SAPIENZA
UNIVERSITÀ DI ROMA

Sapienza University of Rome

Department of Statistical Sciences (DSS)
PhD in School of Statistical Sciences
Curriculum in Actuarial Science

Thesis For The Degree Of Doctor Of Philosophy

Advances in Claims Reserve Modelling

Thesis Advisors
Assoc. Prof. Gian Paolo Clemente
Assist. Prof. Munir Hiabu

Candidate
Gabriele Pittarello
ID Number 1942608

Academic Year 2023-2024 (XXXVI cycle)

Abstract

In the non-life insurance industry, insurance companies are required by local and international regulators to set aside an amount, known as the claims reserve, to meet future payment obligations. The market data shows that the claims reserve represents the main liability of non-life (re)insurance companies (EIOPA, 2023), and we refer to its calculation at a valuation date as reserving. From a statistical perspective, this is a prediction problem. For reserving, the industry relies on simple algorithms that use an aggregated data representation called development triangles. The more detailed databases that make up the development triangles are often held by insurers and contain information about claims at an individual level that could potentially improve reserving. Examples are static covariates like age and insurance type, accident times, payment delays or payment costs.

When the reserve is computed, the data is only partially observed due to so-called right-censoring. At this point, not all the claim histories are observed, so that for some it is only known that payments might occurr in the future; this context of incomplete information makes the statistical models for event history analysis a natural framework for dealing with this type of application.

The purpose of this monograph is to illustrate three models for reserving based on the outlined framework. Our problem formulation allows us to deal with incomplete observations of claims data and to incorporate information on individual covariates. By applying our models to both simulated and real data sets, and comparing them with approaches based on development triangles, we find that these sophisticated models allow for more accurate estimates of the claims reserve.

Keywords: non-life insurance, reserving, survival analysis, machine learning, conditional Aalen-Johanson.

This thesis was defended on 27 May 2024. Board of Examiners:

Prof. Anna Rita Bacinello (University of Trieste, Chairperson)

Prof. Thomas Mikosch (University of Copenhagen)

Assoc. Prof. Massimiliano Menzietti (University of Salerno)

Acknowledgements

L'immaginazione come ho detto
è il primo fonte della felicità umana.

Zibaldone dei Pensieri
Giacomo Leopardi

I would like to thank Gian Paolo Clemente and Munir Hiabu for their supervision during the three years of my doctoral studies. Gian Paolo, thank you for inspiring me, supporting me and launching my career. Munir, thank you for hosting me, shaping my research interests and, most importantly, being a friend.

Thank you to the Department of Statistical Sciences of La Sapienza, University of Rome, for financing my Ph.D. studies. I am grateful to the Department of Mathematical Sciences at the University of Copenhagen for hosting me and making me feel part of their community. I would especially like to thank Mogens Steffensen for making this experience possible. Mogens, that lunch in Milan changed my career. Mange tak. Among the others, I would like to thank Jostein Paulsen for all the hallway chats and Thomas Mikosch for all the lunch conversations about mushrooms. Christian Furrer, thank you for every single coffee we had. Martin Bladt, thank you for our conversations, the research we completed and for keeping the running club alive after I left Copenhagen.

I would like to thank Olivier Lopez and Jostein Paulsen for their careful review of this manuscript. Your comments have greatly improved the quality of this monograph.

Thank you to my co-authors, Munir Hiabu, Andrès Villegas, Edoardo Luini, Emil Hofman, Manfred Marvin Marchione, Martin Bladt. Without your diligent work, this thesis would not exist. Edoardo, I would especially like to thank you for your mentoring during the most difficult moments of my studies.

Alexander, Emil, Julie, Margherita, Nena, Niklas, Phillip, Ulises, and all the friends from Copenhagen. Thank you for making me feel home away from home.

Alberto, Alessandro, Andrea, Davide, Filippo, Massimo, Riccardo, e Stefano. Grazie per essere gli amici di una vita. Alice, grazie per tutte le nostre conversazioni.

Tessa, thank you for the proof reading of this thesis, your unconditionate support, and the time we have had. Most importantly, thank you for the time that we will have.

Mamma e papà, grazie per il vosto amore, supporto, e per avermi insegnato l'etica del lavoro.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In non-life insurance, contracts do not settle when the insured event occurs. At the accident date, the event triggers a claim that might originate payments in the future. For this purpose, insurers need to collect an amount to cover future (possible) losses. This amount is called claims reserve and it assesses claim outstanding loss liabilities (see for example Wüthrich & Merz, 2015, p. 11). We refer to the computation of the claims reserve at a valuation date as reserving. In this manuscript, we present innovative statistical models for reserving that are based on a relatable theory (Andersen, Borgan, Gill, & Keiding, 1993). In this section, we introduce the reserving problem and motivate this manuscript with data and regulatory background on the insurance market.

Let us consider Figure 1.1, where we sketch an example of the timeline evolution of an individual claim. The insured event occurs within the insured period but the claim settles after several years. Several payments occur after the claim is reported to the insurer. Until a claim settle, the insurance company refers to it as an open claim. In certain circumstances, settled claims can be reopened (Friedland, 2010, p. 431).

In this manuscript, we will disregard the so-called claim reopenings. We will then model two groups of claims that are (qualitatively) defined as follows:

- A closed claim is a claim that was settled before the evaluation date and no future payment is expected to occur.
- An open claim is a claim for which the insurance company might pay an amount in the future.

The cash flows are not necessarily known at evaluation date and if they are not, they have to be estimated.

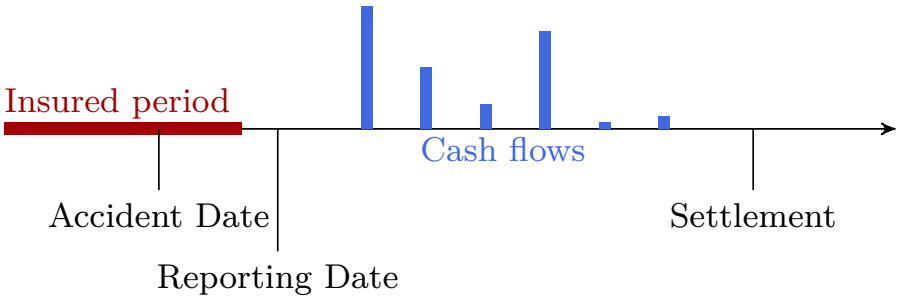


Figure 1.1: Example of an insurance claim history.

As the insurer learns about the claim only after this is reported, we can distinguish between two categories of open claims:

- The Incurred But Not Reported (IBNR) claims. These are claims that occurred during the insured period but the insurer does not know about them at the evaluation date.
- The Reported But Not Settled (RBNS) claims. These are claims that are reported to the insurer but a part (or even the whole) payments history is not yet known at the evaluation date.

The claims reserve is comprehensive of the future cost of IBNR and RBNS claims. The relative impact on the reserve of IBNR and RBNS claims depends on the type of insurance. The data on the US market reported in Friedland (2010, p. 80) show that for these types of policies, with mostly single payments and a short settlement, the reserve consists mainly of RBNS claims. The cost of future IBNR claims will have a greater impact on the reserve for those types of insurance that have a long settlement and several payments. The type of insurance is often referred to with its regulatory term, i.e. line of business (Wüthrich, 2018). The type of insurance is also an example of individual characteristic associated to a claim. The reserving data include several (static) features like age of the claimant (or policyholder), type of cover, and reporting delay that an actuary could potentially use for reserving.

1.1 Best practice in reserving

Data from the European market show that the claims reserve is the largest liability of a non-life insurer (EIOPA, 2023). From a financial perspective, an accurate estimation of the reserve is therefor important for evaluation and accounting. Indeed, insurance companies assess their claims reserve periodically. A survey on U.S. underwriters showed that more than 60% of the companies in the study perform a partial or full analysis of the reserve at least quarterly (Brown, Julga, & Merz, 2023). On the other hand, the same study illustrates that actuaries are using the same tools that they were using ten years ago for this evaluations.

Interestingly, while the methodologies that we propose are part of a prosperous academic research stream (Arjas, 1989; Norberg, 1993, 1997; Lopez, 2019; Antonio & Plat, 2014; Wüthrich, 2018), the insurance industry is still far from having individual reserving models implemented in practice (Brown et al., 2023). Another recent study performed from the Actuarial Studies in Non-life Insurance (ASTIN) section of the International Actuarial Association (IAA) shows that, worldwide, companies mostly rely on simple algorithms with basic software implementation (Dal Moro, Cuypers, & Miehe, 2016). This is motivated by the difficulties in retrieving data of sufficient quality for sophisticated data analysis, and preference for licensed software that do not include literature novelties. In addition, licensed software is considered more secure by the industry as it provides product support and is standard across analysts. In fact, business analysts do not necessarily have expertise on literature novelties that require a consolidated knowledge of statistics for data science and programming.

On the other hand, the industry is recently looking towards data scientists and innovative data oriented projects and companies are trying to experiment the role of new professional figures

in reserving, see again Brown et al. (2023). In their most recent reports, the main consultancy players recognise the importance for underwriters and clients of using innovative approaches to their business (Accenture, 2023; Deloitte, 2023; PWC, 2023). On a larger scale this is a movement that is expected to grow fast in the next few years. The United Nations (UN) forecast that from 2018 to 2025 the general market size of AI and technology on is going from 350 billion dollars to 3.2 trillion dollars (United Nations, 2021). This is accompanied by a trend in education, where university will provide in the next years more and more data scientists that are expected to have interdisciplinary skills (Glantz et al., 2023).

With reference to the models presented in this manuscript, we want to address the shortcomings of the current approaches available in the literature as follows. In order to make the models easily implementable, our methodologies are accompanied by a software implementation (Pittarello, Hiabu, & Villegas, 2022; Hofman, Pittarello, & Hiabu, 2023; Bladt & Furrer, 2023a).

1.1.1 Individual reserving in the European Union regulation landscape

Reserving in the European Union (EU) is regulated both on a national level by local Generally Accepted Accounting Practices (GAAPS) and on an international level by Solvency II (European Parliament, 2009), the European Directive that introduced the idea of a market consistent evaluation of assets and liabilities. The Solvency II directive relies on a three pillars structure: quantitative capital requirements, qualitative supervisory review process and disclosure respectively.

The third pillar provides public disclosure requirements, more specifically an annual report is required on the insurance company's financial and solvency situation. We focus the attention on the first and second pillars. The former regards the market consistent valuation of assets and liabilities for the computation of the Economic Balance Sheet (the main output of the insurer's accounting), capital requirements and validation of internal models. Pillar two is supplementary to pillar one. Indeed, not all the risks can be adequately estimated using pure quantitative measures. With reference to the evaluation of assets and liabilities, the directive asks insurers to carry on a market consistent evaluation of assets and liabilities (article 75). The directive also decomposes the insurance business in modules and sub-modules. As far as concerns the Non-Life module, the most relevant sub-module is the Premium and Reserve Risk. The latter is cited because it outlines the framework within which new and more sophisticated models, such as those presented in this monograph, could reasonably be applied. Our models could be implemented in the aim of a partial or full internal model (article 112) to provide a more accurate estimate of the insurer's capital requirement. The regulator not only wants insurance companies to meet all their obligations, but also wants to encourage companies to develop and use new techniques for internal control and risk management. (Pittarello, 2020).

1.2 Non-life insurance data

In this section, we formulate the reserving problem in continuous time starting from the individual claims. We then connect our framework to the so-called development triangles, a discrete and aggregated data representation that is traditionally used by actuaries for modelling. Development triangles are often referred to as run-off triangles (Hiabu & Pittarello, 2023): in this monograph we use the two terms interchangeably.

At the evaluation date $\mathcal{T}$, we have observed n claim payments. For each payment, we are given with the time delay from accident to payment, T_i, and payment size Z_i, $i = 1, \dots, n$. As illustrated in Figure 1.1, in insurance data there is often a delay between reporting and payment. While a model for the reporting delay will be the discussed in the following chapters, for the moment we restrict our focus on the payments size and their corresponding delay from the insured event.

We only observe (T_i, Z_i) when the payment date is before the evaluation date. Therefore, by design it holds

$$T_i \leq \mathcal{T} - U_i,$$

where U_i is the accident date of claim i. As pointed out in Hiabu (2017) and Bischofberger, Hiabu, and Isakson (2020), statistical inference on (T_i, Z_i) is not directly feasible. In this chapter, we will not discuss a modelling approach for (T_i, Z_i) which will be the focus of the main corpus of this manuscript.

We assume that observations are aggregated in periods of size δ. For notations convenience, we assume that the same aggregation level δ is chosen for both development date, T_i and accident date, U_i and furthermore that $m = \mathcal{T}/\delta$ with $m \in \mathbb{N}_0$. We then work on an equidistant grid $t_0 = 0, \dots, t_m = \mathcal{T}$ and $u_0 = 0, \dots, u_m = \mathcal{T}$, with $t_j - t_{j-1} = u_k - u_{k-1} = \delta$, for $j, k = 0, \dots, m$. In classical run-off triangles we only have information on the sum of payments falling within a parallelogram

$$\mathcal{P}_{kj} = \{(t, u) : t_j + u_k - u \leq t \leq t_{j+1} + u_k - u; u \in [u_k, u_{k+1}), t \geq 0\}.$$

In Figure 1.2 we show an example of a continuous time run-off triangle. In insurance data sets, the individual observations $(Z_i, T_i, U_i), i = 1, \dots n$ are compressed into observations X_{kj}; $j, k = 0, \dots, m$ with

$$X_{kj} = \sum_i Z_i \int I\left((s, U_i) \in \mathcal{P}_{kj}\right) dN_i(s).$$

with $N_i(t) = I(t \geq \mathcal{T} - T_i)$ each with respect to the filtration $\mathcal{F}_{it} = \sigma\left(\left\{\mathcal{T} - T_i \leq s : \ s \leq t\right\} \cup \left\{U_i\right\} \cup \mathcal{N}\right)$, where $\mathcal{N}$ is the set of all zero probability events, as we will discuss more formally in Section 2.2. Throughout the manuscript, we will use I to denote the indicator function.

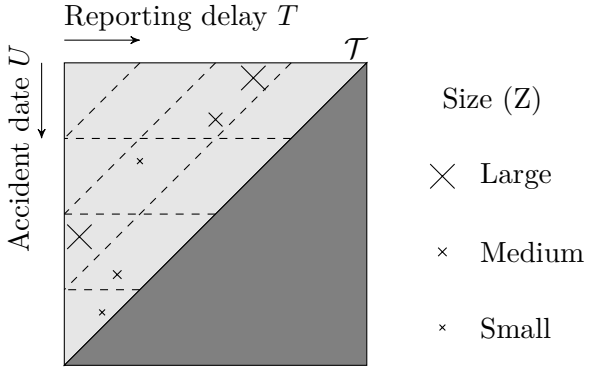


Figure 1.2: Illustration on how individual payments are aggregated into cells with a parallelogram shape.

1.3 The Chain Ladder Method

We start this section by introducing the chain ladder, the most used algorithm for reserving (Dal Moro et al., 2016; Hiabu & Pittarello, 2023). We define the cumulative claim payments in accident period k and development period j by

$$C_{kj} = \sum_{l=0}^{j} X_{kl},$$

which are used to calculate the so-called development factors for $j = 1, \dots, m$:

$$\widehat{f}_j = \sum_{s=0}^{m-j} \frac{C_{s,j-1}}{\sum_{l=0}^{m-j} C_{l,j-1}} \frac{C_{s,j}}{C_{s,j-1}} = \sum_{s=0}^{m-j} \frac{C_{s,j-1}}{\sum_{l=0}^{m-j} C_{l,j-1}} \hat{f}_{s,j} = \frac{\sum_{s=0}^{m-j} C_{s,j}}{\sum_{s=0}^{m-j} C_{s,j-1}}. \quad \text{(Development factors)}$$

In this manuscript, we will refer to $\hat{f}_{k,j} = C_{k,j}/C_{k,j-1}$ as individual development factors.

Assuming no tail is present, we project future observations via a chain principle,

$$\widehat{C}_{k,j} = C_{k,m-k} \prod_{l=m-k+1}^{j} \hat{f}_{kl}, \qquad \widehat{X}_{k,j} = \begin{cases} \widehat{C}_{k,j} - C_{k,j-1} & \text{if } k+j = m+1, \\ \widehat{C}_{k,j} - \widehat{C}_{k,j-1} & \text{if } k+j \geq m+2. \end{cases} \tag{1.1}$$

The claims reserve is $R = \sum_{k=1}^{\mathcal{T}} R_k$, where $R_k = \hat{C}_{k,m} - C_{k,m-k}$.

The actuarial literature has explored several stochastic models that replicate its result for the reserve (Hess & Schmidt, 2002). In this section, we start introducing the stochastic approach in (Mack, 1993), that will be used as a competing model in two of our implementations Hiabu, Hofman, and Pittarello (2023); Bladt and Furrer (2023b). We conclude introducing the age-period-cohort models in Harnau and Nielsen (2018), that was instead using as a competing implementation for Pittarello, Hiabu, and Villegas (2023).

Remark 1.3.1. Development triangles are, historically, the data representation used by actuaries for provisioning (Taylor, 2019). Notably, insurance mathematicians have explored several solutions based on aggregated data that are different from the chain-ladder based approaches that we present in this chapter. We remark that the objective of this thesis is to present new approaches for reserving that integrate the extensive information on individual claims available in company data bases in actuarial models. While a more detailed discussion on aggregated reserving models would provide more insights on how actuaries do reserving in practice, we decided to restrict our presentation to the few models that we used as benchmark in our applications. For a more detailed discussion on aggregate reserving models we refer to Taylor (2012); Wüthrich and Merz (2015).

1.3.1 Mack Model

The Mack Model (Mack, 1993) makes the following assumptions:

Mack 1 Aggregated claims are independent across accident periods: $C_{i,j} \perp\!\!\!\perp C_{k,l}$ for all $i \neq k$.

Mack 2 There exist positive factors $f_1, \dots, f_m$ such that

$$\mathbb{E}\left[C_{k,j} \mid \mathcal{F}_{j-1}\right] = f_j C_{k,j-1}, \quad j = 1, \ldots, m,$$

where $\mathcal{F}_j = \sigma\{C_{i,k} : \ i = 0, \ldots, m; \ k \leq j\}$.

Mack 3 There are positive factors $\sigma_1, \ldots, \sigma_m$ such that

$$\mathrm{Var}\left[C_{k,j} \mid \mathcal{F}_{j-1}\right] = \sigma_j^2 C_{k,j-1}, \quad j = 1, \ldots, m.$$

Under the Mack model, an estimator of σ_j^2, for $j = 1, \ldots, m-1$ is

$$\hat{\sigma}_j^2 = \frac{1}{m-j} \sum_{s=0}^{m-j} C_{s,j-1} \left(\frac{C_{s,j}}{C_{s,j-1}} - \hat{f}_j \right)^2.$$

To get an estimator for σ_m^2, one may need to extrapolate. If claims development is finished after $m-1$ years, one can simply set $\hat{\sigma}_m^2 = 0$. Otherwise, Mack suggests

$$\hat{\sigma}_m^2 = \min \left\{ \frac{\hat{\sigma}_{m-1}^4}{\hat{\sigma}_{m-2}^2}, \hat{\sigma}_{m-2}^2, \hat{\sigma}_{m-1}^2 \right\}.$$

Proposition 1.3.1. Under [Mack 1] – [Mack 3], $\hat{\sigma}_j^2, j = 1, \ldots, m-1$, and $\hat{f}_j, j = 1, \ldots, m$, are unbiased estimators of σ_j^2, and f_j. Furthermore for $j \neq j'$, $\hat{f}_j$ and $\hat{f}_{j'}$ are uncorrelated.

One can show that under Mack's assumption, $\hat{f}_j$ has minimal conditional variance (conditioned on $\mathcal{F}_j$) under all estimators of the form $\sum_{s=0}^{m-j} w_s C_{s,j}/C_{s,j-1}, \sum_s w_s = 1$.

The main purpose in developing a stochastic model that replicates the chain ladder estimates is to be able to discuss estimation uncertainty. Our goal is to derive an estimator for $\mathbb{E}[(\hat{R}_k - R_k)^2 | \mathcal{G}_m]$ where $R_k = \sum_{j=m-k+1}^{m} X_{k,j}$ with plug-in estimator $\hat{R}_k$ and

$$\mathcal{G}_t = \sigma\{C_{k,j} : \ k + j \leq t\}.$$

Note that

$$\mathbb{E}[(\hat{R}_k - R_k)^2 |\ \mathcal{G}_m] = \mathbb{E}[(\hat{C}_{k,m} - C_{k,m})^2 |\ \mathcal{G}_m].$$

Hence,

$$\mathbb{E}[(\hat{R}_k - R_k)^2 |\ \mathcal{G}_m] = \mathrm{Var}(C_{k,m} |\ \mathcal{G}_m) + \left(\mathbb{E}[C_{k,m} |\ \mathcal{G}_m] - \hat{C}_{k,m} \right)^2.$$

The first summand on the right hand side is the process variance and the second summand is the parameter uncertainty.

Straightforward calculations lead to the following estimator for the uncertainty of the reserve in accident period k

$$\hat{\mathbb{E}}[(\hat{R}_k - R_k)^2 | \mathcal{G}_m] = \hat{C}_{k,m}^2 \sum_{s=m-k+1}^{m} \frac{\hat{\sigma}_s^2}{\hat{f}_s^2} \left(\frac{1}{\hat{C}_{k,s-1}} + \frac{1}{\sum_{i'=0}^{m-s} C_{i',s-1}} \right).$$

Since the reserve estimates are correlated, we need to consider a correlation term to estimate the uncertainty of the full reserve, $\mathbb{E}[(\sum_i \hat{R}_k - R_k)^2 |\ \mathcal{G}_m]$,

$$\hat{\mathbb{E}}[(\sum_k \hat{R}_k - R_k)^2 |\, \mathcal{G}_m] = \sum_{k=1}^{m} \left\{ \hat{\mathbb{E}}[(\hat{R}_k - R_k)^2 |\, \mathcal{G}_m] + \hat{C}_{k,m} \left(\sum_{i'=k+1}^{m} \hat{C}_{i',m} \right) \sum_{s=m-k+1}^{m} \frac{2\hat{\sigma}_s^2}{\hat{f}_s^2 \sum_{i'=0}^{m-s} C_{i',s-1}} \right\}.$$

1.3.2 Age-Period-Cohort (APC) models

Claims reserving models replicating the chain-ladder estimates by modelling the incremental claim amounts have been extensively studied in the literature, see e.g. England and Verrall (1999), Pinheiro, Andrade e Silva, and de Lourdes Centeno (2003), Kuang, Nielsen, and Nielsen (2008a), Kuang, Nielsen, and Nielsen (2008b), Kuang, Nielsen, and Perch Nielsen (2011), Björkwall, Hössjer, and Ohlsson (2009), Gabrielli, Richman, and Wüthrich (2020), Avanzi, Taylor, Vu, and Wong (2020), Taylor (2021). We can include in this literature stream the Age-Period-Cohort (APC) models in Harnau and Nielsen (2018). As we will explain in more detail in the next chapter, APC models are a standard approach in the mortality and epidemiology literature, but they can also be used for modelling development triangles. Development triangles are just an age-period-cohort data representation akin to Lexis diagrams. In the sequel, we will borrow the language from mortality (Carstensen, 2007), with k denoting cohort, j age and $k+j$ period. In claims reserving, age j usually corresponds to development period, period $k+j$ to calendar time and cohort j to accident period, see Table 1.1.

Table 1.1: Notation parallel, Lexis dimensions and claims reserving.

Notation	Claims reserving	Lexis dimensions
j	development period	age
k	occurrence or accident date	cohort
$k+j$	calendar date	period

The APC family includes the so-called age-cohort model. Assuming $X_{k,j} \overset{iid}{\sim} \text{Pois}(\exp(\eta_k + \nu_j))$, and $\eta_k, \nu_j \in \mathbb{R}$ we can use a Generalised Linear Model (GLM) with linear predictor

$$\mathbb{E}[\log(X_{k,j})] = \eta_k + \nu_j, \tag{1.2}$$

to model the incremental paid amounts, with $k, j = 0, \ldots, m$. The linear predictor includes

- η_k, an effect that depends on the accident date (cohort).
- ν_j, an effect that depends on the development date (age).

For the model identification we use the canonical parametrization in (Kuang et al., 2008a, 2008b). Note that the same structure in Equation (1.2) is also assumed in the unobserved lower triangle, $k+j > m$. For $k, j = 0, \ldots, m$, given the maximum likelihood estimates for $(\hat{\eta}_k, \hat{\nu}_j)_{k,j}$, one can project into the lower triangle

$$\hat{X}_{k,j} = \exp(\hat{\eta}_k + \hat{\nu}_j).$$

The age-cohort model replicates the point estimates of the chain ladder for the reserve and it was first presented in the actuarial literature in the seminal paper England and Verrall (1999).

The models in (Kuang et al., 2008a; Harnau & Nielsen, 2018) include a more general predictor that incorporates γ_{k+j}, an effect on the calendar period $k+j$

$$\mathbb{E}[\log(X_{k,j})] = \eta_k + \nu_j + \gamma_{k+j}.$$

1.4 Outline of the manuscript

The research on claims reserving included in this manuscript can be divided into two parts. The first part collects the work in Pittarello et al. (2023) and Hiabu et al. (2023) while the second part is based on the work in Bladt and Pittarello (2023). While the two parts show two different approaches for reserving, they both show the application of statistical models for survival analysis to the actuarial science (Andersen et al., 1993). This monograph is organised as follows:

- In Part I, we investigate the relationship between the micro-level structure of claims data and development triangles, with a particular focus on claims counts. Here, we discuss estimation and inference of multiplicative hazard rate models on triangular data. By considering an individual framework for defining our model, we use a counting process on the reverse development time to reformulate the right-censored problem into a feasible left-censored problem and propose an estimator for the claim development based on development triangles. We prove that there is a one-to-one correspondence between the chain-ladder development factors and our estimator. While aggregating the data into triangles the individual information on claims is disregarded, our estimator can be used when the individual observations are not available, as it is often the case in reserving as in epidemiology and mortality studies (Gill, 1986), (Andersen et al., 1993, p. 168). In Chapter 3, we propose a refinement of Chapter 2 for predicting future claims using individual data for model estimation. We now model the individual hazard rates directly and discuss how to project future claim reports using the correspondence to the development factors discussed in Chapter 2. We explain the relationship between development times and survival analysis ties, and elaborate on the relationship between the model baseline and the distribution of claim counts in the tie. We propose a proportional hazards model in which the logarithmic risk function is modelled using various machine learning models: Cox model with splines (Gray, 1992), Neural Networks, and eXtreme Gradient Boosting (Chen & Guestrin, 2016). In this project, we have also extended the state-of-the-art machine learning literature on survival analysis to account for the presence of ties in the data. When individual (continuous) data are available, e.g. on a daily basis, we discuss how to report our predictions on different (discrete) data aggregations, e.g. monthly, quarterly, or yearly. We discuss the invariance of the forecasts for different output aggregations. The usefulness of our approach is demonstrated with an extensive simulation study and a case study on a real portfolio provided by a major insurance company.

- In Part II, we extend the findings of a recent paper (Bladt & Furrer, 2023b), which introduces the conditional Aalen-Johansen estimator. This non-parametric kernel estimator for the occupation probabilities is applicable to any finite-state jump process, and it supports conditioning on feature information. Taking advantage of the panel structure of reserving data, in the manuscript we propose a reinterpretation of a finite state space, where the states

are the discrete times at which each payment was recorded, and an absorbing state represents the final payment. We also propose modelling the counting process in terms of cumulative claim payments rather than time. This allows us to reinterpret the occupancy probabilities of an absorbing state as the cumulative density function of individual claim costs. We derive a nonparametric kernel estimator of individual claim payments as a function of individual characteristics and use it to numerically compute the moments of the reserve distribution. We illustrate the usefulness of our approach on an extensive simulation study and on a real data application. Lastly, we also discuss proper scoring and illustrate how the continuously ranked probability score can be used to assess the performance of our models (Gneiting & Ranjan, 2011).

- Chapter 5 concludes with some final remarks on this monographs and points out that Chapter 3 and Chapter 4 could be joined in future work.

As a last remark of this chapter and before discussing Part I of this manuscript, we would like to highlight that Pittarello et al. (2023) and Hiabu et al. (2023) are accompanied by an open-source software that includes a general implementation of the models that we present in this monograph (Pittarello et al., 2022; Hofman et al., 2023). The results in Bladt and Pittarello (2023) can be replicated using the R package in Bladt and Furrer (2023a).

Part I

Claim development models

Chapter 2

A survival framework for claims reserving

In this research stream, we discuss estimation and inference of multiplicative hazard rate models on development triangles. The included articles answer two very operational questions: How can probabilistic modelling be done when the data is pre-smoothed into development triangles? (Pittarello et al., 2023) What are the potential gains in forecasting when models are estimated on the individual data? (Hiabu et al., 2023) The presented models share a common framework (Hiabu, 2017), but in the first project (Pittarello et al., 2023) we discuss estimation on aggregated data and in the second project (Hiabu et al., 2023) estimation is done entirely at the individual level. In Section 2.2 we describe the theory of reverse time hazard models to introduce this first part of the manuscript (Hiabu & Pittarello, 2023) and, starting from Section 2.3, we discuss and apply to a case study the estimator for the claim development in Pittarello et al. (2023).

2.1 Motivation

In Pittarello et al. (2023), we propose an estimator of the claim development based on run-off triangles. While pre-smoothing into run-off triangles is theoretically less efficient than direct estimation based on individual data, our proposed estimator can be applied on data readily available and most familiar to reserving actuaries. In Section 2.3 we derive an exact and non-asymptotic relationship between time-reversed averaged hazard rates and chain-ladder's development factors. By using discrete estimators, we show how existing GLM-based estimators within the age-period-cohort framework can be employed to model claim development. Interestingly, in Section 2.6 we show that modelling the claim development via an age-model replicates chain-ladder. We find that our proposal of modelling the claim development via an age-period-cohort models seems to outperform age-period-cohort models based on claim amounts in most cases. Our framework can then complement other modelling techniques in the actuary tool-box.

As earlier mentioned, claims reserving models replicating the chain-ladder estimates by modelling the incremental claim amounts have been extensively studied in the literature, see Section 1.3.2. By contrast, given a run-off triangle, we propose to model the claim development instead of the claim amount. It is surprising that despite the idea of modelling the claim development being arguably more in the spirit of the chain-ladder algorithm, there is only little literature

that models the claim development given a run-off triangle. A notable exception is Mack (1993) and extensions thereof to multiple triangles, e.g. Schnieper (1991) and Wüthrich (2018). However, as has been noted in Bischofberger et al. (2020) and Mikosch (2009, Chapter 11.2) the assumptions of the model in Mack (1993) do not conform well with the data-generating mechanism from an individual claim level perspective: claim payments in later development periods are assumed to always stem from earlier payments.

In recent years there have been several papers that model continuous run-off triangles based on individual data. In Miranda, Nielsen, Sperlich, and Verrall (2013) and Hiabu, Mammen, Martínez-Miranda, and Nielsen (2016) the loss-triangle based reserving-problem is translated into a continuous framework and the authors propose to estimate a density function corresponding to claim counts via kernel smoothers. The model has been extended in Lee, Mammen, Nielsen, and Park (2015), Lee, Mammen, Nielsen, and Park (2017), and Mammen, Martínez-Miranda, Nielsen, and Vogt (2021) to account for seasonal effects, operational time and calendar effects, respectively. Instead of modelling the claim counts, Hiabu (2017) and Bischofberger et al. (2020) model the claim development and show an asymptotic relationship between chain-ladder's development factors and a time-reversed hazard rate.

In Pittarello et al. (2023) we elaborate on the one-to-one correspondence between an estimator of the claims development and the chain-ladder development factors. In contrast to our paper, the estimation in Hiabu (2017) and Bischofberger et al. (2020) is based on individual data and the relationship between chain-ladder's development factors and a time-reversed hazard rate is based on continuous observations only.

This chapter comes with the `clmplus` package (Pittarello et al., 2022), an out-of-the box set of tools available in R to practitioners and researchers that may want to apply or build on our framework. The `clmplus` implementation relies on the tools available in the `StMoMo` package (Millossovich, Villegas, & Kaishev, 2018), that made the the age-period-cohort framework easy and ready to use in a mortality context.

2.2 Granular model formulation

Given a cut-off-date $\mathcal{T}$, we have observed n claim payments. For every payment, we are given the time delay from accident until payment, T_i, and payment size Z_i, $i = 1, \dots, n$. We make the following assumption.

[M1] All payments are independent.

As in the age-cohort model (cross-classified GLM in England & Verrall, 1999), assumptions [M1] is rather strong but is made to simplify the mathematical derivations and one may conjecture that it can be relaxed by considering weak dependency between payments. Statistical inference on (T_i, Z_i) is not directly feasible. We only observe (T_i, Z_i) when the payment date is before the cut-off-date. Therefore, by design it holds

$$T_i \leq \mathcal{T} - U_i,$$

where U_i is the accident date of claim i. Hence, we are exposed to a right-truncation problem. A solution to the right-truncation problem is to reverse the time of the counting process leading to a tractable left-truncation problem. To this end we consider the development-time reversed counting

processes

$$N_i(t) = I(t \geq \mathcal{T} - T_i),$$

each with respect to the filtration

$$\mathcal{F}_{it} = \sigma\left(\left\{\mathcal{T} - T_i \leq s : \ s \leq t\right\} \cup \left\{U_i\right\} \cup \mathcal{N}\right),$$

satisfying the *usual conditions* (Andersen et al., 1993, p. 60), and where $\mathcal{N}$ is the set of all zero probability events.

2.2.1 The intensity process

In contrast to Hiabu (2017) and Bischofberger et al. (2020), we allow T_i to depend on U_i. Then, under mild regularity conditions, the intensity process of N_i is

$$\lambda_i(\mathcal{T} - t|u) = \lim_{h \downarrow 0} h^{-1}\mathbb{E}\left[N_i\left\{(\mathcal{T} - t + h)-\right\} - N_i(\mathcal{T} - t-)|\ \mathcal{F}_{i,(\mathcal{T}-t)-}\right] = \alpha(t|U_i)Y_i^R(t),$$

where

$$\begin{aligned}
\alpha(t|u) &= \lim_{h \downarrow 0} h^{-1} P\left(T_i \in (t-h, t]|\ Y_i^R(t) = 1, U_i = u\right), \\
Y_i^R(t) &= I(T_i \leq t < \mathcal{T} - U_i).
\end{aligned}$$

This structure is called Aalen's multiplicative intensity model (Aalen, 1978), and enables non-parametric estimation and inference on the deterministic hazard function α.

Let Z_i denote the payment size of claim i and consider the process $N_i^*(t) = Z_i N_i(t)$. Under mild regularity conditions, it is straightforward to see that

$$\begin{aligned}
\lambda_i^*(\mathcal{T} - t|u) &= \lim_{h \downarrow 0} h^{-1}\mathbb{E}\left[N_i^*\left\{(\mathcal{T} - t + h)-\right\} - N_i^*(\mathcal{T} - t-)|\ \mathcal{F}_{i,(\mathcal{T}-t)-}\right] = \mathcal{R}_n(t)\alpha^*(t|U_i)Y_i^{*,R}(t), \\
\alpha^*(t|u) &= \frac{\mathbb{E}[Z_1|T_1 = t, U_1 = u]}{\mathbb{E}\left[Z_1|\ T_1 \leq t, U_1 = u\right]}\alpha(t|u), \\
Y_i^{*,R}(t) &= Z_i Y_i^R(t), \\
\mathcal{R}_n(t,u) &= \frac{\mathbb{E}\left[Z_1|\ T_1 \leq t, U_1 = u\right]}{Z_i}.
\end{aligned}$$

2.3 Introducing a grid system

We use the notation of Section 1.2 and assume that observations are aggregated into periods of size δ. For notational convenience, we assume that the same aggregation level δ is chosen for both development direction, T_i and accident direction, U_i and furthermore that $m = \mathcal{T}/\delta$ with $m \in \mathbb{N}_0$.

To estimate the development in the run-off triangle, we are interested in the exposure-weighted

average hazard on the parallelogram:

$$\mu_{k,j} = \frac{\delta \int_{\mathcal{P}_{k,j}} \alpha^*(s|u) p_U(u) \gamma(s,u) ds du}{\int_{\mathcal{P}_{k,j}} p_U(u) \gamma(s,u) ds},$$

where $\gamma(s,u) = E[Y_i^{*,R}(s)|U_i = u]$ and p_U is the marginal density of U_i. Under regularity conditions, $\mu_{k,j}$ is, as the number of policies goes to infinity, the asymptotic limit of

$$\frac{\delta \sum_i \int I\left((s,U_i) \in \mathcal{P}_{k,j}\right) \alpha^*(s|U_i) Y_i^{*,R}(s) ds}{\sum_i \int I\left((s,U_i) \in \mathcal{P}_{k,j}\right) Y_i^{*,R}(s) ds}.$$

Hence, if observations on the individual level are available, then a natural estimator of $\mu_{k,j}$ is

$$\begin{aligned}
&\frac{\sum_i Z_i \int I\left((s,U_i) \in \mathcal{P}_{k,j}\right) dN_i(s)}{\delta^{-1} \sum_i \int I\left((s,U_i) \in \mathcal{P}_{k,j}\right) Y_i^{*,R}(s) ds} \\
&= \frac{X_{k,j}}{\delta^{-1} \sum_i \int I\left((s,U_i) \in \mathcal{P}_{k,j}\right) Y_i^{*,R}(s) ds} \\
&= \frac{X_{k,j}}{\sum_{l<j} X_{k,l} + \delta^{-1} \sum_{i:I\left((T_i,U_i)\in\mathcal{P}_{k,j}\right)=1} \int I\left((s,U_i) \in \mathcal{P}_{k,j}\right) Y_i^{*,R}(s) ds}.
\end{aligned}$$

However, if data is given in the form of a run-off triangle, then the second summand in the denominator is not observed and needs to be approximated. Note that the expression

$$\delta^{-1} \sum_{i:I\left((T_i,U_i)\in\mathcal{P}_{k,j}\right)=1} \int I\left((s,U_i) \in \mathcal{P}_{k,j}\right) Y_i^{*,R}(s) ds,$$

can be written as $\eta X_{k,j}$ with $\eta \in [0,1]$. Assuming that (T,U) is uniformly distributed, conditioned on the parallelogram $\mathcal{P}_{k,j}$, an unbiased estimator is given by $\frac{1}{2} X_{k,j}$. To see this note that for those payments that occurred in the parallelogram (j,k), i.e. $I\left((T_i,U_i) \in \mathcal{P}_{k,j}\right) = 1$, the length of the crossing from entering the parallelogram until payment, i.e. in expected terms is equal to $\frac{1}{2}\delta$:

$$\mathbb{E}\left[\int I((s,U_i) \in \mathcal{P}_{k,j}) Y_i^R(s) ds\right] = \delta^{-2} \int_0^{\mathcal{T}} \int_0^{\mathcal{T}} I\left((s,u) \in \mathcal{P}_{k,j}\right) \left(t_{j+1} - s - u + u_k\right) \, ds du = \frac{1}{2}\delta.$$

Hence, we propose to estimate $\mu_{k,j}$ via

$$\widehat{\mu}_{k,j} = \frac{X_{k,j}}{\sum_{l<j} X_{k,l} + \frac{1}{2} X_{k,j}}.$$

In this manuscript, we will denote $E_{k,j} = \sum_{l<j} X_{k,l} + \frac{1}{2} X_{k,j}$. In the next section, we will connect $\widehat{\mu}_{k,j}$ to the chain-ladder's individual development factors $\widehat{f}_{k,j} = \sum_{l\leq j} X_{k,l} / \sum_{l<j} X_{k,l}$.

2.4 Lexis diagrams and development triangles

The link between time-reversed hazard rates and chain-ladder's development factors is akin to the link between central mortality rates and survival probabilities in the context of mortality modelling. Thus, in contrast to (Tsai & Kim, 2022) who leverage claim reserving methods to project mortality rates, we leverage age-period-cohort mortality modelling methods to estimate claim reserves. To this end, we consider a run-off triangle of incremental payments

$$\mathcal{X} = \{X_{k,j} : k, j = 0, \dots, m; k + j \leq m\}, \quad m > 0,$$

where $X_{k,j}$ denotes the observation for accident period k and development period j.

In the case of $\mathcal{X}$ being claim payments, $\widehat{\mu}_{k,j}$ can be motivated as the estimator for $\mu_{k,j}$. $\widehat{\mu}_{k,j}$ is the averaged hazard rate running in reversed development time, additionally weighted by the expected claim size in cell (k, j). This can be seen by noting that $\widehat{\mu}_{k,j}$ is the ratio of occurrence and expected exposure. An explicit and more technical definition of $\mu_{k,j}$ in the latter case will be provided in the next section. We will also prove that a simple rearrangement shows that for $k = 0, \dots, m; j = 1, \dots, m$, the observed claim development, $\widehat{\mu}_{k,j}$, has the following relationship to the individual development factors, $\widehat{f}_{k,j} = \sum_{l \leq j} X_{k,l} / \sum_{l < j} X_{k,l}$:

$$\hat{\mu}_{kj} = 2\frac{\hat{f}_{kj} - 1}{\hat{f}_{kj} + 1},$$

that we rewrite as

$$\widehat{f}_{k,j} = \frac{2 + \widehat{\mu}_{k,j}}{2 - \widehat{\mu}_{k,j}}. \tag{2.1}$$

Equation (2.1) should be seen as general formula to switch between a modeled claim development and the associated development factor; with the latter being subsequently used for prediction. More concretely, starting from the raw observations $\widehat{\mu}_{k,j}$, we will assume some structure on $\mu_{k,j}$ and derive estimates $\tilde{\mu}_{k,j}$ of the claim development. Thereafter one applies the equivalent of Equation (2.1) to derive estimates of the development factors, $\tilde{f}_{k,j}$, which are used to project observations $\mathcal{X}$ into the lower triangle. We will propose to model the observed claim development, $\widehat{\mu}_{k,j}$, for $j = 1, \dots, m$, via a GLM which in its simplest form can replicate chain-ladder's development factors and additionally allows for straightforward variations by changing the underlying distribution or structure.

Interestingly, there is a parallel to draw to mortality modelling in actuarial science and demography. For modelling human mortality, we consider data from large population studies where information about the life histories of individuals is not known (Wilmoth et al., 2021). The observations are aggregated into Lexis diagrams $\mathcal{L}$,

$$\mathcal{L} = \{d_{k,j} : k, j = 0, \dots, q; k + j \leq q\}, \quad q > 0,$$

where $d_{k,j}$ denotes the observed (total) number of deaths for age j and cohort k.

As in the reserving framework, and unlike standard survival studies, we will not propose a

model for individual hazard rates, as the granular data are not available. Instead, we propose an estimator of the so-called central mortality rate, essentially an average hazard rate for age j and cohort k, based on the data aggregation $\mathcal{L}$.

Let $m_{k,j}$ denote the central mortality rate at age j for cohort k, $q_{k,j}$ the probability that an individual aged j in cohort k dies before reaching age $j+1$, and $p_{k,j} = 1 - q_{k,j}$ the probability that an individual in cohort k with age j survives to age $j+1$.

Estimation of $m_{k,j}$ is based on the observed rates $\hat{m}_{k,j} = d_{k,j}/E_{k,j}$, where $E_{k,j}$ denotes the person-years at risk. However, when data are aggregated in Lexis diagrams, the actual number of persons at risk is unknown. The quantity $E_{k,j}$ is then approximated as the number of people of cohort k that died at age $j+1$ or later plus one half times the number of people who died at age j. In mortality modelling, one usually models age-period-cohort effects on the central mortality rate $m_{k,j}$ leading to estimates $\tilde{m}_{k,j}$ (Carstensen, 2007). Afterwards, one calculates the discrete quantities $\tilde{q}_{k,j}$ and $\tilde{p}_{k,j}$ using plug-in estimates in the following standard life-table relationships (Chiang, 1972):

$$q_{k,j} = \frac{2m_{k,j}}{2 + m_{k,j}} \tag{2.2}$$

$$p_{k,j} = 1 - q_{k,j} = 1 - \frac{2m_{k,j}}{2 + m_{k,j}} = \frac{2 - m_{k,j}}{2 + m_{k,j}}. \tag{2.3}$$

As we will discuss in Section 2.5, we will use the same plug-in estimators to show Equation (2.1).

Comparing Equation (2.3) and Equation (2.1) we can summarize the following: in mortality modelling we estimate the central mortality rate $m_{k,j}$ from observations $\hat{m}_{k,j}$ leading to estimates $\tilde{m}_{k,j}$ and use them to derive estimates for one year conditional survival probabilities $\tilde{p}_{k,j}$. For claims reserving we propose to estimate the claim development $\mu_{k,j}$ from observations $\hat{\mu}_{k,j}$ leading to estimates $\tilde{\mu}_{k,j}$ and use it to calculate development factors $\tilde{f}_{k,j}$ which are estimates of inverse conditional probabilities. We will exploit this connection between claim reserving and mortality modelling to provide more flexible claim development models and leverage existing mortality modelling software.

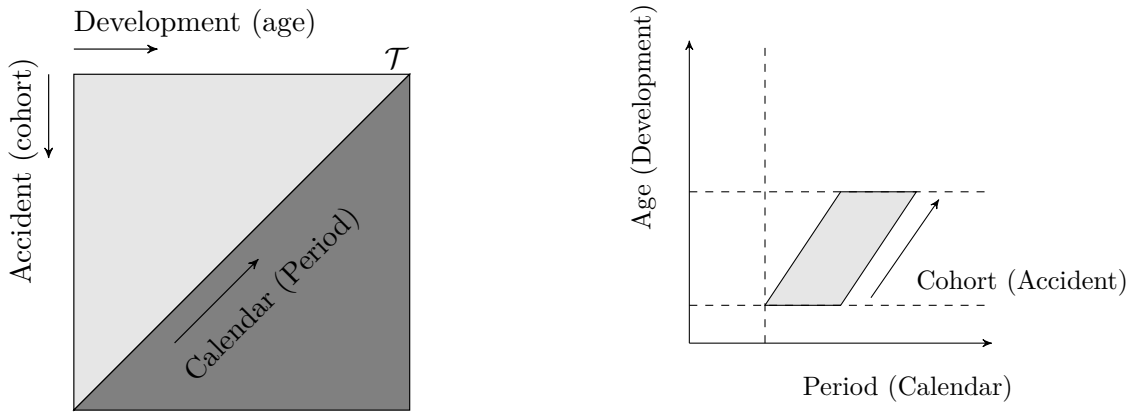
In a similar fashion, we can also obtain an estimator of the survival function. The survival function is in fact related to the development factors with the following relationship $\hat{S}(j|U) = 1/\prod_{l=1}^{j} \hat{f}_{k,l}$.

In Figure 2.1 we show that Lexis diagrams are an age-period-cohort representation of mortality data similar to development triangles represent non-life insurance data.

Remark 2.4.1. In this section, we specified the nature of the triangle and considered incremental claim payments. We could derive the same conclusions for different triangles. A possibility would be $\mathcal{X}$ being an aggregation of incremental claim reportings. In this case, the development period j denotes the delay from accident to report. If we considered claim counts, then $\widehat{\mu}_{k,j}$, can be motivated as the estimator of an averaged hazard rate that runs in reversed development time, denoted $\mu_{k,j}$.

2.5 A model for development triangles

Having established the main results of our proposed modelling approach, we provide in this section further theoretical details underpinning our approach. Let us consider, for $k = 0, \ldots, m$; $j =$



(a) Development triangle. (b) Lexis diagram.

Figure 2.1: Development triangles (left hand side) and Lexis diagrams (right hand side) are common data representations in traditional actuarial mathematics. We leverage their similarities in our modelling framework. In particular, in non-life insurance we denote age as development, cohort as accident and period as calendar.

$1, \dots, m$, $\eta \in [0, 1]$, the following quantity

$$\widehat{\mu}_{k,j}(\eta) = \frac{X_{k,j}}{E_{k,j}(\eta)}, \tag{2.4}$$

where

$$E_{k,j}(\eta) = \sum_{l<j} X_{k,l} + \eta X_{k,j}.$$

Note that in the previous section we only considered the special case $\eta = \frac{1}{2}$, but here we make explicit that $\mu_{k,j}(\eta)$ depends on η. The factor η accounts for how much claims in (k, j) contribute to the exposure in (k, j); it is also known as lost exposure in mortality modelling (Wilmoth et al., 2021, p. 66). The value of η cannot be estimated without additional information and $\eta = \frac{1}{2}$ corresponds to assuming a uniform distribution of claims within a cell. Observe that we are not modelling $\hat{\mu}_{k,j}(\eta)$ for $j = 0$, because a run-off triangle (aggregated data) does not provide any information on the exposure in the first column: $E_{k0}(\eta)$ as fraction of $X_{k,j}$ is completely determined by η. To illustrate the rationality of modelling $\widehat{\mu}_{k,j}(\eta)$, let's assume $\mathcal{X}$ being claim payments and $X_{k,j}$ stemming from iid payments (Z_i, U_i, T_i), $i = 1, \dots, n$, where Z_i is the payment size, U_i accident date and T_i the delay between accident and payment. In Section 2.1, we discussed that $\widehat{\mu}_{k,j}(\eta)$ can be motivated as estimator of $\mu_{k,j}$, an exposure weighted average of the hazard rate in reversed development time $\alpha^*(t|u)$.

For $\eta = \frac{1}{2}$, there is a one-to-one relationship between chain-ladder's individual development factors, $\widehat{f}_{jk}$, and $\widehat{\mu}_{k,j}(\eta)$, see In Equation (2.1). An analogue relationship remains true for general η. For $k = 0, \dots; j = 1, \dots, m$, the individual development factors are defined as $\widehat{f}_{k,j} =$

$\sum_{l\leq j} X_{k,j}/\sum_{l<j} X_{k,l}$, we will show that

$$\widehat{f}_{k,j} = \frac{1+(1-\eta)\widehat{\mu}_{k,j}(\eta)}{1-\eta\widehat{\mu}_{k,j}(\eta)}.$$

Note that by definition the individual development factors do not depend on η, i.e. for two values $\eta, \eta', \eta \neq \eta'$, $\widehat{f}_{k,j} = \{1+(1-\eta)\widehat{\mu}_{k,j}(\eta)\}/\{1-\eta\widehat{\mu}_{k,j}(\eta)\} = \{1+(1-\eta')\widehat{\mu}_{k,j}(\eta')\}/\{1-\eta'\widehat{\mu}_{k,j}(\eta')\} = \sum_{l\leq j} X_{k,j}/\sum_{l<j} X_{k,l}$. However this does not mean that predictions, as will be defined in Section 2.7.2, are not affected by η. The reason for this is that predictions will be based on some predicted development factors $\widetilde{f}_{k,j}$ rather than on $\widehat{f}_{k,j}$ which will in turn be calculated from some predicted claim development $\tilde{\mu}_{k,j}$ rather than $\hat{\mu}_{k,j}(\eta)$. The invariance of $\widehat{f}_{k,j}$ to different choices of η gives hope that $\widetilde{f}_{k,j}$ is not affected much by η. Indeed, in the cases we investigated, we found that changing the value of η only lead to minor changes in the predicted development factors and overall predictions. Therefore, we conjecture that the actual choice of η is of minor importance.

Proof of the correspondence between development factors and average hazard rate.

$$\widehat{f}_{k,j} = \frac{1+(1-\eta)\widehat{\mu}_{k,j}(\eta)}{1-\eta\widehat{\mu}_{k,j}(\eta)},$$

with $k = 0, \ldots, m$, $j = 1, \ldots, m$, and $\eta \in [0, 1]$.

The chain ladder individual development factors are defined as

$$\widehat{f}_{k,j} = \frac{\sum_{l\leq j} X_{k,l}}{\sum_{l<j} X_{k,l}} = 1 + \frac{X_{k,j}}{\sum_{l<j} X_{k,l}}, \tag{2.5}$$

and the estimator for the average (reverse time) hazard rate is

$$\hat{\mu}_{k,j}(\eta) = 1 + \frac{X_{k,j}}{\sum_{l<j} X_{k,l} + \eta X_{k,j}}. \tag{2.6}$$

We can rewrite Equation (2.6) as

$$\sum_{l<j} X_{k,l} = \frac{(1-\eta\widehat{\mu}_{k,j}(\eta))X_{k,j}}{\widehat{\mu}_{k,j}(\eta)}.$$

By plugging in $\sum_{l<j} X_{k,l}$ into (2.5), we obtain

$$\widehat{f}_{k,j} = 1 + \frac{\widehat{\mu}_{k,j}(\eta)}{1-\eta\widehat{\mu}_{k,j}(\eta)}.$$

Then, $\widehat{f}_{k,j}$ can then be simplified to

$$\widehat{f}_{k,j} = \frac{1+(1-\eta)\widehat{\mu}_{k,j}(\eta)}{1-\eta\widehat{\mu}_{k,j}(\eta)}.$$

□

2.5.1 Some practical remarks

One may ask why we propose to model $\mu_{k,j}$ and not the individual development factors $f_{k,j}$ or the actual hazard rate $\alpha^*(t|u)$. We discuss both alternatives consecutively:

- Why not model $f_{k,j}$? The individual development factors stem from a discrete world and are not defined in a continuous setting. The claim development $\mu_{k,j}$ however is up to a constant δ well defined in a continuous setting in the sense that $\mu_{k,j}/\delta$ converges to $\alpha^*(t|u)$ for the grid-size δ converging to zero. This is desirable from an abstract point of view but also from a modelling perspective because it enables to embed the claim development in a wider continuous modelling framework.
- Why not model the continuous hazard rate $\alpha^*(t|u)$ directly? Recall that we start from run-off triangles. While one could still impose a structure on $\alpha^*(t|u)$ first, the implications for $\mu_{k,j}$ are not easy to derive and may complicate things unnecessarily. As mentioned in the previous section, there is an interesting parallel to draw to the field of mortality modelling in actuarial science and demography where usually an averaged version of the hazard rate (force of mortality) is being modeled. The force of mortality is the central object for further modelling and interpretation and it is also used to estimate the conditional probability of dying, see e.g. the methods protocol for the Human Mortality Database (Wilmoth et al., 2021). In a similarly fashion, we propose to estimate the development factors after modelling $\mu_{k,j}$.

2.6 A stochastic model replicating chain-ladder

Let us assume that for $k = 0, \dots, m$ and $j = 1, \dots, m$,

$$\mu_{k,j} = a_j. \qquad \text{[age-model]} \tag{2.7}$$

Furthermore, we assume that the entries $X_{k,j}$ are independent given $E_{k,j}$ and for $k = 0, \dots, m; j = 1, \dots, m$, they follow a Poisson distribution:

$$X_{k,j}|E_{k,j} \sim \text{Pois}(O_{k,j}), \quad O_{k,j} := E_{k,j}\mu_{k,j}.$$

Remark 2.6.1. Analogue to what is done in the classical GLM reserving literature, one can change the Poisson assumption to an overdispersed Poisson assumption without altering the point estimates, see McCullagh and Nelder (2019, p. 323).

This model, which we will refer to as the age-model, assumes that the claim development depends only on age, that is, on the development period of the claim (recall Table 1.1). Then the log-likelihood is given by

$$l\left(a_1, \dots a_m \mid X_{k,j}, E_{k,j}, j = 1, \dots, m; j + k \leq m\right) \propto \sum_{j,k} X_{k,j} \log(a_j E_{k,j}) - E_{k,j} a_j,$$

leading to the first order condition

$$\sum_k \frac{X_{k,j}}{\widehat{a}_j} - E_{k,j} = 0,$$

with minimizer

$$\widetilde{\mu}_{k,j} = \widehat{a}_j = \frac{\sum_k X_{k,j}}{\sum_k E_{k,j}}. \tag{2.8}$$

In particular, for $j = 1, \dots, m$,

$$\tilde{f}_j := \frac{1 + (1-\eta)\widehat{a}_j}{1 - \eta\widehat{a}_j} = \frac{\sum_{l \le j} \sum_{k=0}^{m-j} X_{k,l}}{\sum_{l<j} \sum_{k=0}^{m-j} X_{k,l}} = f_j^{\text{chain-ladder}}, \tag{2.9}$$

showing how the age-model replicates chain-ladder's development factors for any choice of η. The proof can be found below.

Hence, the age-model is a special case where the predicted development factors do not depend on η. From now on, we will assume that claims in the parallelograms occur uniformly, using $\eta = \frac{1}{2}$.

Proof of the age-model replicating the chain-ladder. Under the assumptions in Section 2.6, plugging Equation (2.8) into Equation (2.9), we obtain

$$\widehat{f}_j = \frac{\sum_{k=0}^{m-j} \sum_{l<j} X_{k,l} + \eta \sum_{k=0}^{m-j} X_{k,j} + \sum_{k=0}^{m-j} X_{k,j} - \eta \sum_{k=0}^{m-j} X_{k,j}}{\sum_{k=0}^{m-j} \sum_{l<j} X_{k,l} + \eta \sum_{k=0}^{m-j} X_{k,j} - \eta \sum_{k=0}^{m-j} X_{k,j}}$$

which simplifies to the chain ladder development factors

$$\widehat{f}_j = \frac{\sum_{k=0}^{m-j} \sum_{l \le j} X_{k,l}}{\sum_{k=0}^{m-j} \sum_{l<j} X_{k,l}}$$

□

Remark 2.6.2. Note that minimizing the Poisson likelihood leads to the same minimizer as minimizing the weighted least squares

$$\sum_{j,k} E_{k,j} (\hat{\mu}_{k,j} - \hat{a}_j)^2.$$

2.7 Further models

In the previous sections, and in contrast to existing literature, we modeled the claim development and not the claim amount. A generalized framework for claim amounts via an age-period-cohort construction has been considered in Kuang et al. (2008a, 2008b), Harnau and Nielsen (2018), and Kuang and Nielsen (2020). Apart from its simplicity, one further advantage of our claim development formulation is that we can exploit the connection with mortality rate modelling and extend Equation (2.7) to any model from the family of age-period-cohort (APC) stochastic hazard rate models. We now assume that the claim development has the following form:

$$\log(\mu_{k,j}) = a_j + c_{k+j} + g_k. \tag{2.10}$$

Note that by some misuse of notation, a_j now denotes the age component on the log-scale while in the previous section it was the age component on the original scale. The logarithmic link function, connects the estimator for the hazard $\mu_{k,j}$ (our response variable) to the effects specified in the linear equation on the right-hand side. This equation defines a APC model for the claims development. By doing so, it captures the following components:

- a_j is the age effect (development period) on $\mu_{k,j}$.
- c_{k+j} is the period effect (calendar date) effect on $\mu_{k,j}$.
- g_k is the cohort effect (accident date) on $\mu_{k,j}$.

The theory for age-period-cohort models given our triangular structure is well understood (Kuang et al., 2008b, 2008a). In Table 2.1 we present the effects for the APC models.

APC models are identifiable up to an identification constraint. We show in column four of Table 2.1 the default identification constraints that we used in this paper and in our `clmplus` package. The age-model is identifiable without additional constraints on the (categorical) effect a_j as the linear predictor of the Poisson regression model in Equation (2.10) contains no intercept. For the age-cohort model, we set $g_0 = 0$. For the age-period model, we set $c_1 = 0$. The age-period-cohort model is invariant with respect to the following two parameter transformations:

$$\begin{aligned}(a_j, c_{k+j}, g_k) &\to (a_j + \phi_1 - j\phi_2, c_{k+j} + (k+j)\phi_2, g_k - \phi_1 - k\phi_2)\,,\\ (a_j, c_{k+j}, g_k) &\to (a_j + \psi_1, c_{k+j} - \psi_1, g_k)\,,\end{aligned}$$

with $\phi_1, \phi_2, \psi_1 \in \mathbb{R}$ (Kuang et al., 2008b). Following Haberman and Renshaw (2011), we impose

$$\sum_{s=1}^{m} c_s = 0, \quad \sum_{k=0}^{m-1} g_k = 0, \quad \sum_{k=0}^{m-1} k g_k = 0.$$

The constraints on the cohort effect ($\sum_k g_k = 0, \sum_k k g_k = 0$) imply that g_k, for $k = 0, \ldots, m-1$, fluctuates around zero without any trend.

Table 2.1: In the first column we cite the model label. The second columns provides a description of the Lexis diagram dimensions: age is development period, cohort is accident period and period is calendar time. The effects are displayed in the third columns. In the last column we add the identification constraints on the effects displayed in column three.

`clmplus` (short)	Lexis dimensions	Effects	default identification constraints
a	age (chain-ladder model)	a_j	-
ac	age–cohort	$a_j + g_k$	$g_0 = 0$
ap	age–period	$a_j + c_{k+j}$	$c_1 = 0$
apc	age–period–cohort	$a_j + c_{k+j} + g_k$	$\sum_{s=1}^{m} c_s = \sum_{k=0}^{m-1} g_k = \sum_{k=0}^{m-1} k g_k = 0$

2.7.1 Extrapolation of cohort and period effects

As discussed in Kuang et al. (2008a), it is necessary to have a linear trend or a random walk with a drift to have identification invariant forecasts on the lower triangle. The dynamics of the predicted development factors when using hazard models with a cohort effect will depend on g_k with $k = 0, \ldots, m$. Having no information about the exposure in $j = 0$, we only have predictions $\hat{g}_1, \ldots \hat{g}_{m-1}$. Hence, one needs to extrapolate the cohort effect for the last row, $\widehat{g}_m$. In the case studies that we will present, we will show the results from the models in Table 2.1. Following previous mortality studies (Haberman & Renshaw, 2009; Lovász, 2011), we assume that the cohort component follows an ARIMA (1,1,0) model with drift ν_0:

$$g_k = \nu_0 + g_{k-1} + \phi(g_{k-1} - g_{k-2}) + \xi_k, \qquad \xi_k \sim N(0, \sigma), \tag{2.11}$$

with $\phi \in \mathbb{R}$. Similarly to Haberman and Renshaw (2009); Lovász (2011), for the models in Table 2.1 that include a period component we will need to extrapolate the period effect for the future calendar years, $\widehat{c}_{m+s}, s = 1, \ldots, m-1$. In the next sections we will model the period effects as a random walk with drift ν_1:

$$c_{k+j} = \nu_1 + c_{k+j-1} + \xi_{k+j} \qquad \xi_{k+j}, \sim N(0, \sigma). \tag{2.12}$$

2.7.2 Prediction

Based on Equation (2.10), fitted values are given by

$$\tilde{\mu}_{k,j} = \exp\left(\widehat{a}_j + \widehat{c}_{k+j} + \widehat{g}_k\right).$$

Thus, for the upper triangle, $0 \leq (k, j)$ with $k + j \leq m$, we define the fit as

$$\widehat{X}_{k,j} = E_{k,j}\tilde{\mu}_{k,j}.$$

If one wishes to get predictions for the the lower triangle, $k + j > m$, an obstacle is that the exposure ($E_{k,j}$) is not observed there. A solution is to use the chain principle known from the chain-ladder method. To this end, define the predicted development factors for accident year $k = 0, \ldots, m$ and development year $j = 1, \ldots, m$ as,

$$\tilde{f}_{k,j} = \frac{(2 + \tilde{\mu}_{k,j})}{(2 - \tilde{\mu}_{k,j})}. \tag{2.13}$$

Note that this is analogue to what is done in mortality prediction when the fit for the force of mortality is transformed to conditional survival or death probabilities (cf. Equations 2.2 and 2.2). Next, define the cumulative data as

$$C_{k,j} = \sum_{s \leq j} X_{k,s}.$$

An estimate for the lower triangle ($k + j > m$) is derived via Equation (1.1). The claims reserve

is just the difference between the ultimate cost for the cohort and the cumulative payment available on the last calendar year for the same cohort.

2.8 Data application

In this section we illustrate our framework using the `AutoBI` dataset available in the `R` package `ChainLadder` (Gesmann et al., 2022). For completeness, we include this dataset in Table 2.2.

Table 2.2: Cumulative paid claims of a portfolio of automobile bodily injury liability for an experience period of 1969 to 1976.

	0	1	2	3	4	5	6	7
0	1904	5398	7496	8882	9712	10071	10199	10256
1	2235	6261	8691	10443	11346	11754	12031	
2	2441	7348	10662	12655	13748	14235		
3	2503	8173	11810	14176	15383			
4	2838	8712	12728	15278				
5	2405	7858	11771					
6	2759	9182						
7	2801							

Moreover, Appendix A.1 shows how to use the `clmplus` package to replicate some of the results in this section.

Using the `AutoBI` dataset, we first show that both the GLM approach based on claim amounts (England & Verrall, 1999) and the GLM based on claim developments can replicate chain-ladder. In the latter case, as discussed earlier, this is achieved by specifying the claim development $\mu_{k,j}$ with the age-model in Equation (2.7) and exploiting the relation in Equation (2.9) together with the chain-ladder algorithm to forecast the claims development. After that, we will show the improvement that can be gained by adding a cohort component. Lastly, a discussion on models that require a period component will follow. In this second part, the different claim development fits are compared in terms of scaled deviance residuals on the development. The scaled deviance residuals r_{kj} for cohort k and age j are:

$$r_{k,j} = \text{sign}\left(X_{k,j} - \widehat{X}_{kj}\right)\sqrt{\frac{\text{dev}(k,j)(K-\nu)}{D}}$$

where:

$$\text{dev}(k,j) = 2\left[X_{k,j}\log\left(\frac{X_{k,j}}{\widehat{X}_{kj}}\right) - \left(X_{k,j} - \widehat{X}_{k,j}\right)\right],$$

$D = \sum_k \sum_j \text{dev}(k,j)$, K is the number of observations, and ν the number of parameters in the model. The ratio $D/(K-\nu)$ estimates the dispersion parameter. By providing a heat-map of the scaled deviance residuals $r_{k,j}$ on the upper triangle any undesirable pattern can be detected and removed.

2.8.1 The chain-ladder model

Table 2.3 shows the results of applying three methods to estimate the claims reserve in the `AutoBI` dataset, namely, the chain-ladder reserve as implemented in the `ChainLadder package`, the age-cohort GLM approach based on claim amounts implemented in the `apc` package (Fannon & Nielsen, 2020), and the age-model (Equation (2.7)) within our proposed framework that models the claim development as implemented in our `clmplus` package. Here, we see that all three approaches yield the same reserve estimates in each cohort.

Table 2.3: Comparison between the reserve from the chain-ladder method, the GLM approach based on claim amounts and our proposal, `clmplus`, that models the claim development via a GLM.

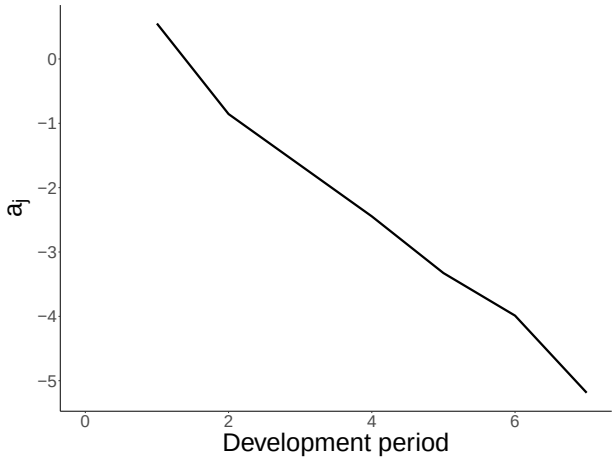
accident year	chain-ladder -	GLM(claim amounts) age-cohort-model	`clmplus` age-model
0	0.00	0.00	0.00
1	67.24	67.24	67.24
2	345.19	345.19	345.19
3	940.69	940.69	940.69
4	2350.86	2350.86	2350.86
5	4466.77	4466.77	4466.77
6	9103.24	9103.24	9103.24
7	14480.44	14480.44	14480.44

In Figure 2.2 we display the parameters estimates for the three different methods, i.e. $\widehat{a}_j$ for `clmplus`, $\widehat{f}_j$ for chain-ladder, and "age-effect" & "cohort effect" for GLM on claim amounts.

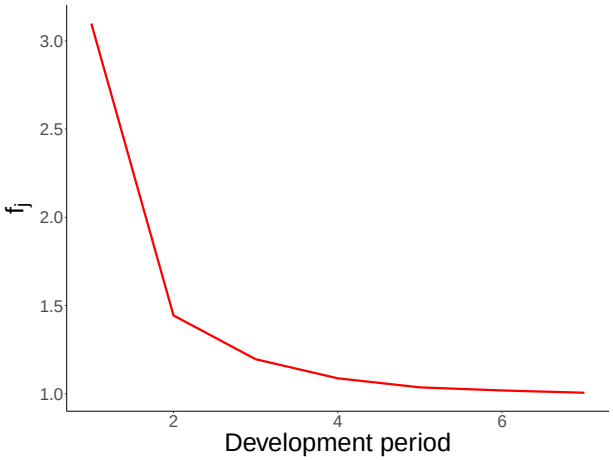
In Figure 2.2a and Figure 2.2b we see, respectively, a decreasing behavior in the age-effect for the `clmplus` and chain-ladder estimates.

The parameters estimates for the GLM approach on claim amounts are displayed in Figure 2.2c and Figure 2.2d.

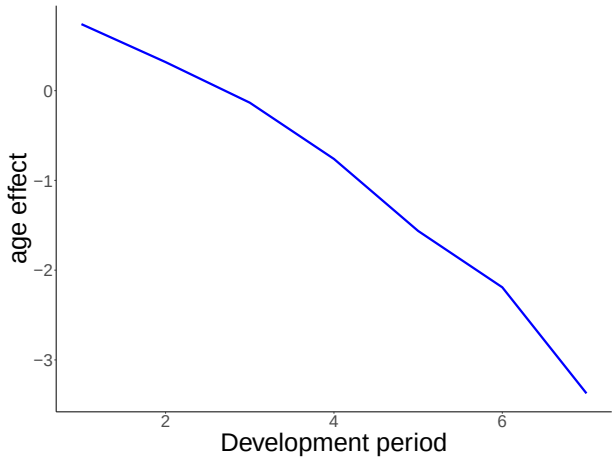
Figure 2.2: Results from the different models fitted on the `AutoBI` dataset. From left to right: the fit of the a_j effect in Equation (2.7), the chain-ladder development factors in Equation (2.9), as well as the age effect and the cohort effect when modelling the claim amount. It follows from the previous sections that the results in Figure 2.2b can be obtain from those in Figure 2.2a with the transformation in Equation (2.9).



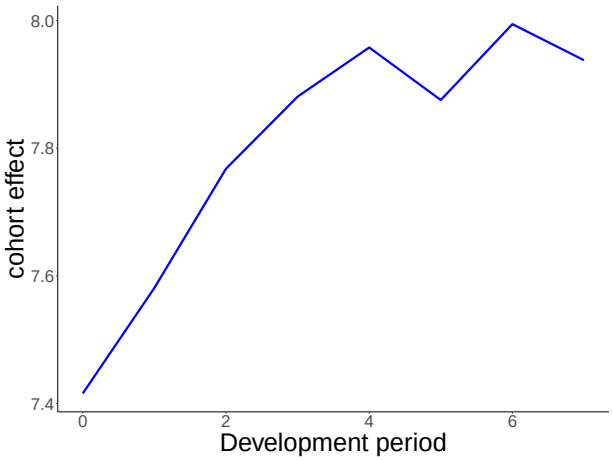
(a) Age effect plot from the age-model chain-ladder plus fit.



(b) Development factors estimated from the chain-ladder method.



(c) Age effect estimated from the GLM approach modelling the claim amount in a cross-classified ODP approach (England & Verrall, 1999). The age effect in the first development period is set to zero to make the model identifiable.



(d) Cohort effect estimated from the GLM approach modelling the claim amount in a cross-classified ODP approach (England & Verrall, 1999).

2.8.2 The benefit of adding the cohort component

It is well known and also seen in the previous section that a GLM with claim amount as response and an age component and a cohort component as predictors can replicate the chain-ladder estimates. We have also seen that we can provide the same results by modelling the claim development via an age component only. This is desirable both from a statistical and a practical perspectives. Indeed, we are able to model less parameters to obtain the chain-ladder reserve and could use the cohort effect as a potential additional improvement to the model fit. To illustrate this, in Figure 2.4 we show heat-maps of scaled deviance residuals for the models in Table 2.1 fitted to the `AutoBI` dataset. For the age-only model in Figure 2.4a, we identify two residuals clusters on the heat-map.

Conversely, the estimates obtained from the age-cohort model in Figure 2.4b show that with no need of adding a period component, we are able to improve the residuals on the fit. Remember that one only needs to extrapolate one point if modelling a cohort effect, while one needs to extrapolate m points ahead if modelling a period effect. The result from the $\widehat{g}_m$ effect extrapolation using an ARIMA(1,1,0) as described in previous sections is displayed in Figure 2.3.

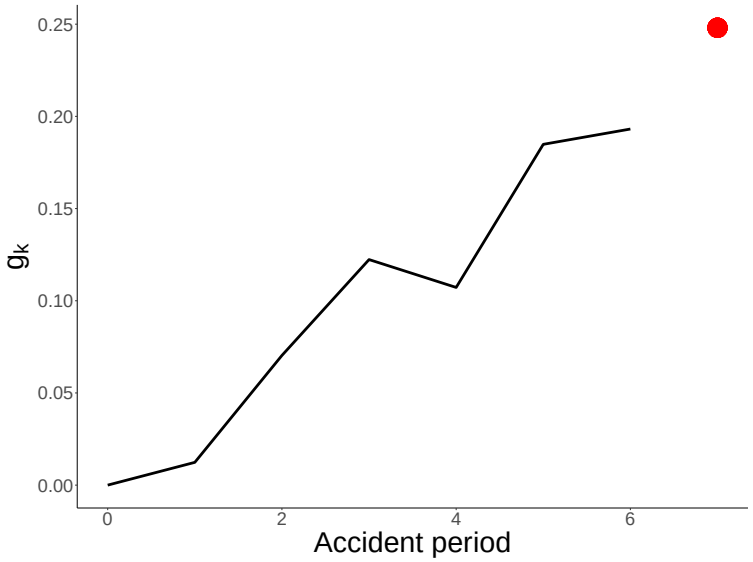
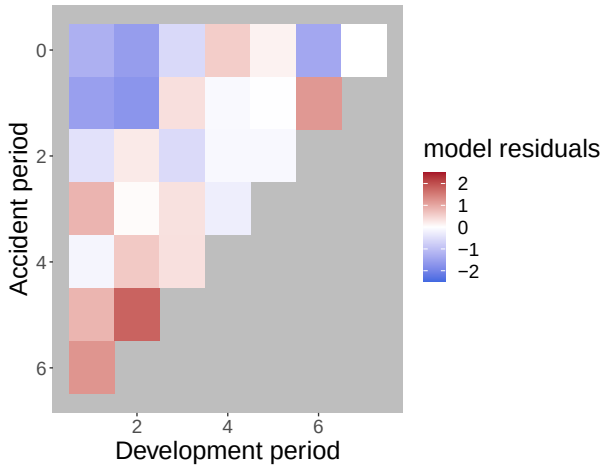


Figure 2.3: Cohort component g_k extrapolated with an ARIMA model (1,1,0) with drift, red dot.

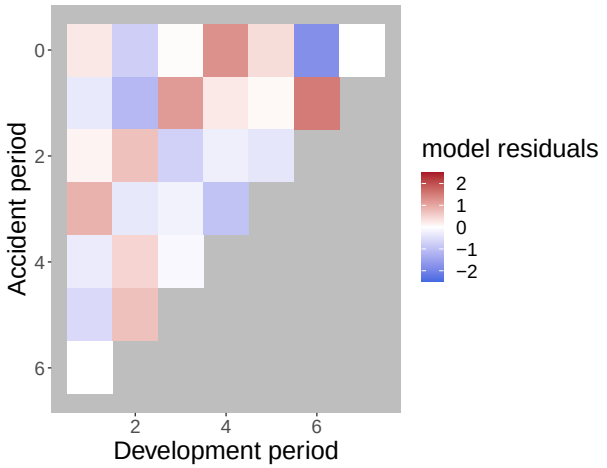
2.8.3 Extrapolating a period component

The flexibility of the chain-ladder plus framework also allows to add a period effect to the claim development; for example via an age-period model or an age-period-cohort model. For the `AutoBI` dataset, Figure 2.4c suggests that there is no clear improvement on the residuals moving from an age-cohort to the age-period model. By contrast, Figure 2.4d indicates that there is a slight improvement when we add a period component to obtain a full age-period-cohort specification.

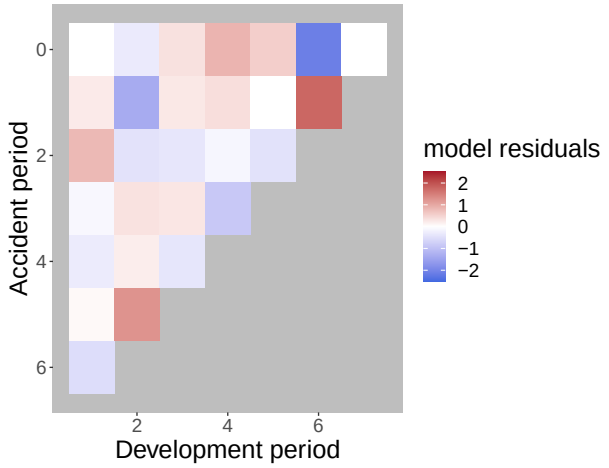
Figure 2.4: Scaled deviance residuals for hazard of the age-model, the age-cohort model, the age-period model and the age-period-cohort model.



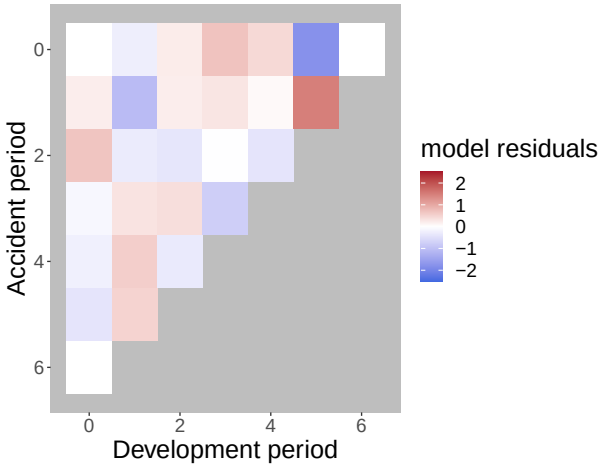
(a) Residuals plot from the age model.



(b) Residuals plot for the age–cohort model.



(c) Residuals plot for the age–period–model.

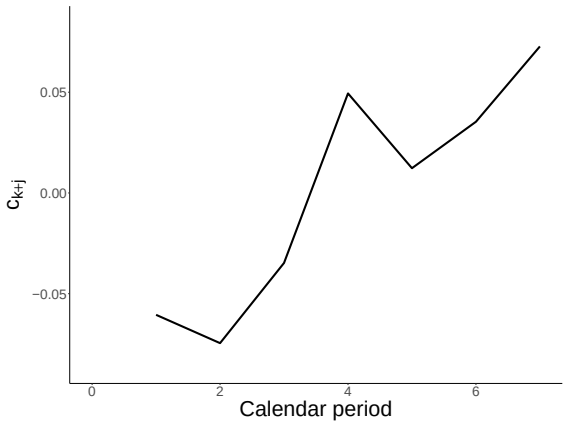


(d) Residuals plot for the age–period–cohort model.

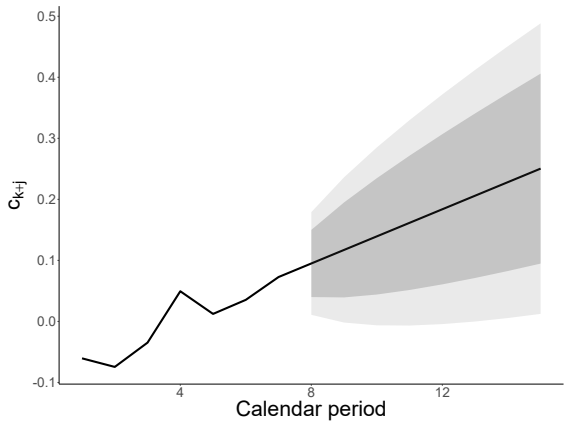
However, we observe once again that additional care is required anytime a period model is extrapolated. In particular, one should be careful in choosing the most suitable extrapolation method for the calendar year component c_{k+j}: this must be a data driven choice.

Figure 2.5 shows the c_{k+j} component in an age-period-cohort fit on the upper triangle (Figure 2.5a) and the extrapolation in the lower triangle (Figure 2.5b). For this figure, we fitted the age-period-cohort model with the constraints $\sum_k g_k = 0$ and $\sum_k k g_k = 0$. Furthermore, we assumed here and in the following sections that the period components follow a random walk with drift for forecasting.

Figure 2.5: Period component extrapolated for the age-period-cohort hazard model on the `AutoBI` dataset.



(a) c_{k+j} fitted on the data. Results on the upper triangle.



(b) c_{k+j} component extrapolated for the lower triangle. We display the 80% and 95% confidence bounds.

Lastly, in Table 2.4 we show the different reserves estimates based on `clmplus` for different cohorts according to the different models we considered in this section. In this example, `ac`, `ap` and `apc` produce more similar results compared to the `a` model. In general, we can see that differences are more pronounced in the later accident years.

Table 2.4: Reserves computed according to different claim development models.

accident year	a	ac	ap	apc
1	0.00	0.00	0.00	0.00
2	67.24	68.20	68.72	68.54
3	345.19	361.77	358.22	359.35
4	940.69	1009.65	992.50	996.34
5	2350.86	2476.54	2503.56	2505.20
6	4466.77	4968.70	4845.14	5006.93
7	9103.24	10052.81	10229.09	10029.15
8	14480.44	19188.40	18377.78	19533.02
Total	31754.43	38126.05	37375.01	38498.54

2.9 Models comparison

While in Section 2.8 we showed a sensible strategy for model selection on a single run-off triangle, within this section we want to assess our model performance on several real datasets. We gathered 30 real and publicly available run-off triangles from the R packages `ChainLadder`, `apc`, `clmplus` and `CASdatasets` (Dutang & Charpentier, 2020). The complete list of the triangles we used is provided in Appendix B. Our objective is to show that practitioners can really benefit from the chain-ladder plus model framework as an additional more versatile tool-box to the already rich literature on claims reserving. After a short discussion on model selection, we compare the performances of chain-ladder plus models in Table 2.1 using the `clmplus` R-package to the performances of the age-cohort and age-period-cohort GLM based on claim amounts using the `apc` R-package, as outlined

in Kuang et al. (2011) and implemented in Fannon and Nielsen (2020).

For model comparison, it is not straightforward to select a good measure of fit.

One issue is that the mean squared error or the mean absolute error would not be comparable across different triangles that exhibit different payments size. Another issue is related to the time series structure of run-off triangles. It is worth noticing again that anytime the claims reserve is computed, the results on the lower triangle are extrapolated. In order to assess the models capability to extrapolate, a reasonable approach is to use the most recent diagonals as test set.

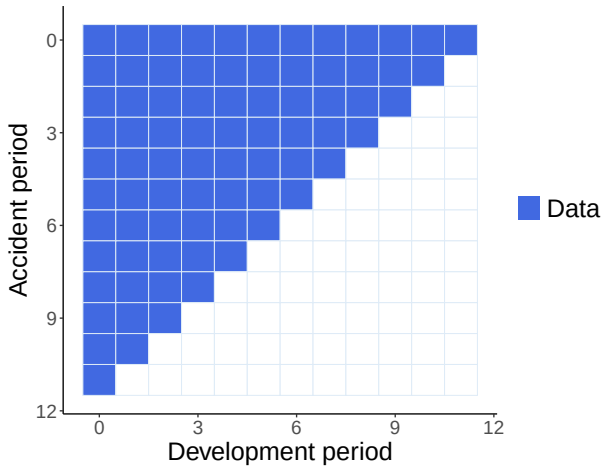
Thus, we will evaluate the performance of different models across diagonals in terms of absolute errors incidence (EI) on a selected diagonal,

$$\mathrm{EI}_j = \left| \frac{\sum_{k+j=m} \widehat{X}_{k,j} - X_{k,j}}{\sum_{k+j \leq m} X_{k,j}} \right|.$$

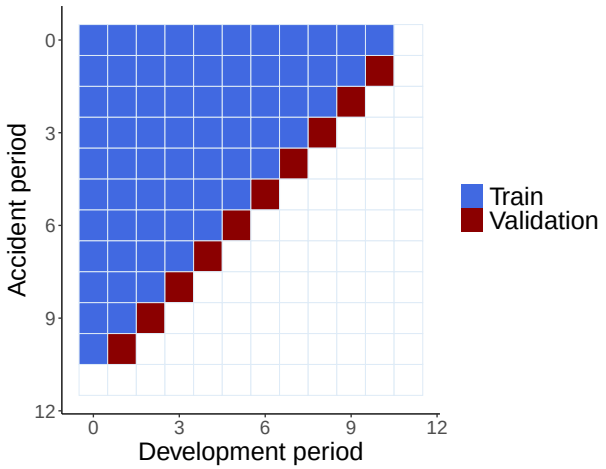
2.9.1 Models ranking

In order to rank the models performances, we evaluated on the calendar year m the best performing model in terms of error incidence over the cumulative payments diagonal. An example for a 12x12 run-off triangle is provided in Figure 2.6.

Figure 2.6: The last diagonal is removed from the triangle and it is used as a validation set to calculate the models rank.



(a) The starting triangle.



(b) The triangle is divided into validation set and training set.

On each dataset, the models were first trained on the training set, see the blue area in Figure 2.6b. The error incidence was then recorded for each model on the last diagonal, see the red area in Figure 2.6b. The models where then ordered in ascending order: the model with the lowest absolute error incidence was selected to be the best performing. Figure 2.7 shows ranking of models for each of the 30 datasets.

To facilitate comparison, Table 2.5 shows the mean rank of each model across the 30 different datasets. In Figure 2.7 and Table 2.5, it is worth noticing once again that the age-cohort model in England and Verrall (1999) and the chain-ladder plus age-model yield the same results on every dataset. In addition, the chain-ladder plus models tend to show the best performance, with particular reference to the age-period and age-period-cohort structure.

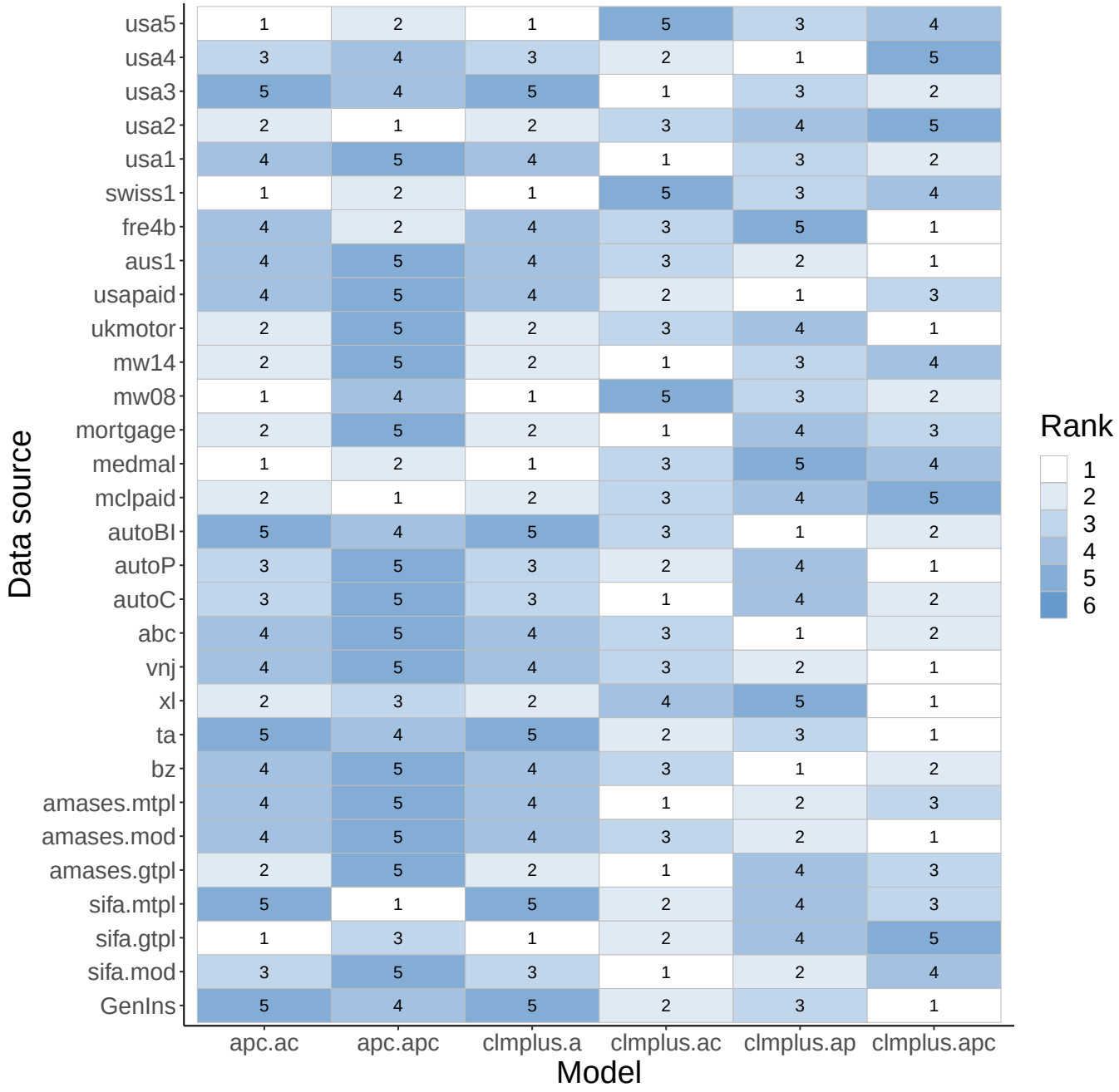


Figure 2.7: Models ranks by dataset.

Model	apc.ac	apc.apc	clmplus.a	clmplus.ac	clmplus.ap	clmplus.apc
Mean rank	3.07	3.87	3.07	2.47	3.00	2.60

Table 2.5: Mean rank for the different models.

2.9.2 Model families comparison

In this section we evaluate whether one would benefit from having at disposal a larger set of models to choose from when carrying out a reserving exercise. To do so, we compare the performance of three sets of potential models on each of the 30 datasets used in the previous section. These sets

of potential models are:

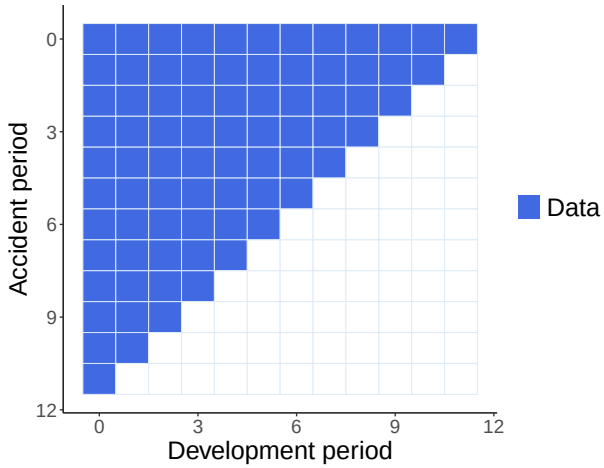
(a) the `apc` package family that models the claim amount,

(b) the `clmplus` package family that models the claim development, and

(c) the union of (a) and (b).

To evaluate the performance of each set we start by splitting the data into training, validation and testing as illustrated in Figure 2.8.

Figure 2.8: The last two diagonals are removed from the triangle and they are used as a test set to evaluate the models performance.



(a) The starting triangle.

(b) The triangle is divided into training set, validation set and test set.

Then, for each dataset the best model within the three different model sets is selected based on the validation set, and, finally, the error incidence (EI) of this best model is then calculated on the test set.

Figure 2.9 shows box plots of the EI for the 30 datasets and each of the three sets of models. We find that the `clmplus` package family (b) has on average a better performance than the `apc` package family (a). However, more importantly, having all models at disposal (set (c)) allows to further improve the forecasting accuracy, demonstrating that it is useful to expand the set of models available in the reserving practice toolkit. It is especially reassuring that picking a model via a validation set seems to generally lead to improved one-year-ahead predictions.

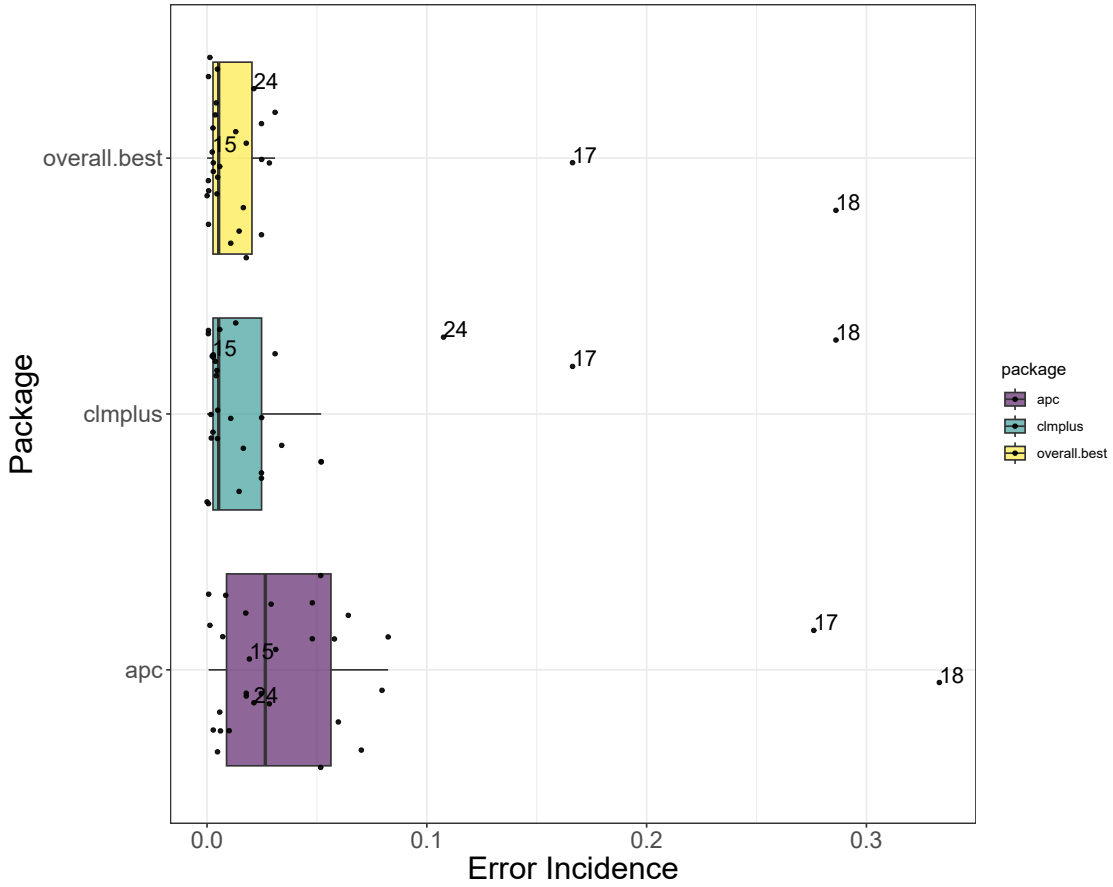


Figure 2.9: Models EI on the test set across the 30 datasets. On each dataset we selected the best performing model via a validation set from the three families `overall.best`, `clmplus` and `apc`.

For illustrative purposes, we marked data sets $15, 17, 18, 24$ in Figure 2.9. In data set 24, using the best model from the `apc` family, considerably reduced the EI compared to picking the best model from the `clmplus` family. The opposite occurs for data set 15.

Chapter 3

Estimation of the claim development using individual data

In this chapter, we propose a model for forecasting IBNR (Incurred But Not Reported) frequencies, refining the reverse-time framework that we introduced in Section 2.2 to include covariates. For the estimation of the hazard rates, we leverage individual claims data. Individual data include accident date, reporting delay, and possibly additional features for every reported claim. A key element of our proposal involves again computing development factors, which may be influenced by both the accident date and other features. These development factors serve as the basis for predictions. While we assume close to continuous observations of accident date and reporting delay, the development factors can be expressed at any level of granularity, such as months, quarters, or year and predictions across different granularity levels exhibit coherence. The calculation of development factors relies on the estimation of a hazard function in reverse development time, and we present three distinct methods for estimating this function: the Cox proportional hazard model, a feed-forward neural network, and XGBoost (eXtreme Gradient Boosting). In all three cases, estimation is based on the same partial likelihood that accommodates left truncation and ties in the data. While the first case is a semi-parametric model that assumes in parts a log-linear structure, the two machine learning approaches only assume that the baseline and the other factors are multiplicatively separable. Through an extensive simulation study and real-world data application, our approach demonstrates promising results. It also allows a seamless comparison with estimations obtained through the chain ladder approach.

3.1 Motivation: prediction of future IBNR

IBNR (Incurred But Not Reported) refers to outstanding claims for which the insurer is liable but which have not been reported at the time the reserve is calculated. Empirical studies on insurance markets show that in many lines of business the estimated cost of IBNR claims is the most important provision for the insurer, see for example Friedland (2010, p. 80). As the number of IBNR claims is not known at the time the reserve is calculated, there is a strong actuarial argument in favour of methods that accurately predict the number of IBNR claims.

In the seminal paper Miranda et al. (2013), the authors discuss how the common run-off triangles encountered in loss reserving in non-life insurance can be understood within a continuous framework

where the goal is to estimate the distribution in the lower triangle. Within this continuous chain-ladder framework, two lines of research emerged: One aims to estimate the underlying density function and with this respect, Lee et al. (2015), Lee et al. (2017), and Mammen et al. (2021) generalize the initial model to account for seasonal effects, operation al time and calendar effects, respectively. In the other line of research, Hiabu et al. (2016) establishes that the observation scheme of a continuous run-off triangle can be understood as right-truncation problem and that by reversing the development time statistical analysis can be conducted under a tractable left-truncation setting. Building on this work, Hiabu (2017) discusses how the hazard function in reverse development time is related to the omnipresent development factors. Hiabu, Mammen, Martínez-Miranda, and Nielsen (2021) extends the framework by allowing for accident date effects and other features via a multiplicative structure. Bischofberger et al. (2020) extends the prediction of claim frequencies to the prediction of claim payments. A simulation study comparing the two lines of research can be found in Bischofberger, Hiabu, Mammen, and Nielsen (2019).

A major drawback of both lines of research so far is that they have not been very practical: estimation is based on kernel smoothers that without further adjustment do not cope well with the sharp patterns often seen in claim developments. Additionally, it is not directly clear how to include categorical features. In fact, with the exception of Hiabu et al. (2021) that allows for continuous features, none of the other work so far considers additional features, neither continuous nor categorical. A limiting factor is that if no structure is imposed a priori, kernel smoother will necessarily suffer from the curse of dimensionality, i.e. exponentially deteriorating estimation performance with every continuous feature added. Machine learning methods on the other hand have shown that they are capable of data-driven dimension reduction and making use of the underlying data structure in order to circumvent the curse of dimensionality. With these considerations, and within the second line of research, we propose three methods to estimate an accident date and other features dependent hazard function: the Cox proportional hazard model, a feed-forward neural network and XGBoost.

In all of the three proposed methods, estimation will be based on the same partial likelihood that accommodates left truncation and ties in the data, see Section 3.3. While there are more and more survival analysis solutions being implemented for common machine learning methods, to the best of our knowledge, neither of the publicly available feed-forward neural network solution nor XGBoost solutions have implementations that can deal with left-truncated data and ties in the data (Wiegrebe, Kopper, Sonabend, & Bender, 2023). Hence, to make our proposal work, we will extend current XGBoost and feed-forward neural networks solutions such that they can handle left truncation and ties under the assumption that ties correspond to intervals in which events happen uniformly.

As mentioned before, we will use the estimated hazard function to calculate development factors which we will subsequently use for prediction in Section 3.4. With this respect, our proposal is related to the approach of Wüthrich (2018) that uses a feed-forward neural network to estimate the development factors directly. One major advantage of our survival analysis approach is that we do not have problems with zero entries in the cumulative run-off triangles. In Wüthrich (2018), the loss function fed into the neural network entails dividing by each cumulative entry, see equation (3.1) in that paper. While zero entries are not common in typical run-off triangles, zero entries are expected to happen often if the features take too many different values resulting in many sparse

triangles. In particular, the approach of Wüthrich (2018) does not allow for continuous features. To circumvent the problem of some zero entries with discrete or categorical features, Wüthrich (2018) proposes some adhoc method that ignores features in those entries. In contrast to that, in our approach we do not have those problems with zero entries and we are able to estimate a conditional hazard function which is possibly dependent on high dimensional feature information.

In a recent preprint, Calcetero-Vanegas, Badescu, and Lin (2023) aim to estimate the same hazard function as we propose to estimate. However, they only apply the Cox-proportional hazard model and not further machine learning methods as we do. Furthermore, they use an inverse probability weighting for prediction. In contrast, we propose to transform the hazard function into development factors which are subsequently used for prediction. While both approaches will produce similar predictions, one advantage of our proposal is the ease of comparison with standard reserving methods based on development factors; see Section 3.6 and following.

3.2 Modelling

At a cut-off-date $\mathcal{T}$, we have observed n claim reports. For each claim i with $i = 1, \dots, n$, we are given the accident date U_i and the time delay from accident until report, T_i^R. Suppose that for each individual we have at our disposal a set of p measurements, i.e. the features $X_i \in \mathbb{R}^p$. We assume that

[A1] All reportings are independent.

Assumptions [A1] is made to simplify the mathematical derivations and we conjecture that it can be relaxed by considering weak dependency between reportings. Direct inference on T_i^R may lead to sampling bias, as T_i^R are observed only if the report happens before the cut-off date $\mathcal{T}$:

$$T_i^R \leq \mathcal{T} - U_i,$$

which is a right-truncation problem. A solution to the right-truncation problem is to reverse the time of the counting process leading to a tractable left-truncation problem (Ware & DeMets, 1976). Concretely, we target $\mathcal{T} - T_i^R$ instead of T_i^R such that the right truncation problem is now a left truncation problem ($\mathcal{T} - T_i^R \geq U_i$) with truncation variable U_i. We consider the development-time reversed counting processes

$$N_i'(t) = I(t \geq \mathcal{T} - T_i^R),$$

each with respect to the filtration $\mathcal{F}_{it} = \sigma\left(\left\{T_i^R - \mathcal{T} \leq s : \ s \leq t\right\} \cup \left\{U_i \leq s : s \leq t\right\} \cup \left\{X_i\right\} \cup \mathcal{N}\right)$, satisfying the *usual conditions* (Andersen et al., 1993, p. 60), and where $\mathcal{N}$ is the set of all zero probability events. In this paper, we will assume that U_i and T_i^R are encoded such that $0 \leq U_i \leq \mathcal{T}; 0 \leq T_i^R \leq \mathcal{T}$ and we are interested in predictions within the set $\{(t, u) : (t, u) \in [0, \mathcal{T}]^2\}$.

3.2.1 The intensity process

Assuming that the intensity λ_i' of the counting process exists and is piecewise continuous, we have

$$\lambda_i'(\mathcal{T}-t|U_i,X_i)=\lim_{h\downarrow 0}h^{-1}E\left[N_i'\left\{(\mathcal{T}-t+h)-\right\}-N_i'(\mathcal{T}-t)-\mid\mathcal{F}_{i,(\mathcal{T}-t)-}\right]=\alpha^R(t|U_i,X_i)Y_i(t), \tag{3.1}$$

where

$$\alpha^R(t|u,x)=\lim_{h\downarrow 0}h^{-1}P\left(T_i^R\in(t-h,t]|\ Y_i(t)=1,X_i=x_i,U_i=u\right),$$
$$Y_i(t)=I(T_i^R\leq t<\mathcal{T}-U_i).$$

Note that $Y_i(t)$ and $\alpha^R(t|U_i,X_i)$ correspond to the intensity $\lambda_i'(\mathcal{T}-t|U_i,X_i)$, meaning the development time input for $Y_i(t)$ and $\alpha^R(t|U_i,X_i)$ is not in reversed direction. The structure, $\lambda_i'(\mathcal{T}-t|U_i,X_i)=\alpha^R(t|U_i,X_i)Y_i(t)$, is called Aalen's multiplicative intensity model (Aalen, 1978), and enables nonparametric estimation and inference on the deterministic hazard function $\alpha^R(t|u,x)$. We propose to model the hazard function as

$$\alpha^R(t|u,x)=\alpha_0^R(t)e^{\phi(x,u;\theta)}, \tag{3.2}$$

where $\alpha_0^R(t)$ is called the baseline hazard and $e^{\phi(x,u;\theta)}$ is the risk score; a component that depends on the features X_i and the accident period U_i and some parameters θ. Assuming that the effects of t and u are multiplicatively separated allows us to have predictions for α^R in the lower triangle, $t>\mathcal{T}-u$, without extrapolation. In the next sections, we will discuss how to specify and model the log-risk function $\phi(x,u;\theta)$. We will consider three different models: the cox model (COX, Cox, 1972), neural networks (NN, Katzman et al., 2018) and gradient boosting (XGB, Friedman, 2001; Chen & Guestrin, 2016).

- In Cox (1972), the log-risk function is assumed to be linear, $\phi(x,u;\theta)=\theta^Tx+\theta_u u$, with $\theta\in\mathbb{R}^p$ and $\theta_u\in\mathbb{R}$. In this paper we will consider the more general log-risk function that includes splines for modelling continuous features.
- In NN, the parameter θ represents the weights of a feed-forward neural network.
- In XGB, the log-risk function is an ensemble of decision trees, i.e. functions piecewise constant on rectangles.

Remark 3.2.1. The notation that we use in this chapter is different compared to the notation in Chapter 2. Our focus in Chapter 2 was in fact modelling the delay between accident and payment that we denoted with the random variable T. In this chapter we will focus on the delay between accident and report and we denote it with the random variable T^R. In this respect, the interest of this chapter is modelling the intensity $\alpha^R(t|u,x)$ of the counting process of the that, differently from the previous chapter, will also dipend on the feature measurements X_i. We recall that Chapter 2 provided a strategy to model the (average) hazard rate $\mu_{k,j}$.

3.3 Estimation

3.3.1 Ties in the data

Some reporting delays T_i^R of different claims could be recorded with the same value, especially when the data records are not very granular. For example data with yearly, quarterly or even daily records will most certainly have multiple occurrences at the same reporting period. In survival analysis, the occurrence of multiple events at a given time point is known as a tie. In this section, we will adopt a general formulation of the problem to account for the incidental presence of ties in the data. We assume that for $i = 1, \dots, n$, reporting times T_i^R are one of

$$t^{(0)} < \dots < t^{(\ell)} < \dots < t^{(m)},$$

with $m \leq n$. The rank of the observations within the tie does not affect the calculations that we will explain in this section. For notational convenience, we assume that observations are ordered from lowest to highest reporting time: $T_i^R \leq T_j^R$ for $i < j$ with $i, j \in \{1, \dots, m\}$. For $\ell = 1, \dots, m$, let us define, at the generic time $t^{(\ell)}$, the exposure set

$$\mathcal{R}(t^{(\ell)}) = \{i \in \{1, \dots, n\} : Y_i(t^{(\ell)}) = 1\} = \left\{i \in \{1, \dots, n\} : T_i^R \leq t^{(\ell)};\ \mathcal{T} - U_i > t^{(\ell)}\right\}$$

and the occurrence set

$$\mathcal{O}(t^{(\ell)}) = \left\{i \in \{1, \dots, n\} : T_i^R = t^{(\ell)}\right\},$$

while we indicate with $O_\ell = \#\mathcal{O}(t^{(\ell)})$ the cardinality of the set $\mathcal{O}(t^{(\ell)})$.

To obtain an estimate of θ, we will first formulate the partial likelihood in a general form and then we will specify the log-risk function. The specifications we use (COX, NN, XGB) were briefly introduced in Section 3.2.1 and they are the main focus of the current section. While the current literature on machine learning for survival analysis is about defining a model for right-censored data (Chen & Guestrin, 2016; Katzman et al., 2018), we contribute by extending the neural network in Katzman et al. (2018) and the gradient boosting machine in Chen and Guestrin (2016) to define a framework for modelling left-truncated data with ties.

3.3.2 Partial likelihood

There are several approaches to account for the presence of ties in the data, see for instance Hosmer, Lemeshow, and May (2008, p. 85). We use the partial likelihood correction for ties presented in Efron (1977, Section 6, point g)

$$\mathcal{L}(X_1, \dots, X_n; U_1, \dots, U_n; \theta) = \prod_{\ell=1}^{m} \prod_{i \in \mathcal{O}(t^{(\ell)})} \frac{e^{\phi(X_i, U_i; \theta)}}{\sum_{k \in \mathcal{R}(t^{(\ell)})} e^{\phi(X_k, U_k; \theta)} - \frac{\psi_i(\ell)}{O_\ell} \sum_{s \in \mathcal{O}(t^{(\ell)})} e^{\phi(X_s, U_s; \theta)}}$$

where $\psi_i(\ell) = i - \sum_{q=1}^{\ell-1} O_q - 1$. Note that $\psi_i(\ell) \in \{0, \dots O_\ell - 1\}$. To ease the notation, let us denote the partial likelihood as $\mathcal{L}(\theta)$. The negative log-likelihood is

$$l(\theta) = -\log(\mathcal{L}(\theta)) = \sum_{\ell=1}^{m} \sum_{i \in \mathcal{O}(t^{(\ell)})} \log \left(\sum_{k \in \mathcal{R}(t^{(\ell)})} e^{\phi(X_k, U_k; \theta)} - \frac{\psi_i(\ell)}{O_\ell} \sum_{s \in \mathcal{O}(t^{(\ell)})} e^{\phi(X_s, U_s; \theta)} \right) - \phi\left(X_i, U_i; \theta\right). \tag{3.3}$$

Cox model (COX)

The Cox proportional model uses a linear function to specify $\phi\left(X, U; \theta\right) = \theta^T X + \theta_u U$, with $\theta \in \mathbb{R}^p$ and $\theta_u \in \mathbb{R}$. To model continuous features that avoid the linear scale, we follow the approach in Gray (1992) and introduce splines in the log-risk function (Eilers & Marx, 1996). Assuming that within the p features we have c features for the linear term $(X_1, \ldots, X_c)$ and $p - c$ features for the splines $(X_{c+1}, \ldots, X_p)$, the log-risk function is

$$\phi\left(X, U; \theta\right) = \sum_{l=1}^{c} \theta_l X_l + \sum_{s=c+1}^{p} \zeta_s\left(X_s\right) + \zeta_u\left(U\right),$$

where, for $V = U$ or X_l, $\zeta_v\left(V\right) = \sum_{k=0}^{\kappa_v} \beta_k^v B_k(V)$ and $\beta^v = \left(\beta_1^v, \ldots, \beta_{\kappa_v}^v\right) \in \mathbb{R}^{\kappa_v}$. Here, $B_k^v(V)$ are basis functions and $\kappa_v \in \mathbb{N}$ is the number of knots in the spline. The smoothing of U and the other $p-c$ features modeled with splines are controlled with an additional penalty term in the log-partial likelihood during the model fitting. In the fitting phase we minimize the penalized likelihood

$$l^p(\theta, \beta^{c+1}, \ldots, \beta^p, \beta^u) = l(\theta) + \frac{1}{2} \sum_s \rho_s \int \left[\zeta_s''(z)\right]^2 dz + \frac{1}{2} \rho_u \int \left[\zeta_u''(z)\right]^2 dz$$

with $\rho_{c+1}, \ldots, \rho_p, \rho_u$ being the parameter controlling the smoothing applied (no penalty for $\rho = 0$ and forcing the spline to a linear form for $\rho = +\infty$). Noting that $\int \left[\zeta_s''(z)\right]^2 dz$ is a quadratic form of β^s, for some definite positive matrix $\mathbf{P}_s$, the penalized likelihood can be rewritten as

$$l^p(\theta, \beta^{c+1}, \ldots, \beta^p, \beta^u) = l(\theta) + \frac{1}{2} \sum_{s=c+1}^{p} \rho_s \beta_s' \mathbf{P}_s \beta_s + \frac{1}{2} \rho_u \beta_u' \mathbf{P}_u \beta_u$$

We minimize $l^p(\theta, \beta^{c+1}, \ldots, \beta^p, \beta^u)$ for the spline parameters and the θ parameters. We fit the COX model using the `R` package `survival` (RStudio Team, 2020; Therneau, 2023), that allows us to specify the smoothing penalty terms and the number of knots in the splines.

Neural Networks (NN)

In this modelling approach, we extend the approach in Katzman et al. (2018) to model left-truncated data using the correction for ties in Efron (1977). We describe the log-risk function with a feed-forward neural network with a vector of parameters $\theta^{NN} \in \mathcal{P}$, being $\mathcal{P}$ the space of the possible neural network parameters (Goodfellow, Bengio, & Courville, 2016, p.168).

In the optimization phase, we minimize the regularized objective function

$$l\left(\theta^{NN}\right) + \rho\left(\epsilon \|\theta^{NN}\|_2^2 + (1-\epsilon)\|\theta^{NN}\|_1\right)$$

where the hyper parameters ρ, ϵ allow for the elastic penalty term and the p-norm is $\|\theta^{NN}\|_p =$

$\left(\sum_v \left|\theta_v^{NN}\right|^p\right)^{\frac{1}{p}}$ with $p \in \mathbb{R}, p \geq 1$, and v is the index of the parameters of the neural network.

eXtreme Gradient Boosting (XGB)

The derivation we present in this section is a modification of Liu et al. (2020), where the authors consider right censoring only and not left-truncation. In gradient boosting the log-risk function is an ensemble of $K \in \mathbb{N}$ functions,

$$\phi\left(X_i, U_i; \theta^{XGB}\right) = f_1(X_i, U_i; \theta_1) + \ldots + f_k(X_i, U_i; \theta_k) + \ldots + f_K(X_i, U_i; \theta_K)$$

with $f_k \in \mathcal{S}$, where $\mathcal{S}$ is the space of all possible CARTs (Hastie, Tibshirani, Friedman, & Friedman, 2009) and $\theta^{XGB} = [K, \theta_1, \ldots, \theta_K]$. Here, $\theta_1, \ldots, \theta_K$ is the number of splits in the k-th regression tree with $k = 1, \ldots, K$.

The XGBoost algorithm, needs as input the gradient and the second order derivatives of the negative log likelihood function (3.3) with respect to $\phi\left(X_i, U_i; \theta^{XGB}\right)$. The gradient is

$$g_i = e^{\phi\left(X_i, U_i; \theta^{XGB}\right)}\left(\upsilon_i - \iota_i\right) - 1,$$

and the second order derivative is

$$h_i = g_i + 1 - e^{2\phi\left(X_i, U_i; \theta^{XGB}\right)}\left(\gamma_i - \omega_i\right),$$

where for $T_i^R = t^{(L)}$ with $L \in \{1, \ldots, m\}$,

$$\upsilon_i = \sum_{l:\ T_i^R \leq t^{(\ell)} < \mathcal{T} - U_i} \sum_{j \in \mathcal{O}(t^{(\ell)})} \frac{1}{\sum_{k \in \mathcal{R}(t^{(\ell)})} e^{\phi(X_k, U_k; \theta^{XGB})} - \frac{\psi_j(\ell)}{O_\ell} \sum_{s \in \mathcal{O}(t^{(\ell)})} e^{\phi(X_s, U_s; \theta^{XGB})}},$$

$$\iota_i = \sum_{j \in \mathcal{O}(t^{(L)})} \frac{\psi_j(L)/O_L}{\sum_{k \in \mathcal{R}(t^{(L)})} e^{\phi(X_k, U_k; \theta^{XGB})} - \frac{\psi_j(L)}{O_L} \sum_{s \in \mathcal{O}(t^{(L)})} e^{\phi(X_s, U_s; \theta^{XGB})}},$$

$$\gamma_i = \sum_{l:\ T_i^R \leq t^{(\ell)} < \mathcal{T} - U_i} \sum_{j \in \mathcal{O}(t^{(\ell)})} \frac{1}{\left(\sum_{k \in \mathcal{R}(t^{(\ell)})} e^{\phi(X_k, U_k; \theta^{XGB})} - \frac{\psi_j(\ell)}{O_\ell} \sum_{s \in \mathcal{O}(t^{(\ell)})} e^{\phi(X_s, U_s; \theta^{XGB})}\right)^2},$$

$$\omega_i = \sum_{j \in \mathcal{O}(t^{(L)})} \frac{\left(1 - \left(1 - \frac{\psi_j(L)}{O_L}\right)^2\right)}{\left(\sum_{k \in \mathcal{R}(t^{(L)})} e^{\phi(X_k, U_k; \theta^{XGB})} - \frac{\psi_j(L)}{O_L} \sum_{s \in \mathcal{O}(t^{(L)})} e^{\phi(X_s, U_s; \theta^{XGB})}\right)^2}.$$

As illustrated in (Chen & Guestrin, 2016), XGB is an iterative algorithm, and at each iteration t, the current tree further split by optimizing the objective function

$$\sum_{i=1}^{n} \left[g_i f_t\left(X_i, U_i; \theta_t\right) + \frac{1}{2} h_i f_t^2\left(X_i, U_i; \theta_t\right) \right] + \gamma \theta_t + \frac{1}{2} \rho \|w_t\|^2,$$

where $\gamma > 0$ is a penalty term on the tree complexity, and $\rho > 0$ is the L2 regularization term for the leaf weights $w \in \mathbb{R}^{\tau_t}$.

3.3.3 Baseline hazard

In this section we discuss the estimation of the distribution of the baseline hazard. Following the discussion in Cox (1972), once an estimate $\hat{\theta}$ of θ is obtained minimising the partial likelihood in Equation (3.3), we can derive an estimator for the baseline using the full-likelihood model. Many implementations of the baseline rely on the approach in N. Breslow (1974). There, the author assumes implicitly that the events $[t^{(\ell)}, t^{(\ell+1)})$ occur simultaneously at $t^{(\ell)}$. In contrast, we will assume that the claims report are uniform distributed within the tie. This makes the estimation of the baseline consistent with the way we will later transform the estimated hazard function into development factors. Our baseline estimator is

$$\hat{\alpha}^R_{0,t^{(\ell)}} = \frac{O_\ell}{\sum_{k\in\mathcal{R}(t^{(\ell)})} e^{\phi\left(x_k,u_k;\hat{\theta}\right)} - 0.5\sum_{s\in\mathcal{O}(t^{(\ell)})} e^{\phi\left(x_s,u_n;\hat{\theta}\right)}}. \tag{3.4}$$

Compared to the baseline estimator in N. Breslow (1974) we add a correction to the denominator to take into account the uniform occurrence of claims in the ties (Development Periods), that makes our baseline estimator consistent with the framework presented in Chapter 2 from Section 2.1, namely the term $0.5\sum_{s\in\mathcal{O}(t^{(\ell)})} e^{\phi\left(x_x,u_n;\hat{\theta}\right)}$. As earlier mentioned in Section 2.5.1, the uniform assumption is common in epidemiology and mortality modelling. For further arguments on the estimation of the baseline that go beyond the scope of this monograph we refer to the discussion on Professor Cox's paper (Cox, 1972), see for instance N. E. Breslow (1972); Oakes (1972).

3.4 Modelling the claims development

In this section, we explain the connection between continuous individual hazard rates and chain ladder development factors. We start the section with the definition of development triangles. We then connect the continuous time framework with the observation of ties that we used for model estimation to the discrete setting that we will construct for model predictions. A byproduct of our proposal is that for different level of aggregation, say yearly or quarterly, the predictions do not change.

3.4.1 Observing development delay and accident date

Usually, development delay and accident date are provided on some level of granularity. To provide a mathematical representation, we start by partitioning the interval $[0, \mathcal{T}]$ corresponding to the support of T^R and U into an equidistant grids $0 = t^{(0)}, \ldots, t^{(m)} = \mathcal{T}$ and $0 = u^{(0)}, \ldots, u^{(m)} = \mathcal{T}$, respectively. Similarly to what we showed in the Introduction, we define the parallelograms for claim reports

$$\mathcal{P}^R_{k,j} = \{(t,u) : t^{(j)} + u^{(k)} - u \leq t \leq t^{(j+1)} + u^{(k)} - u; u \in [u^{(k)}, u^{(k+1)}, t \geq 0\},$$

$k, j = 0, \ldots, m$, and we assume that we don't have access to the continuous observations (T^R_i, U_i), but only know on which parallelogram the observation has fallen. This observation scheme can be seen as observing ties, as described in the previous section. To this end we identify observations

(T_i^R, U_i) as being equal to $(t^{(j)}, u^{(k)})$ if $(T_i^R, U_i) \in \mathcal{P}_{k,j}^R$. In the sequel we will sometime write $\hat{\alpha}^R(j|k,x)$ for the estimator $\hat{\alpha}^R(t^{(j)}|u^{(k)},x)$ derived in the previous section.

3.4.2 From hazard rates to development factors

For $k, j = 0, \ldots, \mathcal{T}$, $k + j \leq M$, the individual reported claims are grouped

$$O_{k,j}(x) = \sum_i \int I\left((s, U_i, X_i) \in \mathcal{P}_{k,j}^R(x)\right) dN_i'(s),$$

where $\mathcal{P}_{k,j}^R(x) = \{(t,u,x) : (t,u) \in \mathcal{P}_{k,j}^R\}$. In reserving, the raw development factors $\hat{f}_{k,j}(x) = \sum_{\ell \leq j} O_{k,\ell}(x) \,/\, \sum_{\ell < j} O_{k,\ell}(x)$ are well-known objects, but they cannot be used for prediction because they are not defined on the un-observed lower triangle $k + j > m$. Even ignoring this problem, $\hat{f}_{k,j}(x)$ is too noisy in describing the development from $j-1$ to j because it is calculated separately for every k and x. We propose to use the hazard rate estimated in the previous section to derive a more stable estimate of the development from $j-1$ to j. For $j, k = 1, \ldots, m$, we propose to estimate the development factors as

$$\tilde{f}_{k,j}(x) = \frac{2 + \hat{\alpha}^R(j|k,x)}{2 - \hat{\alpha}^R(j|k,x)}. \tag{3.5}$$

The formula can be motivated from the fact that Equation (3.5) is true if $\tilde{f}_{k,j}(x)$ is replaced by the raw development factors $\hat{f}_{k,j}(x)$ and $\hat{\alpha}^R(j|u,x)$ is replaced by the raw observations of the ratio of occurrence and exposure. Here, exposure is measured as "claim-years-at-risk" under the assumptions that claim reports are uniformly distributed within a parallelogram; see also (Pittarello et al., 2023) for concrete calculations. In other words, equation (3.5) describes an plug-in estimator.

3.4.3 Forecasting

The cumulative entry at development time $j \leq \mathcal{T} - k$ for accident period k with features x is

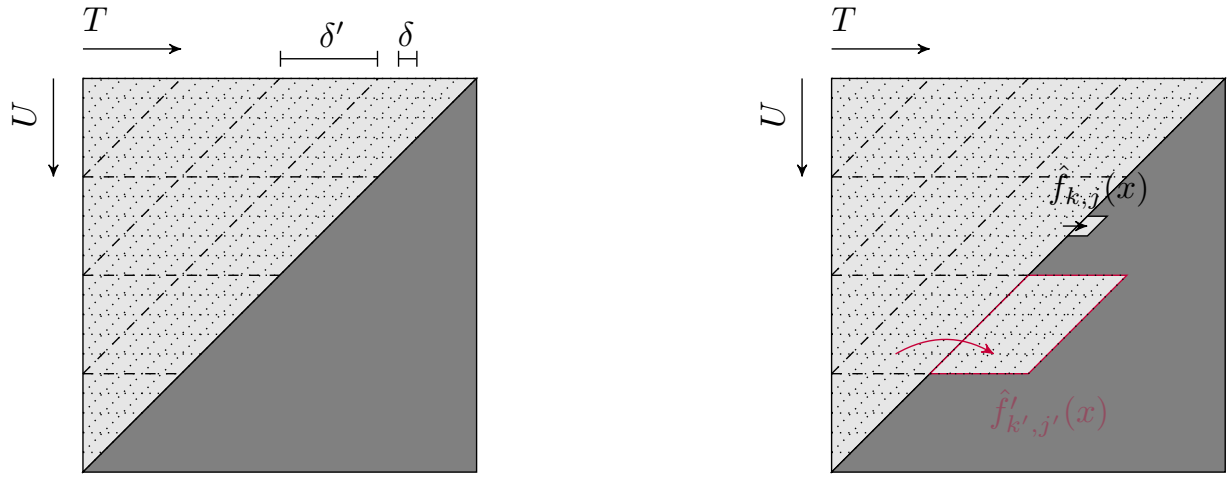
$$C_{k,j}^*(x) = \sum_{s \leq j} O_{k,s}(x).$$

For $j + k > \mathcal{T}$, the expected number of reports at development time j for accident period k is

$$\hat{O}_{k,j}(x) = \begin{cases} C_{k,\mathcal{T}-k}^*(x) \left(\prod_{l=\mathcal{T}-k+1}^{j} \tilde{f}_{kl}(x) - 1\right) & \text{if } j = \mathcal{T} - k + 1. \\ C_{k,\mathcal{T}-k}^*(x) \left(\prod_{l=\mathcal{T}-k+1}^{j} \tilde{f}_{kl}(x) - \prod_{l=\mathcal{T}-k+1}^{p-1} \tilde{f}_{kl}(x)\right) & \text{if } j \geq \mathcal{T} - k + 2 \end{cases}$$

3.4.4 Increasing the granularity of development factors

In the previous sections, we defined the development factors as discrete objects that we use for projecting the future reports given some granularity $\delta = t^{(L)} - t^{(L-1)}$ with $L \in \{1, \ldots, m\}$ and $L \in \mathbb{N}_0$. In this section, we want to elaborate on the possibility of choosing different values of $\delta' > \delta$ for the computation of the development factors. For example in the first data application of this paper we will start from daily data but we will be interested in reporting the results in quarterly and yearly flows. In Figure 3.1a we provide an intuition behind the change of granularity.



(a) development triangle in reverse time showing two levels of granularity (δ and δ'). The data in the parallelograms with granularity δ are included in the parallelograms with granularity δ', with $\delta' > \delta$ and $\lfloor \frac{\delta'}{\delta} \rfloor = 0$.

(b) Using the correspondence between development factors and average hazard rates we can predict the IBNR development on granularity δ (black). Using the probability based approach we can produce development factors for granularity δ' (red).

Figure 3.1: Our approach can handle different levels of granularities (left hand side). Starting from the individual data we can easily produce results for different aggregation levels (right hand side).

For simplicity, let us assume that δ'/δ and $(M+1)(\delta/\delta')$, are integer valued. For $j, k = 1, \ldots, \mathcal{T}$, the original parallelograms $\mathcal{P}^R_{k,j}(x)$ are with respect to some granularity δ.

For $k, j = 0, \ldots, (m+1)(\delta/\delta') - 1 = M'$, we define the parallelogram and fitted occurrences on the granularity δ' as:

$$P'_{k,j}(x) = \bigcup \left\{ \mathcal{P}^R_{h,l}(x) : h \in \left\{ k \cdot \frac{\delta'}{\delta}, ..., (k+1) \cdot \frac{\delta'}{\delta} - 1 \right\}, l \in \left\{ j \cdot \frac{\delta'}{\delta}, ..., (j+1) \cdot \frac{\delta'}{\delta} - 1 \right\} \right\},$$
$$\hat{O}'_{k,j}(x) = \bigcup \left\{ \hat{O}_{h,l}(x) : h \in \left\{ k \cdot \frac{\delta'}{\delta}, ..., (k+1) \cdot \frac{\delta'}{\delta} - 1 \right\}, l \in \left\{ j \cdot \frac{\delta'}{\delta}, ..., (j+1) \cdot \frac{\delta'}{\delta} - 1 \right\} \right\}, \quad k + j > M'.$$

The occurrence for granularity δ' is $O'_{kj}(x) = \sum_i I\left(\left(T_i^R, U_i, X_i \right) \in P'_{kj}(x) \right)$ and the corresponding cumulative entry for granularity δ' becomes $C'_{k,j}(x) = \sum_{s \leq j} O'_{k,s}(x)$. Define

$$\tilde{O}'(x) = \begin{cases} O'_{k,s}(x) & \text{if } k + j \leq M', \\ \hat{O}'_{k,s}(x) & \text{if } k + j > M'. \end{cases}$$

We can use these quantities to obtain development factors with granularity δ':

$$\tilde{f}'_{kj}(x) = \frac{\sum_{l \leq j} \tilde{O}_{kl}(x)}{\sum_{l < j} \tilde{O}_{kl}(x)}.$$

In Figure 3.1b we illustrate the idea behind using the development factors $\hat{f}'_{kj}(x)$ on the granularity δ' to forecast future reports.

3.5 Models comparison

Evaluating the performance of a claims reserving model is not straightforward. Indeed, we need to take into account the time series structure of the data and provide a reasonable performance measure for the different grids that we discuss in this work. In this section we will first propose a strategy to compare the CL to our individual models. Secondly, we will propose two strategies to rank the individual models.

3.5.1 Measuring the performance on the total predictions

We first evaluate the (relative) total absolute errors on the input grid:

$$\texttt{ARE}^{\texttt{TOT}} = \frac{|\sum_{j,k:k+j>M'} \sum_x O_{k,j}(x) - \sum_x \hat{O}_{k,j}(x)|}{\sum_{j,k:k+j>M'} \sum_x O_{k,j}(x)}. \tag{3.6}$$

In this phase, we use the development factors that we obtained to estimate the total future notifications on the lower triangle. The performance measure in Equation 3.6 is then computed to compare the total predicted counts with the actual counts in absolute terms. This is possible as the lower triangle is available from the simulation. The idea is illustrated in Figure 3.2.

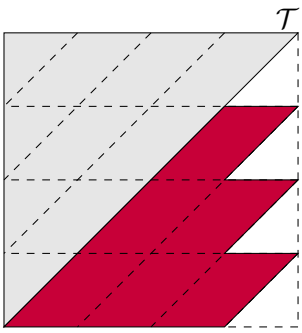
On the one hand, the estimate of the total future counts can be an interesting reference value of the expected evolution of future counts. On the other hand, actuaries predictions are often provided one calendar period ahead. Indeed, the data are updated at the end of each calendar period when new occurrences are reported to the insurance company. We then want to evaluate our models performance a second time with a different performance measure that: a) Considers the new information available at the end of each calendar period until development. b) Measures the performance diagonal-wise. For some granularity δ', we propose:

$$\texttt{ARE}^{\texttt{CAL}} = \frac{\sum_{\tau=M'+1}^{2M'-1} |\sum_{j,k:k+j=\tau} \sum_x O_{k,j}(x) - \sum_x \tilde{f}_{k,j}(x) O_{k,j-1}(x)|}{\sum_{j,k:k+j>M'} \sum_x O_{k,j}(x)} \tag{3.7}$$

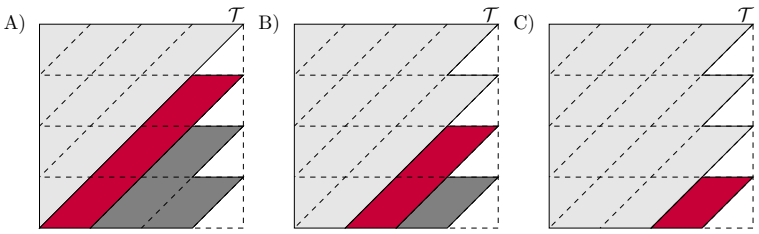
Note that for computational reasons we do not update $\tilde{f}_{k,j}(x)$ and different periods τ, meaning the model is only fitted once. However, predictions use new information on a rolling bases, see (Figure 3.2b). The performance is measured for the calendar periods $\tau = M' + 1, \ldots, 2M' - 1$.

3.5.2 Continuously Ranked Probability Score

When a set of different stochastic models is available, scoring measures can be used as a criterion to assesses the quality of the forecasts by assigning to each model a score. By giving better scores to models that provide better forecasts, we can rank competing forecast procedures, see e.g. Gneiting, Raftery, Balabdaoui, and Westveld (2004); Gneiting, Balabdaoui, and Raftery (2007); Gneiting and Raftery (2007). The $\texttt{ARE}^{\texttt{TOT}}$ and the $\texttt{ARE}^{\texttt{CAL}}$ are chosen as interesting references to compare our models in terms of predicted counts to the realized counts. However, they do only take point estimates into consideration. In this section, we want to introduce a proper scoring rule (Gneiting & Ranjan, 2011). A scoring rule is a function $s(f, y)$ taking values in the real line $[0, \infty)$ where f is a density forecast and $y \in \mathbb{R}$ is a future realization from the conditional sampling distribution Y. The ideal forecast is obtained when f is the true density of Y. A scoring rule is



(a) Data is given on the upper triangle (gray). To calculate $\texttt{ARE}^{\texttt{TOT}}$ in Equation 3.6, we measure model performance error on the total future notifications (red). This is possible as the full triangle was simulated.



(b) Actuaries update their model predictions using the new data they receive at the end of each calendar period. In the example, the future notifications are predicted three times (A), B), and C)). Each time the one year ahead notifications are computed using the new data available to the insurance company (For computational reasons, development factors are only calculated once with the original data) and the error is calculated using the actual realizations. In the picture we show the available data at each evaluation (light gray), the predicted data (red), and the future data (dark gray). The performance measure in Equation 3.7 summarizes the estimation error obtained at A), B), and C).

Figure 3.2: Evaluation of the models performance using $\texttt{ARE}^{\texttt{TOT}}$ (Figure 3.2a) and $\texttt{ARE}^{\texttt{CAL}}$ (Figure 3.2b).

said to be proper if

$$\mathbb{E}_f\left[s(f,Y)\right] = \int f(y)s(f,y)\mathrm{d}y \leq \int f(y)s(g,y)\mathrm{d}y = \mathbb{E}_f\left[s(g,Y)\right]$$

for all densities f and g and with equality only for $f = g$. The Continuously Ranked Probability Score (CRPS) is a proper metric to evaluate the models performance. In the following we re-write the definition such that it depends on a predicted survival function $\hat{S}$:

$$\begin{aligned}
\mathrm{CRPS}(\hat{S}(\cdot|X,U),y) &= \int_0^\infty \mathrm{PS}(\hat{S}(z|X,U), I(y>z))\mathrm{d}z \\
&= \int_0^\infty (\hat{S}(z|X,U) - \mathbb{I}\{y>z\})^2\mathrm{d}z \\
&= \int_0^y \left(1-\hat{S}(z|X,U)\right)^2 dz + \int_y^{+\infty} \left(\hat{S}(z|X,U)\right)^2 dz
\end{aligned}$$

with $\mathrm{PS}(\hat{S}(z|X,U_i), \mathbb{I}\{y>z\}) = (\hat{S}(z|X,U) - \mathbb{I}\{y>z\})^2$ being the Brier Score (Selten, 1998; Gneiting & Raftery, 2007).

As survival function prediction, we will use

$$\hat{S}(j|x,U) = \frac{1}{\prod_{l=1}^{j}\hat{f}_{k,l}(x)}. \tag{3.8}$$

The CRPS is taken to be negatively oriented, meaning that the lowest score indicates the better model. In order to compare different models on a same data set in terms of CRPS we will consider the average CRPS across all the data points.

3.5.3 The partial log-likelihood

We will report the partial log-likelihood of our models for the different data sets that we inspect in this manuscript. The model with the minimum in-sample average negative partial log-likelihood is the model that fits best the training data. In this context, the log-likelihood can only provide a reference value to understand what model is capable to minimize the objective function on the in-sample data during the fitting, i.e. Equation (3.3).

3.6 Data Application on simulated data

In this section we will challenge our modelling approach in five different simulated scenarios. Each scenario is simulated 20 times. For each simulation we will compute our model performances according to the scoring rules that we defined in Section 3.5. The results that we show are averaged over the 20 simulations. We believe 20 simulation is a sufficient number of repetitions to smooth the randomness deriving from the simulations. According to the experience of the authors in the actuarial practice, we generated scenarios that could occur in the real world. Additional details on the simulation algorithm are in Appendix E.

3.6.1 Five simulated scenarios

This section provides an overview of the simulations. The data are simulated on a daily grid on a 4 years time horizon, meaning that we will observe up to 1440 accident days ($\mathcal{T} = 1440$). We provide results on a quarterly and an yearly grid and compare our models to the chain ladder (CL, Mack, 1993). Since CL is an established method in the insurance industry, we chose it as the competing model. The five scenarios have some common characteristics:

- A mix of `claim_type 0` and `claim_type 1`. The parameters are chosen to make `claim_type 0` resemble property damage and `claim_type 1` bodily injuries.
- Bodily injuries (`claim_type 1`) are longer tailed, meaning that their resolution takes more time than property damage (`claim_type 0`), see (Ajne, 1994).

We name the 5 scenarios Alpha, Beta, Gamma, Delta, Epsilon and describe the data composition in Figure E.1.

The simulations were performed using the `R` package `SynthETIC`, see (Avanzi, Taylor, Wang, & Wong, 2021). For each claim type, data is simulated in a continuous setting. To imitate a real world portfolio, we populate the data with an accident day (`AD`) and development day (`DD`) for each observation. In the spirit of (Avanzi et al., 2021), for each claim type, we model occurrence and development of claims in two modules:

- We first fix the total amount of individuals at risk in the portfolio for each accident day. We then let the actual frequency of claims being reported in the accident date be randomly simulated (we draw it from a Poisson distributed random variable with rate 0.2). In scenarios Alpha, Gamma, Delta and Epsilon the individual at risk per `AD` are in the same proportion for each `claim_type`, i.e. in each `claim_type` we have 200 individuals at risk per `AD`. In scenario Beta the individuals at risk are decreasing by one unit every 10 `AD`, e.g. we will have

200 individuals at risk in AD= 1, 199 individuals at risk in AD= 11 and only 56 individuals at risk in AD= 1440.

- We then simulate the reporting time of the individual claims. As we need a proportional hazard structure in reverse development time, we need to be careful when simulating the reporting delays. The proportional hazard is

$$\alpha^R(t|X = \texttt{claim\_type}, U = \texttt{AD}) = \alpha_0^R(t)e^{\phi(\texttt{claim\_type},\texttt{AD};\theta)}.$$

  The parametrization of $\alpha_0^R(t)$ and $\phi\left(\texttt{claim\_type}, \texttt{AD}; \theta\right)$ changes in the different scenarios. In this section we will provide qualitative details on how we designed the different scenarios. We report the parameters we used for the simulation in Appendix E.

To account for the proportional structure of the hazard, we simulate from a Right Truncated Fréchet-Weibull distribution (RTFWD) (Teamah, Elbanna, & Gemeay, 2019). The RTFWD has a four parameter structure ν, π, ξ, k, and is defined with $0 < t \leq b$ with distribution function

$$F\left(t\right) = e^{-\pi^\nu \xi^{\nu k}(t^{-\nu k} - b^{-\nu k})}.$$

With the reverse time hazard

$$\alpha^R(t) = \frac{f(t)}{F(t)},$$

we get, in the RTFWD case,

$$\alpha^R(t) = \frac{\nu k \pi^\nu \xi^{\nu k}(t^{-1-\nu k})e^{-\pi^\nu \xi^{\nu k}(t^{-\nu k} - b^{-\nu k})}}{e^{-\pi^\nu \xi^{\nu k}(t^{-\nu k} - b^{-\nu k})}} = \nu k \pi^\nu \xi^{\nu k}(t^{-1-\nu k}).$$

Let us set

$$\xi = \xi(\texttt{claim\_type}, \texttt{AD}) = \xi_0(e^{\phi(\texttt{claim\_type},\texttt{AD};\theta)})^{\frac{1}{\nu k}},$$

the reverse hazard structure now becomes

$$\alpha^R(t|\texttt{claim\_type}, \texttt{AD}) = \nu k \pi^\nu \xi_0^{\nu k}(t^{-1-\nu k})e^{\phi(\texttt{claim\_type},\texttt{AD};\theta)}. \tag{3.9}$$

The survival function dependency on the features is denoted as $F\left(t|\texttt{claim\_type}, \texttt{AD}\right)$ with $t \geq 0$, similarly to what we did with the hazard models. The survival function is then $S\left(t|\texttt{claim\_type}, \texttt{AD}\right)$ with $t \geq 0$. The baseline hazard is $\alpha_0^R(t) = \nu k \pi^\nu \xi_0^{\nu k}(t^{-1-\nu k})$.

Hence by simulating from Equation 3.9 the reporting times in forward time, we achieve a proportional structure in reverse time. All simulations are truncated at $b = 1440$, resulting in 4 years worth of data. The parameters of each scenario we set up for the simulation are specified in Table E.1. The scenarios Alfa, Beta, Gamma, and Delta have the same proportional hazard structure

$$\alpha^R(t|\texttt{claim\_type}) = \alpha_0^R(t)e^{\phi(\texttt{claim\_type},\texttt{AD};\theta)},$$

with $\texttt{claim\_type} \in \{0, 1\}$ and $\texttt{AD} \in \{1, \ldots, 1440\}$

The scenarios have the following distinctive traits:

- **Scenario Alpha**: this scenario is a mix of `claim_type 0` and `claim_type 1` with same

number of claims at each accident period (i.e. the claims volume). As we have an effect based on the claim type and the reporting of the claims only depends on the baseline, in this scenario the CL assumptions are satisfied.

- **Scenario Beta**: we simulate using the same proportional risk component y as **scenario Alpha**, but the volume of `claim_type 1` is decreasing in the most recent `AD`. When the longer tailed bodily injuries have a decreasing claim volume, aggregated CL methods will overestimate reserves, see (Ajne, 1994).

- **Scenario Gamma**: an interaction between `claim_type 1` and accident period (`AD`) affects the claims occurrence. One could imagine a scenario, where a change in consumer behaviour or company policies resulted in different reporting patterns over time. For the last simulated accident day, the two reporting delay distributions will be identical. The interaction makes the COX model assumption not valid. In addition, the accident period effect makes the CL model assumptions not valid.

- **Scenario Delta**: a seasonality effect dependent on the accident days for `claim_type 0` and `claim_type 1` is present. This could occur in a real world setting with increased work load during winter for certain claim types, or a decreased workforce during the summer holidays. The presence of an accident period effect makes the CL assumptions not satisfied.

- **Scenario Epsilon**: the data generating process violates the proportional hazard assumption, indeed we will generate the data assuming that a) there is an effect of the features on the baseline and b) the proportionality assumption is not valid. Conversely, the CL model is satisfied. In scenario Epsilon, the hazard we simulate from is

$$\alpha^R(t|\texttt{claim\_type}, \texttt{AD}) = \alpha_0^R(t|\texttt{claim\_type}, \texttt{AD})(e^{\phi(\texttt{claim\_type},\texttt{AD};\theta)} + \frac{\phi\left(\texttt{claim\_type}, \texttt{AD}; \theta\right)}{2}),$$

with $\alpha_0^R(t|x) = 0.5\sqrt{0.1\left(0.1(2 + \frac{\phi(\texttt{claim\_type},\texttt{AD};\theta)}{2})t^{-1}\right)}$.

The relevant features on the proportional risk part of the hazard are reported in column two of Table 3.1. In columns four to seven, we indicate with a check mark whether the assumptions for models in the data application (CL, COX, NN, and XGB) are satisfied.

Scenario	Effect(s) on ϕ	CL	COX	NN	XGB
Alpha	`claim_type`	✓	✓	✓	✓
Beta	`claim_type`	✗	✓	✓	✓
Gamma	`claim_type` + `claim_type:`$\sqrt{\texttt{AD}}$	✗	✗	✓	✓
Delta	`claim_type` +`AD`	✗	✓	✓	✓
Epsilon	`claim_type`	✓	✗	✗	✗

Table 3.1: The relevant features affecting the proportional risk component (column two) in the scenarios Alpha, Beta, Gamma, Delta and Epsilon (column one). In order to simplify the reading of this manuscript, we use the notation used in (RStudio Team, 2020). The effect terms are added with the operator + (plus), the interaction terms are added with the operator : (columns). In columns four to seven, we use a check mark if the models assumptions are satisfied in the scenario.

3.6.2 Average negative partial log-likelihood

In this section we report the average negative partial log-likelihood for our models in the 5 scenarios. We train the COX model using all the individual data from calendar periods $\tau = 1, \ldots, \mathcal{T}$. In the training phase of XGB and NN, the input data from calendar periods $\tau = 1, \ldots, \mathcal{T}$ are further split into a main part for training and a smaller part for validation. The split is random and we use 80% of the data for training (the splitting percentage is selected with a rule of thumb). We will then report for XGB and NN the out-of-sample average negative log-likelihood measured on the remaining 20% of the input data. Comparing the in-sample likelihood and the out-of-sample likelihood will tell us whether a model is overfitting the data. In Table 3.2 we show the (average) negative log-likelihood averaged (over the 20 simulations), for each model in each scenario. The results show that XGB and NN seem to best fit the in-sample data compared to COX. Furthermore, XGB is consistently providing a lower likelihood compared to the NN. A similar behavior is reflected in the out-of-sample data.

Model	Scenario	$l(\theta)$ (in-sample)	$l(\theta)$ (out-of-sample)
COX	Alpha	9.238	-
NN		8.648	7.263
XGB		8.634	7.249
COX	Beta	9.112	-
NN		8.512	7.127
XGB		8.502	7.118
COX	Gamma	9.062	-
NN		8.760	7.374
XGB		8.755	7.372
COX	Delta	9.199	-
NN		8.686	7.303
XGB		8.590	7.215
COX	Epsilon	9.121	-
NN		8.563	7.180
XGB		8.556	7.175

Table 3.2: In each scenario, the average log-likelihood $l(\theta)$ is computed in each simulation on each simulated data set. For machine learning models each data set is split in training (in-sample, 80% of the data) and validation (out-of-sample, 20% of the data). The splitting percentage is determined as a rule of thumb. Here, we provide for each scenario the results of the average likelihood $l(\theta)$ (in-sample and out-of-sample) over the 20 simulations.

3.6.3 Modelling the survival function

In the previous sections, we observed that our predictions rely on a discrete time framework meanwhile we defined the hazards in continuous time. In the simulated scenarios we have, for $t \geq 0$, a closed form for the features dependent survival function $S(t|X, U)$. We can compare the survival function with the survival function we estimated with our models, i.e. $\hat{S}(t|X, U)$ in Equation (3.8). Visually inspecting the survival function and comparing it in different scenarios for different models can help in understanding the models fit across different scenarios.

In Figure 3.3 we show the true survival function compared to the fitted survival function in scenario Alpha. It can be noticed that all the three models seem to behave consistently with respect

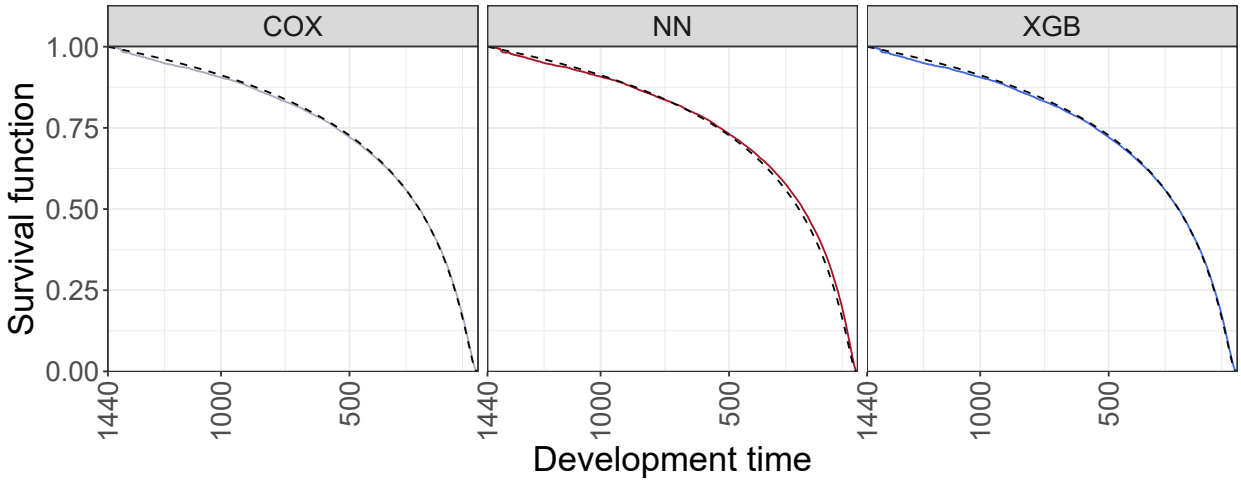


Figure 3.3: Scenario Alpha, `claim_type 1` and `AD 13`. The true survival function (black dotted line) is compared to the fitted survival function with COX, NN, and XGB (left to right).

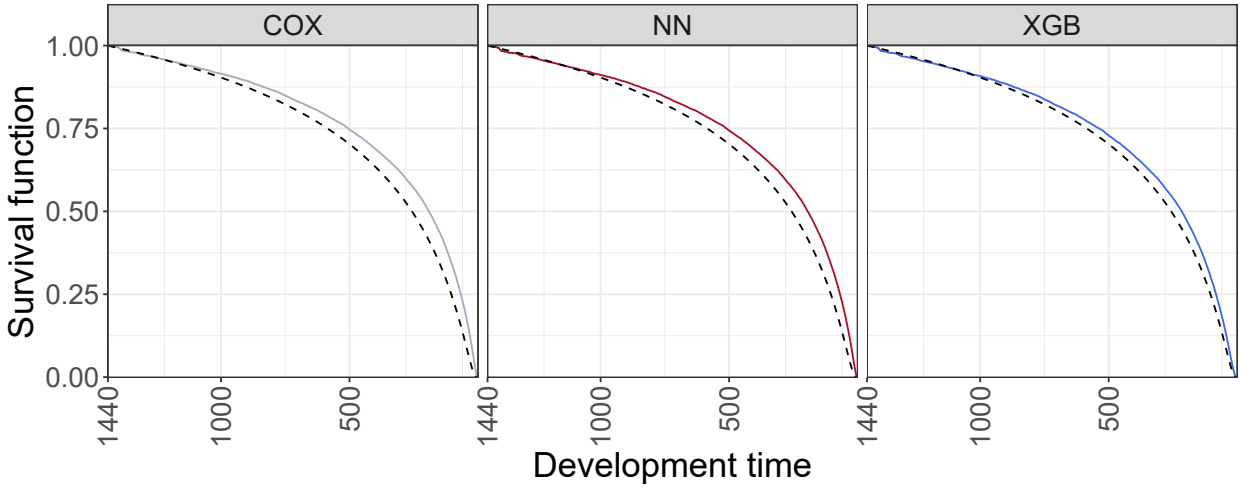


Figure 3.4: Scenario Delta, `claim_type 1` and `AD 691`. The true survival function (black dotted line) is compared to the fitted survival function with COX, NN, and XGB (left to right).

to the true curve. Interestingly, even in the simplest modelling scenario, XGB and COX seem to be better than NN in catching the behavior of the survival function on the right tail.

The same plot can be shown for other scenarios. An interesting case is scenario Delta, where we introduced a seasonality effect dependent on the accident period, see Table 3.1. In this section we first show the feature combination `claim_type 1` and `AD 691` where the three models seem to behave similarly with XGB closer to the true survival function, see Figure 3.4.

3.6.4 Overview of the main model output

Oppositely to the chain ladder model the predicted development factors will depend on accident date and additional feature information; here claim type being equal zero or one. Another important feature of our proposal is that while inputs are assumed continuous, development factors and predictions can be provided on chosen granularity levels. We show a preview of the results that we obtained on a simulated data set coming from scenario Delta, as it was introduced in Section 3.6.

In Figure 3.5 we select one simulation and compare the chain ladder development factors to our output development factors for different granularities (monthly and yearly). The first row shows chain ladder's development factors. The second row shows the output on a monthly level from an XGB fit for three different feature combinations of accident date and claim type. The three plots can be compared to the monthly development factors obtained from chain ladder (first row left panel). The third row is the analogue to the second row, but this time a yearly aggregation has been chosen and it can be compared to the yearly development factors obtained from chain ladder (first row right panel).

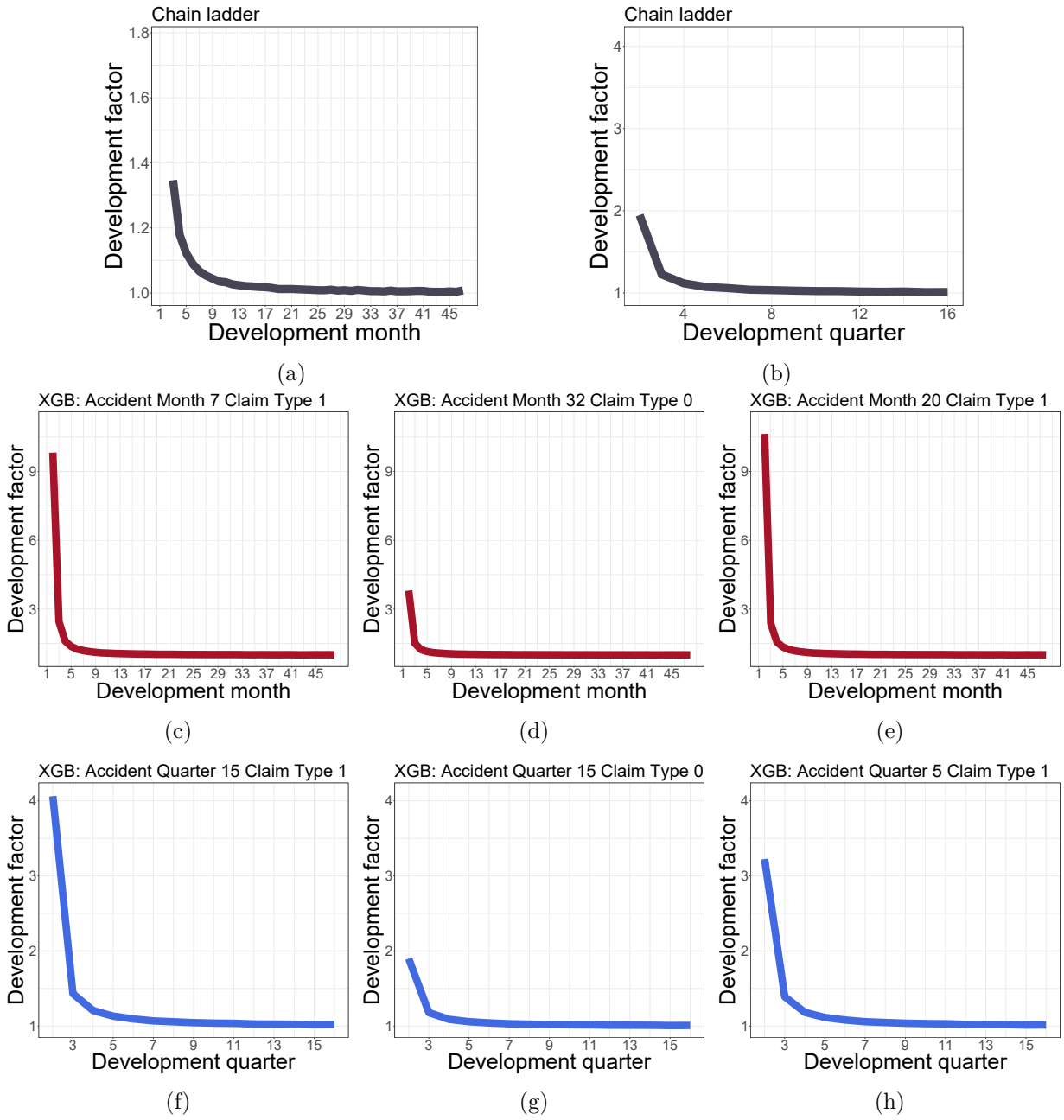


Figure 3.5: The first row, shows the chain ladder development factors fitted on monthly data (Figure 3.5a), and quarterly data (Figure 3.5b). The development factors do not consider additional features we have at our disposal. The second and third row show an XGBoost output from our model proposal that depends on accident date and claim type. The second row, shows monthly development factors for the feature combinations Accident Month 7 and `Claim_type 1` (left panel), Accident Month 31 and `Claim_type 0` (center panel) and Accident Month 20 and `Claim_type 1` (right panel). The third row shows quarterly development factors for the feature combinations Accident Quarter 15 and `Claim_type 1` (left panel), Accident Quarter 15 and `Claim_type 0` (center panel) and Accident Quarter 5 and `Claim_type 1` (right panel).

3.6.5 Results: forecasting using the development factors

In this section we compare the chain ladder with our models in the five simulated scenarios. In each scenario, we simulate 20 data sets to show that our results are not affected from accidental fluctuations of the simulations. Below we describe in brief the procedure that we used.

1. For each scenario (Alpha, Beta, Gamma, Delta, and Epsilon), we simulate 20 data set.
2. On each data set, we find the optimal hyper parameters for NN and XGB using Bayesian optimization (Snoek, Larochelle, & Adams, 2012). More details on the optimization algorithm can be found in Appendix D.
3. On each data set, we fit COX, the optimal NN and XGB.
4. We find the estimated development factors and expected number of future occurrences according to each model.
5. We evaluate the performances for COX, NN, and XGB using the $\texttt{ARE}^{\texttt{TOT}}$, the (average) CRPS and the $\texttt{ARE}^{\texttt{CAL}}$ on a quarterly and yearly basis. The $\texttt{ARE}^{\texttt{TOT}}$ and $\texttt{ARE}^{\texttt{CAL}}$ are also computed for the CL.
6. For each scenario we report the average performance measures of CL, COX, NN, and XGB over the 20 data sets.

Model	Scenario	$\texttt{ARE}^{\texttt{TOT}}$	$\texttt{ARE}^{\texttt{CAL}}$ (quarters)	$\texttt{ARE}^{\texttt{CAL}}$ (years)	CRPS (average)
CL (✓)	Alpha	0.131 (± 0.016)	0.128 (± 0.014)	0.037 (± 0.011)	-
COX (✓)		0.138 (± 0.013)	0.131 (± 0.012)	0.044 (± 0.013)	364.428 (± 5.899)
NN (✓)		0.140 (± 0.021)	0.132 (± 0.015)	0.045 (± 0.015)	366.902 (± 8.567)
XGB (✓)		0.136 (± 0.010)	0.131 (± 0.011)	0.042 (± 0.012)	364.230 (± 5.967)
CL (✗)	Beta	0.215 (± 0.023)	0.194 (± 0.015)	0.122 (± 0.025)	-
COX (✓)		0.162 (± 0.014)	0.161 (± 0.011)	0.050 (± 0.018)	402.794 (± 6.929)
NN (✓)		0.169 (± 0.019)	0.166 (± 0.015)	0.062 (± 0.028)	403.057 (± 7.720)
XGB (✓)		0.162 (± 0.013)	0.162 (± 0.011)	0.048 (± 0.019)	402.343 (± 7.133)
CL (✗)	Gamma	0.260 (± 0.035)	0.230 (± 0.029)	0.191 (± 0.032)	-
COX (✗)		0.147 (± 0.015)	0.149 (± 0.013)	0.057 (± 0.022)	402.917 (± 6.221)
NN (✓)		0.167 (± 0.044)	0.160 (± 0.027)	0.075 (± 0.039)	402.311 (± 9.258)
XGB (✓)		0.154 (± 0.024)	0.147 (± 0.018)	0.053 (± 0.025)	401.996 (± 5.988)
CL (✗)	Delta	0.300 (± 0.024)	0.234 (± 0.018)	0.037 (± 0.012)	-
COX(✓)		0.195 (± 0.022)	0.192 (± 0.018)	0.065 (± 0.020)	386.598 (± 6.667)
NN (✓)		0.213 (± 0.034)	0.204 (± 0.025)	0.051 (± 0.023)	388.707 (± 7.205)
XGB (✓)		0.167 (± 0.019)	0.145 (± 0.011)	0.058 (± 0.025)	369.531 (± 6.599)
CL (✓)	Epsilon	0.119 (± 0.011)	0.115 (± 0.010)	0.035 (± 0.010)	-
COX (✗)		0.135 (± 0.022)	0.127 (± 0.010)	0.060 (± 0.015)	340.267 (± 5.192)
NN (✗)		0.132 (± 0.015)	0.126 (± 0.011)	0.059 (± 0.019)	341.169 (± 5.210)
XGB (✗)		0.149 (± 0.059)	0.127 (± 0.012)	0.057 (± 0.017)	340.344 (± 5.100)

Table 3.3: Results after 20 simulations. We show the different performance measure for our different models (column one). In the different scenarios (column two) the models assumptions might be not satisfied. We indicate with a check mark (✓) the scenarios in which the models assumptions are satisfied. For each model we also display the performance measures we described in the previous sections: absolute total reserving error (column three), absolute error by diagonal for quarterly output (column four) and absolute relative error by diagonal for yearly output (column five). Between brackets, we show the standard deviation of the (average) performance measures over the 20 records. In column six we show the continuously ranked probability score.

The results of the data application on simulated data are summarized in Table 3.3. In column one, we list the models included in the comparison for each scenario (column two). Similarly to Table 3.1, we denote with a check mark or an x mark if the model assumptions are satisfied or not. Let us first consider the comparison with CL in terms of $\texttt{ARE}^{\texttt{TOT}}$ and $\texttt{ARE}^{\texttt{CAL}}$. Our data application shows that CL is outperforming the individual models in scenario Alpha and in scenario Epsilon according to the $\texttt{ARE}^{\texttt{TOT}}$ and the $\texttt{ARE}^{\texttt{CAL}}$. In scenario Alpha, the CL results are very close to those of COX, NN, and XGB. As expected, in scenario Epsilon, breaking down the proportionality assumption, the CL is better than the other models. Interestingly, when we inspect the most complex scenarios (Beta, Gamma, and Delta) our approach (COX, NN, and XGB) provides better scores than the CL model. Notably, we observe that XGB is consistently the best performing model. Notwithstanding the intensive hyper parameters tuning that we performed, NN seem to perform worse than COX in most scenarios. In scenario Delta, where we introduce a seasonality effect we find that according to all the proposed scores XGB is by far the best model. The $\texttt{ARE}^{\texttt{TOT}}$ and $\texttt{ARE}^{\texttt{CAL}}$ are a good benchmark to compare our approach to the CL but they are not proper

scoring rules. Conversely, we can take the average CRPS as the main criterion to rank our models. In the scenarios we considered, the CRPS indicates that XGB is outperforming NN and COX. Interestingly, while in scenarios Alpha, Beta, Gamma, and Epsilon the XGB (average) CRPS is close to COX, in scenario Delta we observe a major drop in the CRPS going from COX to XGB. In scenario Gamma we find that the NN model performs better than COX. Interestingly, in this scenario the COX assumptions are not satisfied.

A cell-wise comparison

In this section, we provide a heat map of the average relative prediction error based on the volumes of the k, j cells over the simulations for each scenario on a quarterly basis. In particular, we want to detect whether the IBNR frequencies predicted for the k, j cells are overestimated or underestimated. In this phase we can detect if paths are present in the models results extrapolation.

The relative error $\texttt{RE}^{k,j}$ for the k, j cell of some development triangle is

$$\texttt{RE}^{k,j} = \frac{\sum_x O'_{k,j}(x) - \sum_x \hat{O}'_{k,j}(x)}{\sum_x O'_{k,j}(x)}. \tag{3.10}$$

In Figure 3.6 we show a heat map of the average $\texttt{RE}^{k,j}$ over the simulated data sets of each model in each scenario. For all the scenarios that we investigated, our models seem to reduce the relative errors within the cells. Cell-wise, XGB models residuals are the closest to zero. An interesting insight comes from scenario Delta. We recall that in scenario Delta we simulated the data with a seasonality effect. Interestingly, the XGB model is the only model able to catch the seasonality effect in scenario Delta. Surprisingly, the NN model is not capable of catching the interactions with the accident period notwithstanding the heavy hyper parameters tuning procedure, see Appendix D.

3.7 Data application on real data

In this section we show a case study based on a real data set from Codan Forsikring. We had at disposal the company complete reserving data from 2012 to 2022 for a short tailed personal line product. Two categorical features are associated to each (claim `claim_type` and `coverage_key`). The complete description of the data set is available in Table 4.3.

Features	Description
`Claim_number`	Policy identifier.
`claim_type` $\in \{1, \dots, 20\}$	Type of claim.
`coverage_key` $\in \{1, \dots, 16\}$	Type of coverage.
`AM`	Accident month.
`CM`	Calendar month of report.
`DM`	Development month.
`incPaid`	Incremental paid amount. We will not use this information in the manuscript.

Table 3.4: Description of the Codan Forsikring data set.

We decided to split the data into 6 chunks of 5 consecutive years each. The splits are reported in Table 3.5, together with the corresponding data size. For each of these splits, we fit our models on the first 4 years (train data) and score our models on the fifth year (test data).

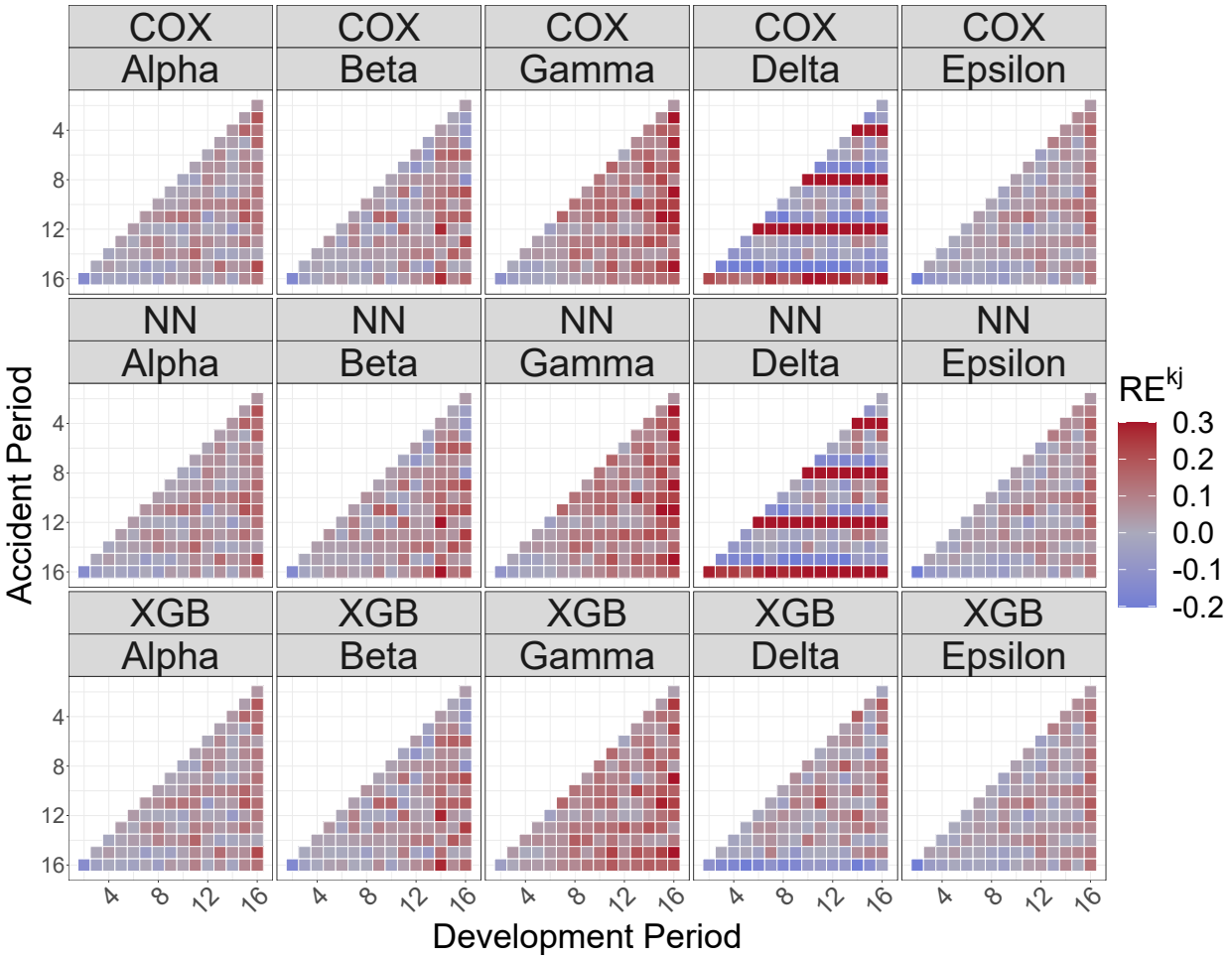


Figure 3.6: The average (over the simulations) of the relative errors: $\frac{\sum_x O_{k,j}(x) - \sum_x \hat{O}_{k,j}(x)}{\sum_x O_{k,j}(x)}$

	Training	Scoring	Train data size (observations number)	Test data size (observations number)	Mean Reporting Delay
Split 1	2012-2016	2017	129381	1474	1.62
Split 2	2013-2017	2018	133367	1553	1.62
Split 3	2014-2018	2019	130405	1566	1.63
Split 4	2015-2019	2020	122842	1394	1.63
Split 5	2016-2020	2021	110884	1225	1.61
Split 6	2017-2021	2022	105618	1069	1.55

Table 3.5: For each split (column one) we reported the accident periods we used to train the models (column two), and the year we use for scoring (column three). In columns four and five we show the train and test data size. In column six we provide the mean reporting delay from accident.

Similarly to the case study on the simulated data, we compare our results with the chain ladder model fitted on a quarterly and yearly grid. We fitted our models on a monthly grid, according to the procedure we described in the manuscript. In Table 3.5, we show the mean reporting delay in each data split. Most of the notifications occurred in the early development periods and the mean reporting delay is between one and two months showing that the data are very stable.

From Figure 3.7 we can see that not all the combinations of the two categorical features (`cover_key` and `claim_type_key`) were observed (see Figure 3.7). Most of the data belong to the combinations `claim_type_key 1` with `cover_key 1` and `cover_key 5` and `claim_type_key 14` with `cover_key 1`.

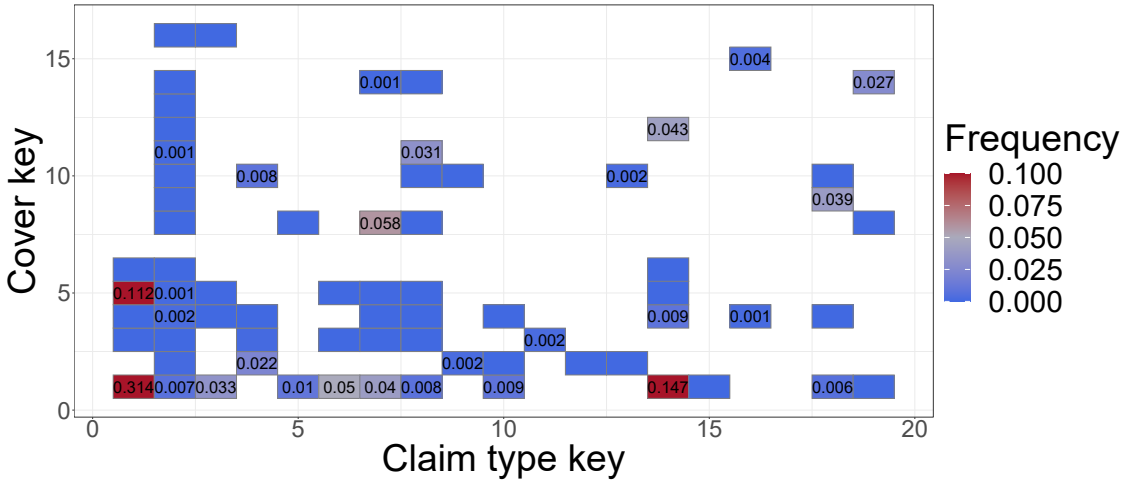


Figure 3.7: Relative frequency on the complete data (2012 – 2021). We do not observe all the combinations of features and most of the data show `claim_type_key` 1, and `cover_key` 1 or `cover_key` 5.

3.7.1 Results

In Table 3.6 we show the our models performance on the real data splits described in Table 3.5. The models are reported in column one and the data splits are reported in column two. In this setting, the chain ladder is expect to perform very well as the data have are reported in the early development months and, from the authors experience on this data sets, there are few reports in the late development periods. Interestingly, from the results of the $\texttt{ARE}^{\texttt{TOT}}$ and $\texttt{ARE}^{\texttt{CAL}}$ (columns three to five) we see that in general we have an improvement in the models performance using COX, NN, or XGB compared to the CL, especially in terms of $\texttt{ARE}^{\texttt{CAL}}$. By comparing COX, NN

and XGB we obtain further insights on the performance of our models. The $l(\theta)$ results (columns six and seven) indicate that XGB is the best model in minimizing the objective function, both in-sample and out-of-sample . Similarly to the simulated scenarios, we randomly split our data into training and validation for fitting NN and XGB. With the same rule of thumb we also kept the 80% of the data for training. As highlighted in the previous sections, the average CRPS is our main criterion to select a model the best performing model. The CRPS suggests that in the first three data splits, NN and XGB outperform the COX model. In general, we observe that COX is the best performing model. This is also enforced from the other performance measures $\texttt{ARE}^{\texttt{TOT}}$ and $\texttt{ARE}^{\texttt{CAL}}$. The motivation for these results is related to the behavior of the data set. We already discussed that there are no complex trends in our data, meaning that there is a benefit in using our modelling approach compared to the CL but the simplest hazard model (COX) is the most accurate based on the performance measures that we defined.

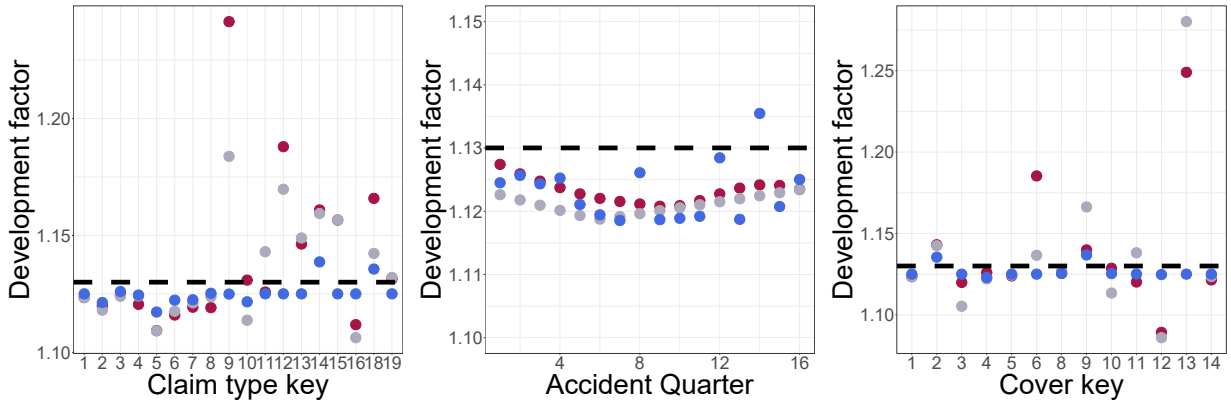
Model	Split	$\texttt{ARE}^{\texttt{TOT}}$	$\texttt{ARE}^{\texttt{CAL}}$ (quarters)	$\texttt{ARE}^{\texttt{CAL}}$ (years)	CRPS (average)	$l(\theta)$ (in-sample)	$l(\theta)$ (out-of-sample)
CL		0.207	0.225	0.264			
COX	Split	0.242	0.208	0.242	6.429	10.781	
NN	1	0.241	0.206	0.237	6.420	10.531	9.144
XGB		0.231	0.202	0.235	6.440	10.529	9.143
CL		0.235	0.323	0.155			
COX	Split	0.307	0.296	0.082	8.061	10.804	
NN	2	0.307	0.296	0.084	8.054	10.551	9.165
XGB		0.287	0.292	0.100	8.074	10.549	9.163
CL		0.262	0.324	0.196			
COX	Split	0.306	0.268	0.069	6.303	10.784	
NN	3	0.297	0.261	0.059	6.496	10.528	9.141
XGB		0.324	0.286	0.081	6.468	10.522	9.138
CL		0.213	0.264	0.181			
COX	Split	0.270	0.248	0.081	6.076	10.725	
NN	4	0.283	0.258	0.082	5.950	10.475	9.089
XGB		0.282	0.254	0.088	6.220	10.463	9.077
CL		0.203	0.193	0.064			
COX	Split	0.160	0.181	0.107	4.855	10.623	
NN	5	0.171	0.188	0.111	5.017	10.371	8.985
XGB		0.169	0.182	0.102	5.012	10.357	8.972
CL		0.198	0.177	0.219			
COX	Split	0.196	0.187	0.176	4.321	10.576	
NN	6	0.196	0.186	0.180	4.342	10.318	8.931
XGB		0.200	0.186	0.168	4.368	10.313	8.926

Table 3.6: Results on the case study on real data. For each model (column one) and in each data split (column two) we show the results of $\texttt{ARE}^{\texttt{TOT}}$ (column three), $\texttt{ARE}^{\texttt{CAL}}$ (columns four and five), the in-sample and out-of-sample likelihood (columns six and seven) and the average CRPS (column eight). The $\texttt{ARE}^{\texttt{CAL}}$ is presented quarterly and yearly in columns four and five respectively.

3.7.2 Sensitivities

Inspired by the analysis of the sensitivities in Wüthrich (2018), we show that for a fixed combination of features and a fixed development period, the marginal effect of the different levels of a model features on the development factors for Spit 1. The benchmark is the development factor modelled with the chain ladder. The first row shows our results on a quarterly grid (Figure 3.8a, Figure 3.8b, Figure 3.8c) for the second development quarter. In Figure 3.8a fix the accident quarter to 16, and `cover_key` to 1 and let `claim_type_key` vary. In Figure 3.8b, we pick `cover_key 1`, `claim_type_key 1` and let the accident quarter change. In Figure 3.8c, we pick accident quarter

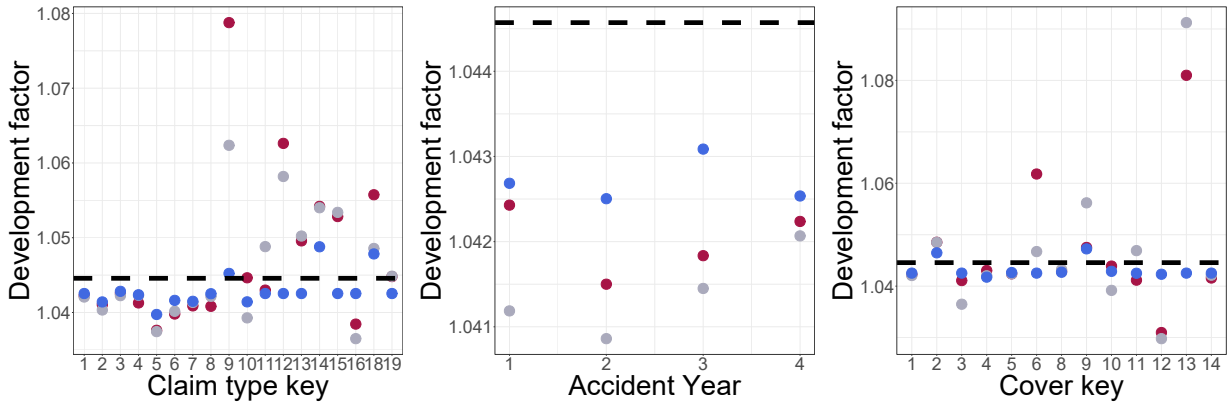
16, `claim_type_key 1` and let the `cover_key`.



(a) (b) (c)

Figure 3.8: Development factor 2 sensitivity for the **quarterly** output in Split 6. The dotted line represents the chain ladder estimate. The dots indicate the estimates from the different models: COX (red), NN (gray) and XGB (blue).

In a similar fashion, in Figure 3.9 we show a similar plot for the yearly results. In Figure 3.9a we fix the accident year to 4, and `cover_key` to 1 and let `claim_type_key` vary. In Figure 3.9b, we pick `cover_key 1`, `claim_type_key 1` and let the accident quarter change. In Figure 3.9c, we pick accident quarter 4, `claim_type_key 1` and let the `cover_key`.



(a) (b) (c)

Figure 3.9: Development factor 2 sensitivity for the **yearly** output in Split 6. The dotted line represents the chain ladder estimate. The dots indicate the estimates from the different models: COX (red), NN (gray) and XGB (blue).

3.8 Concluding remarks

Based on the work of Miranda et al. (2013) and Hiabu (2017), we introduced a survival analysis framework to use machine learning techniques to estimate development factors that can depend on accident date and other features. The approach presented in this paper has been developed with the aim to give higher accuracy in cases where more information in form of individual claims data is available while at the same time conserving the structure reserving actuaries are used to. Our extensive simulation study suggests that our methodology does indeed seem to work well. In

this paper, we have only considered the prediction of IBNR counts and an obvious next step is to integrate our methodology into a wider framework to estimate the outstanding claims amount, which will be the focus of Chapter 4. It could for example be interesting to merge our approach with recent RBNS prediction methods like Crevecoeur, Robben, and Antonio (2022) or Lopez and Milhaud (2021).

Part II

A multi-state model for the cost of individual claims

Chapter 4

An individual model for the claims severity

In the previous sections, we illustrated models for the claim development that allowed us to model feature-based development factors. In this chapter, we propose a multi-state approach for modelling the cost of the individual claims. Our approach builds upon the findings presented in a recent paper (Bladt & Furrer, 2023b), which introduces the conditional Aalen-Johansen estimator. This estimator is a versatile, non-parametric tool for estimating state occupation probabilities conditioned on specific features, and it also discusses its key properties. Note that Aalen-Johansen-type estimators have traditionally found applications in life insurance. However, in this work, we demonstrate how these estimators can also be applied to non-life insurance for reserving purposes. While the three approaches rely on models from survival analysis, we decided to separate this chapter as, similarly to the traditional literature on reserving, we now propose a model for the claim severity.

4.1 Our contribution to multi-state models in insurance

Multi-state models have found widespread use in the domain of life insurance. In the works of Hoem (1969, 1972), they introduce a by-now standard approach, representing biometric states (active, deceased, etc.) as finite states within a spatial model. However, to the best of our knowledge, these models have seen less frequent application in non-life insurance literature. Notable exceptions can be found in the work of Hesselager (1994) and Maciak, Mizera, and Pešta (2022). In Hesselager (1994), the authors model outstanding claim amounts conditionally on the current state, such as Incurred But Not Reported (IBNR), Reported But Not Settled (RBNS), and Settled. In this context, the time variable corresponds to calendar time, similar to that in life insurance. In contrast, our approach in this paper takes a different perspective, replacing time with payments as the fundamental evolution variable. In Maciak et al. (2022), the authors propose modeling the incremental paid amounts as outcomes of a homogeneous Markov chain on a finite state space, though only for aggregated data and in discrete time.

In this manuscript, we propose an individual claims reserving model based on a continuous-time non-explosive pure jump process denoted by J on a finite state space, $\mathcal{S} = \{1, \dots, k\}$, $k \in \mathbb{N}$, where states correspond to the development periods (DP's) within a development triangle, and the "time spent" between state transitions corresponds to the claim size growth between DP's. The state k

serves as an absorbing state, representing claim closure. It is worth noting that, in practice, DPs have a stochastic behavior that leads to variation in the state space dimension across individual claim histories. However, for simplicity, we address this by collapsing long claim developments into a single state, as the specific details are not of primary importance in this context.

In a broader context, the application of counting processes to individual reserving has been a subject of research since the seminal contributions of Arjas (1989); Norberg (1993, 1997), with subsequent refinements and practical considerations detailed in the work of Haastrup and Arjas (1996) and Antonio and Plat (2014). In Antonio and Plat (2014), they assume a framework where claims are generated by a position-dependent marked Poisson process, and they incorporate a severity modelling approach within the mark.

The primary contribution of Bladt and Pittarello (2023) lies in presenting a novel interpretation of state spaces, which enables us to approach the claims reserving problem from a fresh perspective, using a non-parametric estimator from survival analysis. We denote state process by $J(z)$ when the cumulative claim size of an individual reaches z, where $z \in [0, +\infty)$. Note that in this approach, the fundamental variable of interest is the claim size itself, rather than time. The transition probabilities leading to the absorbing state k represent the cumulative distribution of individual claim costs. In other words, as individuals progress through the states, their observations accumulate claim size, reflecting the evolution of their claims. Upon reaching the absorbing state, a claim is considered settled, and no further payments are possible; see Section 4.2.2.

Recent literature on individual reserving has focused on decomposing individual reserving data into different modules that describe the micro-level structure of the portfolio, e.g. payment delay, settlement delay, payment size. By doing so, Huang, Qiu, Wu, and Zhou (2015) and Wang, Wu, and Qiu (2021) extended the discrete model for predicting IBNR claims in Norberg (1986) to a model for RBNS and IBNR claims that discusses, theoretically and with empirical studies, the beneficial effect of including individual information in reserving models. An interesting approach to modelling the joint distribution of reporting delays and claim sizes of RBNS claims under censoring can be found in Lopez (2019). The work in Crevecoeur, Antonio, Desmedt, and Masquelein (2023) embeds Crevecoeur et al. (2022) in a more general context and illustrates how pricing and reserving can be modelled under a unified framework and severity is modelled conditionally on the previous modules. A similar approach has been proposed in Delong, Lindholm, and Wüthrich (2022), wherein the authors incorporate a gamma-distributed severity component into a neural network-based methodology.

In contrast to the aforementioned works, which rely on likelihood-based estimation or parametric assumptions on the claim size distribution, we propose a fully non-parametric estimation of the full cumulative distribution function of claim sizes, building on the results of Bladt and Furrer (2023b). The latter reference proposes a conditional version of the classic Aalen-Johansen estimator, which allows for continuous or discrete covariates, and where minimal assumptions on the process J are imposed, i.e. no Markov assumption is required. In particular, such lax probabilistic framework is desirable when it is unclear whether there exist duration effects in each state, as is the case for claim development processes.

In Section 4.2.4, we also propose a solution for modelling the IBNR and RBNS reserves separately. In Section 4.3 and following, our model is compared with the chain ladder in two case studies, one using simulated data and the other using a recent dataset from a Danish non-life

insurer. In both the simulation case study and the real data case application, we have access to the total future claims costs and can thus compare the predicted amount with the target amount. The predicted curve for the cumulative density function of claims size is compared with the empirical cumulative density function using the Continuously Ranked Probability Score (CRPS), first proposed in Gneiting and Raftery (2007). Interestingly, we show how the CRPS can be used to perform model selection by answering the natural question of how to select the number of states in the state space $\mathcal{S}$.

4.1.1 A comparative literature review

Research papers on loss reserving can be broadly categorized into two main streams of research: models based on development triangles (aggregate loss reserving models), and models based on individual data (individual loss reserving models). The relationship between individual data and aggregate modelling has been extensively studied in Miranda, Nielsen, and Verrall (2012), Hiabu (2017), and Bischofberger et al. (2020), utilizing different survival analysis tools.

The existing body of literature on aggregate reserving has already provided the building blocks for predicting the distribution of loss reserves, with a specific emphasis on models that seek to emulate the empirical chain-ladder algorithm. In this section, we briefly mention some of these models and defer to the comprehensive discussion in Hess and Schmidt (2002) for their detailed categorization. For a distribution-independent approach to computing the mean squared prediction error of the loss reserve, one can refer to Mack (1993), where the author decomposes the prediction error of the loss reserve into its elemental components, namely process variance and estimation variance, as expressed by Wüthrich and Merz (2015, p. 268, eq. 9.10). The model has received several statistical critiques through the years but remains one of the best-performing reserving models in an actuary's toolkit.

In England and Verrall (1999), the authors replicate the point estimates of the chain-ladder method and ingeniously employ bootstrap and generalized linear models (GLM) to compute the estimation variance. For an extension of the model, allowing for negative payments as well as an exploration of the model assumptions underpinning it, one may consult Verrall (2000). An interesting comparison of the fundamental assumptions supporting the aforementioned models is presented in Mack and Venter (2000), where the authors conclude that only the approach outlined in Mack (1993) is the true underlying model of the chain-ladder algorithm. In a more recent development, in Al-Mudafer, Avanzi, Taylor, and Wong (2022), the approach in England and Verrall (1999) is integrated into the framework of mixture density networks Bishop (1994).

Shifting our focus to non-parametric models for individual reserving, the current discussion primarily centers on providing precise point forecasts for the loss reserve. In contrast, Delong and Wüthrich (2023) discusses the role of the variance function in regression problems for pricing.

In another recent development, Wüthrich (2018) incorporates neural networks into the chain-ladder model, introducing features to the existing methodology. Further related work can be found in Lopez, Milhaud, and Thérond (2019) and Lopez and Milhaud (2021), where the authors employ a tree-based algorithm for modelling RBNS claims using censored data. In particular, in Lopez and Milhaud (2021), the use of bootstrapping for estimating the uncertainty in the claims reserve is discussed.

Bladt and Pittarello (2023) introduces a novel approach to the individual reserving problem,

allowing us to explicitly model curves of individual claim costs denoted as $p_k(z|x)$ within our framework.

4.2 The conditional Aalen-Johansen

In this section we introduce our statistical framework, and the quantities of interest, the conditional occupation probabilities. We present the conditional Aalen-Johansen Bladt and Furrer (2023b), a non parametric estimator for the conditional occupation probabilities of an arbitrary jump process. Note that the conditional Aalen-Johansen is closely linked with the literature streams on kernel-based estimation for survival and semi-Markov processes McKeague and Utikal (1990); Nielsen and Linton (1995); Dabrowska (1997), as well as landmarking Van Houwelingen (2007). We conclude the section with some remarks on the interpretation of multi-state models for reserving. As we explain in the next paragraphs, this is a fundamental step to construct our reserving model.

4.2.1 A reserving model based on the conditional Aalen-Johansen

Let $J = (J_z)_{z\geqslant 0}$ be a non-explosive pure jump process on a finite state space $\mathcal{S}$, and take $(\Omega, \mathcal{F}, \mathbb{P})$ to be the underlying probability space. Here, we let $\mathcal{S} = \{1, \dots, k\}$, with $k \in \mathbb{N}$. Denote by Y the possibly infinite absorption time of J. Let X be a random variable with values in $\mathbb{R}^d$ equipped with the Borel $\sigma-$algebra. We assume that the distribution of X admits a density g with respect to the Lebesgue measure λ. Absolute continuity to the case of known atoms can be relaxed as discussed in Bladt and Furrer (2023b).

We introduce a multivariate counting process N with components $N_{j,h} = (N_{j,h}(z))_{z\geqslant 0}$ given by

$$N_{j,h}(z) = \#\left\{s \in (0, z] : J_{s-} = j, J_s = h\right\}, \quad z \geqslant 0,$$

for $j, h \in \mathcal{S}, j \neq h$. Let us introduce the following assumptions.

Assumption 4.2.1. For some $\delta > 0$, it holds that

$$\mathbb{E}\left[N_{j,h}(z)^{1+\delta}\right] < \infty.$$

Define the conditional occupation probabilities according to

$$p_j(z \mid x) = \mathbb{E}\left[I(J_z = j) \mid X = x\right].$$

We denote by $p(z \mid x)$ the row vector with elements $p_j(z \mid x)$. We now introduce the cumulative conditional transition rates and cast the conditional occupation probabilities as a product integral of these cumulative rates. To this end, let

$$p_{j,h}(z \mid x) = \mathbb{E}\left[N_{j,h}(z) \mid X = x\right].$$

Define the cumulative conditional transition rates via

$$\Lambda_{j,h}(z \mid x) = \int_0^z \frac{1}{p_j(s- \mid x)} p_{j,h}(\,\mathrm{d}s \mid x).$$

A cumulative transion hazard matrix, is defined by having the j, j entry $\Lambda_{j,j}(z \mid x) = \sum_{j \neq} \Lambda_{j,h}(z \mid x)$ for $j, h \in \mathcal{S}, j \neq h$ (Overgaard, 2019).

Let $q(\cdot \mid x)$ be the interval function associated with the conditional transition probabilities:

$$q_{j,h}(s, z \mid x) = I(p_j(s) > 0)\frac{\mathbb{E}\left[I(J_s = j)I(J_z = h) \mid X = x\right]}{p_j(s|x)} + I(p_j(s|x) = 0)I(j = h).$$

In Overgaard (2019), it is shown that this assumption ensures that the product integral

$$\prod_0^t (\mathrm{Id} + \Lambda(\mathrm{d}s \mid x))$$

where Id denotes the identity matrix is well-defined. Therefore,

$$p(t \mid x) = p(0 \mid x) \prod_0^t (\mathrm{Id} + \Lambda(\mathrm{d}s \mid x)). \tag{4.1}$$

We are interested in the situation where the observation of J is right-censored. To this end we introduce a strictly positive random variable W describing right-censoring, so that we actually observe the triplet

$$\left(X, (J_z)_{0 \leqslant z \leqslant W}, Y \wedge W\right).$$

Whenever the Markov assumption fails, the following assumption is key for our estimation procedure.

Assumption 4.2.2. Right-censoring is conditionally entirely random:

$$W \perp\!\!\!\perp J \mid X.$$

Remark 4.2.1. From a reserving perspective, the above assumption amounts to stating that, conditional on covariates, closure of a claim is simply specified through an independent random mechanism. One should note that this assumption is not verifiable, despite its widespread use in survival analysis.

We introduce

$$p_j^c(z \mid x) = \mathbb{E}\left[I(J_z = j)I(z < W) \mid X = x\right],$$

and

$$p_{j,h}^c(z \mid x) = \mathbb{E}\left[N_{j,h}(z \wedge W) \mid X = x\right].$$

from which we may conclude that

$$\Lambda_{j,h}(z \mid x) = \int_0^z \frac{1}{p_j^c(s- \mid x)} p_{j,h}^c(\,\mathrm{d}s \mid x), \tag{4.2}$$

within the interior of the support of W.

Estimators

Consider the i.i.d. replicates $\left(X^i, \left(J^i_z\right)_{0\leqslant z\leqslant W^i}, Y^i \wedge W^i\right)_{i=1}^n$. Define $\delta_i := I(Y^i \leq W^i)$, which equals 1 when the observation is absorbed (closed claim), and zero otherwise. In a reserving context, this distinction is usually reflected by specifying two types of claims: closed claims and RBNS (open) claims, and we write respectively $n = n^{\texttt{Closed}} + n^{\texttt{RBNS}}$.

We now construct an estimator for the conditional occupation probabilities $p_j(z \mid x)$, but from Equation (4.1), it essentially suffices to estimate the cumulative conditional transition rates $\Lambda_{j,h}(z \mid x)$, which we do using Equation (4.2). Let K_i be kernel functions of bounded variation and bounded support, and let (a_n) be a bandwidth sequence. We impose the following standard assumption, confering with Stute (1986).

Assumption 4.2.3. The band sequence satisfies:

1. $a_n \to 0$ as $n \to \infty$.
2. $na_n^d \to \infty$ as $n \to \infty$.
3. $\sum_{n=1}^\infty c_n^r < \infty$ for some $r \in (1, 1+\delta)$ where $c_n := \left(na_n^d\right)^{-1} \log n$

One may for $(1+\delta)^{-1} < \eta < 1$ take

$$a_n^d \sim \frac{\log n}{n^{1-\eta}}.$$

We introduce the usual kernel estimator for the density g:

$$\mathrm{g}^{(n)}(x) = \frac{1}{n}\sum_{i=1}^n \mathrm{g}^{(n,i)}(x) = \frac{1}{n}\sum_{i=1}^n \prod_{b=1}^d \frac{1}{a_n} K_b\left(\frac{x_b - X_b^i}{a_n}\right).$$

It is well-known (cf. (Ghosh, 2018)) that

$$\mathrm{g}^{(n)}(x) \overset{\text{a.s.}}{\to} g(x), \quad n \to \infty,$$

and, for any random variable V with i.i.d. replicates (V^i) satisfying $\mathbb{E}[|V|^{1+\delta}] < \infty$ that

$$\frac{1}{n}\frac{\sum_{i=1}^n V^i \mathrm{g}^{(n,i)}(x)}{\mathrm{g}^{(n)}(x)} \overset{\text{a.s.}}{\to} \mathbb{E}[V \mid X = x], \quad n \to \infty. \tag{4.3}$$

Based on the kernel estimator for g, we form the following Nadaraya-Watson type kernel estimators:

$$\mathbb{N}_{j,h}^{(n)}(z \mid x) = \frac{1}{n}\sum_{i=1}^n \mathbb{N}_{j,h}^{(n,i)}(z \mid x) = \frac{1}{n}\sum_{i=1}^n \frac{N_{j,h}^i\left(z \wedge W^i\right)\mathrm{g}^{(n,i)}(x)}{\mathrm{g}^{(n)}(x)} \tag{4.4}$$

$$\mathbb{I}_j^{(n)}(z \mid x) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}_j^{(n,i)}(z \mid x) = \frac{1}{n}\sum_{i=1}^n \frac{I(z < W^i) I(J_z^i = j)\mathrm{g}^{(n,i)}(x)}{\mathrm{g}^{(n)}(x)} \tag{4.5}$$

$$\tag{4.6}$$

Remark 4.2.2. Due to Assumption 1, it follows from Equation (4.3) that

$$\mathbb{N}_{j,h}^{(n)}(z \mid x) \overset{\text{a.s.}}{\to} p_{j,h}^{c}(z \mid x), \quad n \to \infty.$$

Furthermore, it holds that

$$\mathbb{I}_{j}^{(n)}(z \mid x) \overset{\text{a.s.}}{\to} p_{j}^{c}(z \mid x), \quad n \to \infty.$$

Definition 1. The conditional Nelson-Aalen estimator for the cumulative conditional transition rates is defined as

$$\mathbb{A}_{j,h}^{(n)}(z \mid x) = \int_0^z \frac{1}{\mathbb{I}_{j}^{(n)}(s- \mid x)} \mathbb{N}_{j,h}^{(n)}(\mathrm{d}s \mid x).$$

Definition 2. The conditional Aalen-Johansen estimator for the conditional occupation probabilities is defined according Equation (4.1), but with the cumulative conditional transition rates replaced by the conditional Nelson-Aalen estimator, that is

$$\mathbb{p}^{(n)}(z \mid x) = \mathbb{p}^{(n)}(0 \mid x) \prod_0^z \left(\mathrm{Id} + \mathbb{A}^{(n)}(\mathrm{d}s \mid x)\right),$$

where $\mathbb{p}_j^{(n)}(0 \mid x) = \mathbb{I}_j^{(n)}(0 \mid x)$.

Remark 4.2.3. We refer to Bladt and Furrer (2023b) for the proof that the estimators are asymptotically strongly consistent and pathwise Gaussian distributed.

When all the subjects start in state 1 at inception, we can use the estimator of the occupation probabilities of the absorbing state k, $\mathbb{p}_k^{(n)}$ to model the cumulative density function of the individual claims cost.

Remark 4.2.4. The Volterra integral equation defines the product integral of the matrix function $\Lambda(\mathrm{d}s \mid x)$ as the unique solution $p(\cdot \mid x)$ to the equation

$$p(t \mid x) = \mathrm{Id} + \int_{(0,t]} p(s- \mid x)\Lambda(\mathrm{d}s \mid x).$$

However, when a piecewise constant estimator of Λ is given by $\mathbb{A}^{(n)}$, as is described in detail in the previous section, the computation of the product integral simplifies. More precisely, denote $0 = t_0 < t_1 < t_2 < \cdots$ the sequence of time points describing all observed jump times in the sample. Then the recursion

$$\begin{aligned}
&\mathbb{p}_j^{(n)}(t_{\ell+1}|x) - \mathbb{p}_j^{(n)}(t_\ell|x) \\
&= \sum_{\substack{k \in \mathcal{Z} \\ k \neq j}} \mathbb{p}_k^{(n)}(t_\ell|x)\Delta\mathbb{A}_{kj}^{(n)}(t_{\ell+1}|x) - \mathbb{p}_j^{(n)}(t_\ell|x) \sum_{\substack{k \in \mathcal{Z} \\ k \neq j}} \Delta\mathbb{A}_{jk}^{(n)}(t_{\ell+1}|x)
\end{aligned}$$

and starting value $\mathbb{p}^{(n)}(0|x) = \mathbb{I}^{(n)}(0|x)$ completely characterize the product integral $\mathbb{p}_j^{(n)}(\cdot|x)$ at jump times (and constant otherwise).

4.2.2 Practical considerations for claims development

In this section, we offer practical insights into our modelling approach, with the aim of assisting the reader in understanding our multi-state setup for claims reserving. Refer to Figure 4.1, which conceptually illustrates our model featuring a state space $\mathcal{S}$ with $k = 5$ states.

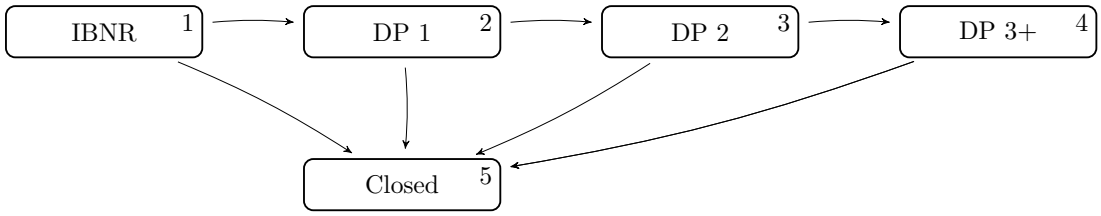


Figure 4.1: Example of multi-state model for claims reserving, with $k = 5$. We interpret time spent in each state as increasing claims size clock, instead of calendar time. The states represent the development periods (DPs) of the individual claims.

In our framework, we represent the active state as IBNR claims, while we treat RBNS claims as transitioning through intermediate states. Any claim that reaches state k is considered closed in our system due to the presence of an absorbing state. Consequently, we do not allow claims to reopen in our framework. Notably, unlike a traditional survival study, we do not explicitly model the time component, and the number of IBNR claims, along with their characteristics, remains unknown at the time of evaluation. In Section 4.2.4, we delve into how to model RBNS and IBNR claims, showcasing how our model can accommodate such scenarios. Furthermore, we have designed our state space so that state DP3+ at step $k-1$ in Figure 4.1 is the final relevant DP that we explicitly model.

Finally, it is worth mentioning that we have not introduced development triangles yet, which are the standard aggregate datasets commonly used in claims reserving applications. The choice of parameter k aligns with the data that, in a conventional modelling approach, would be aggregated into development triangles of varying sizes. Since our proposed framework does not explicitly model the time component, we assume that payments close at the development year following the last payment. This assumption results in a development triangle with $k-1$ development times. This aspect will be particularly relevant for our data application, where we compare our model to the benchmark chain ladder method outlined in Mack (1993), and we need to construct a development triangle for this purpose. Furthermore, we observe that $k - 1 = m$ defined in previous chapters.

Remark 4.2.5. By changing the multi-state model in Figure 4.1, our approach can be modified to take into account additional domain-specific aspects of reserving data. For instance, if we allow transitions back from the state k (closed) to states $3, \dots, k-2$ (DP 2, ...), our framework can be modified to take into account reopenings.

4.2.3 Notation and further connections to reserving

Together with our sample, which is composed of n i.i.d. observations generated from J and its covariates (refer to the previous sections), we additionally observe the following quantities already defined in Chapter 3 that are *not* related to $J^1, \dots, J^n$:

$$\left\{T_i^R, U_i\right\}_{i=1}^n.$$

We recall that U_i and T_i^R represent the accident period and the reporting delay of the i-th claim, with $U_i, T_i^R \in \{1, \ldots, k-1\}$.

We can also provide a definition of the development triangle of cumulative payments with respect to our sample,

$$\mathcal{C}^{(k-1)} = \{C_{\ell,j} : \ell + j \leq k; \ell, j = 1, \ldots, k-1\},$$

where $C_{\ell,j} = \sum_{i=1}^{n} I(U_i = \ell \text{ and } J^i_{z_{i,j}} = j) z_{i,j}$, with $0 \leq z_{i,j} \leq W^i$ such that $J^i_{z^-_{i,j}} \in \{1, \ldots, j-1\}, J^i_{z_{i,j}} = j$ denoting the total cumulative amount paid in accident period ℓ and DP j. We are presently interested forecasting the ultimate cost of our claims,

$$Y^{\texttt{Closed}} + Y^{\texttt{RBNS}} := \sum_{i=1}^{n} \delta_i Y^i + \sum_{i=1}^{n} (1 - \delta_i) Y^i = \sum_{\ell=1}^{k-1} C_{\ell,k-1},$$

The claims reserve is simply $R = \sum_{\ell=1}^{k-1} (C_{\ell,k-1} - C_{\ell,k-\ell})$.

4.2.4 Prediction of RBNS and IBNR claim costs

We can calculate the final cost of RBNS claims numerically using the conditional Aalen-Johansen estimate $\mathbb{p}_k^{(n)}(z \mid x)$ for $p_k(z|x)$. The estimator of the final total cost of RBNS claims is defined by

$$\hat{Y}^{\texttt{RBNS}} = \sum_{i}^{n} I(\delta_i = 0) \left(W^i + \widehat{\mathbb{E}}[Y | Y > W^i, X = x] \right)$$

with

$$\widehat{\mathbb{E}}\left[Y | Y > W^i, X = x\right] = \frac{1}{1 - \mathbb{p}_k^{(n)}(W^i \mid x)} \int_{W^i}^{+\infty} (1 - \mathbb{p}_k^{(n)}(y \mid x)) \mathrm{d}y. \tag{4.7}$$

The general formula for the m-th moment is explicitly given by

$$\frac{1}{m} \widehat{\mathbb{E}}\left[Y^m | Y > W^i, X = x\right] = \frac{1}{1 - \mathbb{p}_k^{(n)}(W^i \mid x)} \int_{W^i}^{+\infty} y^{m-1} (1 - \mathbb{p}_k^{(n)}(y \mid x)) \mathrm{d}y.$$

Strictly speaking, from a reserving standpoint, we also need to propose a strategy for reserving for the (unknown) IBNR claims, which we denote by the random variable $n^{\texttt{IBNR}}$ with values in $\mathbb{N}_0$. The individual cost of the IBNR claims is described by a sequence of iid copies of the random variable $\tilde{Y}$, namely $\left\{\tilde{Y}^{i'}\right\}_{i' \in \mathbb{N}}$. Under the framework of Bladt and Furrer (2023b), $\tilde{Y}$ has cumulative distribution function $p_k(z)$. As we will describe shortly, we specify a model that does not consider the features, since in this case we do not have access to future exposures by feature. We propose to correct the model for the random presence of IBNR using a compound distribution, under the assumptions in Klugman, Panjer, and Willmot (2012, p. 148). We describe the total cost of IBNR claims as follows:

$$Y^{\texttt{IBNR}} = \sum_{i'=1}^{n^{\texttt{IBNR}}} \tilde{Y}^{i'}.$$

The expression for $Y^{\texttt{IBNR}}$ is well known in non-life mathematics and is often referred to as the collective risk model, Parodi (2014). The moments of $Y^{\texttt{IBNR}}$ can be calculated in closed form. In

particular, the expected cost of IBNR claims is

$$\hat{Y}^{\texttt{IBNR}} = \widehat{\mathbb{E}}[Y^{\texttt{IBNR}}] = \hat{n}^{\texttt{IBNR}}\widehat{\mathbb{E}}[\tilde{Y}],$$

with the m-th moment of $\tilde{Y}$ defined by

$$\frac{1}{m}\widehat{\mathbb{E}}[\tilde{Y}^m] = \int_0^\infty y^{m-1}(1 - \mathbb{p}_k^{(n)}(y))\mathrm{d}y.$$

The variance of the IBNR claims is

$$\widehat{\mathrm{VAR}}(Y^{\texttt{IBNR}}) = \hat{n}^{\texttt{IBNR}}\left[\widehat{\mathbb{E}}[\tilde{Y}^2] - \widehat{\mathbb{E}}[\tilde{Y}]^2\right] + \widehat{\mathbb{E}}[\tilde{Y}]^2\widehat{\mathrm{VAR}}(n^{\texttt{IBNR}}).$$

To project the number of IBNR claims, we simply use the chain ladder model on the triangle of counts $\mathcal{D}^{(k-1)}$ and obtain $\hat{n}^{\texttt{IBNR}}$. For each of the future reported claims, we fit the unconditional Aalen-Johansen to the sample n to obtain $\mathbb{p}_k^{(n)}(z)$. The total size of claims is $Y^{\texttt{TOT}} = Y^{\texttt{Closed}} + Y^{\texttt{RBNS}} + Y^{\texttt{IBNR}}$ and we estimate it with $\hat{Y}^{\texttt{TOT}} = Y^{\texttt{Closed}} + \hat{Y}^{\texttt{RBNS}} + \hat{Y}^{\texttt{IBNR}}$. We remark that in several lines of business the contribution of the true cost of IBNR claims ($Y^{\texttt{IBNR}}$) to $Y^{\texttt{TOT}}$ is minor compared to the cost of RBNS claims ($Y^{\texttt{RBNS}}$). This is the case for those lines of business that have data with mostly single payments and a short time lag between reporting and settlement. See for instance the results presented in Friedland (2010, p. 80) on the US market data, comparing for different lines of insurance the size of the estimated cost of IBNR claims to paid and outstanding claims. As we will show in Section 4.4, our real dataset exhibits exactly this behaviour and the IBNR correction will have a minor impact on the estimated final cost.

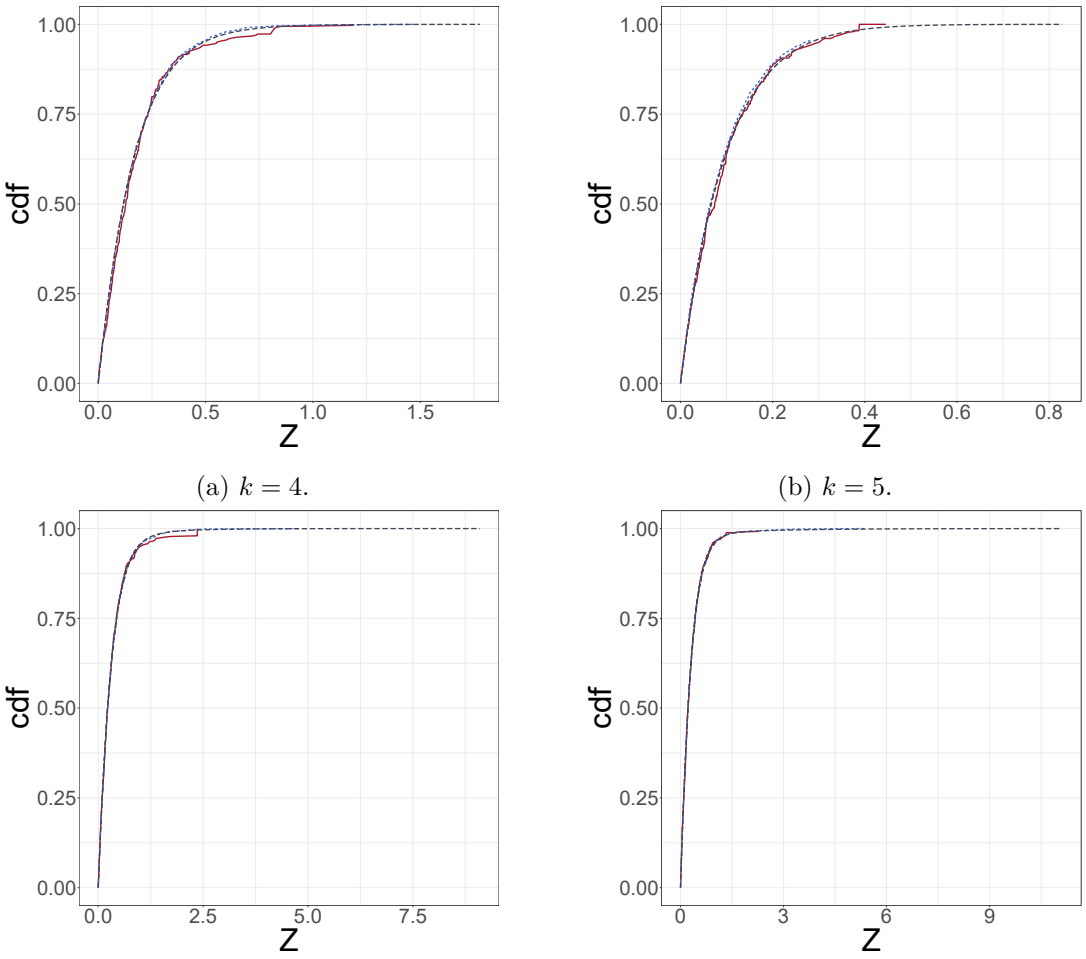
Remark 4.2.6. Relatable approaches to our proposed estimator in Equation (4.7) can be found in the literature on censored regression. Similar techniques are used for the estimation of the linear model in Buckley and James (1979), and for the construction of the *artificial data points* for estimation of the non-linear model in Heuchenne and Van Keilegom (2007).

4.3 Simulation of RBNS claims

In this section, we perform two a simulation studies to provide numerical evidence of the applicability of our models to claims reserving. To demonstrate the potential benefits to actuaries of using our modelling approach, we first compare our results with the chain ladder model presented in Section 1.3.1 (Mack, 1993). A further comparison with the hierarchical reserving model in Crevecoeur et al. (2022) is included in Appendix Appendix G. To facilitate the presentation of our results, we refer to our model as AJ and the chain ladder model in Section 1.3.1 as CL, and to the hierarchical model as `hirem`.

4.3.1 Implementation

It is possible to simulate observations from a jump-process model using the function `sim_path` from the `R` package `AalenJohansen` Bladt and Furrer (2023a), by introducing a matrix of transition intensities. Markov and semi-Markov models are allowed. To obtain a realistic set of examples for the simulation study, we used the Nelson-Aalen implementation of the `R` package `survival` Therneau



(a) $k = 4$. (b) $k = 5$.

(c) $k = 6$. (d) $k = 7$.

Figure 4.2: Comparison of the true severity curve (dark grey dotted line) to the fitted severity curve for different data sizes in the 4 simulated scenarios. The red curve is fitted on $120(k-1)$ RBNS claims, the blue curve is fitted on 1200 RBNS claims. Z is rounded by millions.

(2023) to estimate the (unconditional) cumulative transition intensities $\Lambda_{j,h}^{(n)}(z)$ for different choices of $k = 4, 5, 6, 7$ on the real dataset available to us for writing this manuscript. Then, similar choices of parameters were employed for the present section. See Table 4.3 for the description of the real data. The matrix of intensities can be obtained from the cumulative intensities $\Lambda_{j,h}^{(n)}(z)$. Simulation by knowing the data generation process underlying the data is important to understand model performance compared to the true estimation target. For example, in Figure 4.2 we show an example of the fitted $\mathbb{p}_k^{(n)}(z \mid x)$ on datasets simulated directly with Λ without any feature effect. For each choice of k we simulated two datasets. The first dataset contains $120(k-1)$ observations. The second dataset contains $1200(k-1)$ observations. By comparing the $\mathbb{p}_k^{(n)}(z \mid x)$ fitted to the smallest sample (red) with the curve fitted to the largest sample (blue), we observe that increasing the sample size yields fits that are closer to the true $p_k(z \mid x)$, empirically validating the consistency of the Aalen-Johansen estimator with claim size as operational time. In particular, we have presented a model that can take into account the individual feature information. In the case studies that we present in the next section, we treat the accident period as a feature and add an effect by simply specifying $q_{j,h}(s, z \mid x) = q_{j,h}(s, z)f(x)$, where $f(x)$ is a function of some feature x and we can

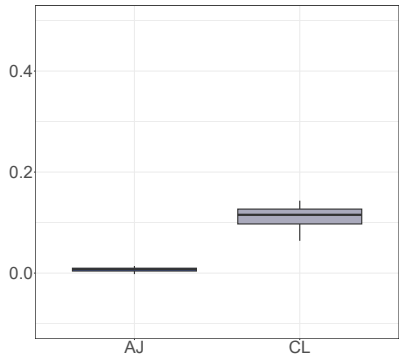
derive $q_{j,h}(s,z)$ from the cumulative transition matrix obtained with the `survival` package. In the simulations we present, we assume that observations are generated from $f(x) = \frac{1}{10+k-x}$, where in place of x we use the accident period, namely U_i introduced in the previous sections. For the simulation of the reserving data sets, we decide arbitrarily the number of observations in each accident period U_i. Each claimant is then associated with a numeric variable representing the accident period U_i, taking values in $\{1\ldots,k-1\}$, from which we can calculate the effect $f(U_i) = \frac{1}{10+k-U_i}$. We also add an effect on the volumes, and generate fewer observations in most recent accident period. Generating decreasing volumes resembles cycles in the underwriting, a case study of an insurer that underwrites less policies in the more recent accident periods, resulting in less claims. In particular, we generated 1200 observations in the first accident period $U_i = 1$ and decreased by 100 the number of observations every accident periods. For instance, we will have 1100 observations with $U_i = 2$ and 1000 observations with $U_i = 3$.

4.3.2 Evaluating the models performance

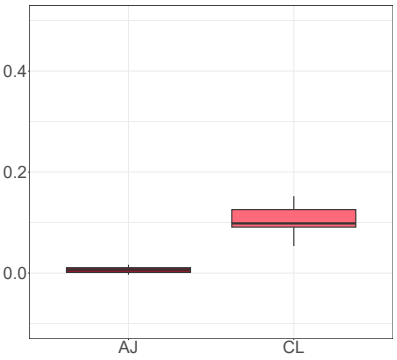
Using the approach introduced in Section 4.3.1, we present two simulated scenarios: Alpha and Beta. In scenario Alpha, we simulate directly with unconditional transition intensities that we have fitted to the real data. In the Beta scenario we introduce the effect of the accident period characteristics and the occupation intensities. In each scenario we simulate, for the different choices of $k = 4, 5, 6, 7$, a total number of 20 datasets to smooth possible fluctuations due to the random generator and report the average results that we obtain. Each dataset in scenario Alpha contains $1200(k-1)$ observations. The dataset in scenario Beta contains $1200(k-1)-100(k-2)$ observations. As motivated in Section 4.3.1, we let the volumes decrease in recent accident periods. We chose arbitrarily to decrease the volume by 100 in every accident period subsequent to the first one. The choice of measure of model performance can be an ardouos task when models are very different, as are AJ and CL. The first performance measure that we propose is the error incidence,

$$\texttt{EI} = \frac{\hat{Y}^{\texttt{TOT}}}{Y^{\texttt{TOT}}} - 1,$$

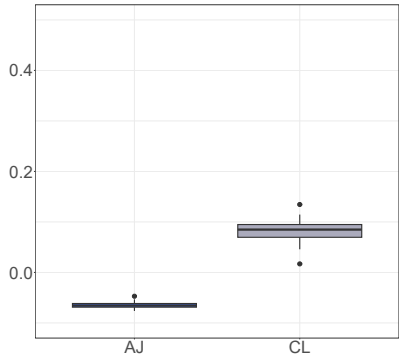
where $\hat{Y}^{\texttt{TOT}}$ is the predicted final claim size and $Y^{\texttt{TOT}}$ is the actual final claim size, which is available from the simulation. The `EI` is an interesting benchmark that provides insight into the relative accuracy of the AJ compared to the CL, though it does not highlight the fact that AJ provides a much more nuanced description of the claim size than point estimates. In Figure 4.3 we provide a box plot of the `EI` for the Alpha scenario (left side) and the Beta scenario (right side) for the different values of k. Compared to the CL, the AJ model performs favourably for each choice of k and in both scenarios. While the CL tends to overestimate $Y^{\texttt{TOT}}$ for all the choices of k, the AJ model tends to provide slightly underestimated results for choices of k bigger than 4. While the `EI` is an interesting benchmark, to better understand the performance of the models compared to the data at hand, we would ideally like to use a proper scoring measure Gneiting and Raftery (2007). In principle, a scoring metric with the property of being proper is able to evaluate the better model by assigning better scores to better models. In this manuscript we consider the Continuously Ranked



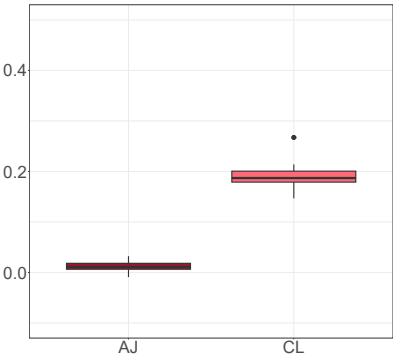
(a) $k = 4$, no features.



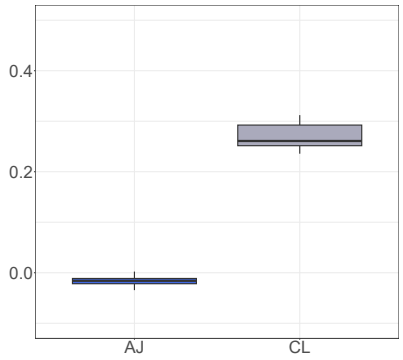
(b) $k = 4$, accident period trend.



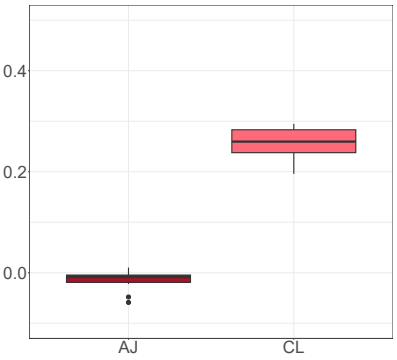
(c) $k = 5$, no features.



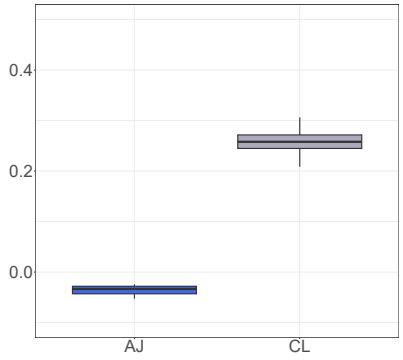
(d) $k = 5$, accident period trend.



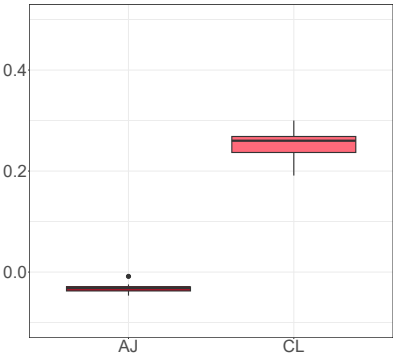
(e) $k = 6$, no features.



(f) $k = 6$, accident period trend.



(g) $k = 7$, no features.



(h) $k = 7$, accident period trend.

Figure 4.3: Box plots of `EI` for AJ and CL over the 20 simulations, for each value of k, in the Alpha scenario (left column) and the Beta scenario (right column).

Probability Score (CRPS) as a proper scoring metric:

$$\begin{aligned}\text{CRPS}(\mathbb{p}_k^{(n)}(\cdot \mid x), y) &= \int_0^{+\infty} (\mathbb{p}_k^{(n)}(z \mid x) - I(y \leq z))^2 \mathrm{d}z \\ &= \int_0^{y} \left(\mathbb{p}_k^{(n)}(z \mid x)\right)^2 \mathrm{d}z + \int_y^{+\infty} \left(1 - \mathbb{p}_k^{(n)}(z \mid x)\right)^2 \mathrm{d}z,\end{aligned}$$

where $\mathbb{p}_k^{(n)}(\cdot \mid x)$ is the predicted curve of the individual claims sizes and y is the (true) value of the final claim size that is available from the simulation. The interested reader can refer to the established statistical literature on proper scoring for a detailed discussion on this topic Gneiting et al. (2004); Gneiting and Raftery (2007); Gneiting and Ranjan (2011).

4.3.3 Empirical analysis

The average results of our 20 simulations for the Alpha scenario are reported in Table 4.1. Each column of the table represents one of the AJ models fitted for the different choices of k (column one). The target $Y^{\texttt{TOT}}$ (average simulated total cost) assuming $Y^{\texttt{IBNR}} = 0$ is in column four. The predicted final cost is $\hat{Y}^{\texttt{TOT}} = Y^{\texttt{Closed}} + Y^{\hat{\texttt{RBNS}}}$. The `EI` results show that, on average, our models outperform the CL for every choice of k except $k = 5$ (columns five and six). For all choices of k, we find that the relative variability of our model is less than that provided by the CL (columns seven and eight). We compute $SD(Y^{\texttt{TOT}}) = \sum_i^{n^{\texttt{RBNS}}} SD(Y^i)$. In this manuscript we will not discuss the estimation error, and the results we present with respect to the variability of the reserve refer only to the process variance. In column nine we report the (average) CRPS. Table 4.2 reports

k	Scenario	U	Actual $Y^{\texttt{TOT}}$ (average)	`EI` (AJ)	`EI` (CL)	$\hat{SD}(Y^{\texttt{TOT}})/\hat{Y}^{\texttt{TOT}}$ (AJ)	$\hat{SD}(Y^{\texttt{TOT}})/\hat{Y}^{\texttt{TOT}}$ (CL)	CRPS (average)
4	Alpha	✗	598.483	0.006	0.111	0.008	0.009	0.145
5		✗	456.746	-0.085	0.082	0.007	0.008	0.095
6		✗	1952.683	-0.016	0.269	0.007	0.009	0.410
7		✗	2359.151	-0.035	0.257	0.007	0.007	0.557

Table 4.1: Results for scenario Alpha. For each value of k (column one) we present the average results over the 20 simulations. Each row of the table corresponds to a different AJ model specification (column three). The table includes the (average) actual $Y^{\texttt{TOT}}$ simulated total cost (column four) and the error incidence for the AJ and the CL (columns five and six). In columns seven and eight we reported the coefficients of variation of $Y^{\texttt{TOT}}$. The results for the CRPS are reported in column nine.

the average results for the 20 simulations of the Beta scenario. For each simulation we fit two models, with and without the use of U as a feature (✓and ✗ in the table). The model with U is correctly specified assuming the Beta scenario. The aim of this experiment is to show that the CRPS is able to judge which is the better model in a context where we know the data generation process. As expected, we obtain a lower average CRPS for the models that include features. The (average) target $Y^{\texttt{TOT}}$ is reported in the fourth column. The results for `EI` indicate that the AJ model generally outperforms the CL model. The relative variability results show comparable results for AJ and CL.

k	Scenario	U	Actual $Y^{\texttt{TOT}}$ (average)	`EI` (AJ)	`EI` (CL)	$\hat{SD}(Y^{\texttt{TOT}})/\hat{Y}^{\texttt{TOT}}$ (AJ)	$\hat{SD}(Y^{\texttt{TOT}})/\hat{Y}^{\texttt{TOT}}$ (CL)	CRPS (average)
4	Beta	✓	42.279	0.006	0.103	0.008	0.008	0.011
		✗		0.005		0.008		0.011
5		✓	94.580	0.011	0.190	0.009	0.008	0.024
		✗		-0.006		0.007		0.025
6		✓	108.423	-0.013	0.256	0.007	0.008	0.028
		✗		-0.016		0.007		0.029
7		✓	116.032	-0.033	0.255	0.007	0.008	0.030
		✗		-0.031		0.008		0.038

Table 4.2: Results for scenario Beta. For each value of k (column one) we present the average results over the 20 simulations. Each row of the table corresponds to a different AJ model specification (column three). The table includes the (average) actual simulated total cost (column four) and the error incidence for the AJ and the CL (columns five and six). In columns seven and eight we reported the coefficients ofvariation of $Y^{\texttt{TOT}}$. The results for the CRPS are reported in column nine.

4.4 A data application on an insurance portfolio

In this section we propose a case study on our model using real data. For this data application, we have at our disposal a recent real dataset from a Danish non-life insurer. The dataset is not publicly available. We provide additional insight in Figure 4.4. The exploratory analysis of the

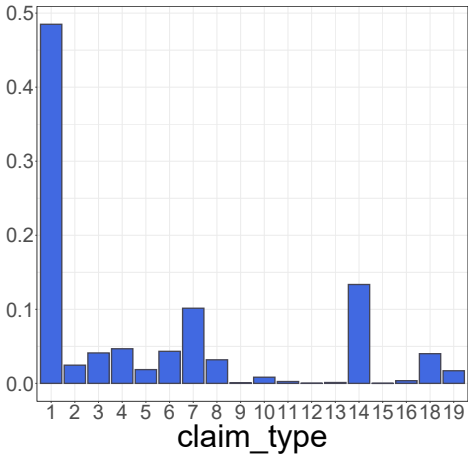
Covariates	Description
`Claim_number`	Policy identifier.
`claim_type` $\in \{1, \dots, 20\}$	Type of claim.
`AM`	Accident month.
`CM`	Calendar month of report.
`DM`	Development month.
`incPaid`	Incremental paid amount.
`Delta`	Indicator, 0 when the claim is open.

Table 4.3: Description of the real dataset.

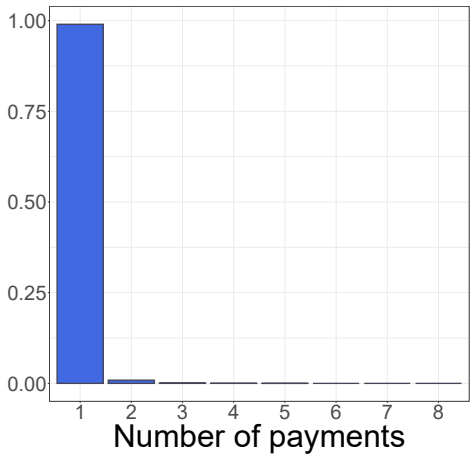
data shows that most of the records have a single payment (Figure 4.4b) and a very fast settlement (Figure 4.4d). The data we present in this manuscript comes from a very stable line of business. Interestingly, we show that we are able to outperform the CL on data where the CL is expected to behave as well as it can. Most of the datasets belong to `claim_type 1`, see Figure 4.4a.

To analyse the real dataset, we propose two strategies that an actuary could follow to calculate the loss reserve. In Section 4.4.1 we show an example where we censor our data by different calendar periods and fit an AJ model for the maximum depth of data available. For example, for a model with $k = 4$ we will have the individual data corresponding to a development triangle with 3 accident periods. The purpose of this application is to show the behaviour of our model compared to the CL on different datasets. In Section 4.4.2 we conclude with an application where we highlight the use of the CRPS measure to perform model selection, after having censored the data at an arbitrary calendar time, i.e. 5 calendar periods. Here we fit different models to the same datasets and select the best performing model in terms of CRPS to calculate the loss reserve. The applications in this section also comprise the IBNR model that we have illustrated in the previous sections. In fact, we predict $\hat{Y}^{\texttt{TOT}} = Y^{\texttt{Closed}} + \hat{Y}^{\texttt{RBNS}} + \hat{Y}^{\texttt{IBNR}}$ and $\widehat{\text{VAR}}(Y^{\texttt{TOT}}) = \widehat{\text{VAR}}(Y^{\texttt{RBNS}}) + \widehat{\text{VAR}}(Y^{\texttt{IBNR}})$.

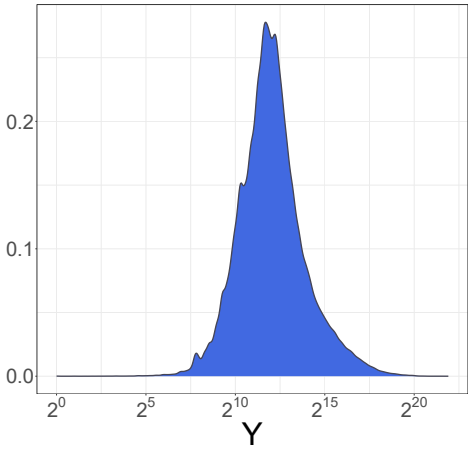
Remark 4.4.1. Notice that the covariate in our application is discrete, which with a uniform kernel effectively amounts to subsampling of data. However, the estimators of the paper provide



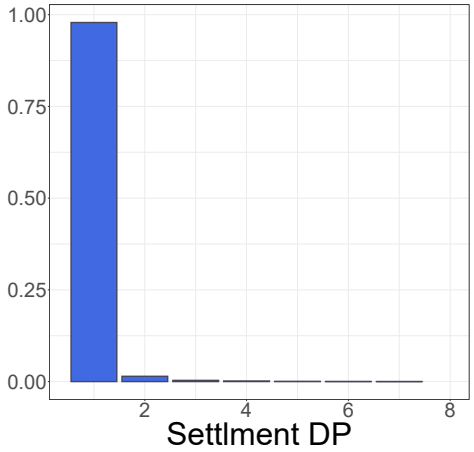
(a) Frequency by `claim_type`.



(b) Distribution of the number of payments.

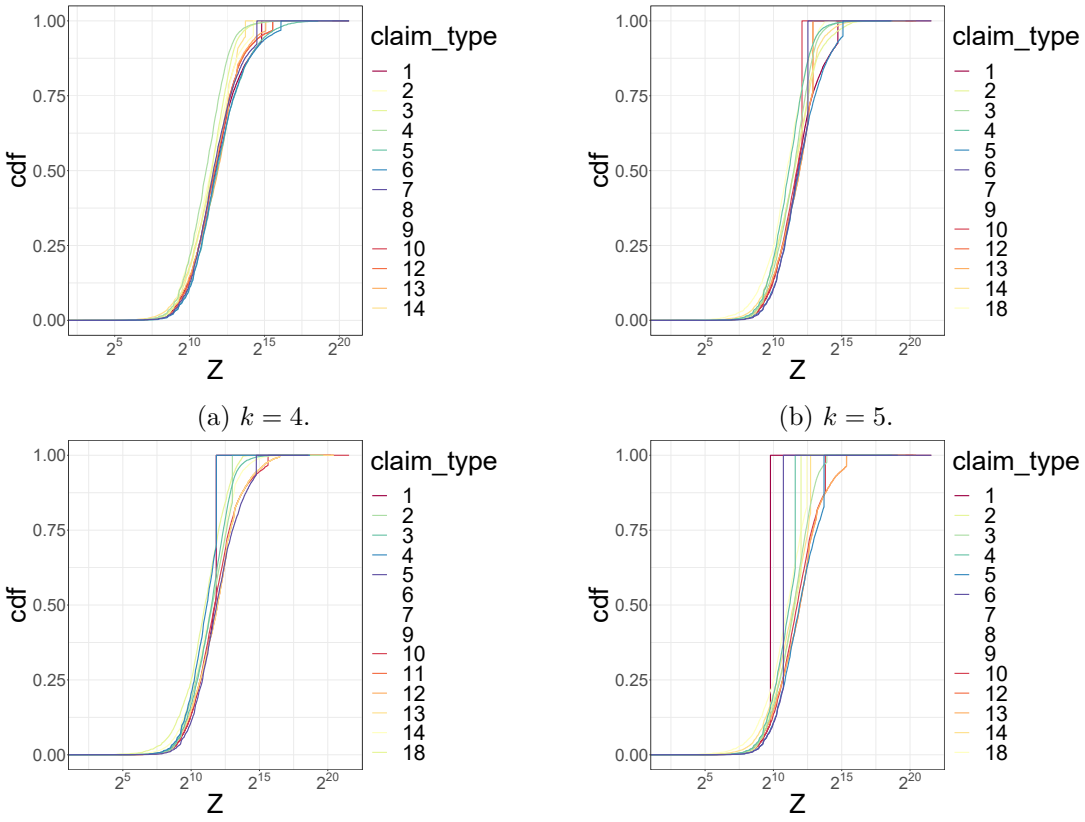


(c) Density plot of the total individual claim size. The x-axis is scaled by `log2`.



(d) Settlement delay from accident.

Figure 4.4: Exploratory data analysis on the real dataset that we use in this section. We show the relative frequencies by type of claim Figure 4.4a, the distribution of the number of payments Figure 4.4b, the density plot of the total individual claim cost Figure 4.4c and the distribution of the settlement delay from accident Figure 4.4d.



(a) $k = 4$. (b) $k = 5$.

(c) $k = 6$. (d) $k = 7$.

Figure 4.5: For each different dimension $k = 4, 5, 6, 7, 8$, we provide the individual claim cost curves for our observations by covariate value `claim_type`. The x-axis represents Z and we scale it by `log2` to ease the plot visualization.

consistent and asymptotically normal estimators even when covariates are continuous. Nonetheless, kernel-based methods are prone to the curse of dimesionality, and hence our method provides the best results when covariate dimensions are small.

4.4.1 Model comparison on different datasets

Consider the first application, where we censor our data at different depths and fit an AJ model using the largest possible state space. Using our model, it is possible to explore and compare the individual claim cost curves for the different values of `claim_type`, as shown in Figure 4.5.

The results are reported in Table 4.4 and they show that in general the AJ models outperform the CL in terms of `EI`, except for the scenario with $k = 7$. For each scenario, we are able to select a model with or without features using the CRPS measure. The rows with a ✓(column two) refer to models that include the `claim_type` feature in the fit. The target $Y^{\texttt{TOT}}$ is shown in column three. While in other scenarios the CRPS for the AJ model that includes the `claim_type` feature is close to or better than the CRPS of the model without the feature, in $k = 5$ the model without the feature has a much lower CRPS. The relative variability results show that AJ and CL have comparable results. In fact, only for $k = 4$ does CL have a lower relative variability.

For each choice of k we can simply select the best performing model in terms of CRPS and compute the claims reserve, the results are displayed in Table 4.5.

k	claim_type	Actual $Y^{\texttt{TOT}}$	EI (AJ)	EI (CL)	$\hat{SD}(Y^{\texttt{TOT}})/\hat{Y}^{\texttt{TOT}}$ (AJ)	$\hat{SD}(Y^{\texttt{TOT}})/\hat{Y}^{\texttt{TOT}}$(CL)	CRPS (average)
4	✓	616.1327	-0.0035	0.0157	0.0029	0.0023	0.0137
	✗		-0.0029		0.0029		0.0156
5	✓	822.5956	-0.0064	0.0209	0.0008	0.0024	0.0131
	✗		-0.0061		0.0007		0.0060
6	✓	999.6005	-0.0059	0.0173	0.0017	0.0017	0.0306
	✗		-0.0052		0.0017		0.0305
7	✓	1190.9112	-0.0146	0.0144	0.0011	0.0017	0.0310
	✗		-0.0142		0.0011		0.0311

Table 4.4: For different choices of k (column one), we fit a model with and without claim_type (column two). The target $Y^{\texttt{TOT}}$ is shown in column three. The EI is shown in columns four and five. The CV of $Y^{\texttt{TOT}}$ are displayed in columns seven and eight for the AJ and CL respectively. The CRPS is shown in column nine.

k	R	$SD(Y^{\texttt{TOT}})$	claim_type
4	10.7832	1.8054	✓
5	7.4981	0.5720	✗
6	10.3268	1.6731	✗
7	5.6317	1.3381	✓

Table 4.5: We selected for each data k (column one), the best performing model in terms of CRPS and present the reserve (column one) and the standard deviation of the reserve (column two).

4.4.2 Model comparison on a single dataset

In the second application, we censor our data after 5 accident period and construct the AJ model for $k = 4, 5, 6$. For each choice of k, we fit a model both with and without the feature claim_type. Table 4.6 shows that in terms of CRPS, the model with $k = 4$ using the claim_type feature is the best model (minimum CRPS). This model is also the best in terms of EI.

k	claim_type	Actual $Y^{\texttt{TOT}}$	EI (AJ)	EI (CL)	$\hat{SD}(Y^{\texttt{TOT}})/\hat{Y}^{\texttt{TOT}}$ (AJ)	$\hat{SD}(Y^{\texttt{TOT}})/\hat{Y}^{\texttt{TOT}}$(CL)	CRPS (average)
4	✗	999.6005	-0.0059	0.0173	0.0016	0.0017	0.0305
	✓		-0.0040		0.0020		0.0301
5	✗		-0.0069		0.0015		0.0307
	✓		-0.0052		0.0018		0.0309
6	✗		-0.0052		0.0017		0.0305
	✓		-0.0059		0.0017		0.0306

Table 4.6: For the dataset with depth 5 accident periods, we fit the AJ model both including and excluding the claim_type feature (column two). The target $Y^{\texttt{TOT}}$ is displayed in column three. The EI for the AJ and the CL can be found in columns four and five. The coefficient of variation of $Y^{\texttt{TOT}}$ is displayed in columns six and seven. The CRPS is displayed in column eight.

Using the CRPS we are able to select the model with $k = 4$ and that uses claim_type as a feature to compute the claims reserve. The model provides a reserve of 11.485 millions and a standard deviation of 2.0385 millions. The strategy presented in this last section is of particular interest to reserving actuaries. While the empirical study in Section 4.4.1 is interesting for understanding the behaviour of our model on different data, in practice the maximum depth of the data is selected using indicators such as claims settlement speed or using the expert judgement of an experienced actuary Wüthrich and Merz (2008, p. 12). The depth of the data is then known and the main interest is simply to select the best model, as shown here. Selecting the most appropriate model can be very difficult, see again Wüthrich and Merz (2008, p. 13), and the strategy presented in this

section provides a mathematically sound approach to doing so.

Chapter 5

Conclusions

In Chapter 2, we introduced a framework to model the one-to-one correspondence between claims development and development factors. We showed that when the micro-level data are aggregated in development triangles, estimation can be performed via a GLM. This is in contrast to existing GLM based reserving literature which models the claim amount. We showed that modelling the claim development via an age model replicates chain-ladder's predictions. The simplicity of the model furthermore invites very naturally to consider model extensions beyond the simple age model. In the data sets considered, including additional effects such as calendar effects and cohort effects often improved the fit. Especially adding a cohort effect seemed to often be beneficial. Interestingly, adding a period effect when modelling the claim amount did often not lead to any improvement. This is possibly due to the fact that inflation is often adjusted before being aggregated into run-off triangles. But there are exceptions and our data study demonstrated that the best results are obtained by having both claim amount models and claim development models as well as various structures at one's disposal. Altough not the primary focus of our contribution, further research on this models could concern the theoretical analysis of our estimators. The nature of the data suggests the use of bootstrapping techiniques to obtain confidence intervals for the reserve.

In Chapter 3, we extend the work in Chapter 2 to model the development of the claim frequency based on individual data and within a recent granular reserving frameworks, see e.g. Pigeon, Antonio, and Denuit (2013), Antonio and Plat (2014), Baudry and Robert (2019), Lopez (2019), Lopez and Milhaud (2021), Okine, Frees, and Shi (2022), Delong et al. (2022), Crevecoeur et al. (2022). Based on the work of (Miranda et al., 2013) and (Hiabu, 2017), we introduced a survival analysis framework to use machine learning techniques to estimate development factors that can depend on accident date and other features. The approach presented in this chapter has been developed with the aim to give higher accuracy in cases where more information in form of individual claims data is available while at the same time conserving the structure reserving actuaries are used to. Our extensive simulation study suggests that our methodology does indeed seem to work well. In this chapter, we have only considered the prediction of IBNR counts and an obvious next step is to integrate our methodology into a wider framework to estimate the outstanding claims amount. It could for example be interesting to merge our approach with the hierarchical reserving model (Crevecoeur et al., 2022).

The methodology described in Chapter 4 presents an improved approach for predicting loss reserves compared to commonly used aggregate loss reserving methods. We also introduce a comprehensive non-parametric estimator for the cumulative density function of individual claim costs, distinguishing it from other individual loss reserving methods that mainly focus on point forecasts of claim reserves. While some questions remain open, particularly regarding the modelling of Incurred But Not Reported (IBNR) claims and their dependence on specific features, practical applications can address these issues by fitting our model with the chain ladder method for future reports and projecting their future costs using the unconditional curve of individual claim costs. Furthermore, this methodology can be extended in various ways. One possibility is to explore the use of reverse time models, as described in Chapter 3, to project the future exposure of IBNR claims. In a broader context, future research could investigate how to model not only the size but also the frequency of individual claims.

References

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, *6*(4), 701–726.

Accenture. (2023). *Total enterprise reinvention: Setting a new performance frontier.* (`https://www.accenture.com/in-en/insights/consulting/total-enterprise-reinvention`, urldate = 2023-11-29)

Ajne, B. (1994). Additivity of chain-ladder projections. *ASTIN Bulletin: The Journal of the IAA*, *24*(2), 311318.

Al-Mudafer, M. T., Avanzi, B., Taylor, G., & Wong, B. (2022). Stochastic loss reserving with mixture density neural networks. *Insurance: Mathematics and Economics*, *105*, 144–174.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1993). *Statistical models based on counting processes.* Springer US.

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

Arjas, E. (1989). The claims reserving problem in non-life insurance: Some structural ideas. *ASTIN Bulletin: The Journal of the IAA*, *19*(2), 139–152.

Avanzi, B., Taylor, G., Vu, P. A., & Wong, B. (2020). A multivariate evolutionary generalised linear model framework with adaptive estimation for claims reserving. *Insurance: Mathematics and Economics*, *93*, 50–71.

Avanzi, B., Taylor, G., Wang, M., & Wong, B. (2021). SynthETIC: An individual insurance claim simulator with feature control. *Insurance: Mathematics and Economics*, *100*, 296–308.

Baudry, M., & Robert, C. Y. (2019). A machine learning approach for individual claims reserving in insurance. *Applied Stochastic Models in Business and Industry*, *35*(5), 1127–1155.

Bischofberger, S. M., Hiabu, M., & Isakson, A. (2020). Continuous chain-ladder with paid data. *Scandinavian Actuarial Journal*, *2020*(6), 477–502.

Bischofberger, S. M., Hiabu, M., Mammen, E., & Nielsen, J. P. (2019). A comparison of in-sample forecasting methods. *Computational Statistics & Data Analysis*, *137*, 133–154.

Bishop, C. M. (1994). *Mixture density networks.* Aston University.

Björkwall, S., Hössjer, O., & Ohlsson, E. (2009). Non-parametric and parametric bootstrap techniques for age-to-age development factor methods in stochastic claims reserving. *Scandinavian Actuarial Journal*, *2009*(4), 306–331.

Bladt, M., & Furrer, C. (2023a). Aalenjohansen: Conditional aalen-johansen estimation [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=AalenJohansen` (R package version 1.0)

Bladt, M., & Furrer, C. (2023b). Conditional aalen-johansen estimation. *arXiv preprint arXiv:2303.02119*.

Bladt, M., & Pittarello, G. (2023). Individual claims reserving using the aalen–johansen estimator. *arXiv preprint arXiv:2311.07384*.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 89–99.

Breslow, N. E. (1972). Professor norman breslow contribution to the discussion on the paper of d. r. cox cox (1972). *JR Stat Soc Ser B*, *34*, 216.

Brown, B. Z., Julga, L., & Merz, J. (2023). Best practices for property and casualty actuarial reserving departments. In *Cas e-forum.*

Buckley, J., & James, I. (1979). Linear regression with censored data. *Biometrika*, *66*(3), 429–436.

Calcetero-Vanegas, S., Badescu, A. L., & Lin, X. S. (2023). Claim reserving via inverse probability weighting: A micro-level chain-ladder method. *arXiv preprint arXiv:2307.10808*.

Carstensen, B. (2007). Age–period–cohort models for the lexis diagram. *Statistics in medicine*, *26*(15), 3018–3045.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/2939672.2939785` doi: 10.1145/2939672.2939785

Chiang, C. L. (1972). On constructing current life tables. *Journal of the American Statistical Association*, *67*(339), 538–541.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202.

Crevecoeur, J., Antonio, K., Desmedt, S., & Masquelein, A. (2023). Bridging the gap between pricing and reserving with an occurrence and development model for non-life insurance claims. *ASTIN Bulletin: The Journal of the IAA*, *53*(2), 185–212.

Crevecoeur, J., & Robben, J. (2024). hirem: Hierarchical reserving models [Computer software manual]. (R package version 0.1.0)

Crevecoeur, J., Robben, J., & Antonio, K. (2022). A hierarchical reserving model for reported non-life insurance claims. *Insurance: Mathematics and Economics*, *104*, 158–184.

Dabrowska, D. M. (1997). Smoothed cox regression. *The Annals of Statistics*, 1510–1540.

Dal Moro, E., Cuypers, F., & Miehe, P. (2016). Non-life reserving practices. *Available online at the address https://www. actuaries. org/ASTIN/Documents/ASTIN_WP_NL_Reserving_Report1. 0_2016-06-15. pdf [accessed 24-Jul-2018]*.

Deloitte. (2023). *2024 global insurance outlook.* (`https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-outlooks/insurance-industry-outlook.html`, urldate = 2023-11-29)

Delong, Ł., Lindholm, M., & Wüthrich, M. V. (2022). Collective reserving using individual claims data. *Scandinavian Actuarial Journal*, *2022*(1), 1–28.

Delong, Ł., & Wüthrich, M. V. (2023). The role of the variance function in mean estimation and validation. *Available at SSRN 4477677*.

Dutang, C., & Charpentier, A. (2020). *Casdatasets: Insurance datasets.* (R package version 1.0-11)

Efron, B. (1977). The efficiency of cox's likelihood function for censored data. *Journal of the American statistical Association*, *72*(359), 557–565.

Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, *11*(2), 89–121.

EIOPA. (2023). *Insurance statistics.* (`https://www.eiopa.europa.eu/tools-and-data/insurance-statistics_en`, urldate = 2023-11-27)

England, P., & Verrall, R. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: mathematics and economics*, *25*(3), 281–293.

European Parliament. (2009). *Directive 2009/138/ec.*

Fannon, Z., & Nielsen, B. (2020). *apc: Age-period-cohort analysis.* Retrieved from `https://CRAN.R-project.org/package=apc` (R package version 2.0.0)

Friedland, J. (2010). Estimating unpaid claims using basic techniques. In *Casualty actuarial society* (Vol. 201).

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Gabrielli, A., Richman, R., & Wüthrich, M. V. (2020). Neural network embedding of the over-dispersed poisson reserving model. *Scandinavian Actuarial Journal*, *2020*(1), 1–29.

Gesmann, M., Murphy, D., Zhang, Y. W., Carrato, A., Wuthrich, M., Concina, F., & Dal Moro, E. (2022). *Chainladder: Statistical methods and models for claims reserving in general insurance.* Retrieved from `https://CRAN.R-project.org/package=ChainLadder` (R package version 0.2.15)

Ghosh, S. (2018). *Kernel smoothing: Principles, methods and applications* (1st ed.). John Wiley & Sons, Inc.

Gill, R. D. (1986). On estimating transition intensities of a markov process with aggregate data of a certain type:" occurrences but no exposures". *Scandinavian Journal of Statistics*, 113–134.

Glantz, M., Johnson, J., Macy, M., Nunez, J. J., Saidi, R., & Velez, C. (2023). Students experience and perspective of a data science program in a two-year college. *Journal of Statistics and Data Science Education*, *31*(3), 248-257. Retrieved from `https://doi.org/10.1080/26939169.2023.2208185` doi: 10.1080/26939169.2023.2208185

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *69*(2), 243–268.

Gneiting, T., Raftery, A., Balabdaoui, F., & Westveld, A. (2004). Verifying probabilistic forecasts: Calibration and sharpness. In *Preprints, 17th conf. on probability and statistics in the atmospheric sciences, seattle, wa, amer. meteor. soc* (Vol. 2).

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, *102*(477), 359–378.

Gneiting, T., & Ranjan, R. (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, *29*(3), 411–422.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press. (`http://www.deeplearningbook.org`)

Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, *87*(420), 942–951.

Haastrup, S., & Arjas, E. (1996). Claims reserving in continuous time; a nonparametric bayesian approach. *ASTIN Bulletin: The Journal of the IAA*, *26*(2), 139–164.

Haberman, S., & Renshaw, A. (2009). On age-period-cohort parametric mortality rate projections. *Insurance: Mathematics and Economics*, *45*(2), 255–270.

Haberman, S., & Renshaw, A. (2011). A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, *48*(1), 35–55.

Harnau, J., & Nielsen, B. (2018). Over-dispersed age-period-cohort models. *Journal of the American Statistical Association*, *113*(524), 1722–1732.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical*

learning: data mining, inference, and prediction (Vol. 2). Springer.

Hess, K. T., & Schmidt, K. D. (2002). A comparison of models for the chain–ladder method. *Insurance: Mathematics and Economics*, *31*(3), 351–364.

Hesselager, O. (1994). A markov model for loss reserving. *ASTIN Bulletin: The Journal of the IAA*, *24*(2), 183–193.

Heuchenne, C., & Van Keilegom, I. (2007). Nonlinear regression with censored data. *Technometrics*, *49*(1), 34–44.

Hiabu, M. (2017). On the relationship between classical chain ladder and granular reserving. *Scandinavian Actuarial Journal*, *2017*(8), 708–729.

Hiabu, M., Hofman, E., & Pittarello, G. (2023). A machine learning approach based on survival analysis for ibnr frequencies in non-life reserving. *Available at ArXiV 2312.14549*.

Hiabu, M., Mammen, E., Martínez-Miranda, M. D., & Nielsen, J. P. (2016). In-sample forecasting with local linear survival densities. *Biometrika*, *103*(4), 843–859.

Hiabu, M., Mammen, E., Martínez-Miranda, M. D., & Nielsen, J. P. (2021). Smooth backfitting of proportional hazards with multiplicative components. *Journal of the American Statistical Association*, *116*(536), 1983–1993.

Hiabu, M., & Pittarello, G. (2023). *Notes on reserving in non-life insurance.* (Department of Mathematical Sciences; University of Copenhagen)

Hoem, J. M. (1969). Markov chain models in life insurance. *Blätter der DGVFM*, *9*(2), 91–107.

Hoem, J. M. (1972). Inhomogeneous semi-markov processes, select actuarial tables, and duration-dependence in demography. In *Population dynamics* (pp. 251–296). Elsevier.

Hofman, E., Pittarello, G., & Hiabu, M. (2023). Resurv [Computer software manual]. Retrieved from `https://github.com/edhofman/ReSurv.git`

Hosmer, D. W., Lemeshow, S., & May, S. (2008). Applied survival analysis. *Wiley Series in Probability and Statistics*, 60.

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

Jones, D. R. (2001). A taxonomy of global optimization methods based on response surfaces. *Journal of global optimization*, *21*, 345–383.

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology*, *18*(1), 1–12.

Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). *Loss models: from data to decisions* (Vol. 715). John Wiley & Sons.

Kuang, D., & Nielsen, B. (2020). Generalized log-normal chain-ladder. *Scandinavian Actuarial Journal*, *2020*(6), 553-576.

Kuang, D., Nielsen, B., & Nielsen, J. P. (2008a). Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika*, *95*(4), 987–991.

Kuang, D., Nielsen, B., & Nielsen, J. P. (2008b). Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika*, *95*(4), 979–986.

Kuang, D., Nielsen, B., & Perch Nielsen, J. (2011). Forecasting in an extended chain-ladder-type model. *The Journal of Risk and Insurance*, *78*(2), 345–359.

Lee, Y. K., Mammen, E., Nielsen, J. P., & Park, B. U. (2015). Asymptotics for in-sample density forecasting. *The Annals of Statistics*, *43*(2), 620–651.

Lee, Y. K., Mammen, E., Nielsen, J. P., & Park, B. U. (2017). Operational time and in-sample density forecasting. *The Annals of Statistics*, *45*(3), 1312–1341.

Liu, P., Fu, B., Yang, S. X., Deng, L., Zhong, X., & Zheng, H. (2020). Optimizing survival analysis of xgboost for ties to predict disease progression of breast cancer. *IEEE Transactions on Biomedical Engineering*, *68*(1), 148–160.

Lopez, O. (2019). A censored copula model for micro-level claim reserving. *Insurance: Mathematics and Economics*, *87*, 1–14.

Lopez, O., & Milhaud, X. (2021). Individual reserving and nonparametric estimation of claim amounts subject to large reporting delays. *Scandinavian Actuarial Journal*, *2021*(1), 34–53.

Lopez, O., Milhaud, X., & Thérond, P.-E. (2019). A tree-based algorithm adapted to microlevel reserving and long development claims. *ASTIN Bulletin: The Journal of the IAA*, *49*(3), 919–919.

Lovász, E. (2011). Analysis of finnish and swedish mortality data with stochastic mortality models. *European Actuarial Journal*, *1*(2), 259–289.

Maciak, M., Mizera, I., & Pešta, M. (2022). Functional profile techniques for claims reserving. *ASTIN Bulletin: The Journal of the IAA*, *52*(2), 449–482.

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin: The Journal of the IAA*, *23*(2), 213–225.

Mack, T., & Venter, G. (2000). A comparison of stochastic models that reproduce chain ladder reserve estimates. *Insurance: mathematics and economics*, *26*(1), 101–107.

Mammen, E., Martínez-Miranda, M. D., Nielsen, J. P., & Vogt, M. (2021). Calendar effect and in-sample forecasting. *Insurance: Mathematics and Economics*, *96*, 31–52.

McCullagh, P., & Nelder, J. A. (2019). *Generalized linear models* (2nd ed.). Routledge.

McKeague, I. W., & Utikal, K. J. (1990). Inference for a nonlinear counting process regression model. *The Annals of Statistics*, *18*(3), 1172–1187.

Meyers, G. G., & Shi, P. (Updated: September 1, 2011)). *Loss reserving data pulled from naic schedule p.* Retrieved 12/04/2024, from `https://www.casact.org/publications-research/research/research-resources/loss-reserving-data-pulled-naic-schedule-p`

Mikosch, T. (2009). *Non-life insurance mathematics. an introduction with the poisson process. 2nd ed.* doi: 10.1007/978-3-540-88233-6

Millossovich, P., Villegas, A. M., & Kaishev, V. K. (2018). Stmomo: An r package for stochastic mortality modelling. *Journal of Statistical Software*, *84*(3).

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

Miranda, M. D. M., Nielsen, J. P., & Verrall, R. (2012). Double chain ladder. *ASTIN Bulletin: The Journal of the IAA*, *42*(1), 59–76.

Nielsen, J. P., & Linton, O. B. (1995). Kernel estimation in a nonparametric marker dependent hazard model. *The Annals of Statistics*, 1735–1748.

Norberg, R. (1986). A contribution to modelling of ibnr claims. *Scandinavian Actuarial Journal*, *1986*(3-4), 155–203.

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

Norberg, R. (1997). Prediction of outstanding liabilities ii: Model variations and extensions. *Insurance Mathematics and Economics*, *3*(21), 251.

Oakes, D. (1972). Mr. david oakes contribution to the discussion on the paper of d. r. cox cox (1972). *Journal of the Royal Statistical Society: Series B*, *34*, 216.

Okine, A. N.-A., Frees, E. W., & Shi, P. (2022). Joint model prediction and application to individual-level loss reserving. *ASTIN Bulletin: The Journal of the IAA*, *52*(1), 91–116.

Overgaard, M. (2019). State occupation probabilities in non-markov models. *Mathematical Methods of Statistics*, *28*, 279–290.

Parodi, P. (2014). *Pricing in general insurance* (first ed.). CRC press. doi: https://doi.org/10.1201/b17525

Pigeon, M., Antonio, K., & Denuit, M. (2013). Individual loss reserving with the multivariate skew normal framework. *ASTIN Bulletin: The Journal of the IAA*, *43*(3), 399–428.

Pinheiro, P. J., Andrade e Silva, J. M., & de Lourdes Centeno, M. (2003). Bootstrap methodology in claim reserving. *Journal of Risk and Insurance*, *70*(4), 701–714.

Pittarello, G. (2020). *Master thesis on modelling claims reserving on individual data,analysis and insights.* (Faculty of Banking, Finance and Insurance Sciences; Catholic University of the Sacred Heart)

Pittarello, G., Hiabu, M., & Villegas, A. (2022). *clmplus: Tool-box of chain ladder + models.* Retrieved from `https://cran.r-project.org/web/packages/clmplus` (R package version 0.0.1.)

Pittarello, G., Hiabu, M., & Villegas, A. M. (2023). Replicating and extending chain-ladder via an age-period-cohort structure on the claim development in a run-off triangle. *arXiv preprint arXiv:2301.03858*.

Pittarello, G., Luini, E., & Marchione, M. M. (2024). Gemact: a python package for non-life (re)insurance modeling. *Annals of Actuarial Science*, 137. doi: 10.1017/S1748499524000022

PWC. (2023). *Insurance 2025 and beyond, insurance reimagined.* (`https://www.pwc.com/gx/en/financial-services/fs-2025/pwc-insurance2025.pdf`, urldate = 2023-11-29)

RStudio Team. (2020). Rstudio: Integrated development environment for r [Computer software manual]. Boston, MA. Retrieved from `http://www.rstudio.com/`

Schnieper, R. (1991). Separating true ibnr and ibner claims1. *ASTIN Bulletin: The Journal of the IAA*, *21*(1), 111–127.

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, *1*, 43–61.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, *25*.

Stute, W. (1986). On almost sure convergence of conditional empirical distribution functions. *The Annals of Probability*, *14*(3), 891–901.

Taylor, G. (2012). *Loss reserving: an actuarial perspective* (Vol. 21). Springer Science & Business Media.

Taylor, G. (2019). Loss reserving models: Granular and machine learning forms. *Risks*, *7*(3), 82.

Taylor, G. (2021). A special tweedie sub-family with application to loss reserving prediction error. *Insurance: Mathematics and Economics*, *101*, 262–288.

Teamah, A. E.-m. A., Elbanna, A. A., & Gemeay, A. M. (2019). Right truncated fréchet-weibull distribution: statistical properties and application. *Delta Journal of Science*, *41*(1), 20–29.

Therneau, T. M. (2023). A package for survival analysis in r [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=survival` (R package version 3.5-5)

Tsai, C. C.-L., & Kim, S. (2022). Model mortality rates using property and casualty insurance reserving methods. *Insurance: Mathematics and Economics*, *106*, 326–340.

United Nations. (2021). *Technology and innovation report 2021.* New York: United Nations Publications.

Van Houwelingen, H. C. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*, *34*(1), 70–85.

Verrall, R. J. (2000). An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: mathematics and economics*, *26*(1), 91–99.

Wang, Z., Wu, X., & Qiu, C. (2021). The impacts of individual information on loss reserving. *ASTIN Bulletin: The Journal of the IAA*, *51*(1), 303–347.

Ware, J. H., & DeMets, D. L. (1976). Reanalysis of some baboon descent data. *Biometrics*, *32*(2), 459–463.

Wiegrebe, S., Kopper, P., Sonabend, R., & Bender, A. (2023). Deep learning for survival analysis: A review. *arXiv preprint arXiv:2305.14961*.

Wilmoth, J. R., Andreev, K., Jdanov, D., Glei, D. A., Riffe, T., Boe, C., ... Barbieri, M. (2021). Methods protocol for the human mortality database. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://mortality. org [version 31/05/2007]*, *6*, 1–78.

Wilson, S. (2022). Parbayesianoptimization: Parallel bayesian optimization of hyperparameters [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=ParBayesianOptimization` (R package version 1.2.6)

Wüthrich, M. V. (2018). Neural networks applied to chain–ladder reserving. *European Actuarial Journal*, *8*(2), 407–436.

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

Wüthrich, M. V., & Merz, M. (2015). *Stochastic claims reserving manual: Advances in dynamic modeling.* Swiss Finance Institute Research Paper No. 15-34. doi: http://dx.doi.org/10.2139/ssrn.2649057

Appendix A

Open source software and reproducible results

Along with the original research papers on which this monograph is based, we have released two open source R software packages, (Pittarello et al., 2022; Hofman et al., 2023). We have also released Python software that is outside the scope of the included papers, but is described in the software article Pittarello, Luini, and Marchione (2024). In this Appendix, we show some small code implementations of our R packages.

A.1 Claims reserving with `clmplus`

In this Section we show the `clmplus` code implementation of some of the models we showed in Sections 2.8.1, 2.8.2 and 2.8.3. The stable version of `clmplus` package is available on CRAN at `https://CRAN.R-project.org/package=clmplus` and the development version and source code are available on GitHub at `https://github.com/gpitt71/clmplus`. Moreover, the full paper replication code can be found at `https://github.com/gpitt71/clmplus-paper-reproducible-results`.

We can install the development version of `clmplus` with the follwing command:

```
R> install.packages("devtools")
R> devtools::install_github("gpitt71/clmplus")
```

After loading the `clmplus` and `ChainLadder` packages, the dataset of cumulative claim payments is transformed into a `RtTriangle` object that extracts from the data the information to model the development factors as we discussed in Chapter 2.

```
R> library(clmplus)
R> library(ChainLadder)

R> data("AutoBI")
R> dataset=AutoBI$AutoBIPaid
R> datapp = AggregateDataPP(cumulative.payments.triangle = dataset)
```

We can see once more the parallel between non-life insurance and mortality modeling by showing the Lexis representation of the triangle of incremental payments in `AutoBI`. In the triangle of

incremental payments each diagonal represents the calendar year (period), the rows represent the accident year (cohort) and the columns are development year (age).

```
R> datapp$incremental.payments.triangle
```

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] 1904 3494 2098 1386  830  359  128   57
[2,] 2235 4026 2430 1752  903  408  277   NA
[3,] 2441 4907 3314 1993 1093  487   NA   NA
[4,] 2503 5670 3637 2366 1207   NA   NA   NA
[5,] 2838 5874 4016 2550   NA   NA   NA   NA
[6,] 2405 5453 3913   NA   NA   NA   NA   NA
[7,] 2759 6423   NA   NA   NA   NA   NA   NA
[8,] 2801   NA   NA   NA   NA   NA   NA   NA
```

In the Lexis representation that clmplus uses for modeling each diagonal represents the cohorts (accident year), the rows represent the age (development year) and the columns are period (calendar year).

```
R> datapp$occurrance
```

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] 1904 2235 2441 2503 2838 2405 2759 2801
[2,]   NA 3494 4026 4907 5670 5874 5453 6423
[3,]   NA   NA 2098 2430 3314 3637 4016 3913
[4,]   NA   NA   NA 1386 1752 1993 2366 2550
[5,]   NA   NA   NA   NA  830  903 1093 1207
[6,]   NA   NA   NA   NA   NA  359  408  487
[7,]   NA   NA   NA   NA   NA   NA  128  277
[8,]   NA   NA   NA   NA   NA   NA   NA   57
```

Starting from the datapp object we can replicate the chain-ladder reserve with an age model.

```
a.model.fit=clmplus(datapp,
                    hazard.model = "a")

a.model<-predict(a.model.fit,
                 hazard.model = "a")
# clmplus reserve (age model)
sum(a.model$reserve)
#31754.43
```

In order to show the comparison, we compute the ultimate cost for the Mack chain-ladder model using the MackChainLadder function from the ChainLadder library. By subtracting the diagonal cumulative amount we find the chain-ladder reserve.

```
#Mack ultimate cost
R> mck.ultimate=ChainLadder::MackChainLadder(dataset)$FullTriangle[,8]
R> sum(mck.ultimate - rtt$diagonal) #Mack reserve
31754.43
```

In a similar fashion to the age-model (a.model), we can define the age-cohort model (ac.model), the age-cohort model (ap.model), and age-period-cohort model (apc.model).

```
R> ac.model.fit=clmplus(datapp,
                  hazard.model = "ac")
R> ac.model=predict(ac.model.fit,
                  gk.fc.model = 'a',
                  gk.order = c(1,1,0))

# clmplus reserve (age-cohort model)
R> sum(ac.model$reserve)
#38126.05

R> ap.model.fit=clmplus(datapp,
                  hazard.model = "ap")

R> ap.model<- predict(ap.model.fit,
                  ckj.fc.model = 'a',
                  ckj.order = c(0,1,0))

# clmplus reserve (age-period model)
R> sum(ap.model$reserve)
#37375.01

R> apc.model.fit=clmplus(datapp,
                   hazard.model = "apc")
R> apc.model<-predict(apc.model.fit,
                   gk.fc.model = 'a',
                   ckj.fc.model = 'a',
                   gk.order = c(1,1,0),
                   ckj.order = c(0,1,0))
# clmplus reserve (age-period-cohort model)
R> sum(apc.model$reserve)
#38498.54
```

The models residuals in Figure 2.4 can be easily obtained with the `plot` function. We show an example for the age model to obtain Figure 2.4a.

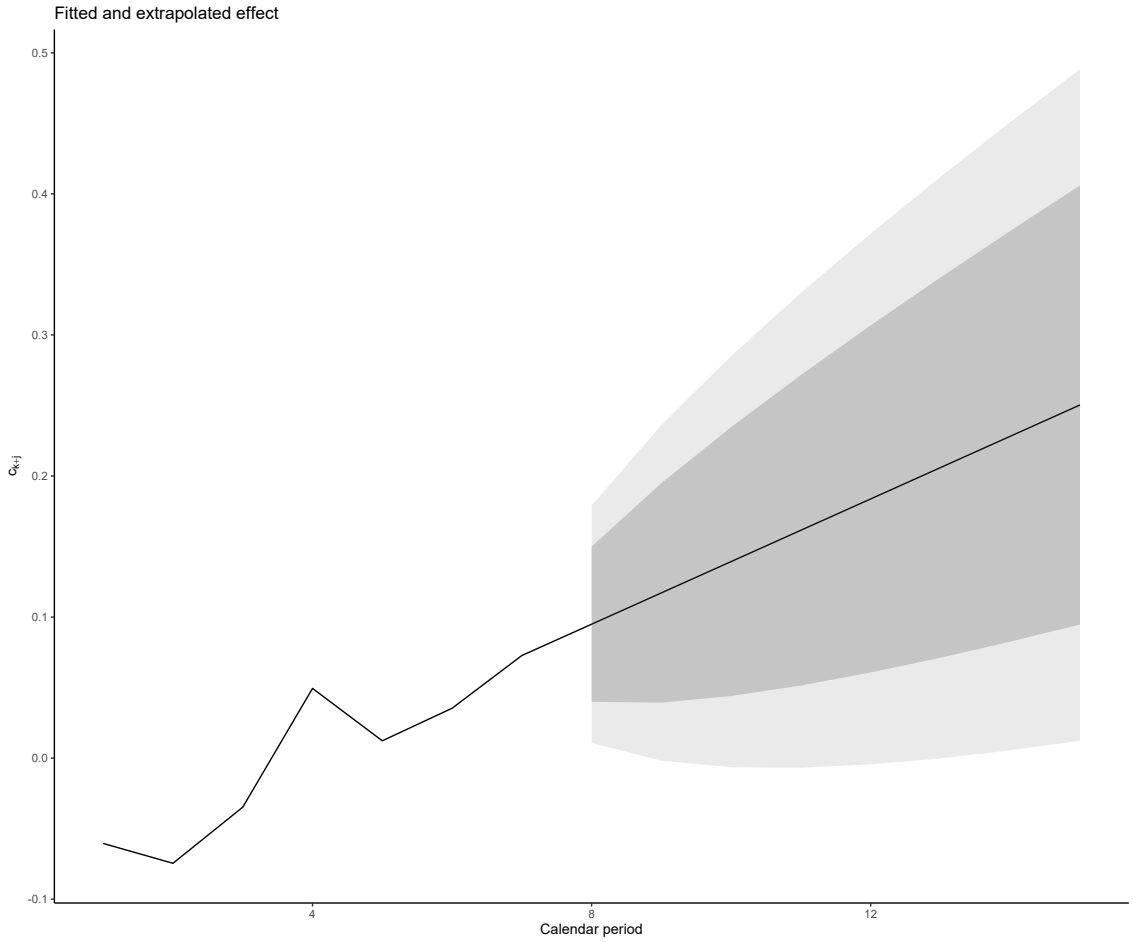
```
R> plot(a.model.fit)
```

The `clmplus` age model effect in Figure 2.2a can be inspected with the `plot` function:

```
plot(a.model)
```

For those models that require to extrapolate a period effect it is possible to inspect the fitted and extrapolated parameters with `clmplus`. Fitted and extrapolated effects for the age-period-cohort model:

```
plot(apc.model)
```



A.2 Individual claims reserving with ReSurv

In order to make the methods presented in Chapter 3 easily accessible, we created ReSurv, an R package that can be used to implement our modeling strategy on new data sets. The ReSurv package has an object-oriented API that we hope will allow the users to break down the individual claims reserving modeling process. The class diagram of our package is illustrated in Figure A.1.

The developer version of the ReSurv package can be downloaded from GitHub https://github.com/edhofman/ReSurv. :

```
R> devtools::install_github('edhofman/ReSurv')
```

In this section, we provide an example of the Cox model implementation. The data are generated using the function data_generation, a wrapper to SynthETIC that allows to replicate the simulation studies we illustrated in the main body of the manuscript. Scenario 0 corresponds to the scenario alpha of this manuscript.

```
R> library(ReSurv)
R> input_data = data_generator(time_unit = 1/360,
    years = 4,
    random_seed=1964,
    period_exposure= 200,
    period_frequency=0.2,
    scenario=0)
```

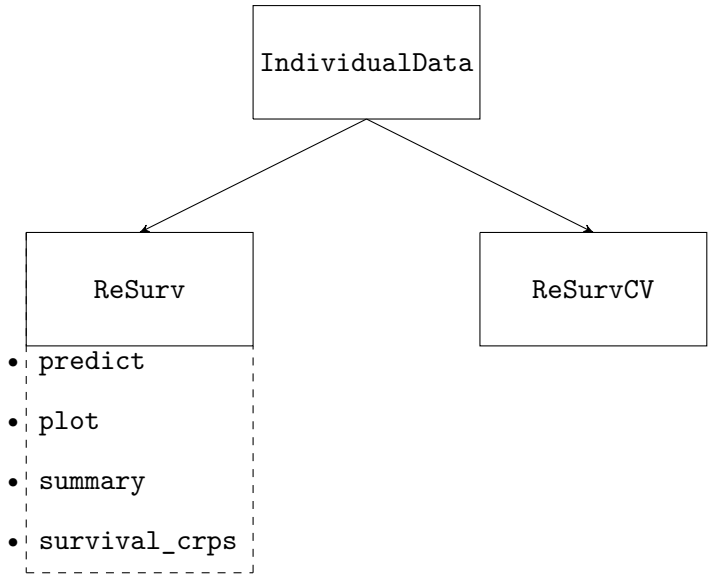


Figure A.1: Class diagram of ReSurv. A rectangle represents a class; an arrow connecting two classes indicates that the target class employs the origin class as an attribute. A ReSurv object entails IndividualData class instances. The methods of the ReSurv class are contained in a dotted box.

The daily frequency of the data is set with the time_unit parameter, the four years time span comes with the years parameter. We set the individuals at risk in each period to 200 and the frequency of report to .2. The individual claims data set is preprocessed within the IndividualData class.

```
R> individual_data <- IndividualData(data= input_data,
    continuous_features="AP",
    categorical_features="claim_type",
    accident_period="AP",
    calendar_period="RP",
    input_time_granularity = "days",
    output_time_granularity = "quarters",
    years=4)
```

In the IndividualData class the input data are passed through the data parameter. The class requires input on the categorical and continuous features to use in the modeling phase, i.e. the categorical_features and continuous_features parameters. The data time formatting is handled with the input_time_granularity and output_time_granularity parameters. The data set time dimensions are handled within the accident_period and calendar_period parameters.

```
R> model_fit <- ReSurv(individual_data,
    hazard_model = "cox")
```

By specifying the hazard model, the ReSurv class takes the preprocessed individual_data to fit the COX model to the data. Below we show an application of predict, one of the methods of the ReSurv class. With predict we can provide the quarterly forecasts.

```
R> model_predictions <- predict(resurv.fit)
```

The predict method can also handle predictions for granularities. In individual_data2 we look for yearly outputs.

```
R> individual_data2 <- IndividualData(data= input_data,
    continuous_features="AP",
    categorical_features="claim_type",
    accident_period="AP",
    calendar_period="RP",
    input_time_granularity = "days",
    output_time_granularity = "years",
    years=4)
```

The predictions can be returned using the `predict` method.

```
R> model_predictions <- predict(resurv.fit, newdata=individual_data2)
```

Appendix B

List of development triangles used for models comparison

	Dataset	Data source (package)
1	GenIns	`ChainLadder`
2	sifa.mod	`clmplus`
3	sifa.gtpl	`clmplus`
4	sifa.mtpl	`clmplus`
5	amases.gtpl	`clmplus`
6	amases.mod	`clmplus`
7	amases.mtpl	`clmplus`
8	bz	`apc`
9	ta	`apc`
10	xl	`apc`
11	vnj	`apc`
12	abc	`ChainLadder`
13	autoC	`ChainLadder`
14	autoP	`ChainLadder`
15	autoBI	`ChainLadder`
16	mclpaid	`ChainLadder`
17	medmal	`ChainLadder`
18	mortgage	`ChainLadder`
19	mw08	`ChainLadder`
20	mw14	`ChainLadder`
21	ukmotor	`ChainLadder`
22	usapaid	`ChainLadder`
23	aus1	`CASdatasets`
24	fre4b	`CASdatasets`
25	swiss1	`CASdatasets`
26	usa1	`CASdatasets`
27	usa2	`CASdatasets`
28	usa3	`CASdatasets`
29	usa4	`CASdatasets`
30	usa5	`CASdatasets`

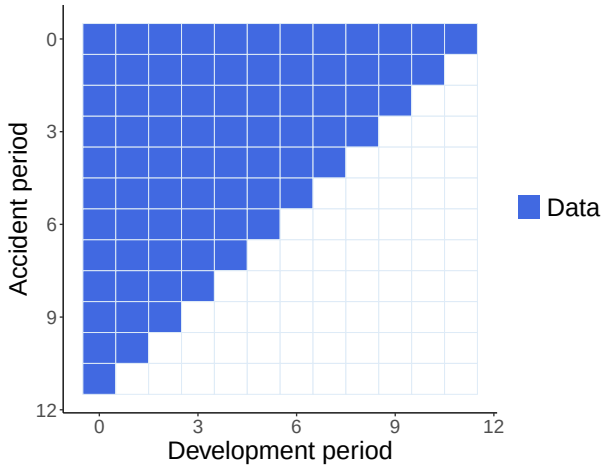
Appendix C

Case study on the National Association of Insurance Commissioners (NAIC) data

In this Section, we consider a case study using the models described in Chapter 2 on historical fully developed market data. The data is publicly available, and it was pulled from *Schedule P - Analysis of Losses and Loss Expenses* in the National Association of Insurance Commissioners (NAIC) database (Meyers & Shi, Updated: September 1, 2011)). The data aggregates run-off triangles of six lines of business for all U.S. property casualty insurers. The triangle data contain the claims of accident years 1988 to 1997 with 10 years development years. In our application, we will consider the following lines of business:Commercial Auto, Medical Malpractice, Other Liability, Private Passenger Auto, Workers Compensation.

We split each data set into training validation and testing as illustrated in Figure C.1.

Figure C.1: The last diagonals is removed from the triangle and the lower triangle is used as a test set to evaluate the models performance.



(a) The starting triangle.

(b) The triangle is divided into training set, validation set and test set.

We repeat, on the NAIC data, the comparison of the three model families illustrated in Section

2.9. In particular, we consider the data from the calendar period $k + j = m$ as the validation set and the data from the subsequent periods $k + j > m$ as the test set (lower triangle). Having now the lower triangle available, allows us to use the future losses to measure the accuracy of our forecasts using the incidence of the absolute prediction error on the true reserve (EI_{R}),

$$\mathrm{EI}_{\mathrm{R}} = \left| \frac{\sum_{k+j>m} \widehat{X}_{kj}}{\sum_{k+j>m} X_{kj}} - 1 \right|.$$

We displayed the EI_{R} for the different lines of business in Figure C.2. The plot shows, for each family of model the EI_{R} on the test of the best performing model among each family for each data set. Interestingly, the union of the `apc` and `clmplus` families of models always obtains a lower prediction error on the reserve compared to using only one of the two families. This indicates that the validation procedure of picking the best model by evaluating the performance on the last diagonal works well in the cases considered.

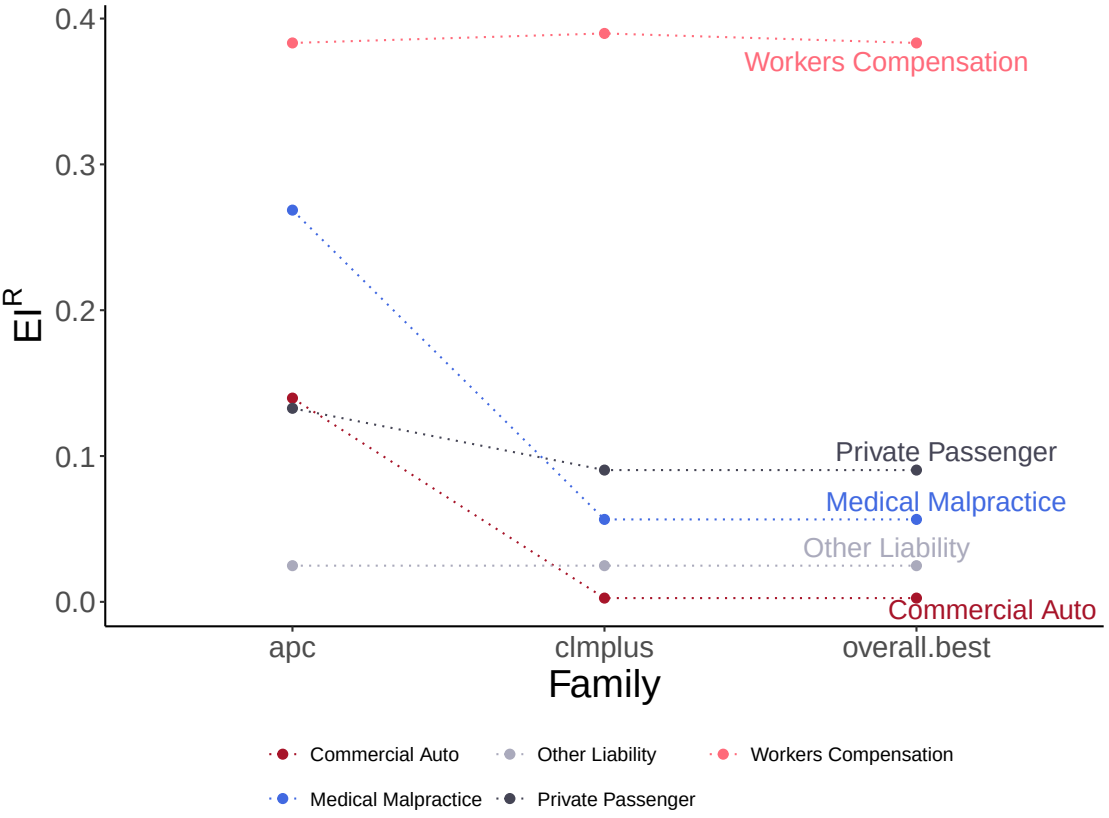


Figure C.2: Models EI^{R} on the test set across the NAIC datasets. On each dataset we selected the best performing model via a validation set from the three families `overall.best`, `clmplus` and `apc`. The dotted line connects error measurements on the same data set.

The results also show that, in the inspected data sets, the models from the `clmplus` family generally perform better than the models from the `apc` family in terms of EI_{R}. The only exception is the Line Workers Compensation where we select the apc model from the `apc` family. In the Line Other Liability, we select the Age Model (Chain Ladder). The performance measurements of the chosen models on the test set are shown, for each Line of Business and each family in Table C.1.

Line of Business	`apc`	`clmplus`	`overall.best`
Commercial Auto	0.140 (ac)	0.003 (ac)	0.003 (ac, `clmplus`)
Medical Malpractice	0.269 (ac)	0.057 (ap)	0.057 (ap, `clmplus`)
Other Liability	0.025 (ac)	0.025 (a)	0.025 (a, `clmplus`)
Private Passenger	0.133 (ac)	0.090 (ap)	0.090 (ap, `clmplus`)
Workers Compensation	0.383 (apc)	0.390 (ac)	0.383 (apc, `apc`)

Table C.1: For each Line of Business and each models set (columns two to four), we show the EI^{R} on test set of the model that we pick on the validation sets.

Appendix D

Bayesian optimization of machine learning algorithms

In Chapter 3, we showed an extended analysis of 5 simulated scenarios and a case study where we use machine learning to process large data sets and catch complex interactions in the data. Machine learning algorithms are very sensitive to the parameters and hyper parameters choice (Hastie et al., 2009, p. 219, Snoek et al., 2012). In this section, we provide general details about the strategy that we used for the hyper parameters selection of NN and XGB. Optimizing the algorithms over many data sets might lead to protracted computational times, (Jones, 2001). As a solution, we use the Bayesian optimization procedure described in Snoek et al. (2012). While grid searches can be cumbersome for big searches, using the approach in (Snoek et al., 2012) we use the information from prior model evaluations to guide the optimal parameters search. Notably, Bayesian optimization methods have shown to be well-performing in challenging optimization problems (Jones, 2001). Let us consider the data $\{(x_i, y_i)\}_{i=1,\dots,n}$. In this setting, the functional relationship between input and output $h : X \to \mathbb{R}$ is modelled as:

$$y_n \sim \mathcal{N}\left(h(x_n), \nu\right),$$

with ν being the variance of noise introduced into the function observations. Furthermore, let us assume that the observations $h\left(x\right)$ are drawn from a Gaussian prior.

Under the Gaussian Process prior, we get a posterior function (being a the acquisition function) $a : X \to \mathbb{R}^+$ that depends on the model solely through its predictive mean function $\mu(x; \{(x_i, y_i)\}_i, \theta)$ and predictive variance function $\sigma^2(x; \{(x_i, y_i)\}_i, \theta)$. Conversely, the acquisition function depends on the previous observations, as well as the Gaussian Process hyperparameters $a(x; \{(x_i, y_i)\}_i, \theta)$. In order to avoid an involved notation we denote

- $a(x; \{(x_i, y_i)\}_i, \theta)$ as $a\left(x\right)$
- $\mu(x; \{(x_i, y_i)\}_i, \theta)$ as $\mu\left(x\right)$
- $\sigma^2(x; \{(x_i, y_i)\}_i, \theta)$ as $\sigma^2\left(x\right)$

There are different definitions for the acquisition function (Snoek et al., 2012), we choose

$$a\left(x\right)=\left(\mu(x)-h_{\max}\right)\Phi\left(\frac{\mu(x)-h_{\max}-\xi}{\sigma(x)}\right)+\sigma(x)\phi\left(\frac{\mu(x)-h_{\max}-\xi}{\sigma(x)}\right), \tag{D.1}$$

where

- $h_{\max}$ is the current maximum value obtained from sampling.
- Φ is the standard normal cumulative density function.
- ϕ is the standard normal probability density function.
- ξ is an exploration parameter (Wilson, 2022).

This approach for Bayesian optimization is described with the following steps:

1. Set the parameters to an initial value.
2. Fit the Gaussian process.
3. Find the parameters that maximize the acquisition function.
4. Score the parameter.
5. Repeat steps 2-4 until some stopping criteria is met (Snoek et al., 2012).

A thorough description can be found in (Snoek et al., 2012). We show the hyperparameters we inspected for each model in Table D.1.

Model	hyperparameter	Range
NN	`num_layers`	$[2, 10]$
	`num_nodes`	$[2, 10]$
	`optim`	$[1, 2]$
	`activation`	$[1, 2]$
	`lr`	$[.005, 0.5]$
	`xi`	$[0, 0.5]$
	`eps`	$[0, 0.5]$
XGB	`eta`	$[0, 1]$
	`max_depth`	$[0, 25]$
	`min_child_weight`	$[0, 50]$
	`subsample`	$[0.1, 1]$
	`lambda`	$[0, 50]$
	`alpha`	$[0, 50]$

Table D.1: The range of hyperparameters we inspected. We set the same ranges for the 5 simulations and the case study on the real data.

Below we disclose the computational times we required for fitting the parameters combinations on the five simulated scenarios (Table D.2) and the real data (Table D.3).

Model	Scenario	Hyperparameter selection	Model fit
COX	Alpha		3.20
NN		52.57	149.31
XGB		3.19	16.77
COX	Beta		2.49
NN		44.00	102.13
XGB		2.83	15.66
COX	Gamma		2.25
NN		68.72	204.58
XGB		5.10	21.91
COX	Delta		2.23
NN		75.05	132.79
XGB		8.72	19.61
COX	Epsilon		1.94
NN		66.33	109.64
XGB		3.51	11.43

Table D.2: Average computational times in minutes, simulated scenarios. Hyperparameters selection is on 3-folded cross validation. Model fit includes development factor fitting.

Model	Split	Hyperparameter selection	Model fit
COX	Split 1		1.53
NN		88.28	27.18
XGB		10.57	21.56
COX	Split 2		1.37
NN		93.30	32.41
XGB		31.25	24.06
COX	Split 3		0.94
NN		161.71	40.91
XGB		19.01	25.86
COX	Split 4		1.80
NN		107.15	38.28
XGB		65.10	28.01
COX	Split 5		1.74
NN		115.49	20.78
XGB		38.12	26.80
COX	Split 6		1.24
NN		87.51	19.99
XGB		35.45	27.38

Table D.3: Computational times in minutes, real data. Hyperparameters selection is on 3-folded cross validation. Model fit includes development factor fitting.

Appendix E

Scenarios simulation

In section 3.6 we illustrated the steps that we followed to generate the five simulated scenarios. We also mentioned that the parametrization of $y = \phi(x)$ changes for the different scenarios. In this section we want to provide extra details on the parameters that we used in the simulation phase. For the data generation we will only need two modules from the `SynthETIC` package, i.e. the number of claims occurring every accident date and the reporting delay. In every scenario for both claim types, the rate of claims occurrence is .2. In scenarios Alpha, Gamma, Delta, and Epsilon the individuals at risk in the portfolio are 200 in each accident date. In scenario Beta the individuals for `claim_type 1` are decreasing. The relative portfolio composition is shown in Figure E.1a and Figure E.1b.

In Table E.1 we report the parameters that we used to simulate from the RTFWD distribution in (Teamah et al., 2019). We recall that the RTFWD has a four parameter structure ν, π, ξ, k, and is defined with $0 < t \leq b$ with cumulative distribution function

$$F(t) = \exp(-\pi^{\nu}\xi^{\nu k}(t^{-\nu k} - b^{-\nu k})).$$

Scenario	ν	π	k	ξ_0
Alfa, Beta, Gamma, Delta	0.5	60	1	0.1
Epsilon	0.5	$60 + (34.5387 I(\texttt{claim\_type ==0}) + 58.6803 I(\texttt{claim\_type ==0}))$	1	0.1

Table E.1: The RTFWD distribution parameters (columns two to four) for the different scenarios (column one).

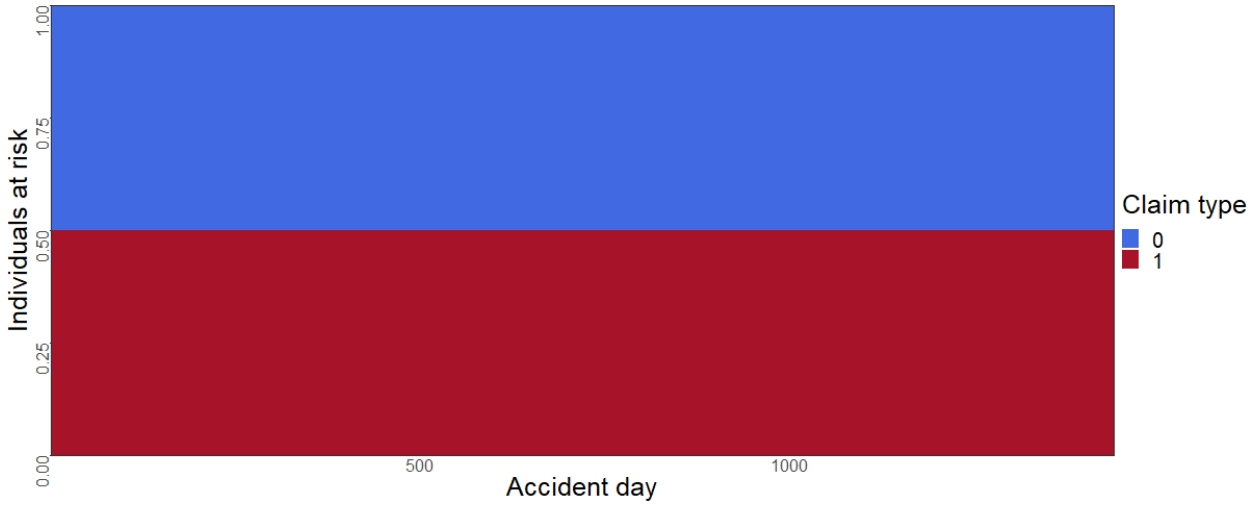
In Table E.2 we reported the parameters that we used for the simulation of the hazard. In the scenarios Alpha, Beta, Gamma, and Delta the reverse time hazard has the form

$$\alpha^R(t|\texttt{claim\_type}, \texttt{AD}) = \alpha_0^R(t) e^{\phi(\texttt{claim\_type},\texttt{AD};\theta)}$$

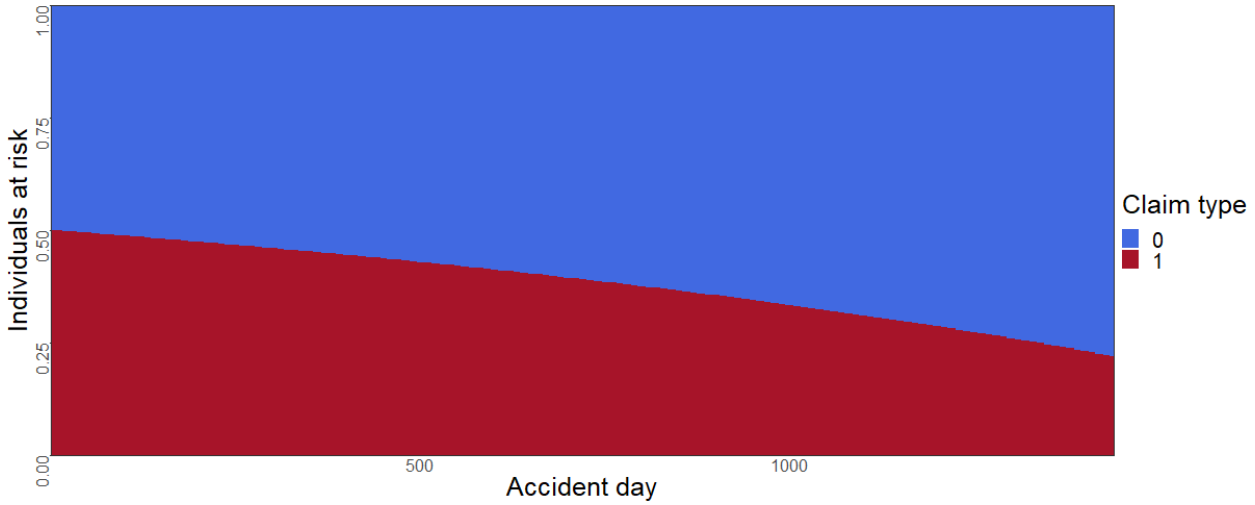
In column two of Panel A of Table E.2 we show the baseline $\alpha_0(t)$. We highlight that $\alpha_0(t)$ is the same for the five scenarios. In column three of panel A we show the different effects of the features on the proportional risk $e^{\phi(\texttt{claim\_type},\texttt{AD};\theta)}$ for the different scenarios (column 1).

In scenario Epsilon the hazard is not generated from a proportional model and it has the form:

$$\alpha^R(t|\texttt{claim\_type}, \texttt{AD}) = \alpha_0^R(t|\texttt{claim\_type})(\exp(\phi\,(\texttt{claim\_type}, \texttt{AD}; \theta)) + f(\texttt{claim\_type})).$$



(a) Portfolio composition by `claim_type` for each accident date (`AD` $\in \{1, \ldots, 1440\}$) in scenarios Alpha, Gamma, Delta and Epsilon.



(b) Portfolio composition by `claim_type` for each accident date (`AD` $\in \{1, \ldots, 1440\}$) in scenario Beta.

Figure E.1: In scenarios Alpha, Gamma, Delta, and Epsilon the portfolio is composed of half `claim_type 0` claims and half `claim_type 1` claims. In scenario Beta the individuals at risk for `claim_type 1` decrease for the most recent accident dates.

Details on the simulation for this scenario are provided in Panel B of Table E.2. In Panel C we show the coefficients values that we chose to simulate the proportional risk for the different scenarios (columns one to seven).

Table E.2: The parameters that we used for the reverse hazard simulation in the different scenarios. Baseline (column two) and effects on the proportional risk (column three) for the five scenarios (column one). Let us define $H(\texttt{AD}) = \texttt{AD} - 30\%(\texttt{AD} - 1)/30\rfloor$

Scenario	$\alpha_0^R(t)$	$\phi\left(\texttt{claim\_type}, \texttt{AD}; \theta\right)$
Alpha	$0.5\sqrt{0.2t^{-1}}$	$\beta_0 I(\texttt{claim\_type ==0}) + \beta_1 I(\texttt{claim\_type ==1})$
Beta		$\beta_0 I(\texttt{claim\_type ==0}) + \beta_1 I(\texttt{claim\_type ==1})$
Gamma		$\beta_0 I(\texttt{claim\_type ==0}) + \beta_1 I(\texttt{claim\_type ==1}) + \beta_2 I(\texttt{claim\_type ==1})\sqrt{\texttt{AD}}$
Delta		$\beta_0 I(\texttt{claim\_type ==0}) + \beta_1 I(\texttt{claim\_type ==1}) + \beta_3 I(H(\texttt{AD}) \in \{2,3,4\})$ $+\beta_4 I(H(\texttt{AD}) \in \{5,6,7\}) + \beta_5 I(H(\texttt{AD}) \in \{8,9,10\}) + \beta_6 I(H(\texttt{AD}) \in \{11,0,1\})$

Scenario	$f(\texttt{claim\_type})$	$\alpha_0^R(t\|\texttt{claim\_type})$	$\phi\left(\texttt{claim\_type}, \texttt{AD}; \theta\right)$
Epsilon	$0.5\phi\left(\texttt{claim\_type}, \texttt{AD}; \theta\right)$	$0.5\sqrt{0.1(2 + f(x))t^{-1}}$	$\beta_0 I(\texttt{claim\_type ==0}) + \beta_1 I(\texttt{claim\_type ==1})$

Table E.3: Panel B: Baseline (column three), effects on $\phi\left(\texttt{claim\_type}, \texttt{AD}; \theta\right)$ (column four) and $f(\texttt{claim\_type})$ for scenario Epsilon (column one).

β_0	β_1	β_2	β_3	β_4	β_5	β_6
1.1512	1.95601	-0.021206	-0.3	0.4	-0.7	0.1

Table E.4: Panel C: Features effects, $\beta_0, \ldots \beta_6 \in \mathbb{R}$

Appendix F

Minimizing the log-likelihood

In this Section we show average in-sample negative partial log-likelihood of the hazard models in Chapter 3, i.e. the loss function we minimize during the models training computed on the data that we used to fit the models. In order to ease our notation, we will indicate the average negative log-likelihood in Equation 3.3 as l. The model with the minimum in-sample average negative partial log-likelihood is the model that fits best the training data. We train the COX model using all the individual data from calendar periods $\tau = 1, \dots, \mathcal{T}$. In the training phase of XGB and NN, the input data from calendar periods $\tau = 1, \dots, \mathcal{T}$ are further split into a main part for training and a smaller part for validation. The split is random and we use 80% of the data for training (the splitting percentage is selected with a rule of thumb). We will then report for XGB and NN the out-of-sample average negative log-likelihood measured on the remaining 20% of the input data. Comparing the in-sample likelihood and the out-of-sample likelihood will tell us whether a model is overfitting the data. Indeed, we expect that a) the models can be ordered in the same way based on their descending score in-sample and out-of-sample and b) the magnitude of the scores in-sample and out-of-sample is comparable. In Table 3.2 we show the (average) negative log-likelihood averaged (over the 20 simulations), for each model in each scenario. The results show that XGB and NN seem to best fit the in-sample data compared to COX. Furthermore, XGB is consistently providing a lower likelihood compared to the NN. A similar behavior is reflected in the out-of-sample data.

A similar table can be provided for the real world data application. In Table F.1 we show for each model in each split the minimized log-likelihood.

Model	Split	l (in-sample)	l (out-of-sample)
COX	Split 1	10.78	-
NN		10.54	9.151
XGB		10.53	9.143
COX	Split 2	10.80	-
NN		10.55	9.165
XGB		10.55	9.162
COX	Split 3	10.78	-
NN		10.53	9.141
XGB		10.52	9.138
COX	Split 3	10.73	-
NN		10.48	9.089
XGB		10.46	9.077
COX	Split 5	10.62	-
NN		10.37	8.985
XGB		10.36	8.973
COX	Split 6	10.58	-
NN		10.32	8.931
XGB		10.31	8.926

Table F.1: Minimum log-likelihood for each model in each split of the real world data.

Appendix G

Comparison with `hirem`

We perform a first comparison with the `hirem` model on the scenario Beta that we introduced in Chapter 4. The results are reported in Table G.1. For the different values of k (column one), we compare the AJ model (with feature U) to the CL and `hirem` in terms of `EI` and CRPS. Each row corresponds to a different model (column two). The (average) target $Y^{\texttt{TOT}}$ is reported in column three. The `EI` (column four) shows that the AJ model is predicting the target $Y^{\texttt{TOT}}$ more accurately than the CL and `hirem`. We also observe that the `hirem` model seems to be less accurate for higher values of k. Consistently with the data generating assumptions, the CRPS indicates that the AJ model better describes the curve of the claim size. The relative variation is reported from Table 4.2 for the AJ model and the CL model (column five). We observe a similar relative variation of $Y^{\texttt{TOT}}$ on our models compared to `hirem`.

Scenario	Model	$Y^{\texttt{TOT}}$ (average)	`EI`	$\hat{\text{sd}}(Y^{\texttt{TOT}})/\hat{Y}^{\texttt{TOT}}$	CRPS (average, relative)
Beta (k=4)	AJ	42.279	0.006	0.008	1.000
	CL		0.103	0.008	-
	`hirem`		-0.028	0.009	1.582
Beta (k=5)	AJ	94.580	0.011	0.009	1.000
	CL		0.190	0.008	-
	`hirem`		0.071	0.009	1.680
Beta (k=6)	AJ	108.423	-0.013	0.007	1.000
	CL		0.256	0.008	-
	`hirem`		0.148	0.009	1.697
Beta (k=7)	AJ	116.032	-0.033	0.007	1.000
	CL		0.255	0.008	-
	`hirem`		0.168	0.008	1.790

Table G.1: Results for scenario Beta. For each value of k (column one) we present the average results over the 20 simulations. Each row of the table corresponds to a different model (column two). The table includes the (average) actual simulated total cost (column three) and the error incidence for the AJ, the CL, and `hirem` (columns four). In columns five we show the average relative variation and in column six we reported the average CRPS relative to the AJ CRPS.

A second comparison is performed on synthetic data generated from the individual claims simulators available in the R package Crevecoeur and Robben (2024). The package includes the `hirem` models implementation and four individual claims simulators for reported claims for $k = 10$. While the four scenarios are briefly introduced in this section, we refer to the package documentation

and the main manuscript for a detailed description. The generated features that are relevant for our application are reported in Table G.2.

The data include one categorical feature, the type of claim (`Type`), which we use in our application and a hidden feature (`Hidden`) that is not known to the insurer at the evaluation. The package includes a baseline simulated scenario (Baseline) and three scenarios that include modifications compare to the Baseline. In the scenario Claim Mix, there is a change in the portfolio distribution with respect to the feature `Type`. In the scenario Extreme, a seasonality effect in calendar year 9 simulates an increase in the claim occurrences. Lastly, the Settlement scenario simulates a change of the claim settlement process to quicker settlement after calendar period 7.

The results of this second application are reported Table G.3. We find that, both in terms of `EI` and CRPS and in the four simulated scenarios, our models always outperform the CL, while `hirem` performs better than the AJ (columns five and seven). The CRPS in column seven are relative to the CRPS of the AJ model with features. These results are expected, as the data are generated using assumptions that are consistent with the `hirem` model, see Appendix A of Crevecoeur et al. (2023). Notwithstanding, the AJ model is performing comparatively well. The results are averaged over ten simulations.

We remark that the applications reported in Table G.1 and Table G.3 indicate that, in general, for the scenarios that we inspected, individual models always outperform the chain ladder benchmark. However, as expected, the models perform best when the data generating process is consistent with the models assumptions.

Covariates	Description
`claim.nr`	Policy identifier
`Type` $\in \{T1, T2, T3\}$	Type of claim
`Hidden` $\in \{I, M, H\}$	Covariate unknown to the insurer (disregarded for modeling)
`occ.year`	Accident year
`rep.year`	Calendar year of reporting
`dev.year`	Development year
`size`	Incremental paid amount
`settlement`	Indicator, 1 in the development year of settlement

Table G.2: Description of the `hirem` data.

Scenario	Type	Model	$Y^{\texttt{TOT}}$(average)	`EI`	$\hat{\text{sd}}(Y^{\texttt{TOT}})/\hat{Y}^{\texttt{TOT}}$	CRPS (average, relative)
Baseline	✓	AJ	50.240	0.070	0.004	1.000
	✗			0.067	0.004	1.001
	✓	`hirem`		0.012	0.002	0.785
	✗	CL		0.232	0.010	
Claim Mix	✓	AJ	57.880	0.071	0.004	1.000
	✗			0.066	0.004	1.002
	✓	`hirem`		0.014	0.002	0.792
	✗	CL		0.240	0.010	
Extreme	✓	AJ	48.812	0.076	0.005	1.000
	✗			0.069	0.004	1.002
	✓	`hirem`		0.039	0.002	0.755
	✗	CL		0.231	0.010	
Settlement	✓	AJ	50.241	0.069	0.004	1.000
	✗			0.065	0.004	1.001
	✓	`hirem`		0.011	0.002	0.766
	✗	CL		0.231	0.001	

Table G.3: The `hirem` package includes four data generators in four different scenarios with $k = 10$ (column one). We compare our models (AJ with and without features, column two) to the `hirem` model in and the CL (column three). The actual $Y^{\texttt{TOT}}$ target is reported in column four. We show the `EI` and the CRPS results in columns five and seven respectevely. The predicted relative variation of of the total cost is shown in column six.