \cdot \alpha_{s,s-t}^{\texttt{rep.\,dow,\,first week}} \cdot \alpha_{s-t}^{\texttt{delay}}$$

$$(2.10)$$

$$= \exp\Big(\big(\boldsymbol{x}_t^{\texttt{occ.\,dom}}\big)' \cdot \gamma^{\texttt{occ.\,dom}} + \big(\boldsymbol{x}_t^{\texttt{occ.\,month}}\big)' \cdot \gamma^{\texttt{occ.\,month}}$$

$$+ \big(\boldsymbol{x}_s^{\texttt{rep.\,holiday}}\big)' \cdot \gamma^{\texttt{rep.\,holiday}} + \big(\boldsymbol{x}_s^{\texttt{rep.\,month}}\big)' \cdot \gamma^{\texttt{rep.\,month}}$$

$$+ \big(\boldsymbol{x}_{s,s-t}^{\texttt{rep.\,dow,\,first week}}\big)' \cdot \gamma^{\texttt{rep.\,dow,\,first week}} + \big(\boldsymbol{x}_{s-t}^{\texttt{delay}}\big)' \cdot \gamma^{\texttt{delay}} \Big).$$

We model the impact of the occurrence date on the reporting delay by incorporating effects for the day of the month $\alpha_t^{\texttt{occ.\,dom}}$ and the month $\alpha_t^{\texttt{occ.\,month}}$ on which the accident occurs. The holiday effect in Figure 2.5a is modeled by $\alpha_s^{\texttt{rep.\,holiday}}$, which distinguishes between national and unofficial holidays. Seasonal variations in reporting are captured by $\alpha_s^{\texttt{rep.\,month}}$, which scales reporting exposure based on the month in which the claim is reported. An interaction effect $\alpha_{s,s-t}^{\texttt{rep.\,dow,\,first week}}$ estimates the reporting exposure for combinations of a reporting delay in the first week ($s - t = 0, 1, \ldots, 6$) and the day of the week on which the claim is reported. Separate weekday parameters are estimated for delays of more than one week, $s - t \geq 7$. As such, we capture the weekday effect from Figure 2.5a with additional flexibility in the first week after the claim occurs. Finally, $\alpha_{s-t}^{\texttt{delay}}$ partitions the time elapsed since the accident occurred in 23 bins according to the strategy specified in online Appendix 2.8. These bins adapt the tail of the distribution as well as increase the probability of reporting after 14, 30 and 365 days.

2.3.3   Results

Parameter estimates

We estimate the model parameters by maximizing the loglikelihood in (2.8) using 8 years of data i.e. all accidents that occurred and were reported between July 1, 1996 and September 5, 2004. The resulting training data set contains 274 187 reported claims, for which we model the reporting process using 125 parameters. Figure 2.8 shows the maximum likelihood estimates for the reporting exposure parameters $\exp(\boldsymbol{\gamma})$ in (2.10). Together with these point estimates we plot 95%-confidence intervals derived from the Fisher information matrix for $\boldsymbol{\gamma}$.

Occurrence day of month   Figure 2.8a shows the effect of the day of the month on which the accident occurred. Reporting exposure is lower for accidents that
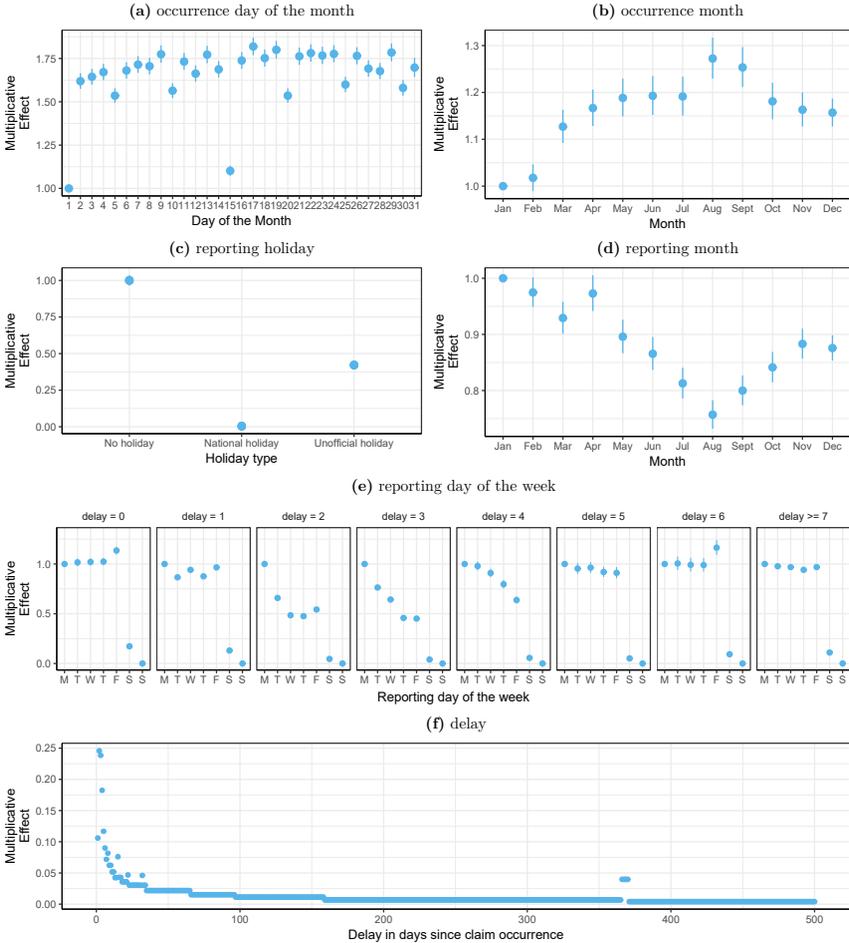
Figure 2.8: Maximum likelihood estimates with 95%-confidence intervals for the reporting exposure parameters $\exp(\boldsymbol{\gamma})$ in (2.10).

occur on the first or fifteenth of the month, which implies that accidents from these days have a longer reporting delay. This is most likely the result of data quality issues. Insureds who report a claim with a long reporting delay might no longer remember the exact occurrence date of the corresponding accident, which leads them to register the occurrence date at the start (first) or middle (fifteenth) of the month. This creates an increase in the average reporting delay for events that occurred on the first and fifteenth of the month. The same effect

to a lesser degree is visible on the 5th, 10th, 20th, 25th and 30th of the month.

Month   Two month effects are included in the reporting exposure structure. Figure 2.8b shows the effect for $\exp(\gamma^{\texttt{occ} \cdot \texttt{month}})$ which considers the month in which the accident occurs. These parameters indicate that reporting is slower for accidents that occurred around the beginning of the year (January, February) and faster in the summer. Figure 2.8d visualizes the parameters for the reporting month, $\exp(\gamma^{\texttt{rep} \cdot \texttt{month}})$. We observe a reduction in reporting exposure during the summer months. Slightly counterintuitive, we find that the parameters $\gamma^{\texttt{occ} \cdot \texttt{month}}$ and $\gamma^{\texttt{rep} \cdot \texttt{month}}$ largely offset each other for accidents that occur and get reported in the same calendar month. When combining these effects, the reduction in reporting exposure during the summer is mostly noticeable for claims that occurred before the summer months.

Holiday   Figure 2.8c shows the effect of holidays on reporting exposure. Hardly any claim gets reported on national holidays and the reporting probability is reduced by more than 50% on unofficial holidays (Good Friday and New Year's Eve). These estimates are of the same magnitude as the effects found in the empirical analysis in Figure 2.5.

Reporting day of the week   We include the day of the week effect in the reporting exposure specification (2.10) through an interaction between the time elapsed after the accident occurred $s - t$ and the day of the week on which the claim is reported. Figure 2.8e shows a grouping of the estimated coefficients based on the time elapsed since the occurrence of the accident. For all delays we notice a reduction in reporting exposure during the weekend, with few reports on Saturday and almost no reports on Sunday. This interaction is important as the estimated parameters differ strongly based on the delay considered. For example, accidents that occur on Friday or Saturday are often reported on the next Monday, which corresponds to a delay of two and three days respectively. Since Monday is the reference level, the fitted parameters for other weekdays are lower at these delays. The right most panel in Figure 2.8e shows the effect of the reporting day of the week for delays beyond one week. For these longer delays, all working days (`Mon - Fri`) have a similar reporting exposure.

Delay   Figure 2.8f shows the evolution of the reporting exposure component $\exp(\gamma^{\texttt{delay}})$ in (2.10) as a function of the time elapsed since the accident occurred. This effect scales the reporting probability at specific delays such that the time-changed reporting delay $\tilde{U}$ better resembles an exponential distribution. We

identified 23 bins upfront based on the strategy of online appendix 2.8. The first eight days after occurrence end up in separate bins. These short delays are important, since many claims get reported soon after their occurrence date. Moreover, Figure 2.8f shows that the calibrated effect changes strongly for these delays. The model also contains bins to capture the increase in reporting probability for delays of exactly 14, 21 and 31 days as well as for reporting after one year. The bin size widens when reporting delay increases. The final two bins $[158, 364]$ and $[370, \infty)$ let the model capture the tail of the distribution.

2.3.4 Out-of-time predictions

We predict the number of hidden events, i.e. the IBNR claim count, following the strategy outlined in Section 2.2.3. Because the non-parametric occurrence estimators are unreliable for recent event dates for which few events are observed, we propose a pragmatic approach to get around this drawbacks. Insurance companies use very specific evaluation dates when calculating reserves, such as the end of a quarter, semester or financial year. Typically the calculations are not performed at those exact evaluation dates, but a couple of days later (at the so-called computation date). Accordingly we predict the number of hidden events on August 31, 2004 using data until September 5, 2004. As such, the granular model predicts 2012.7 unreported claims on August 31, 2004, whereas the true number of IBNR claims (based on data until August 31, 2009) was 2049.

Future observation of hidden events Our daily model splits the total IBNR point estimate of 2012.7 claims by future reporting date. Figure 2.9a shows the estimated number of daily reported claims in September and October, 2004 for accidents that occurred before August 31, 2004. The dashed line in Figure 2.9a indicates the computation date. We do not make predictions for dates falling before the computation date as this data is observed. The model accurately predicts the low report counts during the weekend. This is the merit of adding the day of the week effect in the reporting exposure model. Also the overall reporting pattern closely matches the observed values. Figure 2.9b aggregates these daily report counts by month. This figure shows the estimated number of reported claims in the first twelve months following August, 2004. In these months the observed and predicted IBNR counts are very similar.

Evolution of the number of hidden events The primary focus of our granular model is estimating the total IBNR count. The top panel of Figure 2.10 plots the predicted number of unreported claims on each evaluation date between
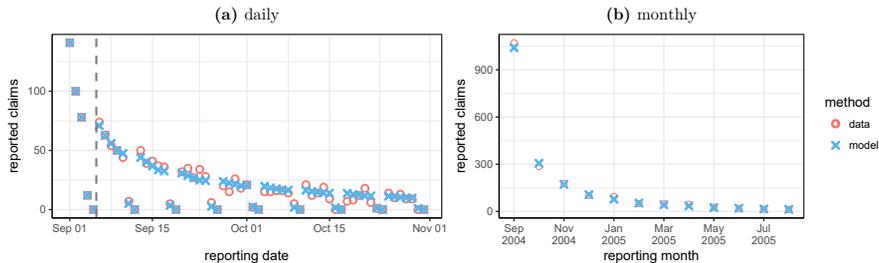
Figure 2.9: Out-of-time prediction of the number of reported claims for accidents that occurred before August 31, 2004. These predictions are compared with the actual number of reported claims. **(a)** Estimated at a daily level for the next two months. The dashed line indicates the last observed date (September 5, 2014). **(b)** Estimates aggregated by reporting month for the next twelve months.

September, 2003 and August, 2004. Each point estimate is an out-of-time IBNR estimate obtained from the granular model calibrated on the historical data available five days after the corresponding evaluation date. We compare these estimates with the actual number of IBNR claims computed from the data until August 31, 2009. Our model recognizes the trend in IBNR counts with more unreported claims during the summer compared to the winter months. The model also correctly predicts an increase in IBNR claims at the start of the year (here: January 1, 2004) as a result of the holidays in this period. The middle panel of Figure 2.10 shows the prediction error, i.e. the difference between the predicted number of IBNR claims and the actual count. The prediction error for the granular model is centred around zero and there are no large outliers. The bottom panel of Figure 2.10 zooms in on the estimates for dates in October and November, 2003. This figure shows that the day of the week parameters allow the model to accurately capture the weekday pattern in IBNR counts.

Benchmark with a model for aggregate data We benchmark our granular approach to Mack's chain ladder method Mack [1993] on aggregated data, which is the industry standard in claims reserving. This method discretizes time and aggregates the observed events into a two dimensional table based on the occurrence period and the discretized reporting delay. A Poisson generalized linear model (GLM) then models the effect of the occurrence and reporting period on these aggregated records. We investigate two aggregation levels, namely aggregating based on a yearly as well as a 28 day grid. We refer to Huang et al. [2015] for a more detailed discussion on reserving with granular
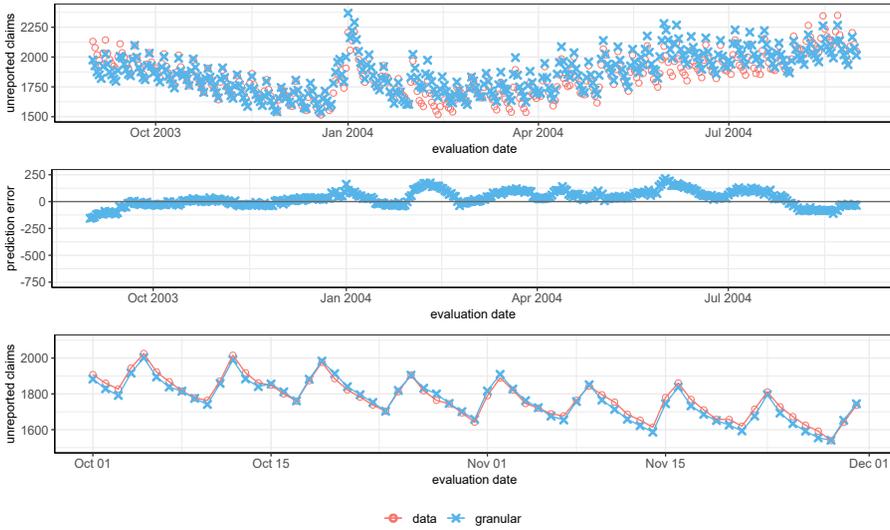
Figure 2.10: Out-of-time prediction of the total IBNR count by the granular reserving method for each evaluation date between September 2003 and August, 2004. These estimates are compared with the observed values using data until August, 2009. The middle panel shows the difference between the predicted and actual IBNR count. The bottom panel zooms in on the estimates in October and November, 2003.

data versus data aggregated in two dimensional tables. Figure 2.11 shows the estimated IBNR counts under both chain ladder implementations evaluated on each date between September, 2003 and August, 2004. Both versions of the chain ladder detect the seasonal pattern in unreported claim counts, which is related to seasonality in the occurrence process. The end of the year holidays and corresponding increase in IBNR counts is a yearly seasonal effect in the reporting process. The chain ladder assumptions allow for seasonal effects when the period of seasonality coincides with the discretized time periods. For this reason, the yearly chain ladder method correctly predicts an increase in IBNR counts around the end of the year, whereas the 28 day chain ladder method severely underestimates IBNR counts for these dates. The bottom panel of Figure 2.11 zooms in on the period October to November 2003. The 28 day chain ladder method retrieves the day of the week effect, since the length of every bin is a multiple of 7 and therefore contains the same weekdays. The yearly chain ladder method has bins with either 365 or 366 days. Since both bin sizes are not divisible by 7, the yearly chain ladder method is unable to recognize the day of the week effect. This results in a systematic overestimation
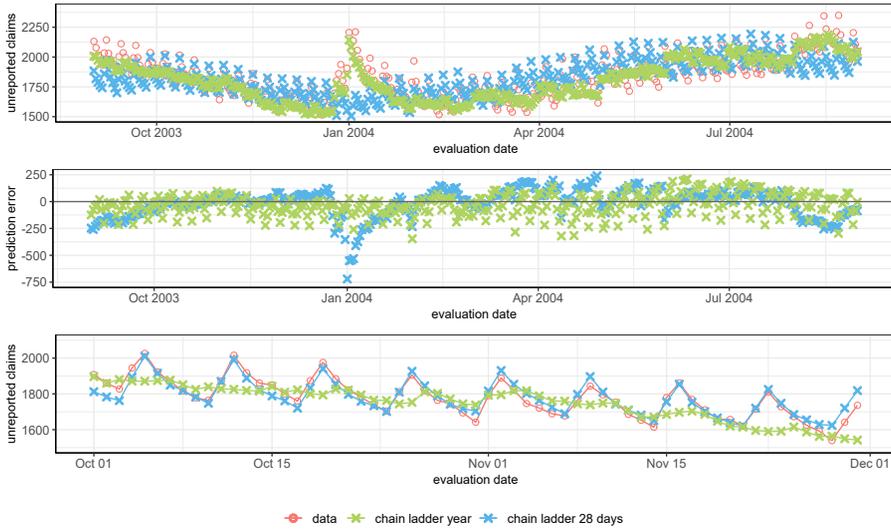
Figure 2.11: Out-of-time prediction of the total IBNR count by the yearly and 28 day chain ladder methods for each evaluation date between September 2003 and August, 2004. These estimates are compared with the observed values using data until August, 2009. The middle panel shows the difference between the predicted and actual IBNR count. The bottom panel zooms in on the estimates in October and November, 2003.

of IBNR counts on Fridays and an underestimation on Sunday. The middle panel of Figure 2.11 shows the difference between the predicted and actual IBNR count. The inability of the 28 day chain ladder to capture the holiday effect results in large underestimations around this time of the year. The yearly chain ladder overall performs better, but the prediction error is sensitive to the day of the week on which the reserve is calculated. Capturing the holiday and the day of the week effect simultaneously requires a model specified at the daily level. The chain ladder method assumes independence between the reporting delay distribution and the occurrence period of the claim. Since Figure 2.5 and 2.6 indicate that this assumption is not valid at the daily level, a daily chain ladder would not perform well. Our granular method explains both phenomena together by abandoning this independence assumption.

2.3.5   Scenario testing

Investigated scenarios

We further evaluate our approach with portfolios simulated along four different scenarios. Each scenario generates data from an insurance portfolio from January 1, 1998 onwards. Figure 2.12 outlines the structure of these data sets. The insurer observes the claims that are reported before the computation date (the gray area in Figure 2.12) and predicts the number of claims that were not yet reported on the evaluation date (the hatched area in Figure 2.12). We consider two evaluation dates (December 31, 2003 and August 31, 2004) to visualize the impact of holidays near the end of the year on the accuracy of IBNR claim count predictions. The four scenarios focus on characteristics of the portfolio or the claim handling process that have an impact on the total IBNR count. Figure 2.13 visualizes the occurrence, reporting and IBNR processes for a single simulated data set from each of the four scenarios.
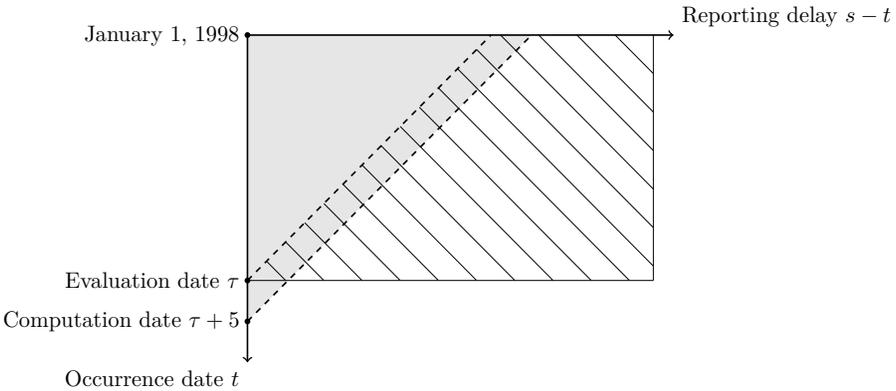
Figure 2.12: Structure of a simulated data set. We simulate accidents that occur between the first of January, 1998 and the computation date, together with their associated reporting delay. The gray area shows the data that is used to fit the model and to predict the hatched area, which consists of the number of unreported claims at the evaluation date τ. We obtain perfect predictions for the intersection of the gray area and the hatched area, since in this region the reported counts are observed.

Scenario 1: Baseline scenario   This is the basic scenario from which the other three scenarios will slightly deviate. The occurrence of insured events follows a Poisson distribution with an average of 100 claims on each occurrence date. For

these occurrences the reporting delay is simulated along the model specification outlined in Section 2.2, i.e. the distribution of the time-changed reporting delay $\tilde{U}$ follows a lognormal distribution with density

$$f_{\tilde{U}}(u) = \frac{1}{u\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\cdot\left(\frac{\ln(u)-\mu}{\sigma}\right)^2},$$

where $\mu = 0$ and $\sigma = 1$. The daily reporting exposure depends only on the reporting date and is given by

$$\alpha_{t,s} = 0.10 \cdot (0.20)^{\mathbb{1}_{s\in\text{Sat}}+\mathbb{1}_{s\in\text{unofficial-holiday}}} \cdot (0.01)^{\mathbb{1}_{s\in\text{Sun}}+\mathbb{1}_{s\in\text{national-holiday}}},$$

where `Sat`, `Sun`, `national-holiday` and `unofficial-holiday` are the sets of all Saturdays, Sundays, national holidays and unofficial holidays respectively. As such, the reporting probability is reduced by 80% on Saturdays and unofficial holidays and by 99% on Sundays and national holidays. These effects are of the same order as those found in the exploratory data analysis, see e.g. Figure 2.5 in Section 2.3.1 and result in an average reporting delay of slightly more than three weeks. The top row of Figure 2.13 visualizes a simulation from this baseline scenario. The middle panel shows two regimes of reporting, where the days with few reported claims correspond to the weekend and holidays.

Scenario 2: Volatile occurrences In this scenario external causes, such as the weather, have a large effect on the number of accidents that occur on a given date. The environment can be in two states, a good state with an average of 100 accidents per day and a bad state in which there are on average 400 accidents. The transitions between these states follow a Markov process with transition matrix

$$\begin{array}{cc} \text{from/to} & \begin{array}{cc} \text{good} & \text{bad} \end{array} \\ \begin{array}{c} \text{good} \\ \text{bad} \end{array} & \left(\begin{array}{cc} 0.9 & 0.1 \\ 0.6 & 0.4 \end{array}\right). \end{array}$$

The model starts in the good state and then occasionally moves to the bad state. From this bad state there is a large probability of returning to the good state with less occurrences on average. The second row of Figure 2.13 (lhs) visualizes the impact of this bad state on the occurrence process. The reporting delay distribution is the one described in the baseline scenario.

Scenario 3: Low claim frequency This scenario illustrates the effect of a strong reduction in the number of occurred accidents. The occurrence process is modeled by a Poisson distribution with a daily average of two claims. The reporting model from the baseline scenario is used. This scenario is visualized in the bottom row of Figure 2.13. We observe that a low number of accidents leads to more volatility in the IBNR process.

Scenario 4: Online reporting   In this scenario the insurer introduces an online tool for claim reporting. This online tool is launched at January 1, 2003 and increases the number of reports in the weekend and on holidays. The new reporting exposures become

$$
\alpha_{t,s} = \begin{cases} 0.10 \cdot (0.20)^{\mathbb{1}_{s \in \mathtt{Sat}} + \mathbb{1}_{s \in \mathtt{Unofficial\text{-}holiday}}} \cdot (0.01)^{\mathbb{1}_{s \in \mathtt{Sun}} + \mathbb{1}_{s \in \mathtt{Holiday}}} & s < 01/01/2003 \\ 0.10 \cdot (0.50)^{\mathbb{1}_{s \in \mathtt{Sat}} + \mathbb{1}_{s \in \mathtt{Unofficial\text{-}holiday}}} \cdot (0.20)^{\mathbb{1}_{s \in \mathtt{Sun}} + \mathbb{1}_{s \in \mathtt{Holiday}}} & s \geq 01/01/2003 \end{cases}.
$$

This reporting model is combined with the same occurrence process as in the baseline model, that is a Poisson process with a constant intensity of 100 claims each day. The bottom row of Figure 2.13 visualizes a simulation from this scenario. A vertical black line indicates the breakpoint on January 1, 2003. After the introduction of online reporting we no longer observe dates with zero reports.
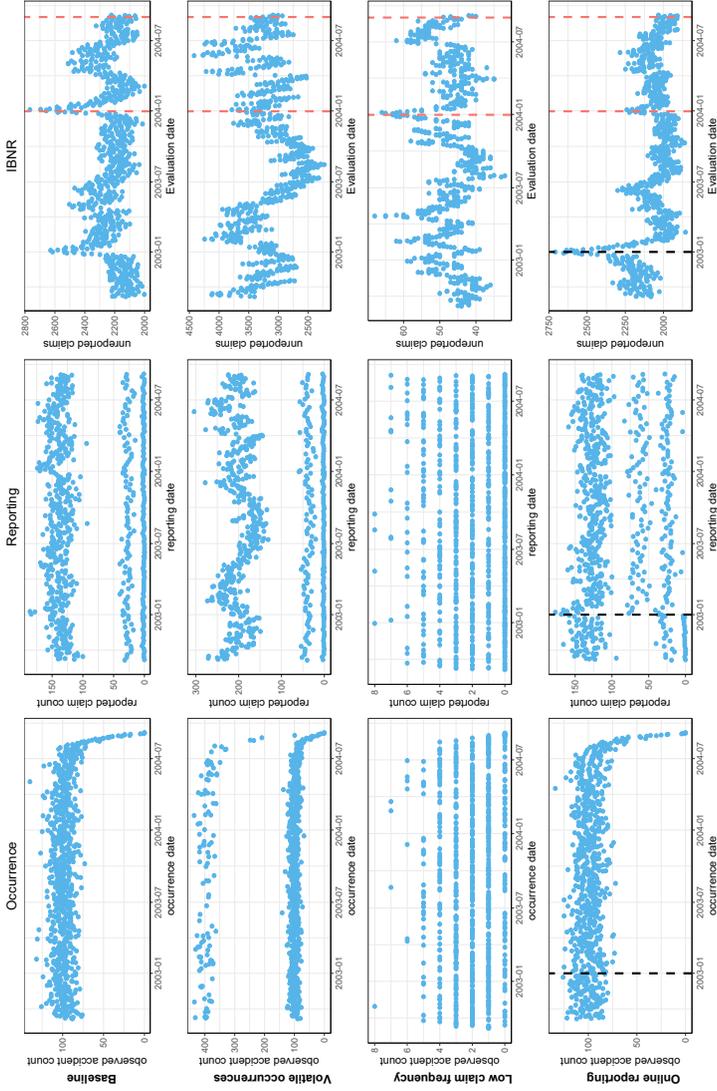
Figure 2.13: Each row visualizes a simulated data set from one of the four scenarios. The left column shows the daily number of accidents that were reported by August 31, 2004 (cf. Figure 2.3). The middle column shows the daily number of reported claims (cf. Figure 2.4). The right column visualizes the number of unreported accidents using a rolling evaluation date (cf. Figure 2.7). The red dashed lines in the IBNR plots indicate the evaluation dates of December 31, 2003 and August 31, 2004.

Calibrated models: granular versus aggregate

We compare the accuracy of the predictions of the hidden event counts using three models, namely the exact granular model from which we simulated the data, an approximate granular model and a model for yearly aggregated data. The historical information (gray area in Figure 2.12) is used to predict the number of IBNR claims (hatched area in Figure 2.12). Under the granular approach these predictions naturally extend to delays beyond those yet observed, whereas in the aggregate approach we limit the prediction window to the longest observed delay. We consider a gap of five days between the computation and the valuation date. The observations from these five days improve the prediction of the occurrence intensities λ_t and the reporting probabilities $p_{t,s}$, whereas there is no straightforward way to incorporate this data in the method for yearly, aggregated data. The ability to use this additional data is one of the advantages of the granular approach.

Exact granular model We use our knowledge of the shape of the distribution and reporting exposure structure behind the various scenarios and calibrate the exact same model for reporting delay on the historical data. Hence we estimate the variance parameter in the lognormal distribution for the smoothed reporting delay $\tilde{U}$ and the parameters $\boldsymbol{\gamma}$ for the covariate effects in the reporting exposures $\alpha_{t,s}$. The reporting exposure $\alpha_{t,s}$ changes the scale of the time axis which is similar to the effect of the scale parameter $\exp(\mu)$ of the lognormal distribution. We avoid identifiability issues by setting μ equal to zero. The occurrence process is modeled non-parametrically as described in Section 2.2.

Approximate granular model This model considers the more realistic situation where the insurer wants to fit the model of Section 2.2, but is unaware of the exact underlying distribution. Motivated by computational benefits the insurer chooses an exponential distribution for the smoothed reporting delay $\tilde{U}$, and structures the reporting exposures as

$$\alpha_{t,s} = \alpha_s^{\texttt{dow}} \cdot \alpha_s^{\texttt{holiday}} \cdot \alpha_{s-t}^{\texttt{delay}} \tag{2.11}$$

$$= \exp((\boldsymbol{x}_s^{\texttt{dow}})' \cdot \gamma^{\texttt{dow}} + (\boldsymbol{x}_s^{\texttt{holiday}})' \cdot \gamma^{\texttt{holiday}} + (\boldsymbol{x}_{s-t}^{\texttt{delay}})' \cdot \gamma^{\texttt{delay}}).$$

In this specification $\alpha_s^{\texttt{dow}}$ captures the day of the week effect, $\alpha_s^{\texttt{holiday}}$ identifies national and unofficial holidays and $\alpha_{s-t}^{\texttt{delay}}$ adapts reporting exposure based on the time elapsed since the claim occurred. For a single simulated data set we bin reporting delay in 13 bins according to the strategy outlined in online appendix 2.8. These same bins are then reused to construct the delay covariate

for all other simulations. In the fourth scenario (online reporting), we estimate different parameter values for the parameters $\gamma^{\texttt{dow}}$ and $\gamma^{\texttt{holiday}}$ for reporting dates before and after January 1, 2003.

A model for aggregated data: the chain ladder The chain ladder method described in Section 2.3.4 is the industry standard for predicting the number of unreported claims. We aggregate the simulated data by calendar year and benchmark our granular approach to the chain ladder method on this aggregated data.

Results and discussion

We evaluate the performance of the reserving models by predicting the total number of IBNR claims at the evaluation date, which corresponds to the hatched area in Figure 2.12. This prediction is compared with the actual number of unreported claims as observed in the simulated data set. We simulate 1000 data sets and calibrate the three models outlined in Section 2.3.5 on each of these. The prediction accuracy is measured by the percentage error (PE), i.e.

$$\text{PE} = 100 \cdot \frac{N^{\text{IBNR}}(\tau) - \widehat{N^{\text{IBNR}}}(\tau)}{N^{\text{IBNR}}(\tau)}.$$

Positive percentage errors reflect underestimation, whereas negative values indicate an overestimation of IBNR counts. Table 2.1 shows the mean and standard deviation of the percentage error for the two granular models and the chain ladder method. In Figure 2.14 boxplots of the percentage error visualize the model performance across the four scenarios.

Impact of evaluation date We observe in all four scenarios an increase in unreported claims on New Year's Eve (see the last column in Figure 2.13). This is the result of multiple holidays at the end of the year, which prevents clients from reporting their claim. We compare the average percentage error in Table 2.1 on December 31, 2003 and August 31, 2004 to quantify the impact of these holidays on prediction accuracy. The exact granular model fits the distributional specification that was used in the simulation. Therefore this model can perfectly capture the effect of holidays and has an average error close to zero on both dates. Seasonal effects do not violate the chain ladder assumptions when their seasonal cycle coincides with the chain ladder period. Since the end of the year holidays can be seen as a yearly seasonal event they do not affect the prediction accuracy in the yearly chain ladder method. This explains the fairly similar

Scenario	Eval. date	exact granular		approx. granular		chain ladder	
		μ	σ	μ	σ	μ	σ
Baseline	31 Dec 2003	-0.09	3.17	4.85	2.75	2.70	2.17
	31 Aug 2004	-0.01	2.75	-0.18	2.82	1.20	2.36
Volatile occurrences	31 Dec 2003	0.11	2.64	5.01	2.93	0.16	15.52
	31 Aug 2004	-0.04	2.27	-0.20	2.51	-0.82	14.90
Low claim frequency	31 Dec 2003	-0.69	23.89	4.42	20.85	1.65	16.25
	31 Aug 2004	-2.30	20.19	-2.52	20.72	-1.33	17.96
Online reporting	31 Dec 2003	-0.13	3.12	2.93	3.07	-12.46	2.91
	31 Aug 2004	0.02	2.80	0.73	2.89	-7.00	2.68

Table 2.1: Evaluation of the mean and standard deviation of the percentage error of the exact granular model, the approximate granular model and the chain ladder method across four different scenarios and two evaluation dates.
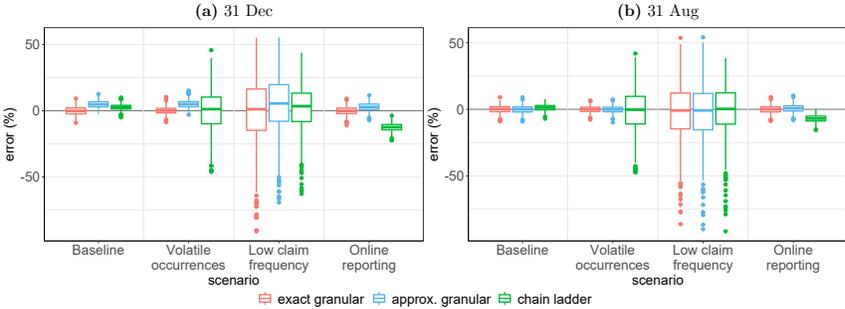

Figure 2.14: Boxplots of the Percentage Error (PE) of the IBNR estimate across the four scenarios and on both evaluation dates.

errors on both evaluation dates for the chain ladder method. Table 2.1 reveals an underestimation of IBNR counts for the approximate granular model on December 31 across all four scenarios. The data is simulated with a lognormal distribution for the smoothed reporting delay, whereas in the approximate granular model we fit an exponential distribution. Since these distributions are quite different, we include a delay effect $\alpha_{s-t}^{\texttt{delay}}$ in (2.11). This effect can increase the reporting probability at specific delays, hereby moving the time-changed data closer to an exponential distribution. However, the delay covariate can not remove all differences between these distributions and this leads to a small underestimation on December 31, 2004 in all scenarios. For all three models the choice of evaluation date does not influence the standard deviation of the

percentage error.

Baseline   The top row of Figure 2.13 visualizes a single data set from the baseline scenario. Both the occurrence and reporting process are stable. This leads to a yearly periodical pattern in IBNR counts, which is easy to predict. Since all three models perform well (see Figure 2.14), there is no reason to replace the chain ladder method by a granular model in this scenario.

Volatile occurrences   The range of IBNR values encountered throughout a year is much wider in this scenario compared to the other three scenarios. Table 2.1 and Figure 2.14 show that the performance of the granular models is in line with their performance in the baseline scenario. The occurrence process has little effect on the prediction accuracy, since we model the occurrence process non-parametrically. The chain ladder method performs well on average, but the standard deviation has risen compared to the baseline scenario. In over half of the cases the chain ladder produces an error of more than 10% when predicting the number of unreported claims. The chain ladder method aggregates claims by occurrence year, hereby losing the exact occurrence information. When the model was in the bad state on the evaluation date, this leads to large underestimations of total IBNR counts. This scenario identifies an unstable accident occurrence process as a reason for considering a granular model.

Low claim frequency   The occurrence frequency is reduced from an average of hundred daily claims to only two claims. The third row of Figure 2.13 visualizes a data set from this scenario. Since on average only two accidents occur per day, our predictions for the intensities λ_t in the occurrence process are less reliable. As seen in Figure 2.14 this leads to large prediction errors for all models. This uncertainty follows mostly from the Poisson assumption (A1) in the data generation process. The coefficient of variation $\frac{\sigma}{\mu}$ for a Poisson distribution with intensity λ is given by $\frac{1}{\sqrt{\lambda}}$. A lower intensity in the Poisson proccess corresponds with a larger coefficient of variation and thus more uncertainty in the data. We conclude that accurate estimation of the number of hidden events is only possible when the expected number of events is sufficiently large.

Online reporting   On January 1, 2003 the insurer introduces an online tool to report claims, which creates a breakpoint in the reporting process. The granular model performs well on both evaluation dates, since we estimate different exposure parameters after the breakpoint. Both evaluation dates correspond with around one year of post breakpoint data, which is insufficient for applying

the chain ladder method. Therefore, we calibrate the chain ladder method on all the available data, which leads to an overestimation of the IBNR counts. This scenario illustrates the benefits of a granular reserving model, when breakpoints can be identified in the data.

2.4 Conclusion

We propose a new method to model the number of events that occurred in the past, but which are not yet registered due to an observation delay. Our approach provides an elegant and flexible framework for modeling the observation delay subject to calendar day covariates by introducing the concept of observation exposure. This framework can be applied for predicting the future cost of warranties, pricing maintenance contracts and many other applications in operational research where events are observed with a delay. We illustrate our method in an extensive insurance case-study. Compared to methods designed for aggregated data our granular approach has three advantages. First of all, introducing covariates gives insight into the observation process. Second, our granular model can predict the expected number of observations for each future date. This enables the detection of changes in the reporting process in a fast way. Third, by introducing covariates the predictive performance is less sensitive to the chosen evaluation date. The simulation study further identifies a volatile occurrence process and breakpoints in the event observation process as important arguments for choosing a data driven, granular model as developed in this chapter.

2.5 Acknowledgments

2.6 Appendix A: Maximum likelihood estimation of observation exposure parameters

We model a parameter vector $\boldsymbol{\gamma}$ which structures the observation exposures.

$$\ell(\boldsymbol{\gamma}; \boldsymbol{\chi}) = \sum_{t=1}^{\tau} \sum_{s=t}^{\tau} N_{t,s} \cdot \log(p_{t,s}) - \sum_{t=1}^{\tau} N_t^{\mathrm{R}}(\tau) \cdot \log(p_t^{\mathrm{R}}(\tau)) \qquad (2.12)$$

$$= \sum_{t=1}^{\tau} \sum_{s=t}^{\tau} N_{t,s} \cdot \log \left(F_{\tilde{U}} \left(\varphi_t(s-t+1) \right) - F_{\tilde{U}} \left(\varphi_t(s-t) \right) \right)$$

$$- \sum_{t=1}^{\tau} N_t^{\mathrm{R}}(\tau) \cdot \log \left(F_{\tilde{U}} \left(\varphi_t(\tau - t + 1) \right) \right),$$

where

$$\varphi_t(d) = \sum_{v=t}^{t+d-1} \exp(\boldsymbol{x}_{t,v}' \boldsymbol{\gamma}).$$

No analytical solution exists for the optimal parameters $\boldsymbol{\gamma}$ and numerical optimization is required. We use the Newton-Raphson algorithm to maximize the likelihood (2.12). The Newton-Raphson algorithm updates the parameter estimates iteratively as follows

$$\hat{\boldsymbol{\gamma}}^{(k+1)} = \hat{\boldsymbol{\gamma}}^{(k)} - \boldsymbol{H}^{-1}(\hat{\boldsymbol{\gamma}}^{(k)}) \cdot \boldsymbol{S}(\hat{\boldsymbol{\gamma}}^{(k)}). \qquad (2.13)$$

In this formula $\boldsymbol{S}$ denotes the score vector and $\boldsymbol{H}$ is the Hessian of the loglikelihood in (2.12), i.e. the vector of first order and the matrix of second order partial derivatives respectively. Below we derive the expression for the first and second order derivatives of the loglikelihood when $F_{\tilde{U}}$ is a known twice continuously differentiable distribution function. The components of the score vector $\boldsymbol{S}$ are

$$\frac{\partial \ell(\boldsymbol{\gamma}, \boldsymbol{\xi}; \boldsymbol{\chi})}{\partial \gamma_i} = \sum_{t=1}^{\tau} \sum_{s=t}^{\tau} \frac{N_{t,s}}{p_{t,s}} \cdot f_{\tilde{U}} \left(\varphi_t(s-t+1) \right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t+1)$$

$$- \sum_{t=1}^{\tau} \sum_{s=t}^{\tau} \frac{N_{t,s}}{p_{t,s}} \cdot f_{\tilde{U}} \left(\varphi_t(s-t) \right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t)$$

$$- \sum_{t=1}^{\tau} \frac{N_t^{\mathrm{R}}(\tau)}{p_t^{\mathrm{R}}(\tau)} \cdot f_{\tilde{U}} \left(\varphi_t(\tau - t + 1) \right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(\tau - t + 1),$$

where $f_{\tilde{U}}(\,\cdot\,)$ denotes the density function of $F_{\tilde{U}}(\,\cdot\,)$ and

$$p_{t,s} = F_{\tilde{U}}\left(\varphi_t(s - t + 1)\right) - F_{\tilde{U}}\left(\varphi_t(s - t)\right)$$

$$p_{t,s}^{\mathrm{R}}(\tau) = F_{\tilde{U}}\left(\varphi_t(\tau - t + 1)\right).$$

The derivatives of the time change operator φ_t with respect to $\boldsymbol{\gamma}$ are

$$\frac{\partial}{\partial \gamma_i}\varphi_t(s - t + 1) = \sum_{v=t}^{s} x_{t,v,i} \cdot \alpha_{t,v}$$

where $x_{t,s,i}$ is the covariate value of the i-th parameter for reporting on date s for a claim that occurred on date t. The Hessian $\boldsymbol{H}$ is given by

$$\frac{\partial \ell(\boldsymbol{\gamma}; \boldsymbol{\chi})}{\partial \gamma_i \partial \gamma_j} =$$

$$\sum_{t=1}^{\tau} \sum_{s=t}^{\tau} \frac{N_{t,s}}{p_{t,s}} \cdot \left[f_{\tilde{U}}' \left(\varphi_t(s-t+1) \right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t+1) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(s-t+1) \right.$$

$$- f_{\tilde{U}}' \left(\varphi_t(s-t) \right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(s-t)$$

$$\left. + f_{\tilde{U}} \left(\varphi_t(s-t+1) \right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i \partial \gamma_j}(s-t+1) - f_{\tilde{U}} \left(\varphi_t(s-t) \right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i \partial \gamma_j}(s-t) \right]$$

$$- \sum_{t=1}^{\tau} \sum_{s=t}^{\tau} \frac{N_{t,s}}{p_{t,s}^2} \cdot \left[f_{\tilde{U}} \left(\varphi_t(s-t+1) \right)^2 \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t+1) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(s-t+1) \right.$$

$$+ f_{\tilde{U}} \left(\varphi_t(s-t) \right)^2 \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(s-t)$$

$$- f_{\tilde{U}} \left(\varphi_t(s-t+1) \right) \cdot f_{\tilde{U}} \left(\varphi_t(s-t) \right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t+1) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(s-t)$$

$$\left. - f_{\tilde{U}} \left(\varphi_t(s-t+1) \right) \cdot f_{\tilde{U}} \left(\varphi_t(s-t) \right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(s-t) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(s-t+1) \right]$$

$$- \sum_{t=1}^{\tau} \frac{N_t^{\mathrm{R}}(\tau)}{p_t^R(\tau)} \cdot \left[f_{\tilde{U}}' \left(\varphi_t(\tau - t + 1) \right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(\tau - t + 1) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(\tau - t + 1) \right.$$

$$\left. + f_{\tilde{U}} \left(\varphi_t(\tau - t + 1) \right) \cdot \frac{\partial \varphi_t}{\partial \gamma_i \partial \gamma_j}(\tau - t + 1) \right]$$

$$+ \sum_{t=1}^{\tau} \frac{N_t^{\mathrm{R}}(\tau)}{p_t^{\mathrm{R}}(\tau)^2} \cdot f_{\tilde{U}} \left(\varphi_t(\tau - t + 1) \right)^2 \cdot \frac{\partial \varphi_t}{\partial \gamma_i}(\tau - t + 1) \cdot \frac{\partial \varphi_t}{\partial \gamma_j}(\tau - t + 1),$$

where the second order derivatives of φ_t with respect to $\boldsymbol{\gamma}$ are

$$\frac{\partial}{\partial \gamma_i \partial \gamma_j} \varphi_t(s-t+1) = \sum_{v=t}^{s} x_{t,v,i} \cdot x_{t,v,j} \cdot \alpha_{t,v}$$

The Newton-Raphson algorithm in (2.13) models the observation exposure parameters $\boldsymbol{\gamma}$. Together with the observation parameters, the simulation study

of Section 2.3.5 estimates the variance parameter σ in the lognormal time-changed distribution. The Newton-Raphson algorithm in (2.13) can easily be extended to this case, where the distribution function of $F_{\tilde{U}}$ depends on parameters.

2.7 Appendix B: Simulation procedure

We outline the algorithm that was used to generate data sets from the four scenarios specified in Section 2.3.5. This algorithm combines a model for the occurrence of events with a model for the observation delay as described in Section 2.2. We divide the algorithm in three steps.

Step 1. Occurrence We first generate the number of occurred events. The number of daily events follows a Poisson distribution

$$N_t \sim \text{Poisson}(\lambda_t),$$

where the intensity λ_t is obtained from the occurrence process specification for the scenarios in Section 2.3.5.

Step 2. Observation We now simulate the observation date for each occurred event. Combining equation (2.6) and (2.7), we can write the probability that an event from date t is observed on date s as

$$p_{t,s} = P\left(\tilde{U} \in \left[\sum_{v=t}^{s-1} \alpha_{t,v}, \sum_{v=t}^{s} \alpha_{t,v}\right)\right).$$

We define the observation date random variable

$$S_t = \min_s \left\{ s \in \mathbb{N} \,\Big|\, \sum_{v=t}^{s} \alpha_{t,v} > \tilde{U} \right\}. \tag{2.14}$$

This expression transforms the time-changed observation delay random variable into the associated observation date. Consequently S_t satisfies $P(S_t = s) = p_{t,s}$. For each event that occurred on date t we generate a realization from the distribution of $\tilde{U}$. We obtain the corresponding observation date by replacing the random variable $\tilde{U}$ in (2.14) by this sampled value.

Step 3. Truncation With steps 1 and 2 we have simulated an observation date for each occurred event. We split this data set into observed and hidden events. We use the data set with observed events to calibrate the model and to predict the number of hidden events. The hidden events are kept only for evaluating the prediction accuracy.

2.8 Appendix C: A standard distribution for the time changed observation delay

Modeling the time-changed observation delay with an exponential distribution has significant computational benefits. Therefore, this section puts focus on the use of the exponential distribution as a standard distribution for modeling the time-changed observation delay $\tilde{U}$. Since the exponential distribution is light-tailed it is less suited for long or heavy-tailed delays. We outline a strategy for addressing this weakness of the exponential distribution.

Our strategy bins the possible observation delays ($s-t = 0, 1, \ldots$) and categorizes these bins with a delay covariate $x_{s-t}^{\texttt{delay}}$. This covariate is then included in the observation exposure specification. For each bin we estimate a parameter to capture its effect on observation exposure. These parameters can strongly reshape the distribution, hereby overcoming many of the disadvantages of the exponential distribution. We present a maximum likelihood driven binning strategy in Appendix 2.8.1 and then Appendix 2.8.2 derives the same bins by linking our approach to the non-parametric Kaplan-Meier estimator [Kaplan and Meier, 1958].

2.8.1 Binning observation delay

Our binning strategy maximizes the loglikelihood in (2.8) when the observation exposures depend only on the time elapsed since the event occurred, i.e.

$$\alpha_{t,s} = \exp(\boldsymbol{\gamma}^{\texttt{delay}} \cdot x_{s-t}^{\texttt{delay}}) = \exp(\gamma^{\texttt{s-t}}),$$

where we estimate for each delay $s - t$ a separate parameter $\gamma^{\texttt{s-t}}$. Furthermore we neglect the last term in (2.8), capturing the effect of the right truncation.

Under these restrictions, the loglikelihood to optimize is

$$\ell(\boldsymbol{\gamma}; \boldsymbol{\chi}) = -\sum_{t=1}^{\tau} \sum_{v=t}^{\tau-1} \left(\sum_{s=v+1}^{\tau} N_{t,s} \right) \cdot \exp(\gamma^{\mathtt{v-t}})$$

$$+ \sum_{t=1}^{\tau} \sum_{s=t}^{\tau} N_{t,s} \cdot \log \left(1 - \exp(-\exp(\gamma^{\mathtt{s-t}})) \right)$$

We compute the derivatives of $\ell(\boldsymbol{\gamma}; \boldsymbol{\chi})$ with respect to the observation exposure parameter $\gamma^{\mathtt{d}}$ for positive delays $d \in \mathbb{N}$

$$\frac{\partial \ell(\boldsymbol{\gamma}; \boldsymbol{\chi})}{\partial \gamma^{\mathtt{d}}} = -\exp(\gamma^{\mathtt{d}}) \cdot \sum_{t=1}^{\tau-d-1} \sum_{s=t+d+1}^{\tau} N_{t,s} + \frac{\exp(\gamma^{\mathtt{d}})}{\exp(\exp(\gamma^{\mathtt{d}})) - 1} \cdot \sum_{t=1}^{\tau-d} N_{t,t+d}.$$

Both sums in this expression have a logical interpretation. The first sum $\left(\sum_{t=1}^{\tau-1-d} \sum_{s=d+t+1}^{\tau} N_{t,s} \right)$ counts the number of observed events with a delay longer than d days, whereas the second sum $\left(\sum_{t=1}^{\tau-d} N_{t,t+d} \right)$ counts all events with a delay of exactly d days. These derivatives are zero when

$$\exp(\gamma^{\mathtt{d}}) = -\log \left(1 - \frac{|\mathtt{delay = d}|}{|\mathtt{delay \geq d}|} \right), \tag{2.15}$$

where $|\mathtt{delay = d}|$ denotes the number of events observed with a delay of d days and $|\mathtt{delay > d}|$ the number of events with a delay of more than d days.

We propose to bin the observation delay by grouping delays for which (2.15) is approximately constant. Figure 2.15 visualizes this approach for the liability insurance data set discussed in Section 2.3. This figure shows in red the estimated delay parameters using approximation (2.15). The top panel shows the estimates for delays up to 31 days, whereas the parameters for larger delays (up to 400 days) are shown in the bottom panel. Based on this knowledge observation delay is grouped in 23 bins, separated by vertical gray bars in Figure 2.15. We use more bins for short delays, since for these delays (2.15) differs strongly. Moreover, many accidents have a short observation delay, which makes these first delays more important. As expected, this binning strategy identifies an increase in observation probability after exactly one year. In Section 2.3 we structure these bins in a categorical delay covariate $x_{s-t}^{\mathtt{delay}}$ and estimate observation delay in a maximum likelihood framework. In Figure 2.15 the fitted parameters are plotted in blue. These parameters deviate from those found using approximation (2.15), since other covariate effects were estimated simultaneously. However, the maximum likelihood estimates are close to the approximate values which makes this approximation suitable for choosing initial values in the calibration.
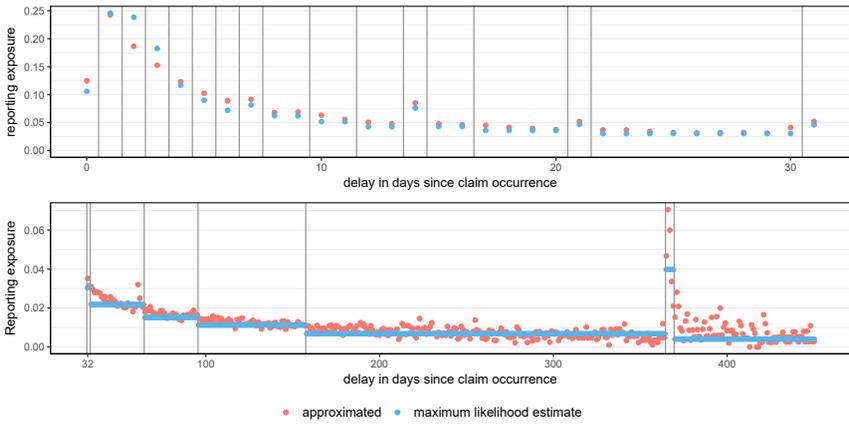
Figure 2.15: Observation exposure estimates for the delay effect during the first month after the accident occurrence (top) and longer delays (bottom). In red, we show estimates obtained for each delay using (2.15). The vertical lines indicate the chosen bins. Maximum likelihood estimates for the observation delay parameter corresponding to each bin in the regression structure proposed in Section 2.3.2 are plotted in blue.

2.8.2 A link with the Kaplan-Meier estimator

We show that under the binning strategy of Appendix 2.8.1 the time changed model has the same flexibility as the Kaplan-Meier estimator and is as such suitable for modelling a wide range of portfolios.

The Kaplan-Meier estimator for the survival function of the observation delay random variable is

$$P(\widehat{\mathtt{delay} > d}) = \prod_{i=0}^{d} \left(1 - \frac{|\mathtt{delay} = i|}{|\mathtt{delay} \geq i|} \right), \qquad (2.16)$$

When we model the time-changed observation delay distribution $\tilde{U}$ using an exponential distribution then the survival probability for an event from

occurrence day t is

$$P\left(\texttt{delay} > d \mid \texttt{occ. day} = t\right) = P\left(\tilde{U} \geq \varphi_t(d+1)\right) \qquad (2.17)$$

$$= 1 - F_{\tilde{U}}\left(\sum_{i=1}^{d+1} \alpha_{t,t+i-1}\right)$$

$$= \prod_{i=0}^{d} \exp\left(-\alpha_{t,t+i}\right).$$

Notice the similarity between this expression and the Kaplan-Meier estimator in (2.16). When the observation exposure only depends on the time passed since the occurrence of the event, i.e. $\alpha_{t,t+i} := \alpha_i$, then

$$P\left(\texttt{delay} > d\right) = \prod_{i=0}^{d} \exp\left(-\alpha_i\right),$$

where α_i is the observation exposure at delay i. This expression no longer depends on the occurrence date t of the event. The Kaplan-Meier estimator is retrieved when

$$\alpha_i = -\log\left(1 - \frac{|\texttt{delay} = i|}{|\texttt{delay} \geq i|}\right). \qquad (2.18)$$

Since $\alpha_i = \exp(\gamma^i)$, this is the same estimator we found in (2.15) through maximum likelihood estimation. This show that by estimating a separate delay parameter for each delay $(d = 0, 1, \ldots)$ we obtain a model with the same flexibility as the non-parametric Kaplan-Meier estimator.

Chapter 3

A hierarchical reserving model for reported non-life insurance claims

Abstract

Traditional non-life reserving models largely neglect the vast amount of information collected over the lifetime of a claim. This information includes covariates describing the policy (e.g. the value of the insured risk), claim cause (e.g. hail) as well as the detailed claim's history (e.g. settlement, payment, involvement lawyer). We present the hierarchical reserving model as a modular framework for integrating a claim's history and claim-specific covariates into the development process. Hierarchical reserving models decompose the joint likelihood of the development process over time. Moreover, they are tailored to the portfolio at hand by adding a layer to the model for each of the registered events (e.g. settlement, payment). Layers are modelled with classical techniques (e.g. generalized linear models) or machine learning methods (e.g. gradient boosting machines) and using claim-specific covariates. As a result of its flexibility, this framework incorporates many existing reserving models, ranging from aggregate models designed for runoff triangles to individual models using claim-specific covariates. This connection allows us to develop a data-driven strategy for choosing between aggregate and individual reserving in the presence of covariates; an important decision for reserving practitioners that is largely left unexplored in scientific literature. We illustrate our method with a case study on a real life insurance data set. This case study provides new insights in

the covariates driving the development of claims and demonstrates the flexibility and robustness of the hierarchical reserving model over time.

This chapter is based on Jonas Crevecoeur and Katrien Antonio. A hierarchical reserving model for non-life insurance claims. 2020a. Available at arXiv: `https://arxiv.org/abs/1910.12692`.

3.1   Introduction

Insurers set aside funds, the so-called reserve, for covering claims from past exposure years. This reserve is often split into a reserve for Incurred, But Not yet Reported (IBNR) claims and a reserve for Reported, But Not yet Settled (RBNS) claims. These separate reserves differ in the range of statistical tools that are available for modelling them. Since the claims that compose the IBNR reserve are not yet reported, claim-specific and policy(holder) covariates are unavailable for differentiating the cost per claim. Therefore, IBNR reserving mostly focuses on accurately estimating the number of unreported claims, followed by allocating a fixed cost per unreported claim. In RBNS reserving, the insurer is aware of the number of open claims as well as their characteristics and development so-far. This opens the possibility for reserving models that predict the future cost on a per claim basis. This chapter focuses on predicting the RBNS reserve by modelling the development of reported claims.

Traditionally, the non-life insurance literature has been dominated by analytic models designed for aggregated data, such as the chain ladder method [Mack, 1993, 1999]. These models compress the historical data on the development of claims over time in a two dimensional table, the so-called runoff triangle, by aggregating payments by occurrence and development year. Low data requirements, implementation simplicity and a straightforward interpretation of the predicted reserve justify the popularity of these models. However, by compressing the data valuable insights into the risk characteristics of individual claims are lost. This makes the reserve less robust against changes in the portfolio composition and extreme one-time events. In response to this, individual reserving methods designed for granular data available at the level of individual claims, have first been proposed in the nineties. Individual reserving remained largely unexplored for about two decades, with revived interest in recent years thanks to an increased focus on big data analytics.

We identify three streams in the current literature on individual reserving. Following Norberg [1993, 1999], a first stream analyzes the events registered during a claim's development in continuous time. Lopez et al. [2016, 2019] adapt regression trees to the right-censoring present in continuous time reserving data.

Covariates in these trees capture the heterogeneity in the claim size as well as in the time to settlement of reported claims. In Antonio and Plat [2014] hazard rates drive the time to events in the development of claims (e.g. a payment, or settlement) and a lognormal regression model is proposed for the payment size. Reserving in continuous time requires a time to event model that allows for multiple payments and multiple types of (recurrent) events. Since such models are complicated, many individual reserving models are defined in a more convenient discrete time framework, where the events in a claim's lifetime are registered in discrete time periods. A second stream of reserving methods models the reserve in discrete time by adapting models from insurance pricing, as such taking advantage of the detailed covariate information available within insurance companies. Since these covariates only become available at reporting, such models focus on the reserve for reported, but not settled (RBNS) claims, while using techniques from aggregated reserving to estimate the reserve for unreported claims. Larsen [2007] focuses on Generalized Linear Models (GLMs), Wüthrich [2018] considers regression trees and Wüthrich [2018] looks at neural networks for reserving. A third stream of papers aggregates the data into multiple runoff triangles. Martínez Miranda et al. [2012], Wahl et al. [2019] and Denuit and Trufin [2017, 2018] consider two, three and four triangles respectively. While the aggregation of the data makes these models easy to implement, covariate information of individual claims can not be used. The recent expansion in (individual) reserving methodology has resulted in a fragmented literature with few comparative studies and no unified approach with proven robustness and general applicability. The lack of a solid modelling framework hinders the implementation of individual reserving in insurance practice. Moreover, providing data driven guidance on the choice between aggregate and individual reserving is a very important question that is largely unexplored in the reserving literature.

We aim to fill this gap in the literature by presenting the hierarchical reserving model as an intuitive framework for RBNS reserving with a focus on applicability in practice. This framework decomposes the joint likelihood of the development process of individual claims after reporting in discrete time. Hierarchical reserving models are tailored to the portfolio at hand by adding layers, which represent the events (e.g. settlement, positive or negative payments, changes in the incurred, . . .) registered over the lifetime of a claim. This modular approach enables us to restate many existing reserving models, including models based on data aggregated into a runoff triangle, as special cases of the hierarchical reserving model. This intuitive model building process allows us to concentrate on the decisions made during the modelling process, such as model calibration and evaluation. These aspects of the modelling process have received little attention in individual reserving literature up to now with many papers following the model building steps applied in pricing or aggregate reserving. This results

in a loss of performance as such methods do not consider the individual and censored structure of the data.

This chapter is organized as follows. Section 3.2 introduces the hierarchical reserving model, presents best practices for calibrating this model to insurance data and explains how this model can be used to predict the future reserve. Section 3.3 investigates the connection between hierarchical reserving models proposed at individual claim level and certain aggregate reserving models. This results in a data driven strategy for choosing between aggregate and individual reserving. Section 3.4 demonstrates this methodology in a case study on a home insurance data set. This is a novel data set, which has not been used before in the literature on reserving. An R package accompanies this chapter enabling researchers and practitioners to directly apply the hierarchical reserving model to their portfolios.

3.2   A hierarchical reserving model

It is common in insurance pricing to decompose the joint likelihood into a frequency and severity contribution [Henckaerts et al., 2018]. Frees and Valdez [2008] extend this idea by splitting the severity contribution per claim type. In this spirit, we propose a hierarchical reserving model, which decomposes the joint likelihood of the claim development process over time and registered events (e.g. settlement, payment).

3.2.1   Notation and statistical model

We record the development of reported claims in discrete time over a period of τ years. For each reported claim k, r_k denotes the reporting year and the vector $\boldsymbol{x}_k$ denotes the claim information available at the end of the reporting year. This information vector is static and consists of the circumstances of the claim, policy(holder) covariates and the claim development (e.g. initial reserve, payments) in the reporting year. In the years after reporting, so-called update vectors, denoted $\boldsymbol{U}_k^j$, describe the change in the claim development information in year $r_k + j - 1$. The length and components of $\boldsymbol{U}_k^j$ depend on the events (e.g. claim settlement, change in the incurred, involvement of a lawyer) registered in the portfolio at hand. This chapter defines a modular model building approach that can be tailored to the chosen structure in $\boldsymbol{U}_k^j$. These models, called hierarchical reserving models, are based on two fundamental assumptions

Hierarchical model assumptions

(A1) All claims settle within d years after reporting;

(A2) The development of a claim is independent of the development of the other claims in the portfolio.

Although upper limit d on the the settlement delay is not necessarily limited to the length of the observation window, we implicitly assume $d = \tau$ for notational convenience. Given these assumptions, the set of observed claim updates after reporting, say $\mathcal{R}^{\texttt{Obs}}$, for a portfolio of n claims is

$$\mathcal{R}^{\texttt{Obs}} = \{ \boldsymbol{U}_k^j \mid k = 1, \ldots, n, j = 2, \ldots, \tau_k \},$$

with $\tau_k = \min(d, \tau - r_k + 1)$ the number of observed development years since reporting for claim k. The associated likelihood is

$$\mathcal{L}\left(\mathcal{R}^{\texttt{Obs}}\right) = \prod_{k=1}^{n} f\left(\boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{\tau_k} \mid \boldsymbol{x}_k\right),$$

where we use assumption (A2) to write the likelihood as a product of claim-specific likelihood contributions. Inspired by Frees and Valdez [2008], we introduce a hierarchical structure in this likelihood by applying the law of conditional probability twice. First, we include the temporal dimension by splitting the likelihood in chronological order

$$\mathcal{L}\left(\mathcal{R}^{\texttt{Obs}}\right) = \prod_{k=1}^{n} \prod_{j=2}^{\tau_k} f\left(\boldsymbol{U}_k^j \mid \boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{j-1}, \boldsymbol{x}_k\right).$$

By conditioning on past events, we acknowledge that the future development of a claim depends on its development in previous years. Second, we split the likelihood by the events registered in the vector $\boldsymbol{U}_k^j$

$$\mathcal{L}\left(\mathcal{R}^{\texttt{Obs}}\right) = \prod_{k=1}^{n} \prod_{j=2}^{\tau_k} \prod_{l=1}^{s} f\left(U_{k,l}^j \mid \boldsymbol{U}_k^{(2)}, \ldots, \boldsymbol{U}_k^{j-1}, U_{k,1}^j, \ldots, U_{k,l-1}^j, \boldsymbol{x}_k\right), \quad (3.1)$$

where s is the length of the update vector $\boldsymbol{U}_k^j$. In the remainder of this chapter, we refer to these events registered over the lifetime of a claim as the layers of the hierarchical model. The order of the layers is an important model choice, since the outcome of a layer becomes a covariate when modelling higher indexed layers. Since the assumptions (A1-A2) are common in reserving literature, most discrete time reserving models can be seen as a special case of our hierarchical reserving framework. Notice that in contrast with the chain ladder method, the

hierarchical framework includes the full history of the claim and thus allows for non-Markovian models.

When applying the hierarchical claim development model to a specific portfolio, we extend assumptions (A1-A2) with an additional assumption, which tailors the structure of the update vector $\boldsymbol{U}_k^j$ to the portfolio at hand. For example, in the case study covered in Section 3.4, we model $\boldsymbol{U}_k^j$ with a three-layer hierarchical model.

Hierarchical layers

(A3) The update vector $\boldsymbol{U}_k^j$ for claim k in development year j has three layers $\boldsymbol{U}_k^j = (U_{k,1}^j, U_{k,2}^j, U_{k,3}^j) = (C_k^j, P_k^j, Y_k^j)$:

- C_k^j is the settlement indicator which is one when claim k settles in development year j and zero otherwise. Conditional on past events, the settlement indicator follows a Bernoulli distribution with

$$C_k^j \mid_{\boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{j-1}, \boldsymbol{x}_k} \sim \texttt{Bernoulli}\left(p\left(\boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{j-1}, \boldsymbol{x}_k\right)\right).$$

- P_k^j is the payment indicator which is one when there is a payment for claim k in development year j and zero otherwise. Conditional on past events, the payment indicator follows a Bernoulli distribution with

$$P_k^j \mid_{\boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{j-1}, C_k^j, \boldsymbol{x}_k} \sim \texttt{Bernoulli}\left(q\left(\boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{j-1}, C_k^j, \boldsymbol{x}_k\right)\right).$$

- Y_k^j is the payment size, given that there was a payment in development year j. Conditional on past events, the payment size is gamma distributed with mean

$$E(Y_k^j \mid \boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{j-1}, C_k^j, P_k^j, \boldsymbol{x}_k, j) = \mu(\boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{j-1}, C_k^j, P_k^j, \boldsymbol{x}_k)$$

and variance

$$\sigma^2(Y_k^j \mid \boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{j-1}, C_k^j, P_k^j, \boldsymbol{x}_k) = \theta \cdot \mu(\boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{j-1}, C_k^j, P_k^j, \boldsymbol{x}_k).$$

As such, we structure the development of claims with a simple three-layer hierarchical model. Conditioning on the settlement status in past years, allows us to train the model on the development of open claims only, whereas without settlement indicator, the model would predict new payments for already settled claims. Moreover, by choosing settlement as the first layer of the hierarchical

model, settlement becomes a covariate when modelling later layers. The gamma distribution for the sizes is frequently used in insurance pricing literature when modelling attritional losses [Henckaerts et al., 2020]. Choosing a strictly positive distribution assumes that there are no recoveries in the portfolio. In portfolios in which recoveries are common, additional layers should be added to the hierarchical model to allow for negative payments.

3.2.2 Hierarchical model calibration

The hierarchical claim development framework makes no assumption with respect to the statistical modelling technique that is used to model the individual layers. The case-study in Section 3.4 illustrates the proposed hierarchical reserving model by calibrating both a Generalized Linear Model (GLM) as well as a Gradient Boosting Model (GBM) to the layers outlined in (A3). Although standard procedures are available for calibrating these models, special attention is required for the variable selection process or (hyper) parameter turning steps. In reserving the historical, observed data contains mainly records from early development years, whereas the future predictions are more oriented towards later years. This imbalance between the training and prediction data set poses a model risk when covariates exhibit a different effect on the first development years versus the later development years. In machine learning literature this phenomenon is known as a covariate shift [Sugiyama et al., 2007b]. Following Sugiyama et al. [2007a], we correct for a potential covariate shift by maximizing a weighted likelihood in which weights depend on the development year, i.e.

$$\mathcal{L}^{\texttt{weighted}}\left(\mathcal{R}^{\texttt{0bs}}\right) = \prod_{k=1}^{n} \prod_{j=2}^{\tau_k} w_j \prod_{l=1}^{s} f\left(U_{k,l}^{j} \mid \boldsymbol{U}_{k}^{(2)}, \ldots, \boldsymbol{U}_{k}^{(j-1)}, U_{k,1}^{j}, \ldots, U_{k,l-1}^{j}, \boldsymbol{x}_k\right),$$
$$(3.2)$$

where w_j is the weight assigned to an observation from development year j. Following Sugiyama et al. [2007a], we define these weights as the ratio of the number of records from development year j in the prediction data set to the number of records from development year j in the training data set. For typical reserving data sets this ratio is observed and can be computed as

$$w_j = \frac{\sum_{i=d-j+2}^{d} n_i}{\sum_{i=1}^{d-j+1} n_i},$$

where n_i is the number of reported claims in reporting year i. These weights increase in j assigning more weight to observations from later development years. When selecting covariates or tuning (hyper) parameters, we maximize (3.2) in a 5-fold cross validation scheme. For this, we calibrate predictive models per layer l and allocate observations at the level of a claim k and a development year j (see (3.2)) to different folds.

3.2.3 Predicting the future development of claims

Algorithm 1 simulates the development of reported claims beyond the observation window τ. In line with the hierarchical structure of the model, development years are simulated in chronological order and within a development year, this simulation algorithm respects the order of the layers. The simulation order is important, since simulated values from previous development years and lower indexed layers become inputs for later development years and higher indexed layers.

Algorithm 1: Simulating the future development of reported claims

---

Input: the observed development of reported claims
Output: simulation of the future development of reported claims
foreach *claim k* **do**
    for *development year j* **in** $\tau + 1 - r_k \ldots d$ **do**
        for *hierarchical layer l* **in** $1 \ldots s$ **do**
            Simulate $U_l^j \mid \boldsymbol{U}_k^{(2)}, \ldots, \boldsymbol{U}_k^{(j-1)}, U_{k,1}^j, \ldots, U_{k,l-1}^j, \boldsymbol{x}_k$
        end
    end
end

Following this algorithm, the simulated data has the same hierarchical layered structure as the input data set, which enables us to derive aggregated quantities for the events registered in the update vector $\boldsymbol{U}_k^j$. For example, given the specific hierarchical structure in assumption (A3), we obtain estimates for the number of open claims, the number of payments and the total payment size. Prediction intervals for these reserving quantities are obtained by running Algorithm 1 many times.

3.2.4 Implementation in R

We have developed a package called `hirem` [Crevecoeur, 2020] for defining and calibrating hierarchical reserving models as well as simulating the future development of claims. In this package, layers can be estimated with generalized linear models (GLMs) or gradient boosting models (GBMs). The case-study of Section 3.4 uses the implementation from Southworth [2015] of the `gbm` package, which adds the gamma loss function to the original package developed by Greenwell et al. [2018].

3.3 Bridging aggregate and individual reserving

Most claim reserving models used in insurance companies are based on data aggregated into runoff triangles. We start from data registered at the level of individual claims and illustrate how aggregate reserving models can be retrieved as special cases of the hierarchical reserving model. Section 3.3.1 investigates the simplified case of a hierarchical model with independent layers. Section 3.3.2 extends these results and allows a simple, but common dependency structure between the layers. The results of these sections offer valuable insights and statistical tools for choosing between aggregate and individual reserving. Section 3.3.3 demonstrates the universality of our framework by constructing hierarchical reserving models inspired by recent literature contributions on aggregate reserving with multiple runoff triangles. The hierarchical reserving model as a unified framework for RBNS reserving facilitates model comparison and offers new insights as to how these models could be extended to data registered at the level of individual claims.

3.3.1 From individual hierarchical reserving models with independent layers to aggregate reserving models

In contrast with traditional reserving models, the hierarchical reserving model proposed in Section 3.2.1 analyses the development of claims from development year two since reporting onwards. For this, our approach collects all information registered during the reporting year of a claim k in a vector $\boldsymbol{x}_k$. This vector not only includes claim covariates (e.g. the cause of the accident), but also covariates structuring the development of the claim in the reporting year (e.g. the amount paid during the reporting year). In this section, we denote by $\boldsymbol{U}_k^1$ the claim development information that becomes available during the reporting year. Next to this, $\boldsymbol{x}_k$ refers in this section to the remaining static claim covariates that become available at reporting. Introducing $\boldsymbol{U}_k^1$ brings our notation more in line with traditional reserving practice and enables us to model the development of claims from development year one onwards.

We construct for each of the layers l in the update vector $\boldsymbol{U}_k^j$ a runoff triangle $(X_l^{ij})_{1 \leq i,j \leq d}$ with cells

$$X_l^{ij} = \sum_{k:r_k=i} U_{k,l}^j.$$

In line with our focus on modelling RBNS claims, we aggregate by reporting year and development year since reporting instead of the traditional set-up where aggregation goes per occurrence year and development year since occurrence.

As such, we model the development of claims since reporting. Although, we denote the reporting year of claim k by r_k, we keep the traditional index i for the rows, i.e. the reporting years, in the runoff triangle.

Let us now assume that the individual updates depend multiplicatively on the reporting year and the development year, i.e.

$$\mathrm{E}(U^j_{k,l}) = \alpha_{r_k,l} \cdot \beta_{j,l} \quad \text{and} \quad \sum_{j=1}^{d} \beta_{j,l} = 1, \tag{3.3}$$

for each layer l and where $\alpha_{r_k,l}$ is the effect of reporting year r_k and $\beta_{j,l}$ is the effect of development year j. When we aggregate these individual updates into a runoff triangle, the cell values follow a similar multiplicative relation, i.e.

$$\mathrm{E}(X^{ij}_l) = \mathrm{E}\left(\sum_{k:r_k=i} U^j_{k,l}\right) = n_i \cdot \alpha_{i,l} \cdot \beta_{j,l} := \tilde{\alpha}_{i,l} \cdot \beta_{j,l}, \tag{3.4}$$

where n_i, the number of reported claims in reporting year i, is observed. As a result, we can calibrate individual hierarchical reserving models that only depend multiplicatively on reporting year and development year using data aggregated into runoff triangles.

Matching (3.3) with the original hierarchical reserving model specification in (3.1), we rephrase the expected value for the updates $\boldsymbol{U}^j_k$ at the individual level in full generality as

$$\mathrm{E}(U^j_{k,l}) = \alpha_{r_k,l} \cdot \beta_{j,l} \cdot \phi\left(\boldsymbol{U}^1_k, \ldots, \boldsymbol{U}^{j-1}_k, U^j_{k,1}, \ldots, U^j_{k,l-1}, \boldsymbol{x}_k\right),$$

where $\phi(\cdot)$ represents the effect of all other covariates. When we add a distributional assumption for $\boldsymbol{U}^j_k$, choosing between an aggregate or individual reserving model reduces to testing for $\phi(\cdot) = 1$. Since the models with and without $\phi(\cdot)$ are nested, a likelihood ratio test can be used for this.

3.3.2 From individual hierarchical reserving models with dependent layers to aggregate reserving models

The reserving models constructed in Section 3.3.2 treat each layer independent of the others. This results in simple aggregated models, where each layer is estimated from a single runoff triangle, independent from the other layers. However, in most multi-layer hierarchical structures some dependence between the layers is inevitable and offering a simple framework to include these dependencies is one of the main motivations for the hierarchical reserving

model. This section investigates the special, but common setting of a two-layer hierarchical model in which layer one is a binary random variable and layer two is zero whenever layer one equals zero. As an example, think of layer one as a payment indicator and layer two as the payment size. When there is no payment, the payment size is zero.

Again focusing on the multiplicative structure of reporting and development year, we structure the expected values of the layers as

$$\mathrm{E}(U_{k,1}^j) = \alpha_{r_k,1} \cdot \beta_{j,1} \cdot \phi\left(\boldsymbol{U}_k^1, \ldots, \boldsymbol{U}_k^{j-1}, \boldsymbol{x}_k\right)$$

$$\mathrm{E}(U_{k,2}^j) = \begin{cases} \alpha_{r_k,2} \cdot \beta_{j,2} \cdot \psi\left(\boldsymbol{U}_k^1, \ldots, \boldsymbol{U}_k^{j-1}, \boldsymbol{x}_k\right) & U_{k,1}^j = 1 \\ 0 & U_{k,1}^j = 0 \end{cases},$$

where $\sum_{j=1}^d \beta_{j,l} = 1$ for $l \in \{1, 2\}$. When $\phi(\cdot)$ and $\psi(\cdot)$ are both equal to one, the claim development depends only on reporting year and development year in a multiplicative way. We then retrieve

$$\mathrm{E}(X_1^{ij}) = E\left(\sum_{k:r_k=i} U_{k,1}^j\right) = n_i \cdot \alpha_{i,1} \cdot \beta_{j,1} \coloneqq \tilde{\alpha}_{i,1} \cdot \beta_{j,1}$$

$$\mathrm{E}(X_2^{ij} \mid X_1^{ij}) = \mathrm{E}\left(\sum_{k:r_k=i} U_{k,2}^j\right) = X_1^{ij} \cdot \alpha_{i,2} \cdot \beta_{j,2}. \tag{3.5}$$

When calibrating the model for the second layer, the observed upper triangle of the first layer acts as an exposure term. When estimating the future reserve, this exposure term, X_1^{ij}, should be estimated using the model proposed for the first layer. If we interpret the first layer as a payment indicator and the second layer as the payment size then the number of payments becomes the exposure for the total payment size. Similar to Section 3.3.1, statistical tests for $\phi(\cdot) = \psi(\cdot) = 1$ offer data driven tools for choosing between individual and aggregate reserving.

3.3.3 Hierarchical reserving models inspired by aggregate reserving models proposed for multiple runoff triangles

As a result of the weak assumptions underlying the hierarchical reserving model, many existing reserving models can be restated as special cases of our framework. As a unifying framework, the hierarchical reserving model facilitates model comparison and allows extending the calibration and simulation strategy

developed in this chapter to other models. In the case of models designed for aggregate data, the hierarchical reserving framework in addition offers insights as to how these models could be extended to data registered at the level of individual claims.

As an illustration of the generality of our framework, we construct hierarchical reserving models inspired by recent contributions on aggregate reserving using data structured in multiple triangles. We discuss two examples of such models, namely the double chain ladder [Martínez Miranda et al., 2012] and the collective reserving model [Wahl et al., 2019]. As motivated in Section 3.1, we limit our analysis to the RBNS part of these aggregate models.

Double chain ladder The double chain ladder (DCL) [Martínez Miranda et al., 2012] extends the chain ladder method to two runoff triangles to obtain separate estimates for the IBNR and RBNS reserve. Since we only consider the development of claims after reporting, we focus on the triangle of claim sizes and construct a one-layer hierarchical model. DCL structures the expected payment size for a claim k in development year j since reporting, denoted $U_{k,1}^{j}$, as

$$\mathrm{E}(U_{k,1}^{j}) = \tilde{\pi}_j \cdot \tilde{\mu}_j \cdot \gamma_{i_k},$$

where i_k denotes the occurrence year of the claim and $\tilde{\pi}_j$ and $\tilde{\mu}_j$ are the payment probability in development year j and average payment size in development year j respectively. The coefficient γ_{i_k} adjusts the size of the payments from occurrence year i_k for inflation. Letting inflation depend on the occurrence year is natural in DCL, which aggregates runoff triangles by occurrence year and development year since occurrence. Since runoff triangles based on the hierarchical reserving model aggregate by reporting year and development year since reporting, it is in our framework more natural to model inflation per reporting year. If we change the occurrence year effect γ_{i_k} by a reporting year effect γ_{r_k}, the individual updates become

$$\mathrm{E}(U_{k,1}^{j}) = \tilde{\pi}_j \cdot \tilde{\mu}_j \cdot \gamma_{r_k}.$$

Aggregating these individual updates, we retrieve

$$\mathrm{E}(X_1^{ij}) = \mathrm{E}\left(\sum_{k:r_k=i} U_{k,1}^{j}\right) = n_i \cdot \tilde{\pi}_j \cdot \tilde{\mu}_j \cdot \gamma_i.$$

This is the same model as (3.3), when we rewrite $\alpha_i = n_i \cdot \gamma_i$ and $\beta_j = \tilde{\pi}_j \cdot \tilde{\mu}_j$.

The collective reserving model Extending the earlier work of Verrall et al. [2010] and Martínez Miranda et al. [2012], the collective reserving model [Wahl

et al., 2019] structures the claim development after reporting in two layers. These layers represent the number of payments and the size per payment. Inspired by Wahl et al. [2019]'s aggregate model, we structure the individual updates as

$$\mathrm{E}(U_{k,1}^j) \sim \texttt{Poisson}(\lambda_j), \qquad \text{(number of payments)}$$

$$\mathrm{E}(U_{k,2}^j \mid U_{k,1}^j) = \mu(i_k, r_k, j) \cdot U_{k,1}^j, \qquad \text{(payment size)}$$

where a claim can have multiple payments in the same year, each with an average size $\mu(i_k, r_k, j)$, which depends on the occurrence year, reporting year and development year since reporting. When $\mu(i_k, r_k, j) = \alpha_{r_k} \cdot \beta_j$, the model aggregates to

$$\mathrm{E}(X_1^{ij}) = n_i \cdot \lambda_j,$$

$$\mathrm{E}(X_2^{ij} \mid X_1^{ij}) = X_1^{ij} \cdot \alpha_i \cdot \beta_j,$$

where n_i denotes the number of claims reported in reporting year i. This representation is almost identical to (3.5), with the estimated effect of reporting year i for the number of payments replaced by the observed count n_i.

3.4 Case study: European home insurance portfolio

This case study models the RBNS reserve for a European home insurance portfolio. This insurance reimburses damages to the insured property and its contents resulting from a wide range of causes including fire damage, water damage and theft. For reasons of confidentiality we can not disclose the size of the portfolio and the associated reserve. Therefore, we express the performance of the investigated reserving methods via a percentage error measure, comparing the actual and predicted reserve.

3.4.1 Data characteristics

We observe the development of individual claims over a seven year period from January, 2011 until December, 2017. Figure 3.1 structures individual payments by the reporting date of the claim (vertical axis) and the number of days elapsed since reporting (horizontal axis). Every dot represents a single payment and one claim can have multiple payments. A triangular structure

appears, since the claim development after December, 2017 is censored. Home insurance is a short tailed business line, with many payments in the first years after reporting. The black grid in Figure 3.1 visualizes how individual payments would be aggregated when constructing a yearly runoff triangle. As shown in this triangle, extreme weather events cause sudden spikes in the number of reported claims. This has a large impact on the stability of the runoff triangle in classical, aggregate reserving. Therefore, insurers most often reserve these claims separately based on expert opinion. In this chapter we analyze the robustness of various hierarchical reserving methods by predicting the future reserve with and without extreme weather events. Table 3.1 provides a detailed description of the available covariates. We group these covariates into four categories. Policy covariates identify the policy or policyholder entering the claim. These covariates are available when pricing the contract. Claim covariates describe the static characteristics of the claim. These covariates become available in the reporting year of the claim. Development covariates describe the yearly evolution of the claim and layer covariates constitute the layers of the hierarchical reserving model.
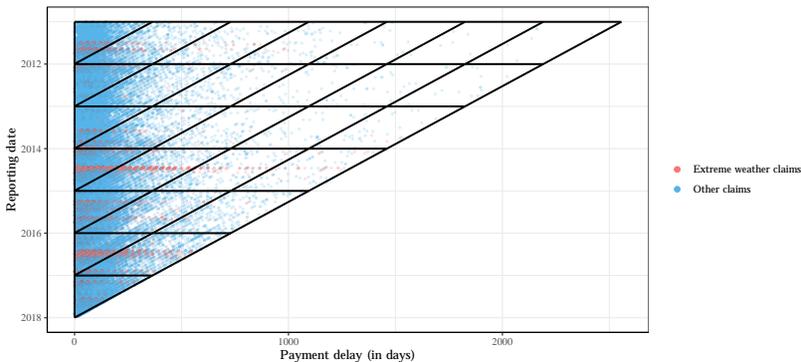
Figure 3.1: Payments structured by reporting date and payment delay in days. Every dot represents a single payment and one claim can have multiple payments. A grid indicates how individual payments would be aggregated when constructing a yearly runoff triangle. Claims resulting from extreme weather (e.g. a storm) are colored red.

Policy covariates

valuables	Objects were declared with a value exceeding the standard cover: yes or no
age.insured	The age of the policyholder
profession	Profession of the policyholder, 16 categories
sex	Gender of the policyholder: male or female
construction.year	The year in which the building was constructed
property.value	The value of the property in Euro

Claim covariates

acc.date	Date on which the accident occurred
rep.date	Date on which the claim was reported to the insurer
rep.delay	Delay in days between the occurrence and reporting of the claim
rep.month	Calendar month in which the claim was reported (Jan - Dec)
coverage	The main coverage applicable to the claim: theft, building or contents
catnat	The cause of the claim, grouped in 12 categories
extreme.weather	Claim is the result of extreme weather (e.g. storm): yes or no
initial.reserve	Expert estimate of the initial reserve at the end of the reporting year

Development covariates

dev.year	The number of years elapsed since the reporting of the claim
calendar.year	Number of years elapsed between the start of the portfolio and dev.year
size.last.year	Total amount paid in the previous development year
total.amount.paid	Total amount paid in all previous development years

Layer covariates

close	The claim closes in the current development year: yes or no
payment	A payment occurs in the current development year: yes or no
size	Total amount paid in the current development year

Table 3.1: List of covariates available in the home insurance data set. A level NA (not available) identifies the records with no registered value for a covariate.

In Figure 3.2 a treemap visualizes the available claims grouped into the 12 risk categories as coded in the covariate `catnat`. Each claim is represented by a rectangle, where the size of this rectangle visualizes the amount paid for that claim by the end of December, 2017. Water and fire damage are the most important insurance covers in this portfolio. Together these risks generate more than half of the total claim cost. Fire claims are typically larger than non-fire claims. Although less than 5% of all claims are related to fire, these claims represent more than 25% of the total cost. The large difference between the average size of fire claims versus the average size of non-fire claims, motivates us to build separate reserving models for fire claims on the one hand and non-fire claims on the other hand. Estimating separate reserves for risks with a different development pattern is a common approach in traditional reserving. Alternatively, we can distinguish fire and non-fire claims by including a covariate in the hierarchical reserving model. However, this latter approach would result in an unfair comparison between individual models, which can use this covariate, and traditional reserving methods for aggregate data, which can not use this covariate.
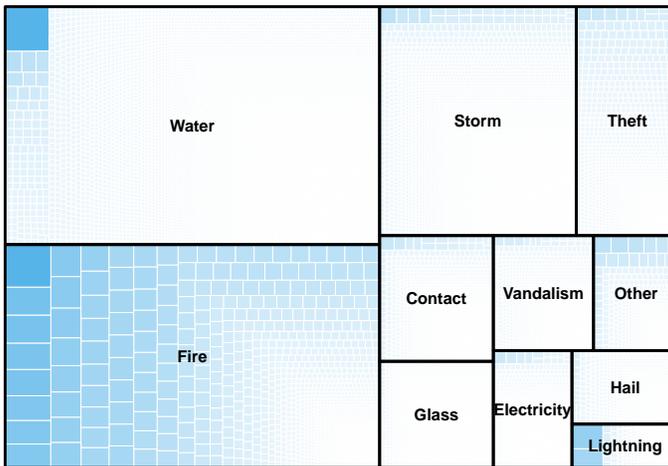
Figure 3.2: Treemap of individual claims observed on 31 December, 2017 grouped into the 12 risk categories present in the portfolio as coded in the covariate `catnat`. Each claim is represented by a rectangle, where the size of this rectangle visualizes the amount paid for that claim by the end of December, 2017.

3.4.2 Hierarchical reserving models for fire and non-fire claims

We analyse the performance of hierarchical reserving models based on GLMs, GBMs and the chain ladder method on 365 evaluation dates between January 1, 2015 and December 31, 2015. Instead of a single out-of-time evaluation (as e.g. in Antonio and Plat [2014], Wüthrich [2018]) the moving window evaluation enables a more thorough assessment of the sensitivity and general applicability of the model. On each evaluation date τ we train the models on the observed data (January, 2011 until τ) and compare the out-of-sample reserve estimate with the actual claim development over the next two development years.

Hierarchical reserving models

Hierarchical GLM   The hierarchical GLM follows the three layer structure close, payment and size defined in assumption (A3) and models each of these layers with a Generalized Linear Model (GLM). Actuaries are familiar with GLMs, given the long tradition of using GLMs in insurance pricing and reserving. Therefore, GLMs are the most likely candidate for supporting the transition from aggregate to individual reserving in practice. As is common in insurance pricing, we bin the continuous variables `age.insured`, `construction.year`, `property.value` and `rep.delay`. Table 3.2 shows the chosen bins for each covariate. We do not include the continuous development covariates `size.last.year` and `total.amount.paid` in the hierarchical GLM, since these covariates are highly correlated with the development year. On the first evaluation date, January 1, 2015, we select the optimal set of covariates for each of the three GLMs (`close`, `payment` and `size`) using forward selection with 5-fold cross validation, i.e. we iteratively add the covariate that results in the largest increase in the weighted likelihood (3.2) over all hold-out folds. In the moving window evaluation, we do not reselect the covariates on the other 364 evaluation dates, but recalibrate the parameters on each evaluation date using the most recent data.

Figure 3.3a shows the selected covariates in each GLM as well as a measure of the importance of each selected covariate. We compute covariate importance as the increase in the weighted likelihood (3.2) over all hold-out folds when sequentially adding covariates using forward selection. These increases are rescaled per GLM and sum to 100. For non-fire claims the set of selected covariates changes only slightly when we omit extreme weather events. This is in line with the low importance assigned to the covariate `extreme.weather` when these claims are included. The interaction `dev.year * rep.month` allows a more accurate determination of a claim's age, while still reserving in a yearly framework. This is by far the most important determinant for the settlement and payment process of non-fire claims. The importance of development year

Variable	Bins
`age.insured`	$[0, 39]$, $[40, 49]$, $[50, 64]$, 65+, NA
`construction.year`	$1950-$, $[1950, 1969]$, $[1970, 1984]$, 1985+, NA
`property.value`	$150\,000-$, $(150\,000, 200\,000]$, $(200\,000, 250\,000]$, $250\,000+$, NA
`rep.delay`	$5-$, $[5, 21]$, 21+

Table 3.2: List of chosen bins for the continuous covariates in the hierarchical GLM.

as a covariate for individual reserving is a strong validation for aggregate reserving models, which cannot use other covariates. Surprisingly, `dev.year` and `dev.year * rep.month` have little effect on the size of non-fire claims. The most important determinants for the payment size are the claim type as coded in `catnat` and the `initial.reserve` set by the expert. The data set contains less fire claims and as a result fewer covariates are selected in the corresponding GLMs. Although these GLMs might be less predictive, the few selected covariates obtain high importance scores, since scores are always scaled to 100. In particular, `valuables` has an importance of 100, since it is the only covariate selected in the GLM for the settlement of fire claims. As with non-fire claims the `initial.reserve` is an important predictor for the payment size.

Hierarchical GBM The hierarchical GBM follows the same three layer structure as the hierarchical GLM, but models each layer with a tree based Gradient Boosting Model (GBM). GBMs, as introduced by Friedman [2001], model the data with a sequence of shallow decision trees, in which each tree improves the fit of the previous trees. The GBM has three major advantages. First, through a sequence of trees a non-linear effect can be estimated for continuous covariates, thus removing the need to bin continuous variables. Second, automatic feature selection is integrated in the calibration process. Third, simple interaction effects between the covariates are automatically modelled. As a result of these advantages, the covariates `age.insured`, `construction.year`, `property.value`, `rep.delay` and `initial.reserve` can be included as continuous covariates. Furthermore, we do not include the interaction `dev.year * rep.month` as the model will automatically construct the relevant interactions. In return, a number of tuning parameters such as the number of trees and the depth of each tree have to be tuned. We tune these parameters on January 1, 2015 using the cross validation strategy of

(a) variable importance GLM

	non–fire			non–fire, exclude extreme weather			fire		
	close	payment	size	close	payment	size	close	payment	size
valuables							100		
age.insured								37	
profession		1			3				
sex					1				
construction	1			1				5	
property.value		5	9		5	2			
rep.delay			9			6		34	
rep.month	–	–	–	–	–	–	–	–	–
coverage			1						
catnat		4	54	4	1	75	×	×	×
global	5			×	×	×	×	×	×
initial.reserve		1	17		10	9			82
dev.year			10			6		24	18
dev.year * rep.month	94	88		95	79				
close	×	1		×	1	2	×		
size.last.year	–	–	–	–	–	–	–	–	–
total.amount.paid	–	–	–	–	–	–	–	–	–

(b) variable importance GBM

	non–fire			non–fire, exclude extreme weather			fire		
	close	payment	size	close	payment	size	close	payment	size
valuables	0	0	0	1	0	0	4	1	1
age.insured	6	6	6	5	5	5	9	11	14
profession	9	10	6	9	9	8	9	7	6
sex	0	1	0	0	1	0	1	1	0
construction	8	8	6	8	7	4	9	9	6
property.value	10	13	11	11	11	6	12	12	11
rep.delay	5	5	8	5	4	4	4	5	5
rep.month	18	22	10	17	23	10	22	24	15
coverage	0	0	2	0	0	0	0	0	0
catnat	6	6	11	6	7	7	×	×	×
global	0	0	0	×	×	×	×	×	×
initial.reserve	7	9	25	8	9	38	8	8	23
dev.year	13	8	1	11	10	1	3	6	3
dev.year * rep.month	–	–	–	–	–	–	–	–	–
close	×	1	1	×	1	3	×	1	4
size.last.year	6	5	6	7	5	8	7	7	4
total.amount.paid	12	6	7	12	8	6	12	8	8

× covariate unavailable   – covariate not included

Figure 3.3: Relative importance of the selected covariates in **(a)** the hierarchical GLM and **(b)** the hierarchical GBM. Relative importance is computed as the increase in likelihood attributed to a single covariate relative to the total increase in likelihood caused by all covariates.

Section 3.2.2. Once tuned, these parameters remain fixed throughout the 364 remaining evaluation dates.

Figure 3.3b shows the relative importance of the covariates in the various GBMs. The importance of a specific covariate is expressed as the total improvement of the loss function over all splits including that covariate averaged over the 365 evaluation dates and scaled to 100. Since there is no explicit variable selection, importance is distributed over all covariates, which complicates the interpretation. For recent claims, `rep.month` allows for a more granular expression of the time elapsed since reporting the claim, which is important when modelling the target variables close, payment and size. `initial.reserve`

is the most important covariate when predicting the size of payments. This shows that claim experts base their reserve estimate on information of the claim beyond the covariates available in model building. Similarly, the importance of other covariates shows that the practice of determining an initial reserve can be further improved by using a statistical model. The claim type `catnat`, which was important in the hierarchical GLM, is less important in the GBM.

Chain ladder method We compare the previous individual hierarchical reserving models with the classical chain ladder method based on yearly aggregated data. As indicated in Section 3.3.1, the chain ladder method can be rephrased as a hierarchical reserving model with a single layer, i.e. the payment size. On each evaluation date, we compute the RBNS reserve by applying the classical chain ladder method to a runoff triangle of payment sizes aggregated by reporting and development year. The choice for aggregating by reporting year results in an estimate for the RBNS reserve, as motivated in Section 3.3. Confidence bounds for the reserve estimate will be derived from a normal assumption combined with the standard errors under the Mack model [Mack, 1999].

Evaluation of the RBNS reserve

On each evaluation date we predict the expected RBNS reserve for the open claims over the next two years. We measure model performance via the percentage error of the predicted reserve compared to the actual reserve, that is

$$\text{percentage error} = \frac{\texttt{predicted} - \texttt{actual}}{\texttt{actual}} \cdot 100\%.$$

Figure 3.4 shows the evolution of the percentage error between January 2015 and December 2015 as obtained with the three hierarchical reserving models. The percentage error is capped at 100% for improved readability of the figures. Table 3.3 summarizes the daily errors by calculating the average percentage error and the average absolute percentage error over the 365 evaluation dates.

The reserve for non-fire claims (Figure 3.4a) combines the outstanding amounts on many small claims, which provides a sufficiently rich data set for accurately training the individual hierarchical reserving models. This results in a similar performance for the hierarchical GLM and GBM. Extreme weather events in past years, produce outliers in the cells in the runoff triangle. This has a large impact on the chain ladder method, which fails to provide reasonable reserve estimates. This is a well known weakness of the chain ladder method and it is interesting that the individual models do not have this weakness,

Portfolio	hierarchical GLM		hierarchical GBM		chain ladder	
	$\mu(PE)$	$\mu(|PE|)$	$\mu(PE)$	$\mu(|PE|)$	$\mu(PE)$	$\mu(|PE|)$
non-fire claims	0.92	**7.32**	-1.80	10.23	33.89	51.31
non-fire claims, exclude extreme weather	-9.76	**14.90**	-14.28	20.18	-18.10	19.07
fire-claims	-20.82	**26.44**	-16.42	26.50	-28.41	29.76

Table 3.3: Evaluation of the average performance of the hierarchical GLM, hierarchical GBM and chain ladder method over 365 evaluation dates between January 1, 2015 and December 31, 2015. Average performance is expressed as the mean percentage error and the mean absolute percentage error.

since they scale the reserve estimate automatically with the number of claims. Furthermore, Figure 3.3 shows that the covariate `extreme.weather` is rarely selected in the hierarchical models, which indicates that the development of these extreme weather claims does not fundamentally differ from regular claims. Not having to separate these extreme weather events from the other claims is a major advantage of individual reserving. When we remove extreme weather events (Figure 3.4b), performance across all three models becomes relatively competitive. Both the chain ladder method and the individual hierarchical models benefit from a data set with a large number of claims. Table 3.3 shows that performance is slightly better for the individual models, which surprisingly perform even better when we would not exclude extreme weather events from the data set. We observe higher prediction errors for all three models when predicting the reserve for fire claims (Figure 3.4c). The combination of a low claim frequency and potentially high costs makes the reserve for fire claims difficult to predict.

3.5 Conclusion

We propose the hierarchical reserving model as a general framework for RBNS reserving in discrete time. By adding layers and choosing predictive models this framework can be tailored to any insurance portfolio. At the same time, our approach enables the development of best practices for calibration (see Section 3.2.2) and offers statistical tools for comparing hierarchical models. Model comparison extends to many existing reserving models, which can be restated as hierarchical reserving models. Moreover, Section 3.3 presents a connection with aggregate reserving models, allowing a data driven choice between aggregate and individual reserving. We illustrate our framework on
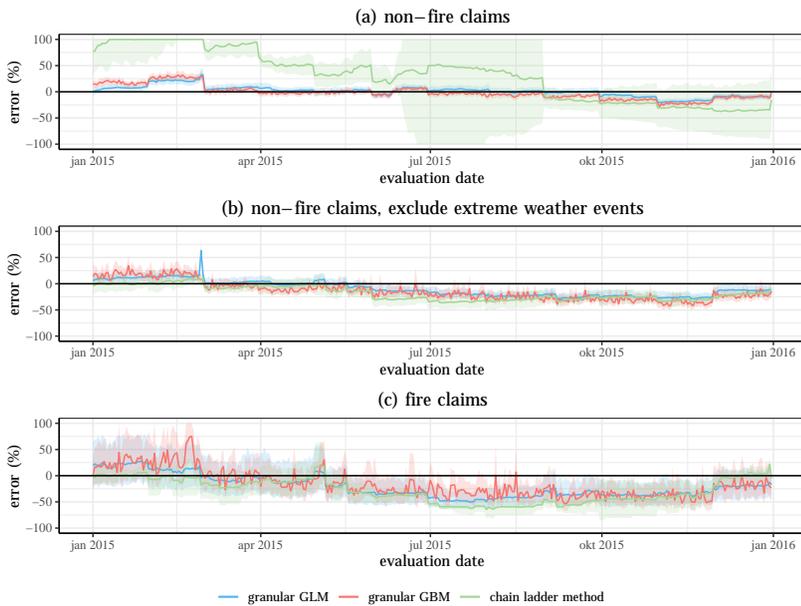
Figure 3.4: Percentage error in the prediction of the RBNS reserve on evaluation dates between January 1, 2015 and December 31, 2015 under the hierarchical GLM, hierarchical GBM and the chain ladder method. Errors are limited to 100%. **(a)** shows the reserve for non-fire claims, **(b)** the reserve for non-fire claims, when extreme weather events are excluded and **(c)** the reserve for fire claims.

a detailed case study with a home insurance data set. The flexibility of the framework is demonstrated by calibrating the same three layer structure with generalized linear models and gradient boosting models. As a best practice, we minimize the effect of day-to-day volatility when comparing our reserving models, by evaluating the performance over 365 evaluation days. The individual hierarchical models consistently outperform the classical chain ladder method on aggregated data and have the additional benefit that extreme weather events do not have to be removed prior to reserving.

3.6 Acknowledgements

Chapter 4

Bridging the gap between pricing and reserving with an occurrence and development model for non-life insurance claims

Abstract

Due to the presence of reporting and settlement delay, historical claim data sets in non-life insurance are typically incomplete. As a result observed claim counts and claim severities are right censored. Therefore, non-life insurance pricing is currently approached via a two-step procedure. First, insurers compute best estimates for claim frequency and severity at the level of individual policies based on the incomplete, historical claim data. Second, pricing actuaries build predictive models to estimate technical, pure premiums for new policies by treating these best estimates as actual observed outcomes, hereby neglecting the uncertainty present in them. We propose an alternative one-step approach for non-life pricing by analysing the incomplete information registered during the development of claims. The granularity of our model allows it to be applied to both pricing and reserving, hence bridging two key actuarial tasks that have traditionally been discussed in silos. We illustrate our proposed model on a reinsurance portfolio, where large uncertainties in the best estimates result from

long reporting and settlement delays, low claim frequencies and extreme claim sizes.

This chapter is based on Jonas Crevecoeur and Katrien Antonio. Bridging the gap between pricing and reserving with an occurrence and development model for non-life insurance claims. 2020b. Working paper.

4.1 Introduction

The insurance industry is characterized by an inverted production cycle in which the premium for a new policy is determined before observing the associated loss. Estimating these losses is the task of pricing actuaries. In non-life insurance, the total loss L on a new policy is often estimated with a frequency-severity decomposition [Denuit et al., 2007, Frees and Valdez, 2008], which models the expected loss as the product of the expected number of claims $E(N)$ (frequency) and the expected cost per claim $E(Y)$ (severity), i.e.

$$E(L) = E(N) \cdot E(Y).$$

Typically, independence is assumed between the frequency and severity component of the loss. Personalized, risk-based premiums are obtained by analysing the risk characteristics of individuals while building predictive models for claim frequency and severity. Pricing requires a data set with claim counts at the level of individual polices and claim sizes at the level of individual claims. However, an aspect completely ignored by pricing literature is that claim counts and claim sizes are rarely observed due to reporting and settlement delays in the claim development process. This is particularly relevant in long-tailed business lines (e.g. workers' compensation and reinsurance) where claim settlement can take several years.

Figure 4.1 visualizes the development process of a single claim. This process starts with the occurrence of an accident, which is reported to the insurer after some delay. If the claim is eligible for compensation under the insurance policy, a number of payments follow. Finally, the claim settles and we observe its total cost. Depending on the insurer and line of business other relevant events (e.g. the involvement of a lawyer) will be recorded during the lifetime of a claim. For settled claims, we observe the full development process and thus the total claim size. However, the development process is only partially observed for reported, but not yet settled claims. For claims that occurred in the past, but are not yet reported the development process is completely missing from the database maintained by the insurance company.
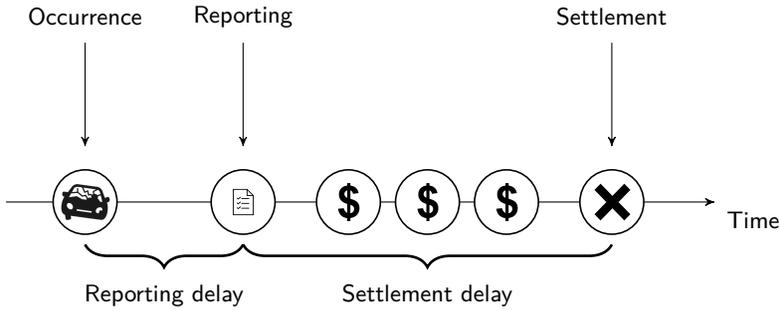
Figure 4.1: Development process of a single claim

Due to the delays present in the claim development process, we observe the number of reported claims instead of the number of actually occurred claims, which also includes claims that are not (yet) reported. Similarly, the observed claim sizes for open claims underestimate actual losses, since future payments are missing. As a result of the incomplete claim history, pricing is in practice a two step approach. First, claim counts and sizes are estimated per policy and per claim, respectively, based on the available claim history, i.e.

$$\hat{N} = E(N \mid \mathcal{F}_\tau) \quad \text{and} \quad \hat{Y} = E(Y \mid \mathcal{F}_\tau),$$

where $\mathcal{F}_\tau$ denotes the information available at the observation date τ. In a second step, these estimates, so called best-estimates, are treated as actual observations when the pricing actuary constructs predictive models for claim frequency and severity.

Best estimates as constructed in the first step of this pricing procedure, can be obtained in several ways. In the most common approach, claim handlers estimate the future claim cost based on their expert opinion. Combined with the amount already paid, this estimate constitute the expert's best estimate of the total claim size, which is also called the incurred. Alternatively, methods from non-life reserving can be adapted to estimate the total, ultimate cost of individual claims. The literature on non-life reserving unravels along two axes: aggregate and individual reserving models. Aggregate reserving models (e.g. the chain ladder method [Mack, 1993, 1999]) ignore individual claim characteristics and model a single claim development process for all claims that occur within an accident year. We obtain best estimates for pricing by applying this development process to individual, reported claims. Constructing best estimates based on an aggregate reserving model has two important disadvantages. First, most aggregate reserving models do not distinguish between open and settled claims.

In this case, consistency with the reserving model requires pricing actuaries to replace claim sizes for both open and settled claims by best estimates. Second, ignoring risk characteristics when constructing best estimates reduces the effect of these characteristics when modelling claim severity. Following Norberg [1993, 1999] individual reserving models have emerged, which construct best estimates at the level of individual claims. In the recently developed literature on individual reserving, we see most potential in a stream of individual reserving models in discrete time adopting techniques from pricing literature. Larsen [2007], Wüthrich [2018], Crevecoeur and Antonio [2020a] and Łukasz Delong et al. [2020] focus on Generalized Linear Models (GLMs), regression trees, gradient boosting models and neural networks, respectively. Claim covariates in these models result in best estimates tailored to individual claims.

Pricing literature mainly focuses on the second step of the pricing procedure, where a statistical model is fitted to the best estimates. Although actual observations and best estimates follow different statistical distributions, the frequency-severity decomposition still holds, i.e.

$$E(L) = E(N) \cdot E(Y) = E\big(E(N \mid \mathcal{F}_\tau)\big) \cdot E\big(E(Y \mid \mathcal{F}_\tau)\big),$$

as a result of the tower rule for conditional expectations. This property is essential for traditional pricing, since it enables an unbiased estimation of the loss based on best estimates. However, many other properties of the loss (e.g. the variance) are not preserved when treating best estimates as actual observations. In particular, as a result of Jensen's inequality [Jensen, 1906], severity is underestimated for policies covering losses above a deductible d, i.e.

$$E((Y - d)_+) \geq E\left(((E(Y \mid \mathcal{F}_\tau) - d)_+\right).$$

This is especially relevant in excess-of-loss reinsurance pricing, where deductibles are high and long settlement delays result in many open claims. Moreover, when pricing based on best estimates, risk characteristics selected when modelling frequency and severity should be interpreted as effects on the best estimate rather than effects on the actual observations. Parameters capturing these effects are likely affected by the method used for constructing these best estimates.

We propose a novel, one-step approach for non-life insurance pricing by analysing the potentially incomplete information registered during the development of claims. This approach resolves the inconsistencies between actual observations and best estimates in traditional pricing. Moreover, by modelling the occurrence and development process of claims, our proposed model can be used for both non-life pricing and non-life reserving, hereby bridging two key actuarial tasks of a non-life actuary. We demonstrate our methodology with a case-study on pricing and reserving based on a reinsurance data set. This is one of the first works applying techniques from individual reserving on a reinsurance portfolio.

The reinsurance industry is characterized by low claim frequencies and high claim severities [Albrecher et al., 2017], which demands special attention when building predictive models for the development of individual claims.

This chapter is organized as follows. Section 4.2 introduces a model for the occurrence and development of non-life insurance claims. Section 4.3 illustrates how this model can be used for pricing and reserving non-life policies. Section 4.4 demonstrates this methodology in a case-study on a reinsurance data set. Section 4.5 concludes this chapter.

4.2 An occurrence and development model for non-life claims

We present a discrete time occurrence and development model (ODM) at the level of individual policies and claims, respectively. Our ODM consists of two parts. Section 4.2.1 analyses the occurrence and reporting of claims, while Section 4.2.2 analyses the development of claims after reporting. Together these two parts model the full claim history.

4.2.1 Occurrence and reporting of non-life claims

We consider a portfolio with historical data on n policies, each with a coverage period contained within a single calendar year. In case the historical data set contains policies covering multiple calendar years, these policies will be split by calendar year and treated as separate policies. Let N_i denote the claim frequency of policy i, i.e. the total number of claims that occur during the coverage period for policy i. Due to a possible delay in reporting (see Figure 4.1), these counts N_i are not directly observable. Instead we observe counts $N_{i,j}$, which register the number of claims from policy i that are reported in the j-th year after occurrence, i.e. year $\texttt{occ(i)} + j - 1$, where $\texttt{occ(i)}$ is the occurrence year covered by policy i. At the observation date τ, we observe the counts $N_{i,j}$ with $\texttt{occ(i)} + j - 1 \leq \tau$, whereas the claims counted in $N_{i,j}$ with $\texttt{occ(i)} + j - 1 > \tau$ are not yet reported. We propose a model for predicting these unreported counts $N_{i,j}$. Following Jewell [1990] and Norberg [1993], our model is based on the following assumptions:

(F1) Claims are reported with a maximal delay of d years. This maximal delay d is at most the length of the observation window τ of the portfolio, i.e. $d \leq \tau$.

(F2) Claim counts N_i are independent and follow a Poisson distribution with intensity $\lambda_i(\boldsymbol{x}_i)$, which depends on policy covariates $\boldsymbol{x}_i$.

(F3) Conditional on the total number of claims N_i, the reported counts $N_{i,j}$ are multinomially distributed with reporting probabilities $p_{i,j}(\boldsymbol{x}_i)$, which depend on policy covariates $\boldsymbol{x}_i$.

Assumption (F1) limits the reporting delay and allows us to retrieve the total claim frequency on policy i as

$$N_i = \sum_{j=1}^{d} N_{i,j}.$$

Assumption (F2) follows the insurance pricing literature by modelling claim frequency with a Poisson distribution. Assumption (F3) is very general, but requires each claim to be reported independent of the other claims in the portfolio. The independence assumptions in (F2-F3) are similar to those in classical insurance pricing, but might be violated in case of high impact events, e.g. extreme weather, where claims occur in clusters. We choose these assumptions, nevertheless, since they are essential for modelling occurrence and reporting at the level of individual policies. As a result of the thinning property for Poisson distributions, assumption (F2-F3) implies

$$N_{ij} \sim \texttt{Poisson}(\lambda_i(\boldsymbol{x}_i) \cdot p_{i,j}(\boldsymbol{x}_i)).$$

The set of observed claims consists of $\{N_{ij} \mid i = 1, \ldots, n, j = 1, \ldots, \tau_i\}$, where $\tau_i := \max(d, \tau - \texttt{occ}(i) + 1)$ is the number of observed reporting years for policy i. The associated log-likelihood is

$$\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{p}) = \sum_{i=1}^{n} \sum_{j=1}^{\tau_i} -\lambda_i(\boldsymbol{x}_i) \cdot p_{ij}(\boldsymbol{x}_i) + N_{ij} \cdot \log(\lambda_i(\boldsymbol{x}_i)) + N_{ij} \cdot \log(p_{ij}(\boldsymbol{x}_i)) - \log(N_{ij}!).$$

(4.1)

We extend the work of Verbelen et al. [2019] for optimizing this likelihood with respect to chosen structures for $\boldsymbol{\lambda}$ and $\boldsymbol{p}$. Our main contribution relates to specifying this likelihood at the level of individual policies and the chosen structure for the reporting process driven by $\boldsymbol{p}$. Verbelen et al. [2019] argue that the joint estimation of $\boldsymbol{\lambda}$ and $\boldsymbol{p}$ in (4.1) is complicated due to the presence of the interaction term $\lambda_i(\boldsymbol{x}_i) \cdot p_{ij}(\boldsymbol{x}_i)$. Using an EM-algorithm [Dempster et al., 1977], which treats the observations $\{N_{ij} \mid i \leq n, \tau_i < j \leq d\}$ as hidden observations, the authors decouple the occurrence, $\lambda_i(\boldsymbol{x}_i)$, and reporting, $p_{ij}(\boldsymbol{x}_i)$, parameters in (4.1). Following Verbelen et al. [2019], we predict in the E-step the hidden observations as

$$\tilde{N}_{ij}(\boldsymbol{x}_i) = \begin{cases} N_{ij} & j \leq \tau_i \\ \lambda_i(\boldsymbol{x}_i) \cdot p_{ij}(\boldsymbol{x}_i) & \tau_i < j \leq d \end{cases},$$

and maximize in the M-step the completed log-likelihood

$$\mathcal{L}_c(\boldsymbol{\lambda}, \boldsymbol{p}) = \sum_{i=1}^{n} -\lambda_i(\boldsymbol{x}_i) + \tilde{N}_i \cdot \log(\lambda_i(\boldsymbol{x}_i)) + \sum_{j=1}^{d} \tilde{N}_{i,j} \cdot \log(p_{i,j}(\boldsymbol{x}_i)) - \log(\tilde{N}_{i,j}!),$$

where $\tilde{N}_i = \sum_{j=1}^{d} \tilde{N}_{i,j}$. The likelihood in (4.1) now splits in an occurrence and reporting contribution. For the occurrence process, we maximize

$$\mathcal{L}_c^{\texttt{occ}}(\boldsymbol{\lambda}) = \sum_{i=1}^{n} -\lambda_i(\boldsymbol{x}_i) + \tilde{N}_i \cdot \log(\lambda_i(\boldsymbol{x}_i)).$$

This likelihood is proportional to the Poisson likelihood optimized in insurance pricing when modelling claim frequency. The partially observed claim counts N_i, however, are replaced by counts $\tilde{N}_i$, which adjust the data for unreported claims. For the reporting process, we maximize

$$\mathcal{L}_c^{\texttt{rep}}(\boldsymbol{p}) = \sum_{i=1}^{n} \sum_{j=1}^{d} \tilde{N}_{ij} \cdot \log(p_{ij}(\boldsymbol{x}_i)), \quad \text{subject to} \quad \sum_{j=1}^{d} p_{ij}(\boldsymbol{x}_i) = 1, \forall i. \quad (4.2)$$

The estimation of a distribution for the reporting probabilities p_{ij} in this multinomial likelihood is complicated by the sum-to-one restriction on the reporting probabilities for each policy i. For this reason, we reparametrize the d probabilities $(p_{ij})_{j=1...,d}$ into $d-1$ probabilities $(q_{ij})_{j=1...,d-1}$ as follows

$$p_{ij} = \begin{cases} \prod_{\kappa=1}^{d-1} q_{i,\kappa} & j = 1 \\ (1 - q_{i,j-1}) \cdot \prod_{\kappa=j}^{d-1} q_{i,\kappa} & 1 < j < d \\ (1 - q_{i,d-1}) & j = d \end{cases} . \quad (4.3)$$

The one-to-one correspondence between the probabilities $p_{i,j}$ and $q_{i,j}$ allows us to switch from the p probabilities, which are restricted by the sum-to-one condition, to the unrestricted q probabilities. Combining (4.3) with (4.2), the likelihood for the reporting process becomes

$$\mathcal{L}_c^{\texttt{rep}}(\boldsymbol{q}) = \sum_{i=1}^{n} \sum_{\kappa=1}^{d-1} \left(\sum_{j=1}^{\kappa} \tilde{N}_{i,j} \right) \cdot \log(q_{i,\kappa}(\boldsymbol{x}_i)) + \sum_{i=1}^{n} \sum_{j=1}^{d-1} \tilde{N}_{i,j+1} \cdot \log(q_{i,j}(\boldsymbol{x}_i)). \quad (4.4)$$

This likelihood is a sum of binomial likelihood contributions and can be optimized with standard statistical modelling techniques.

In pricing, the occurrence process is used when estimating the expected claim frequency of specific insurance policies. This occurrence process is adjusted for the presence of not yet reported claims. In reserving, the estimated occurrence and reporting processes allow us to estimate the number of unreported claims and their associated reporting delays.

4.2.2 A hierarchical model for the development of reported non-life claims

Insurers track many dynamic claim characteristics (e.g. amount paid, settlement status, involvement of a lawyer) over the lifetime of a claim. Extending Chapter 3, we construct a hierarchical model to predict the joint evolution of these claim characteristics. Our approach differentiates between the initial state of the claim characteristics as observed in the reporting year of the claim and the updates in later years. We let the vector $\boldsymbol{I}_k$ structure the initial claim characteristics for claim k at the end of the reporting year, denoted $\texttt{rep(k)}$. In later years, update vectors $\boldsymbol{U}_k^j$ structure the evolution of claim k in the j-th year since reporting, i.e. year $\texttt{rep(k)} + j - 1$. The choice of the claim characteristics captured by the vectors $\boldsymbol{I}_k$ and $\boldsymbol{U}_k^j$ depends on the portfolio at hand. The case-study in Section 4.4 illustrates a possible set up in which the joint evolution of the settlement status, amount paid and incurred are tracked over the lifetime of a claim. From now on we refer to these chosen characteristics as the layers of our model. Let the vector $\mathcal{X}_k$ store the observed development of claim k, i.e.

$$\mathcal{X}_k := \{I_k, U_k^2, \ldots, U_k^{\tau_k}\},$$

with $\tau_k = \tau - \texttt{rep(k)} + 1$ the number of observed years since reporting for claim k. Our approach models the full claim evolution recorded in $\mathcal{X}_k$ based on a single assumption.

(S1) The development of a claim is independent of the development of the other claims in the portfolio.

This independence assumption is essential for modelling the development at the level of individual claims. As a result of (S1) we can write the likelihood for a portfolio with m reported claims as

$$\mathcal{L} = \prod_{k=1}^{m} f(\boldsymbol{I}_k, \boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{\tau_k} \mid \boldsymbol{x}_k),$$

where $f(\boldsymbol{I}_k, \boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{\tau_k} \mid \boldsymbol{x}_k)$ is the joint likelihood of the observed development process of claim k and $\boldsymbol{x}_k$ denotes the static claim covariates for claim k. Our hierarchical approach decomposes this joint likelihood over time as well as over the layers of $\boldsymbol{I}_k$ and $\boldsymbol{U}_k^j$ by applying the law of conditional probability twice. First, the likelihood is split in chronological order

$$\mathcal{L} = \prod_{k=1}^{m} f(\boldsymbol{I}_k \mid \boldsymbol{x}_k) \cdot \prod_{j=2}^{\tau_k} f(\boldsymbol{U}_k^j \mid \boldsymbol{I}_k, \boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{j-1}, \boldsymbol{x}_k).$$

By conditioning on past events, we allow our model to use the historical development of a claim (e.g. total amount paid, reserve, settlement status in previous years) when predicting the development in future years. Second, we decompose the likelihood over the layers of $\boldsymbol{I}_k$ and $\boldsymbol{U}_k^j$

$$\mathcal{L} = \prod_{k=1}^{m} \prod_{l=1}^{v} f(I_{k,l} \mid I_{k,l}, \ldots, I_{k,l-1}, \boldsymbol{x}_k) \times$$

$$\prod_{k=1}^{m} \prod_{j=2}^{\tau_k} \prod_{l=1}^{w} f(U_{k,l}^j \mid \boldsymbol{I}_k, \boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^{j-1}, U_{k,1}^j, \ldots, U_{k,l-1}^j, \boldsymbol{x}_k).$$

where v and w denote the length of the initial vector $\boldsymbol{I}_k$ and update vector $\boldsymbol{U}_k^j$ respectively. Through conditioning on previous layers, we allow for dependencies in the development of the claim characteristics within a year. We model this decomposed likelihood by specifying a statistical model for the likelihood contribution related to each layer leading to a total of $v + w$ statistical models.

We use the proposed hierarchical model to estimate the claim severity of individual policies in pricing. In reserving, this model allows us to estimate the future cost of reported as well as not yet reported claims.

4.3 Pricing and reserving with the occurrence and development model

Our ODM provides insights in the occurrence and development of individual claims. In pricing, we use these insights to predict the future loss for new policies. In reserving, our ODM estimates the reporting and future development of claims that occurred in the past.

4.3.1 Non-life pricing with the occurrence and development model

Following the frequency-severity decomposition we estimate the pure premium π_i for policy i as the product of the expected claim frequency, $E(N_i)$, and the expected claim severity, $E(Y_i)$, i.e.

$$\pi_i = E(N_i) \cdot E(Y_i).$$

By following this decomposition we assume independence between claim frequency and severity. Claim frequency estimates follow immediately from the

claim count model proposed in Section 4.2.1. In contrast with traditional claim frequency models, our approach adjusts the estimated claim frequency for the presence of unreported claims.

We consider two approaches for modelling claim severity based on our ODM. The first approach simulates new claims for a given policy from ground up, whereas the second approach simulates the future development of open claims.

Simulating new claims   This approach uses our fitted ODM to simulate a large number of new claims for policy i, which are then averaged to obtain an estimate of its expected severity. Algorithm 2 outlines the procedure for simulating a new claim from a policy with characteristics $\boldsymbol{x}$.

Algorithm 2: Simulating the severity of a new claim

Input: policy with characteristics $\boldsymbol{x}$
Output: simulation of the claim severity

Simulate rep.delay given $\boldsymbol{x}$.
Simulate $\boldsymbol{I}$ given $\boldsymbol{x}$, rep.delay.
Set $s = 1$.
if *not.settled($\boldsymbol{I}$)* **then**
   do
      Simulate $\boldsymbol{U}^{s+1}$ given $\boldsymbol{x}$, rep.delay, $\boldsymbol{I}, \boldsymbol{U}^2, \ldots, \boldsymbol{U}^s$.
      Set $s = s + 1$.
   while *not.settled($\boldsymbol{I}, \boldsymbol{U}^2, \ldots, \boldsymbol{U}^s$)*
end
Evaluate $Y = Y(\boldsymbol{I}, \boldsymbol{U}^2, \ldots, \boldsymbol{U}^s)$.

It is essential that the amount paid or the incurred is tracked within $\boldsymbol{I}$ and $\boldsymbol{U}^j$, such that the claim severity at settlement can be computed as a function of the simulated development process. By simulating a large number of claims for a given policy we obtain an empirical distribution for its severity from which we compute the expected severity. Besides total claim severity, this approach estimates the future cash flows on a policy, which is important for matching durations in asset-liability management and facilitates discounting when computing net present values. In this strategy, claim severity follows directly from our estimated ODM. However, many layers in the hierarchical development model of Section 4.2.2 may result in a less transparent severity estimate, which limits the applicability of this approach for regulatory and commercial purposes.

Simulating future paths for open claims This alternative approach for severity modelling complements the observed development of claims up to the present moment with a large number of simulated future paths, say $n_{\texttt{path}}$, for each open claim. For each claim k and path p we compute the total claim size $Y_{k,p}$ to obtain a distribution for the total claim size per open claim. In a second step, we fit a severity distribution by assigning a weight of one to actual observations from closed claims and a weight of $\frac{1}{n_{\texttt{path}}}$ to all simulated paths from open claims, i.e. we maximize the following log-likelihood

$$\mathcal{L}^{\texttt{ODM}}(f_Y) = \sum_{k=1}^{m} \delta_k \cdot \log(f_Y(Y_k)) + (1 - \delta_k) \cdot \frac{1}{n_{\texttt{path}}} \cdot \sum_{p=1}^{n_{\texttt{path}}} \log(f_Y(Y_{k,p})), \quad (4.5)$$

where f_Y is the proposed parametric severity distribution and δ_k is one when claim k settles before the evaluation date τ and zero otherwise. This likelihood includes all possible paths for open claims, whereas traditional severity models average these paths to obtain a best estimate and would maximize

$$\mathcal{L}^{\texttt{trad}}(f_Y) = \sum_{k=1}^{m} \delta_k \cdot \log(f_Y(Y_k)) + (1 - \delta_k) \cdot \log\left(f_Y \left(\sum_{p=1}^{n_{\texttt{path}}} \frac{1}{n_{\texttt{path}}} \cdot Y_{k,p} \right) \right).$$

By focusing on reported claims, this approach for severity modelling stays close to traditional pricing practice, while resolving the contradiction between best estimates and actual observations in traditional pricing.

4.3.2 Non-life reserving with the occurrence and development model

Reserving models estimate the aggregated future cost for all claims from a past exposure period. We divide the total claims reserve in a reserve for incurred, but not (yet) reported claims, i.e. the IBNR reserve, and a reserve for reported, but not (yet) settled claims, i.e. the RBNS reserve. The total reserve, denoted $\mathcal{R}$, is the sum of these two reserve contributions, i.e.

$$\mathcal{R} = \mathcal{R}^{\texttt{IBNR}} + \mathcal{R}^{\texttt{RBNS}}.$$

We compute the IBNR reserve by aggregating the expected severity from unreported claims, i.e.

$$E(\mathcal{R}^{\texttt{IBNR}}) = \sum_{i=1}^{n} \sum_{j=\tau_i+1}^{d} \mathrm{E}(N_{i,j}) \cdot \mathrm{E}(Y_i \mid \texttt{rep.delay} = j).$$

Similar to the frequency-severity decomposition in pricing, this formula assumes independence between the number of claims and the claim severity. Estimates

for the number of reported claims per year, $N_{i,j}$, follow immediately from the occurrence and reporting processes proposed in Section 4.2.1. Claim severity is estimated with the techniques outlined in Section 4.3.1 for pricing.

For the RBNS reserve, we compute the future cost for all reported, but not yet settled claims. Following Chapter 3, we simulate future paths for the development of all open claims and estimate the RBNS reserve by aggregating the future costs from these simulated paths.

4.4 Case-study on pricing and reserving large motor claims

We illustrate our method on a Belgian motor third party liability (MTPL) reinsurance data set registering the detailed development of 4277 large motor insurance claims between 2000 and 2017. These claims originate from 21 underlying MTPL insurance portfolios, which act as the clients from the reinsurance perspective. We label these portfolios A, B, . . . , U. This case-study analyses the claims from these portfolios for pricing and reserving excess-of-loss reinsurance contracts. In this type of contract, the reinsurer reimburses the costs for individual claims exceeding a deductible D and up to a limit L [Albrecher et al., 2017].

For large claims, insurers carefully track the evolution of the expert's judgement of its expected total cost, the so-called incurred, which is the sum of the amount already paid and the expected future payments. For the purpose of pricing excess-of-loss reinsurance contracts, insurers are obliged to report a claim to the reinsurer once its incurred exceeds a predefined threshold, the so-called reporting priority. The reporting priority is determined upfront and depends on both the underlying portfolio and the occurrence year. Usually reporting priorities are set significantly below the deductible D of the excess-of-loss contract, hereby providing the reinsurer with data on sufficiently many claims to estimate the cost above the deductible. Figure 4.2 visualizes the thresholds (priority, deductible and limit) for our excess-of-loss contract. In this example, claim 1 is reported to the reinsurer in year 2 when the incurred first exceeds the reporting priority. Even when the incurred of claim 2 falls below the priority in year 4, the reinsurer keeps receiving yearly updates on this claim. At settlement, the amount incurred and paid are equal and the reinsurer covers the losses between the deductible and the limit (region III), while the insurer covers the remaining losses (regions I, II and IV).

For evaluating model performance, we split the data and train our model on
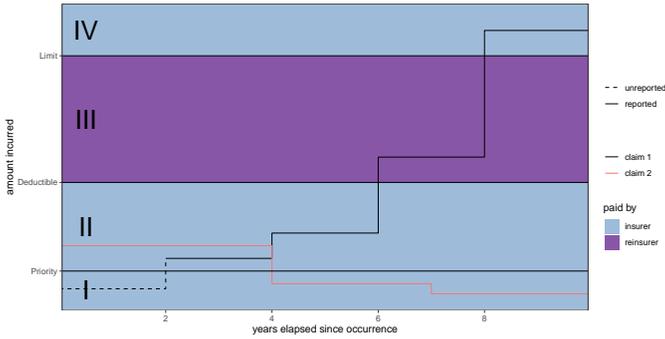
Figure 4.2: Illustration of the different thresholds in an excess-of-loss contract. Claims are reported when the incurred first exceeds the reporting priority. At settlement, the loss between the deductible and the limit is covered by the reinsurer.

the years 2000-2014. The remaining years 2015-2017 constitute the out-of-time test data set. Before fitting our ODM we apply three modifications to the data. First, we remove negative payments. Since the data set contains only a small number of negative payments ($< 2\%$ of the total amount paid), we believe that the potential gain in model accuracy by incorporating negative payments does not outweigh the increase in model complexity and uncertainty when doing so. Second, we smooth the data by removing changes in the incurred and amount paid of less than 100 euro. Finally, we deflate the data to the level of 2014 using the inflation curve provided by the reinsurer. After modelling the payment patterns, we reinflate these payments to the corresponding payment years when calculating prices and reserves.

4.4.1   Occurrence and reporting of large claims

We slightly adapt Section 4.2.1 to our reinsurance setting. We denote by policy a reinsurance contract on an insurance portfolio covering a single underwriting year and index these policies by i. In our data set, a claim from policy i is reported when the incurred amount exceeds the reporting priority, denoted `priority(i)`. These priorities are policy-specific, which complicates the comparison of occurrence intensities and reporting delays across policies with different priorities. Therefore, we choose a new, common priority P and let N_{ij}^P denote the number of claims from policy i for which the incurred first exceeds the priority P in year `occ(i)` $+ j - 1$. The total number of claims from

policy i that exceed the priority P at least once is

$$N_i^P := \sum_{j=1}^{d} N_{ij}^P.$$

Since long reporting delays are common in reinsurance, we set the maximal delay d equal to 15, i.e. the length of the observation window of our data set. Choosing a common reporting priority P naturally restricts the available data for fitting our model to policies for which we observe all claims exceeding P, i.e. `priority(i)` $\leq P$. Only for these policies, we observe the reported claim counts N_{ij}^P. Since we want to investigate the effect of choosing the priority P on the computed price, we model the occurrence intensity and reporting delay at three common priorities, $750\,000$, $1\,000\,000$ and $1\,250\,000$. From the 21 portfolios in our data set, we observe at these priorities claims from 9, 15 and 15 portfolios, respectively.

Following Section 4.2.1, we model the occurrence process with a Poisson distribution with intensity

$$\lambda_i = e_i \cdot \lambda_{\texttt{portfolio(i)}},$$

where e_i is the exposure expressed as the number of vehicles insured by policy i and $\lambda_{\texttt{portfolio(i)}}$ is the portfolio-specific effect on the claim intensity. We model the reporting probabilities $p_{i,j}$ via their one-to-one connection to the probabilities $q_{i,j}$ introduced in (4.3). The q probabilities are estimated by maximizing likelihood (4.4) via a binomial GLM with logit link function and

$$q_{i,j} = 1 - \exp(-\exp(\gamma_j + \gamma_{\texttt{portfolio(i)}})),$$

where γ_j is the effect of the reporting year and $\gamma_{\texttt{portfolio(i)}}$ captures reporting delay variations across portfolios.

Figure 4.3 visualizes the estimated occurrence intensity and reporting delay distribution above $P = 750\,000$ for the 9 portfolios available at this priority. Figure 4.3a shows the occurrence intensity per $100\,000$ insured vehicles in the underlying portfolio. We clearly distinguish two regimes in the occurrence intensity: low occurrence intensities ($2.17 - 2.51$ large claims per $100\,000$ vehicles) in portfolio A, H, K and O and high occurrence intensities ($3.18 - 3.54$ large claims per $100\,000$ vehicles) in portfolio B, I, J, M and S. This split in two regimes could indicate a different share of large vehicles (e.g. buses and trucks) insured in these portfolios.

Figure 4.3b shows the estimated reporting delay distribution per portfolio. The incurred amount, as the sum of the amount paid so far and the insurer's expert estimate of the outstanding claim amount, is volatile in the first years after
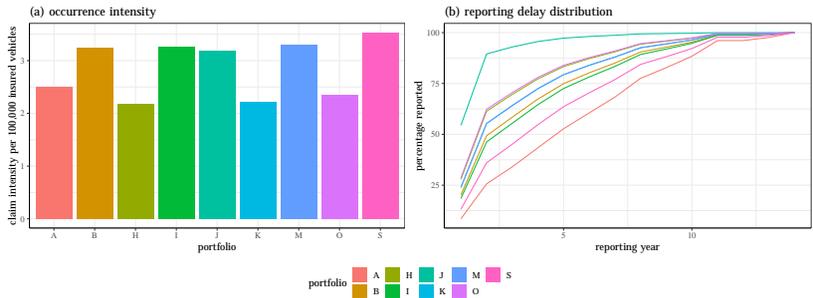
Figure 4.3: **(a)** Estimated number of claims exceeding the priority of 750 000 per 100 000 insured vehicles in 9 portfolios and **(b)** fitted reporting delay distribution for the reporting of claims exceeding the priority of 750 000 by insurer

the occurrence of claims, when it is not yet determined which insurer(s) should reimburse the claim. As a result, long reporting delays between the occurrence of a claim and its incurred exceeding the priority for the first time are common. Moreover, since each insurer has its own reserving policy, we find considerable differences in reporting delay across portfolios.

The insights revealed in Figure 4.3 are important for reinsurers when pricing contracts on these portfolios. Reinsurers can consider sharing these insights with their policyholders, i.e. the insurers. This would allow insurers to compare the reporting delay for their portfolio to the rest of the market and motivate insurers with long reporting delays to improve their reserving strategy for large claims.

4.4.2 A hierarchical model for the development of large claims after reporting

For each reported claim, our data set tracks the evolution of the settlement status, the amount paid and the amount incurred per year. Since these events in a claim's development process are clearly dependent (e.g. no payments for settled claims, low settlement probability when the reserve (incurred - paid) is large), we use the hierarchical model of Section 4.2.2 to model the joint evolution of these claim characteristics.

We choose a reporting priority, $P = 750\,000$, and interpret $\boldsymbol{I}_k$ as the status of claim k when its incurred first exceeds 750 000. The top panel of Figure 4.4a visualizes our 3-layer hierarchical model for $\boldsymbol{I}_k$. At reporting, the incurred

exceeds the reporting priority of 750 000. Layer 1 models the excess incurred, i.e. the difference between the initial incurred and this reporting priority. The outcome of this first layer is an input when modelling payments. Layer 2 models whether a part of the incurred has already been paid at reporting. In this case, layer 3 models the amount paid at reporting as a percentage of the total incurred. We do not model the settlement status in the year of reporting, since in our data set large claims never settle immediately at reporting.

Figure 4.4b visualizes our 8-layer hierarchical model for the updates $\boldsymbol{U}_k^j$ in the years after reporting. First, layer 1 models the settlement status of a claim. Settlement status is an input when modelling payments. Layer 2 models the presence of a payment and layer 3 models the size of a payment conditional on the presence of a payment. Note that we only take payments above 100 into account. Following a payment, we deterministically decrease the reserve by the payment size. When the claim settles, the incurred is set equal to the total amount paid. This is a deterministic operation and no modelling is required. However, when a claim does not settle, layers 4 to 8 model reserve changes. These five layers let our model capture the reserve dropping to zero, increases in the reserve and decreases in the reserve expressed as percentages of the outstanding reserve in previous years.

We present a simple and automizable procedure for fitting a statistical model to each of the 11 (3+8) layers of the hierarchical model. We model each of the layers with a tree-based Gradient Boosting Model (GBM), which fits the data via a chain of shallow decision trees where each tree improves the fit of the previous trees. Three properties make GBMs interesting for automatization. First, automatic binning of continuous covariates allows for capturing non-linear effects. Second, interaction effects are automatically detected when using shallow trees with multiple splits. Third, covariate selection is integrated in the calibration process. For each GBM, we tune five hyperparameters using five-fold cross validation on our training data set.

Table 4.1 specifies the distributional assumption per layer. We distinguish three types of outcome variables: binary outcomes, percentages and numeric outcomes not bounded to the interval $(0, 1)$. We model binary outcomes (e.g. `settlement`) with a binomial GBM with logit link function, i.e. we minimize the loss

$$\mathcal{L}(f) = \sum_i y_i \cdot f(\boldsymbol{x}_i) - \log(1 + f(\boldsymbol{x}_i)),$$

where y_i are the observed outcomes and $\boldsymbol{x}_i$ denotes the available covariates for the i-th observation. Percentage outcomes (e.g. `pct_paid`) are first transformed to the domain $(-\infty, \infty)$ using a logit transform and then modelled using a

(a) Initial claim status $\boldsymbol{I}$
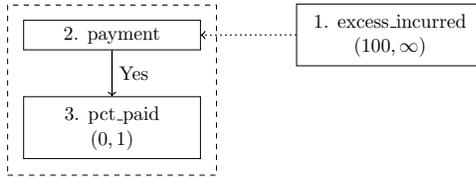
(b) Updates $\boldsymbol{U}^j$
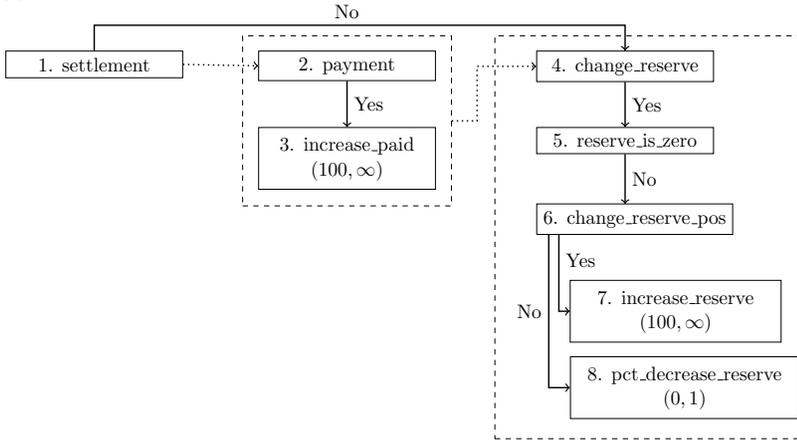
Figure 4.4: Flowchart illustrating the various layers in the claim development model and their connections. Dotted lines indicate that the outcome of one layer serves as an input for a later layer. Solid lines indicate that a layer is modelled conditional on the outcome of a previous layer. The numbers indicate the order in which the layers are modelled.

Gaussian GBM, i.e. we minimize the loss

$$\mathcal{L}(f) = \sum_i (\text{logit}(y_i) - f(\boldsymbol{x}_i))^2.$$

The variance σ^2 of the Gaussian distribution is estimated as the mean squared error of the residuals, i.e.

$$\hat{\sigma}^2 = \frac{1}{n} \cdot \sum_i (\text{logit}(y_i) - \hat{f}(\boldsymbol{x}_i))^2,$$

where n is the number of observations. Other numeric outcomes are left-truncated at 100 due to the smoothing applied to the data. Moreover, these outcomes are heavily right skewed given our reinsurance context. Therefore,

component	distribution	transform	link
Initial claim status			
excess_incurred	trunc.Gaussian	$(.)^{0.117}$	.
payment	binomial	.	logit
pct_paid	Gaussian	logit	.
Updates			
settlement	binomial	.	
payment	binomial	.	logit
increase_paid	trunc.Gaussian	$(.)^{0.155}$	.
change_reserve	binomial	.	logit
reserve_is_zero	binomial	.	logit
change_reserve_pos	binomial	.	logit
increase_reserve	trunc.Gaussian	$(.)^{0.105}$	.
pct_decrease_reserve	Gaussian	logit	.

Table 4.1: Distributional specification for the model components in the hierarchical claim development model visualized in Figure 4.4.

we first normalize these outcomes by applying a power transform, i.e. replace the random variable X by X^p for some power p, and then estimate a truncated Gaussian GBM for the normalized outcomes, i.e. we minimize the loss

$$\mathcal{L}(f, \sigma, p) = \sum_i \log(\sigma) + \frac{(y_i^p - f(\boldsymbol{x}_i))^2}{2\sigma^2} + \log(\Phi(100^p \mid f(\boldsymbol{x}_i), \sigma)) - \log(p) - p \cdot \log(y_i),$$

(4.6)

where p is the exponent in the power transform and $\Phi(\cdot \mid \mu, \sigma)$ is the cdf of the Gaussian distribution with mean μ and standard deviation σ. As a result of the flexibility of the GBM, this loss function is not lower bounded when we simultaneously optimize for $f(\cdot), \sigma$ and p. We opt for a two step approach. First, we minimize (4.6) with respect to σ and p and a constant $f(\cdot)$. Figure 4.5 shows QQ-plots of the normalized outcome variables after the first step with respect to the truncated Gaussian distribution. Second, we re-estimate $f(\cdot)$ and σ using a truncated Gaussian GBM, while keeping the power p fixed.

Figure 4.6 shows for each fitted GBM the relative importance of the included covariates, where we define variable importance as the decrease in the loss function of the GBM over all tree splits including the covariate under consideration. `portfolio` is an important covariate for almost all layers, which
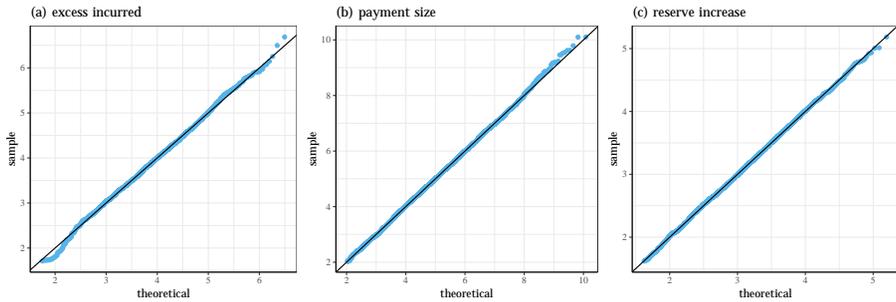
Figure 4.5: Truncated normal qq-plot of the outcome variables `excess_incurred` (a), `increase_paid` (b) and `reserve_increase` (c) after applying a power transformation.

indicates that there are clear differences in how insurers handle their large claims. Most noteworthy is the effect of the insurer on the layer `change reserve`. In some portfolios experts re-evaluate their large claims almost every year, whereas other insurers rarely update their large claims. Both `reserve` (incurred - paid) and `ratio paid incurred` ($\frac{paid}{incurred}$) describe a relationship between the incurred and amount paid. Together these covariates are for many layers the most important determinant for the evolution of claims. In traditional, aggregated reserving models, claim development depends only on the number of `years elapsed since reporting`, i.e. the development year. Surprisingly, this covariate becomes irrelevant when other claim characteristics such as the reserve are available.

We select some important covariates in the hierarchical model and investigate in Figure 4.7 their marginal effect on the outcome variable with partial dependence plots. The incurred is smaller for claims reported after a long delay (4.7a), but the fraction of the incurred that has already been paid is larger for these claims (4.7b). This is intuitive, since the insurer has observed and made payments for these claims over multiple years before adjusting the incurred to a level above the reporting priority. As expected, Figure 4.7c shows that claims settle faster when the reserve is near zero. Slightly surprising, a payment in the previous development year, irrespective of the payment size, increases the payment probability in the current year (4.7d). This could indicate the presence of claims with recurrent payments. Figure 4.7e shows the effect of portfolio on the probability of making a reserve adjustment. Some portfolios revise their large claims almost yearly, whereas update probabilities are much lower for smaller portfolios. Increases in the reserve are more likely when less than half of the incurred has been paid (4.7f).

Figure 4.6: Available covariates and their relative importance on the layers of the hierarchical model for the initial status $\boldsymbol{I}_k$ and updates $\boldsymbol{U}_k^j$. Relative importance is computed as the decrease in the loss function over all splits including a specific covariate relative to the total decrease in the loss function caused by all covariates.
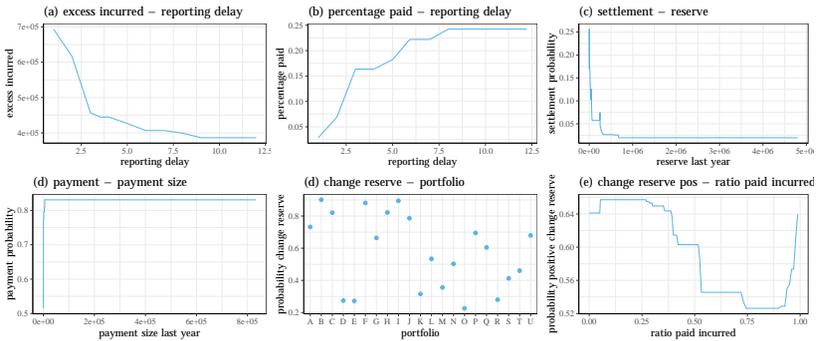
Figure 4.7: Selection of partial dependence plots in the hierarchical claim development model. The vertical axis has been transformed to an interpretable scale.

4.4.3   Pricing an excess-of-loss reinsurance contract

We price an excess-of-loss reinsurance contract covering losses from claims exceeding a deductible $D = 2\,500\,000$ up to a limit $L = 5\,000\,000$. Following the frequency-severity decomposition the pure premium π is

$$\pi^P = E(N^P) \cdot E(((Y^P \wedge L) - D)_+),$$

N^P and Y^P are the frequency and severity, respectively, of claims reported above a priority P and $(Y^P \wedge L)$ denotes the minimum of Y^P and L. Premiums are obtained by estimating the expected claim frequency and claim severity of new policies with the formulas of Section 4.3.1 based on the ODM defined in Section 4.2 and calibrated in Section 4.4.1 and 4.4.2.

We have calibrated the claim frequency as

$$N_i \sim \texttt{Poisson}(e_i \cdot \lambda_{\texttt{portfolio(i)}}).$$

We obtain an estimate N_{i^*} for a new policy i^* from an existing portfolio by multiplying the exposure e_{i^*}, expressed as the number of vehicles, for this policy with the portfolio specific intensity $\lambda_{\texttt{portfolio}(i^*)}$. For better comparison of prices across portfolios, we compute in this chapter premiums per insured vehicle, i.e. with $e_{i^*} = 1$.

Section 4.3.1 proposes two strategies for simulating the claim severity distribution. The first strategy simulates a large number of paths for a new claim from ground up, whereas the second strategy simulates the future development of open claims. We illustrate both simulation strategies for modelling the severity distribution of a new policy from portfolio A that occurred in 2015 and was reported at a reporting priority of $750\,000$.

Simulating paths for a new claim   We simulate $20\,000$ paths for the development of a new claim from policy A that occurred in 2015. Figure 4.8 visualizes the evolution of the amount paid and incurred calculated over these $20\,000$ paths. Solid lines indicate the average amount paid and incurred, whereas dashed lines show the 95% confidence intervals for these amounts. At reporting the incurred exceeds $750\,000$ for all simulated paths. However, soon after reporting the lower bound for the incurred drops to zero as some of these paths will settle without payment. This represents the case where the claim is not eligible for compensation within the portfolio. This is a common scenario for large motor insurance claims, where often many parties and hence insurers are involved in an accident and it is initially not always clear which insurer should reimburse the claim. After 15 years have elapsed since reporting, i.e. the observation window of our training data set, many simulated paths have not yet

settled. In Figure 4.8 this is seen by the large difference between the amount paid and amount incurred after 15 years. Supported by the low importance of the covariate `number of elapsed years since reporting` (Figure 4.6), we extrapolate our hierarchical model and simulate the development up to 60 years after the reporting of the claim. After 60 years almost all paths have settled and the amount paid has converged towards the amount incurred. The distribution of the amount paid after 60 years is our simulated severity distribution for a new claim from policy `A` that occurs in 2015 and is reported with a priority of 750 000.
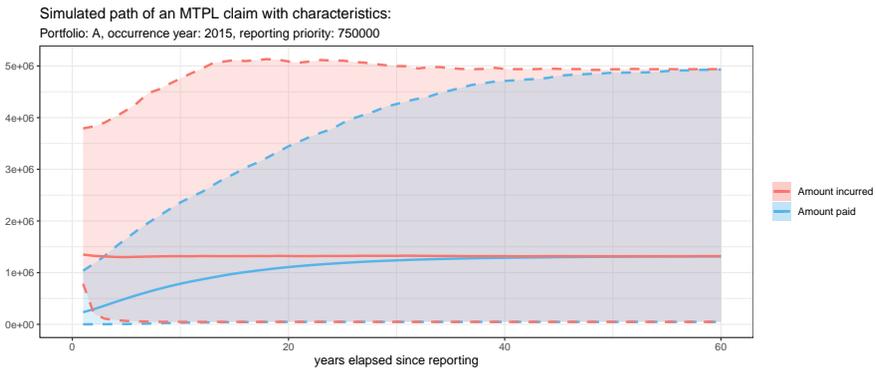
Figure 4.8: Simulated evolution of the amount incurred and amount paid for a new claim from portfolio `A` that occurs in year 2015 and is reported at a priority of 750 000. Solid lines show the average amount paid and incurred, while dashed lines indicate the 95% confidence intervals for these amounts.

Simulating future paths for open claims   Alternatively, we simulate 200 paths for the future development of each open claim. Figure 4.9 shows simulated 80% confidence intervals for the amount paid at settlement for the 401 observed claims from portfolio `A`. For open claims these confidence bounds are calculated based on the simulated paths in which the amount incurred breaches the priority of 750 000 over the lifetime of the claim. Compared to Figure 4.8, we choose smaller confidence intervals (80% instead of 95%), since outcomes for individual claims are heavy tailed and we only use 200 simulations per claim when constructing these intervals. Only 33 of these 401 claims have settled in our training data set. Claims are sorted by median severity, which is indicated with a solid black line. Distributions per claim are heavily right skewed with the median near the lower end of the confidence interval. A severity distribution can be estimated based on observed claims and simulated future paths by maximizing the likelihood in (4.5). In this case-study, we put focus on the

simulations by the ODM and use the empirical cumulative distribution function (ECDF) as a non-parametric estimator for claim severity. In the construction of the ECDF we assign a weight of $\frac{1}{200}$ to each simulated path and a weight of 1 to each settled claim. Appendix 4.7 investigates the application of EVT to replace the tail of this ECDF with an extreme value distribution.
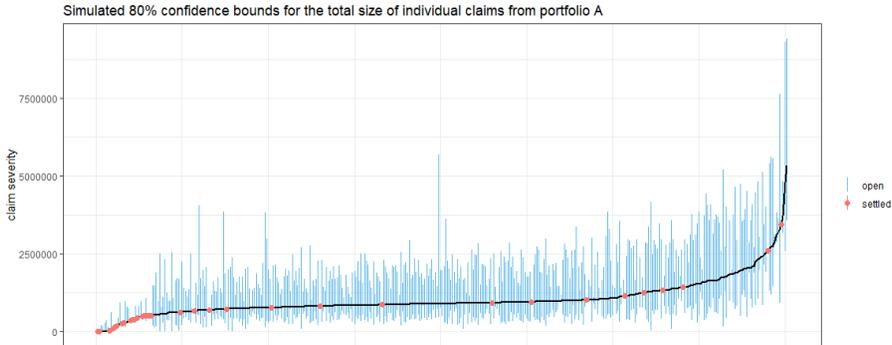
Figure 4.9: 80% confidence intervals for 401 observed claims from portfolio `A` based on 200 simulations per open claim. Only paths that exceed the incurred of 750 000 are retained. Observed claims are sorted by median loss, which is indicated with a solid black line.

Comparing simulated severity distributions Figure 4.10 compares the simulated claim severity distribution based on simulated paths from ground up (blue) and simulated future paths for observed claims (red). Since we are pricing an excess-of-loss contract with a limit of 5 000 000, we improve the readability of this figure by showing the distribution for losses below 5 000 000. For portfolio `A` both simulation strategies result in nearly identical severity distributions. Repeating the same approach for portfolio `B`, we retrieve a more heavy tailed severity distribution when simulating future paths for observed claims. Portfolio `B` contains several of the largest claims in our data set. Comparing data across all insurers, the hierarchical claim development model recognizes these extreme claims as outliers, which results in a more moderate severity distribution when simulating new claims from ground up. Figure 4.10 compares the claim severity distributions proposed in this chapter with the empirical cdf based on best estimates (green), where for each open claim the best estimate is calculated by averaging claim severity over the 200 simulated paths. This distribution has the same mean, but a lower variance than the distribution based on simulated paths for observed claims. As a result of Jensen's inequality, the empirical cdf

underestimates the actual claim severity for excess-of-loss contracts, i.e.

$$E((Y - d)_+) \overset{Jensen}{\geq} E\left(\left((E(Y \mid \mathcal{F}_\tau) - d)_+\right)\right).$$
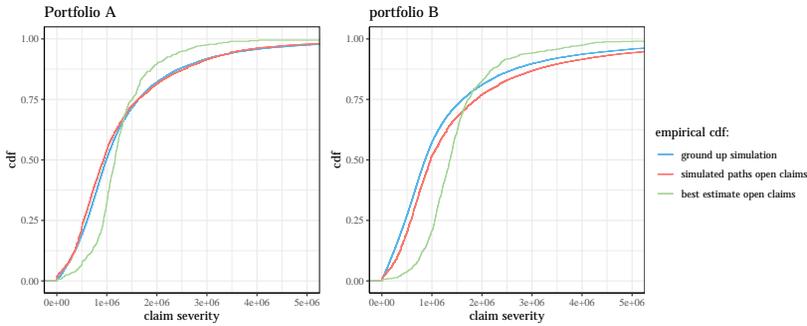
Figure 4.10: Simulated severity distribution of MTPL claims from portfolio A and B with a reporting priority of 750 000. For each portfolio, we show the severity distribution based on 20 000 from group up simulated new claims (blue), observed claims complemented with 200 simulated paths per open claim (red) and observed claims with open claims replaced by best estimates (green).

Pricing an excess-of-loss policy   Figure 4.11 shows the estimated loss per insured vehicle for an excess-of-loss policy with severity estimated based on **(a)** simulating 20 000 new claims from ground up and **(b)** observed claims complemented with 200 simulated paths per open claim. In theory, the choice of priority should not influence the price. In practice differences in the estimated pure premium arise since the priority determines the available historical claims when modelling the ODM. We investigate the sensitivity of the pure premium with respect to the priority by modelling frequency with a priority of 750 000, 1 000 000 and 1 250 000. For most portfolios, the price remains relatively constant when changing priorities, but larger variations are observed for some small portfolios (e.g. portfolio S). These variations mainly result from our claim frequency model for which the priority determines the available claims when training the model. Since we detect two regimes in the occurrence intensity in Figure 4.3a, our frequency model could be made more robust by estimating a single occurrence intensity parameter per regime. Estimated prices when **(a)** simulating new claims and **(b)** simulating paths for open claims are comparable. Price differences are often the result of realised extreme claims, which more heavily influence the estimated cost based on observed claims.
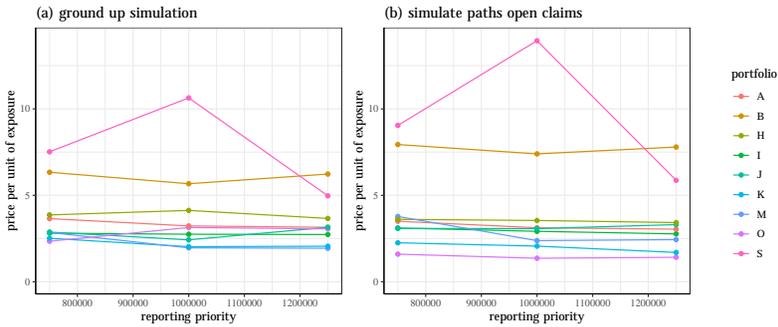
Figure 4.11: Estimated cost per insured vehicle for an excess-of-loss policy with deductible $2\,500\,000$ and limit $5\,000\,000$. Claim severity is estimated based on **(a)** simulating $20\,000$ new claims from ground up and **(b)** observed claims complemented with 200 simulated paths per open claim. Prices are computed at reporting priorities: $750\,000$, $1\,000\,000$ and $1\,250\,000$.

4.4.4 Reserving for reinsurance

Reserving actuaries estimate the aggregated, future costs for claims from past exposure years. In reinsurance, these costs depend on the structure of the contract sold. We estimate the reserve under two contracts. The first contract (see Figure 4.12 below) covers all losses for claims for which the incurred at least once exceeds the reporting priority of $750\,000$. Although this contract is not sold in practice, we investigate it, since it mimics the classical insurance setting as close as possible given our data. For accurately reserving this contract, it is important that our ODM captures the average development pattern of claims over time well. The second contract (see Figure 4.14 below) covers losses between $2\,500\,000$ and $5\,000\,000$, i.e. the policy that we priced in Section 4.4.3. This contract focuses on the performance of our ODM for large claims. For convenience, we assume that these contracts were sold for occurrence years 2000-2015 and the nine portfolios with a reporting priority below $750\,000$ in our data set.

Reserving is straightforward and reuses the techniques developed for pricing in Section 4.4.3. For the IBNR reserve, we predict the number of unreported claims per calendar year from our occurrence and reporting model and their severity by simulating new claims from ground up. These simulations account for the effect of long reporting delays for IBNR claims on the claim development process (Figure 4.7a and 4.7b). For the RBNS reserve, we simulate the future development of open claims.
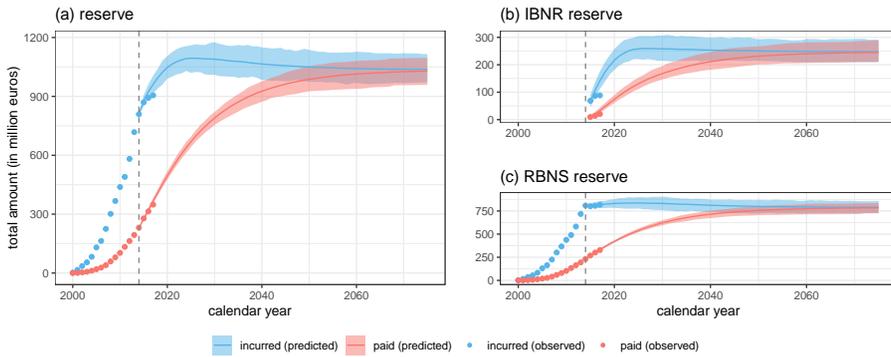
Figure 4.12: Evolution of the total amount incurred and paid for claims that occurred between 2000 and 2014 and that exceed the reporting priority of 750 000 during their lifetime. The **(a)** total reserve is split into the **(b)** IBNR and **(c)** RBNS reserve. Simulated 95% confidence intervals are shown for these amounts, with solid lines indicating expected values. Points indicate for calendar years 2015-2017 the actual out-of-time observations.

Covering all losses for reported claims   Figure 4.12 shows the estimated evolution of the total amount incurred and paid for claims that occurred between 2000 and 2014 and that exceed the reporting priority of 750 000 during their lifetime. For calendar years 2015-2017, we compare the estimated evolution with actual observations from the out-of-time data set. Figures 4.12b and 4.12c split the total reserve into the IBNR and RBNS reserve. For the RBNS reserve, the total amount incurred decreases slightly over time. This indicates that claim experts overestimate the expected cost of large claims when setting incurred amounts. For the total reserve, we estimate a sharp increase of the incurred in the first calendar years following 2014 as new claims get reported. Figure 4.12b shows that our model overestimates the increase in the incurred, which is due to an overestimation of the number of unreported claims (not shown). In Belgium, judges use indicative tables based on mortality and interest rates to determine the compensation for bodily injury claims. In 2012, interest rates for these tables were updated from 2% to 1%, which led claim experts to sharply increase the incurred amounts in 2013 and 2014. This initially led to an increase in the number of reported claims, as suddenly more claims exceeded the reporting priority, followed by a decrease in reported claim counts in later years. Since these external effects can not be predicted by data driven models, expert judgement will always remain important for reinsurance. Our model estimates that the total amount paid continues to slowly increase towards the total incurred.

Long delays in our reinsurance data set compel us to use most of the observed
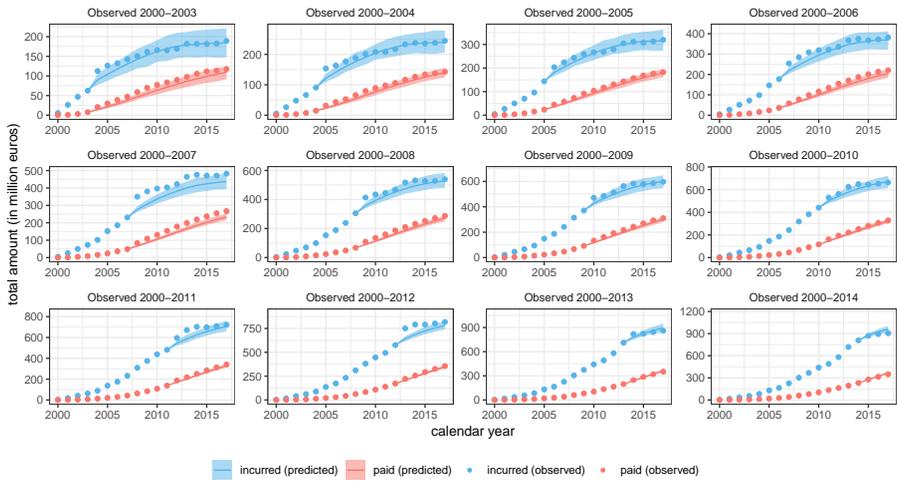
Figure 4.13: Panels show, for different observation windows, the evolution of the total amount incurred and paid until 2017 for claims that have occurred within the observation window. 95% confidence intervals are shown for these amounts, with solid lines indicating expected values for years outside the observation window. Points show the actual amount incurred and paid extracted from the data until 2017.

calendar years (2000-2014) for training our model, leaving only three years (2015-2017) for an out-of-time evaluation. For longer reserve evaluations, we choose an earlier observation date τ and use our fitted ODM and the observed claim history at time τ to predict the future evolution of the amount incurred and paid for claims that occurred before τ. This is, however, no out-of-time evaluation, since we still train our ODM on the years 2000-2014. Figure 4.13 shows these evaluations of the total reserve (IBNR + RBNS) for τ ranging from 2003 to 2014. Overall, the estimated evolution of the amount incurred and paid roughly follows the evolution recorded in our data set. The interest rate in the indicative table changed in 2002 (4% to 3%), 2008 (3% to 2%) and 2012 (2% to 1%), these changes cause sudden sudden shocks in the amount incurred for all claims which are not captured by our model.

Covering losses from an excess-of-loss contract We focus on the reserve of the excess-of-loss contract priced in Section 4.4.3 covering losses between 2 500 000 and 5 000 000. Figure 4.14 shows the estimated evolution of the amount incurred and paid within the layer of our excess-of-loss contract. Although we have recorded only few payments within the layer yet, we can accurately infer

the payment pattern from the general dynamics estimated in the hierarchical model of Section 4.4.2. This illustrates the importance of calibrating models at a lower reporting priority in reinsurance, such that sufficient data regarding the development of large claims is available. Where the incurred for reported claims, i.e. the RBNS reserve, remained more or less constant when reserving from ground up (Figure 4.12c), we observe an initial increase followed by a decrease for the total incurred within the layer of our excess-of-loss contract (Figure 4.14c). This behaviour can be explained by applying Jensen's law twice. The incurred is the insurer's best estimate of the total claim size given the current information, i.e. $E(Y \mid \mathcal{F}_\tau)$. Initially, when it is unlikely that a claim will exceed the limit of $5\,000\,000$, the contract behaves as a deductible. This is a convex loss function, such that

$$E((E(Y \mid \mathcal{F}_\tau) - 2\,500\,000)_+) \overset{\text{Jensen}}{\leq} E(E((Y - 2\,500\,000)_+ \mid \mathcal{F}_\tau))$$

$$= E((Y - 2\,500\,000)_+).$$

As more information becomes available, the difference between both sides in this inequality decreases. Since the right hand side (rhs) is time independent, the lhs increases over time, i.e. the expected amount incurred within the layer of our contract increases. Once claim experts become confident that a claim will breach the deductible, the contract starts to behave more like a limited loss. This is a concave loss function with

$$E(E(Y \mid \mathcal{F}_\tau) \wedge 5\,000\,000) \overset{\text{Jensen}}{\geq} E(E(Y \wedge 5\,000\,000 \mid \mathcal{F}_\tau)) = E(Y \wedge 5\,000\,000).$$

Following the same reasoning, the expected incurred within the layer decreases for these claims. As a result of the overestimation of the number of reported claims, our ODM overestimates the IBNR reserve in Figure 4.14b.

4.5   Conclusion

We propose an occurrence and development model (ODM) for analysing the detailed claim information registered in non-life insurance portfolios. Our ODM can be used for non-life pricing as well as non-life reserving, hereby bridging two key actuarial tasks. For pricing, we present a one-step approach, which resolves the contradictions between observations and best estimates in traditional pricing literature. For reserving, we model the cost of unreported claims constituting the IBNR reserve at the level of individual policies and the future payments constituting the RBNS reserve at the level of individual claims. An extensive case-study illustrates our methodology on a motor reinsurance portfolio.
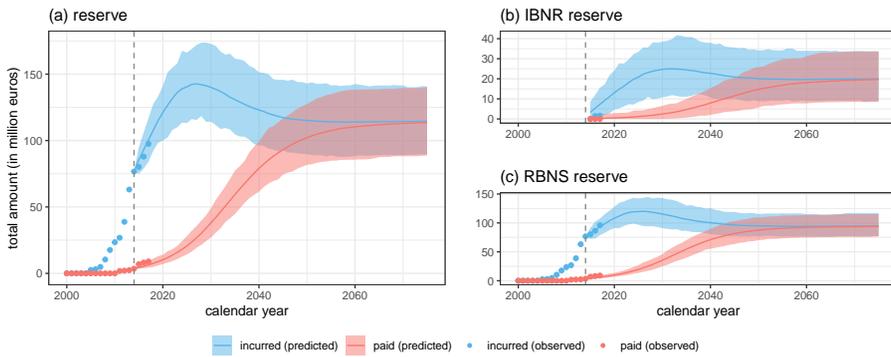
Figure 4.14: Evolution of the aggregated amount incurred and paid between 2 500 000 and 5 000 000 for claims that occurred between 2000 and 2014. The **(a)** total reserve is split into the **(b)** IBNR and **(c)** RBNS reserve. 95% confidence intervals are shown for these amounts, with solid lines indicating expected values. Points indicate for calendar years 2015-2017 the actual out-of-time observations.

Constructing best estimates for open claims is complicated in reinsurance, where reporting and settlement delays are long and claim development is uncertain. This is a situation in which our ODM, which does not rely on best estimates for pricing, clearly outshines traditional methodology. Using Jensen's inequality we prove that the empirical distribution based on best estimates underestimates the variance of the claim severity distribution. This is best illustrated in Figure 4.10, where the claim severity distribution modelled by our ODM has a significantly larger variance than the empirical claim severity distribution based on best estimates. For reserving, we present an individual reserving model using paid and incurred data. Despite large uncertainties governing the development of reinsurance claims, our model is able to accurately predict the joint evolution of the paid and incurred amounts.

4.6   Acknowledgements

4.7 Appendix A: Estimating an extreme value distribution for claim severity using stochastic input data

Section 4.3.1 proposes two approaches for modelling claim severity. This appendix focuses on the second approach in which we first replace each open claim by a simulated distribution and then fit a claim severity distribution by maximizing log-likelihood (4.5). Section 4.4.3 demonstrates this approach on a reinsurance data set and fits an empirical cdf to the simulated paths for open claims. We now extend the severity model of Section 4.4.3 by estimating an extreme value distribution to the simulated paths exceeding a predefined threshold T. We limit our analysis to the Pareto distribution with density

$$f_Y(y \mid \alpha) = \frac{\alpha T^\alpha}{y^{\alpha+1}} \quad \text{for} \quad y > T$$

which depends on a single parameter α. Filtering likelihood (4.5) for severities exceeding T, we maximize

$$\mathcal{L}(\alpha \mid \boldsymbol{y}, T) = \sum_k \delta_k \mathbb{1}_{y_k > T} \cdot (\log(\alpha) + \alpha \log(T) - (\alpha+1) \log(y_k)) +$$

$$\sum_k (1 - \delta_k) \frac{1}{n} \sum_{p=1}^n \mathbb{1}_{y_{k,p} > T} \cdot (\log(\alpha) + \alpha \log(T) - (\alpha+1) \log(y_{k,p})),$$

where δ_k is one when claim k has settled and zero otherwise and n denotes the number of simulations for each open claim. The maximum likelihood estimator for $\frac{1}{\alpha}$ is

$$\widehat{\alpha^{-1}}(T) = \frac{\sum_k \delta_k \log(y_k) \mathbb{1}_{y_k > T} + (1 - \delta_k) \frac{1}{n} \sum_{p=1}^n \log(y_{k,p}) \mathbb{1}_{y_{k,p} > T}}{\sum_k \delta_k \mathbb{1}_{y_k > T} + (1 - \delta_k) \frac{1}{n} \sum_{p=1}^n \mathbb{1}_{y_{k,p} > T}}. \tag{4.7}$$

We call this estimate for α^{-1} the weighted Hill estimator, since it reduces to the classical Hill estimator [Hill, 1975] when the input data is deterministic, i.e. all claims are settled. Figure 4.15 shows $\widehat{\alpha^{-1}}(T)$ as a function of the threshold T for the nine portfolios priced in Section 4.4.3. The weighted Hill plot becomes horizontal for large thresholds T, which corresponds to a Pareto tail. Since all portfolios correspond to motor insurance, we find similar values for $\frac{1}{\alpha}$ around $(0.20, 0.35)$ in the tail of the hill plot. In practice, the similarity between motor insurance claims across portfolios would prompt reinsurers to merge these portfolios and estimate a single extreme value distribution for their motor insurance claims.
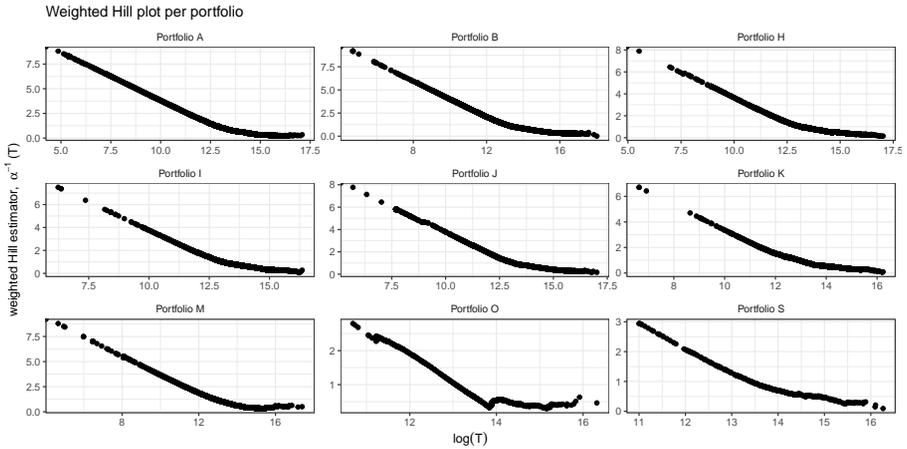
Figure 4.15: Weighted Hill plots for claim severity. Each panel corresponds to
one of the portfolios priced in Section 4.4.3.

In Figure 4.16, we further examine the hypothesis of a Pareto tail by constructing
a Pareto QQ-plot for each of these nine portfolios and for losses exceeding
5 000 000, i.e. exp(15.42). In general these QQ-plots support the assumption of
a Pareto tail. For portfolio B and I we identify some regions with a different
tail behaviour, which should be examined more carefully before implementation
in practice.
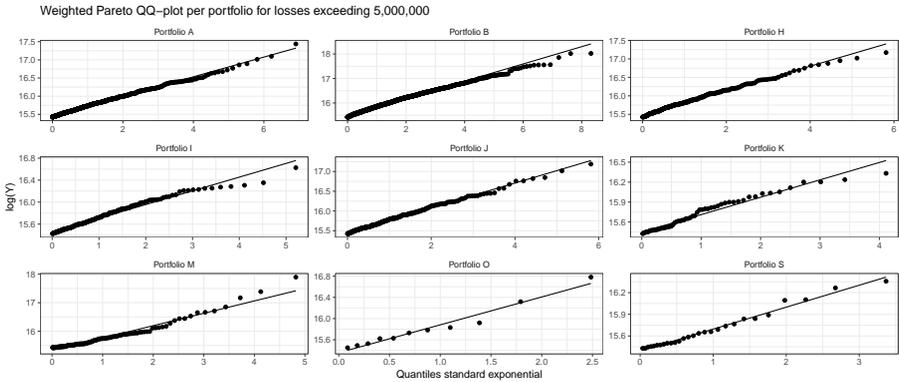
Figure 4.16: Pareto QQ-plot for claim severity above 5 000 000. Each panel
corresponds to one of the portfolios priced in Section 4.4.3.

Chapter 5

Outlook

The reserve is typically the largest element on an insurance balance sheet. New regulations, i.e. IFRS 17 and Solvencey II, motivate insurers nowadays to gain more insight into their reserves. Combined with an increase of computational power, this has in recent years sparked a boom in research on individual reserving. Despite this surge in interest, it still takes substantial effort by insurers to implement individual reserving methods in practice. The research in this PhD thesis results from an intense collaboration with several insurers to develop new methods that could facilitate the implementation of individual reserving in practice. This chapter concludes this work by listing some further directions for research expanding on this work.

5.1 Further developments in modelling the occurrence and reporting of claims

Chapter 2 illustrates that due to delayed reporting the number of unreported claims increases during the weekend and on holidays. Following these findings, we develop a new model for the number of unreported claims using daily data and with a focus on incorporating calendar day effects (e.g. weekend, holiday) in the reporting process. A simulation study compares this approach with a traditional model based on aggregated data and concludes that using granular, daily data has significant advantages for portfolios with volatile occurrence process (e.g. as a result of extreme weather events). A scenario with low daily claim counts is challenging for both aggregate and granular models.

In Chapter 4 we return to yearly data and model the claim occurrence process at the level of individual policyholders. We apply this approach to a small reinsurance portfolio, where differences in claim frequency between policyholders, i.e. insurers, can be large.

A straightforward extension of this work would be to combine both approaches and model the occurrence and reporting process for individual policyholders using daily data. Such an approach could assist insurers in determining which policyholders or claims are now typically reported with a delay. Insurers could use these insights when reviewing their reporting processes in an attempt to reduce reporting delay. The developed methodology can also be applied to other areas (e.g. warranty contracts, epidemiology) where events are observed after a delay. Currently in the COVID-19 pandemic, the effect of reporting delay on the number of reported cases has received considerable media attention. Similar to the data in our insurance portfolios, less infections are reported during the weekend and holidays as the result of a weekend effect.

5.2 Further developments in modelling the development of reported claims

Chapter 3 and Chapter 4 present the hierarchical reserving model as a novel and intuitive approach for jointly modelling all events registered during the development of a claim. This approach focuses strongly on implementability in practice. To further facilitate this, a part of our code was released in the form of an R package.

In this work we primarily focused on accurately predicting the expected reserve. Less attention has been given to prediction uncertainty and case-studies indicate that the current confidence bounds tend to be too narrow. This appears to be a general trend in individual reserving, which demands further research. A potential reason for this underestimation could be found in the typical independence assumptions in individual reserving, which assume that each claim develops independently of all others. In practice this independence assumption is often violated. In particular, we expect the development process to be dependent for claims resulting from the same extreme weather event or for several claims from the same policyholder. In the latter case, methods from credibility theory could be investigated to include dependency between the different claims from a policyholder. Another domain for further research relates to the development of data driven tools for selecting the components in the hierarchical reserving model. The hierarchical reserving model could also contribute to other fields where observations can be expressed as incomplete

time series and the objective is to predict the evolution of these time series multiple steps into the future in the presence of time varying covariates. This is for example the case in patient data, when we are interested in following the evolution of symptoms in a single patient over multiple years.

5.3 Further developments in insurance pricing

Non-life insurance pricing typically follows a two-step approach. First, censored claim counts and claim severities are replaced by best estimates. Second, a frequency and severity model is fitted to these best estimates. In Chapter 4 we develop a one step approach for insurance pricing, which directly uses the censored observations as inputs. Moreover, this approach unifies pricing and reserving methodology, hereby opening a new area of research in non-life insurance.

In Chapter 4 we demonstrate our method on a small reinsurance data set. A logical next step would be to apply this method to an insurance portfolio and to compare the resulting prices with prices found using traditional techniques. We see two phenomena that could result in an underestimation of variable importance in traditional two-step pricing approaches. First, policyholder covariates are often not included in the first step to remove the censoring from the data. Using a non-parametric model for this first step would smooth potential covariate effects. Second, by replacing open claims by best estimates the uncertainty present in the data is underestimated. Given these phenomena, our one-step approach could provide new insights into risk classification. This approach has most potential in severity modelling for which more observations are censored and for which traditional techniques often identify few relevant covariates.

List of Figures

List of Tables

Bibliography

Artur Akbarov and Shaomin Wu. Warranty claims data analysis considering sales delay. *Quality and Reliability Engineering International*, 29:113–123, 2012. doi: https://doi.org/10.1002/qre.1302.

Hansjörg Albrecher, Jan Beirlant, and Jozef L. Teugels. *Reinsurance: Actuarial and Statistical Aspects*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., 2017. URL `https://doi.org/10.1002/9781119412540`.

K. Antonio and R. Plat. Micro–level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, 7:649–669, 2014. doi: https://doi.org/10.1080/03461238.2012.755938.

S. Apeland and P.A. Scarf. A fully subjective approach to modelling inspection maintenance. *European Journal of Operational Research*, 148:410–425, 2003. doi: https://doi.org/10.1016/S0377-2217(02)00356-9.

Benjamin Avanzi, Bernard Wong, and Xinda Yang. A micro–level claim count model with overdispersion and reporting delays. *Insurance: Mathematics and Economics*, 71:1–14, 2016. doi: https://doi.org/10.1016/j.insmatheco.2016.07.002.

Andrei L Badescu, X Sheldon Lin, and Dameng Tang. A marked Cox model for the number of IBNR claims: Theory. *Insurance: Mathematics and Economics*, 69:29–37, 2016. doi: https://doi.org/10.1016/j.insmatheco.2016.03.016.

R.D. Baler and W. Wang. Developing and testing the delay-time model. *The Journal of the Operational Research Society*, 44:361–374, 1993. doi: https://doi.org/10.2307/2584414.

M.D. Berrade, P.A. Scarf, and C.A.V. Cavalcante. Conditional inspection and maintenance of a system with two interacting components. *European Journal of Operational Research*, 268:533–544, 2018. doi: https://doi.org/10.1016/j.ejor.2018.01.042.

Marco Bonetti, Pasquale Cirillo, Paola Musile Tanzi, and Elisabetta Trinchero. An analysis of the number of medical malpractice claims and their amounts. *PLOS ONE*, 11(4):1–30, 04 2016. doi: https://doi.org/10.1371/journal.pone.0153362.

A.H. Christer. Innovatory decision making. In D.J. White and K.C. Brow, editors, *The Role and Effectiveness of Theories of Decision in Practice*, pages 369–377. Hodder and Stoughton, London, 1973.

Jonas Crevecoeur. *hirem: Hierarchical Reserving Models*, 2020. URL `https://github.com/jonascrevecoeur/hirem`. R package version 1.0.

Jonas Crevecoeur and Katrien Antonio. A hierarchical reserving model for non-life insurance claims. 2020a. Available at arXiv: `https://arxiv.org/abs/1910.12692`.

Jonas Crevecoeur and Katrien Antonio. Bridging the gap between pricing and reserving with an occurrence and development model for non-life insurance claims. 2020b. Working paper.

Jonas Crevecoeur, Katrien Antonio, and Roel Verbelen. Modeling the number of hidden events subject to observation delay. *European Journal of Operational Research*, 277(3):930 – 944, 2019. ISSN 0377-2217. URL `https://doi.org/10.1016/j.ejor.2019.02.044`.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977. URL `https://doi.org/10.1111/j.2517-6161.1977.tb01600.x`.

Michel Denuit and Julien Trufin. Beyond the Tweedie reserving model: The collective approach to loss development. *North American Actuarial Journal*, 21(4):611–619, 2017. URL `https://doi.org/10.1080/10920277.2017.1353428`.

Michel Denuit and Julien Trufin. Collective loss reserving with two types of claims in motor third party liability insurance. *Journal of Computational and Applied Mathematics*, 335:168 – 184, 2018. URL `https://doi.org/10.1016/j.cam.2017.11.044`.

Michel Denuit, Xavier Marechal, Sandra Pitrebois, and Jean-Francois Walhin. *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems.* John Wiley & Sons Ltd, West Sussex, 1 edition, 2007. ISBN 9780470026779.

Edward W. Frees and Emiliano A. Valdez. Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, 103(484):1457–1469, 2008. URL `https://doi.org/10.1198/016214508000000823`.

Jerome H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 29(5):1189–1232, 2001. URL `https://doi.org/10.1214/aos/1013203451`.

Els Godecharle and Katrien Antonio. Reserving by conditioning on markers of individual claims: a case study using historical simulation. *North American Actuarial Journal*, 19(4):273–288, 2015. doi: https://doi.org/10.1080/10920277.2015.1046607.

Brandon Greenwell, Bradley Boehmke, Jay Cunningham, and GBM Developers. *gbm: Generalized Boosted Regression Models*, 2018. URL `https://github.com/gbm-developers/gbm`. R package version 2.1.4.9000.

Jeffrey E. Harris. Reporting delays and the incidence of AIDS. *Journal of the American Statistical Association*, 85(412):915–924, dec 1990. ISSN 0162-1459 (print), 1537-274X (electronic). doi: https://doi.org/10.2307/2289588.

Roel Henckaerts, Katrien Antonio, Maxime Clijsters, and Roel Verbelen. A data driven binning strategy for the construction of insurance tariff classes. *Scandinavian Actuarial Journal*, 2018(8):681–705, 2018. URL `https://doi.org/10.1080/03461238.2018.1429300`.

Roel Henckaerts, Marie-Pier Côté, Katrien Antonio, and Roel Verbelen. Boosting insights in insurance tariff plans with tree-based machine learning methods. *North American Actuarial Journal*, 2020. In press.

Bruce M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3(5):1163–1174, 09 1975. doi: 10.1214/aos/1176343247. URL `https://doi.org/10.1214/aos/1176343247`.

Jinlong Huang, Chunjuan Qiu, Xianyi Wu, and Xian Zhou. An individual loss reserving model with independent reporting and settlement. *Insurance: Mathematics and Economics*, 64:232 – 245, 2015. ISSN 0167-6687. doi: https://doi.org/10.1016/j.insmatheco.2015.05.010.

J. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennese. *Acta Mathematica*, 30:175–193, 1906. URL `https://doi.org/10.1007/BF02418571`.

William S. Jewell. Predicting IBNYR events and delays. *ASTIN Bulletin*, 20:93–111, 1990. doi: https://doi.org/10.2143/AST.19.1.2014914.

E. L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958. doi: https://doi.org/10.1080/01621459.1958.10501452. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1958.10501452.

J. F. C. Kingman. *Poisson Processes*. Oxford Studies in Probability. Oxford University Press, 1993. ISBN 0198536933,9780198536932.

D. Kuang, B. Nielsen, and J. P. Nielsen. Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika*, 95(4): 979–986, 2008. doi: https://doi.org/10.1093/biomet/asn026. URL +http://dx.doi.org/10.1093/biomet/asn026.

C. R. Larsen. An individual claims reserving model. *ASTIN Bulletin*, 37(1):113 – 132, 2007.

Olivier Lopez, Xavier Milhaud, and Pierre-E. Thérond. Tree-based censored regression with applications in insurance. *Electron. J. Statist.*, 10(2):2685–2716, 2016. URL https://doi.org/10.1214/16-EJS1189.

Olivier Lopez, Xavier Milhaud, and Pierre-E. Thérond. A tree-based algorithm adapted to microlevel reserving and long development claims. *ASTIN Bulletin*, 49(3):741–762, 2019. URL https://doi.org/10.1017/asb.2019.1.

Thomas Mack. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, 23:213–225, 1993. doi: https://doi.org/10.2143/AST.23.2.2005092.

Thomas Mack. The standard error of chain ladder reserve estimates: Recursive calculation and inclusion of a tail factor. *ASTIN Bulletin*, 29:361–366, 1999. doi: https://doi.org/10.2143/AST.29.2.504622.

María Dolores Martínez Miranda, Jens Perch Nielsen, and Richard J. Verrall. Double Chain Ladder. *ASTIN Bulletin*, 42:59–76, 2012. doi: https://doi.org/10.2143/AST.42.1.2160712.

Michael J. Mortenson, Neil F. Doherty, and Stewart Robinson. Operational research from taylorism to terabytes: A research agenda for the analytics age. *European Journal of Operational Research*, 241(3):583 – 595, 2015. ISSN 0377-2217. doi: https://doi.org/10.1016/j.ejor.2014.08.029. URL http://www.sciencedirect.com/science/article/pii/S037722171400664X.

R. Norberg. Prediction of outstanding liabilities in non-life insurance. *ASTIN Bulletin*, 23(1):95–115, 1993. doi: https://doi.org/10.2143/AST.23.1.2005103.

R. Norberg. Prediction of outstanding liabilities II. Model variations and extensions. *ASTIN Bulletin*, 29(1):5–27, 1999. doi: https://doi.org/10.2143/AST.29.1.504603.

Angela Noufaily, Paddy Farrington, Paul Garthwaite, Doyo Gragn Enki, Nick Andrews, and Andre Charlett. Detection of infectious disease outbreaks from laboratory data with reporting delays. *Journal of the American Statistical Association*, 111(514):488–499, 2016. doi: https://doi.org/10.1080/01621459.2015.1119047.

Mathieu Pigeon, Katrien Antonio, and Michel Denuit. Individual loss reserving with the multivariate skew normal distribution. *ASTIN Bulletin*, 43:399–428, 2013. doi: https://doi.org/10.1017/asb.2013.20.

Mathieu Pigeon, Katrien Antonio, and Michel Denuit. Individual loss reserving using paid–incurred data. *Insurance: Mathematics and Economics*, 58:121–131, 2014. doi: https://doi.org/10.1016/j.insmatheco.2014.06.012.

Maëlle Salmon, Dirk Schumacher, Klaus Stark, and Michael Höhle. Bayesian outbreak detection in the presence of reporting delays. *Biometrical Journal*, 57:1051–1067, 2015. doi: https://doi.org/10.1002/bimj.201400159.

Harry Southworth. *gbm: Generalized Boosted Regression Models*, 2015. URL `https://github.com/harrysouthworth/gbm`. R package version 2.1.

Masashi Sugiyama, Matthias Krauledat, and Klaus-Robert Müller. Covariate shift adaptation by importance weighted cross validation. *J. Mach. Learn. Res.*, 8:985–1005, December 2007a. ISSN 1532-4435.

Masashi Sugiyama, Shinichi Nakajima, Hisashi Kashima, Paul von Bünau, and Motoaki Kawanabe. Direct importance estimation with model selection and its application to covariate shift adaptation. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, NIPS'07, pages 1433–1440, USA, 2007b. Curran Associates Inc. ISBN 978-1-60560-352-0. URL `http://dl.acm.org/citation.cfm?id=2981562.2981742`.

Anatoliy Swishchuk. *Change of time methods in quantitative finance.* SpringerBriefs in Mathematics. Springer, 2016. ISBN 978-3-319-32406-7; 978-3-319-32408-1. doi: https://doi.org/10.1007/978-3-319-32408-1. URL `http://dx.doi.org/10.1007/978-3-319-32408-1`.

Gregory Taylor. *Loss reserving: an actuarial perspective.* Kluwer Academic Publishers, 2000.

Łukasz Delong, Mathias Lindholm, and Mario V. Wüthrich. Collective reserving using individual claims data. 2020. Available at SSNR: `http://dx.doi.org/10.2139/ssrn.3582398`.

Roel Verbelen, Katrien Antonio, Gerda Claeskens, and Jonas Crevecoeur. Modeling the occurrence of events subject to a reporting delay via an EM algorithm. 2019. Available at arXiv: `https://arxiv.org/abs/1909.08336`.

Richard Verrall, Jens Perch Nielsen, and Anders Hedegaard Jessen. Prediction of RBNS and IBNR claims using claim amounts and claim counts. *ASTIN Bulletin*, 40(2):871–887, 2010. URL `https://doi.org/10.2143/AST.40.2.2061139`.

Richard J. Verrall and Mario V. Wüthrich. Understanding reporting delay in general insurance. *Risks*, 4(3), 2016. doi: https://doi.org/10.3390/risks4030025.

Felix Wahl, Mathias Lindholm, and Richard Verrall. The collective reserving model. *Insurance: Mathematics and Economics*, 87:34 – 50, 2019. URL `https://doi.org/10.1016/j.insmatheco.2019.04.003`.

W. Wang. Subjective estimation of the delay time distribution in maintenance modelling. *European Journal of Operational Research*, 99:516–529, 1997. doi: https://doi.org/10.1016/S0377-2217(96)00318-9.

W. Wang. A model for maintenance service contract design, negotiation and optimization. *European Journal of Operational Research*, 201:239–246, 2010. doi: https://doi.org/10.1016/j.ejor.2009.02.018.

Mario V. Wüthrich. Machine learning in individual claims reserving. *Scandinavian Actuarial Journal*, 2018(6):465–480, 2018.

Mario V. Wüthrich. Neural networks applied to chain–ladder reserving. *European Actuarial Journal*, 8(2):407–436, 2018. URL `https://doi.org/10.1007/s13385-018-0184-4`.

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

Mario V. Wüthrich and Michael Merz. *Stochastic Claims Reserving Manual: Advances in Dynamic Modeling*. 2015. Available at SSRN: `https://ssrn.com/abstract=2649057`.

Zhisheng Ye and Hon Keung Tony Ng. On analysis of incomplete field failure data. *The Annals of Applied Statistics*, 8:1713–1727, 2014. doi: https://doi.org/10.1214/14-AOAS752. URL `https://doi.org/10.1214/14-AOAS752`.