Mathematical Statistics
Stockholm University

Explicit moments for a class of micro-models in non-life insurance

Felix Wahl

Research Report 2019:6

ISSN 1650-0377

Postal address:
Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:
http://www.math.su.se

Mathematical Statistics
Stockholm University
Research Report **2019:6**,
http://www.math.su.se

Explicit moments for a class of micro-models in non-life insurance

Felix Wahl*

May 2019

Abstract

This paper considers properties of the micro-model analysed in Antonio and Plat (2014). The main results are analytical expressions for the moments of the outstanding claims payments subdivided into IBNR claims and individual RBNS claims. These moments are possible to compute explicitly using the discretisation scheme for estimation and simulation used in Antonio and Plat (2014) since the expressions then do not involve any integrals that, typically, would require numerical solutions. Other aspects of the model that are investigated are properties of the maximum likelihood estimators of the model parameters, such as bias and consistency, and a way of computing prediction uncertainty in terms of the mean squared error of prediction that does not require simulations. Moreover, a brief discussion is given on how to compute moments or risk-measures of the claims development result (CDR) using simulations, which based on the results of the present paper can be done without any nested simulations. Based on this it is straightforward to compute the one-year Solvency Capital Requirement, which corresponds to the 99.5% Value-at-Risk of the one-year CDR. Finally, a brief numerical illustration is used to show the theoretical performance of the maximum likelihood estimators of the parameters in the claims development process under this model using a realistic set-up based on the case-study of Antonio and Plat (2014).

JEL: G22.

Keywords: Stochastic claims reserving; risk; solvency; loss reserving; Poisson process.

1 Introduction

In the present paper, we consider properties of the model analysed in Antonio and Plat (2014). In that paper, the authors perform an extensive case study of a model that falls under the general class of models introduced in Norberg (1993), which is a class of models for individual claims in non-life insurance. These types of models are usually referred to as *micro-models*. They carry out full likelihood estimation within this framework and simulate the reserve together with the full reserve loss distribution. An observation regarding the reserve part of their work is that there is no need to simulate to obtain the reserve. Our main results in the present paper are analytical expressions for the moments of the outstanding claims payments under this particular model, which is valuable in computing,

* Department of Mathematics, Stockholm University, Sweden

for instance, the best estimate reserve. The moments are computed using the discretisation scheme used for estimation and simulation in Antonio and Plat (2014) since one then acquires analytical expressions not involving any integrals. It is of course also possible to compute moments without the discretisation, although the moments are then in terms of integrals that, typically, need to be solved numerically.

Other aspects of the model that we investigate are properties of the parameter estimators, such as bias and consistency, and a way of computing prediction uncertainty in terms of the mean squared error of prediction (MSEP) that does not require simulations. We also give a brief discussion on how to compute moments or risk-measures of the claims development result (CDR) using simulations, which based on the results of the present paper can be done without any nested simulations. This would be necessary if one had also to simulate to compute the moments themselves, see e.g. Sigmundsdóttir and Lindholm (2017) where this type of nested simulations are used to assess one-year non-life insurance risks. From the simulations, it is straightforward to compute the one-year Solvency Capital Requirement (SCR), which corresponds to the 99.5% Value-at-Risk of the one-year CDR.

Several papers have been written on micro-models — for instance, the seminal works of Arjas (1989) (using martingales) and Norberg (1993) (using a marked Poisson process approach). The approach of Norberg (1993) is further investigated in, for instance, Norberg (1999), Antonio and Plat (2014) and, in a Bayesian framework, in Haastrup and Arjas (1996). For a textbook introduction to this class of models, see Mikosch (2009).

When using micro-models, it is challenging to acquire easily computable moments of the outstanding claims payments. Perhaps partly for this reason, quite a body of work has accumulated in the area of discrete-time micro-models (usually modelled on the aggregate level), see, e.g. Verrall et al. (2010), Miranda et al. (2011, 2012) and Wahl et al. (2019). Discrete-time here meaning that the assumptions are made in discrete time for the individual claims. Similar to these works, other authors use aggregated data based on models built from the ground up, but not necessarily based on the individual claims, see, e.g. Bühlmann et al. (1980), Norberg (1986) and Lindholm et al. (2017). The present paper is the first paper giving easily computable, analytical moments, that do not involve integrals, of a quite general class of (continuous time) micro-models. Hopefully, these moments will help make micro-models more accessible.

The remainder of the paper is organised as follows: In Section 2 we give a brief introduction to the model class analysed in this paper. Section 3 describes the assumption of piecewise constant hazard rates for the event generating processes that are part of the development of a claim, i.e. the discretisation scheme. Section 4 and its subsections contain the main results consisting of moments of the outstanding claims payments split on IBNR and RBNS claims. In Section 5 we take a closer look at the maximum likelihood estimation of the parameters in the model. In particular, given some weak assumptions, we give a proper motivation for the use of the normal distribution as an approximation when considering parameter uncertainty. Based on Section 5, we move on in Section 6 to show how to assess prediction uncertainty (together with parameter uncertainty) using a semi-analytical approximation of the conditional MSEP. Section 7 gives a brief discussion on the computation of the moments and risk-measures of the CDR, and also the SCR. Finally, in Section 8, we provide a short numerical illustration of the performance of the estimators of the rates of the development process using some realistic parameter values and sample sizes based on the case study of Antonio and Plat (2014).

2 Model

For a given accident year, we assume that there is some exposure measure $w(t)$, given which, claims are collectively generated by a non-homogeneous Poisson process with rate $w(t)\lambda(t)$. The claims occurrence times generated by this process will be denoted by T_i. Each claim, once occurred, is then reported to the insurance company a random time U_i later according to a distribution $\mathbb{P}_{U|t}$.

Where the set-up of Antonio and Plat (2014) differs from the more general set-up in Norberg (1993) is in regards to the development of a claim. Once reported, a claim develops according to a process constructed by three mutually independent non-homogeneous Poisson processes: one generating payments at rate $h_p(t)$, one generating settlements at rate $h_{se}(t)$, and one generating settlements with payments at rate $h_{sep}(t)$. The development process records the payment events that occur up until the first settlement event, at which point the processes are all stopped. For the ith claim, this yields payment times V_{ij} and a settlement time V_i. For each of these payment times, a distribution $\mathbb{P}_p$ generates payment severities. In the present paper, as in Antonio and Plat (2014), this distribution is allowed to depend on the time of the accident and the reporting delay. More generally, we allow it to depend on some set of covariates that are known, at the latest, at reporting. Further, it is assumed that all claims are independent and, moreover, all payment severities are mutually independent and independent of the number of payments.

3 Assumption on the rates

As stated in Section 2, let $h_e(t)$ denote the rate with which events of type 'e' occur, where $e \in \{se, sep, p\}$, corresponding to 'settlement without payment', 'settlement with a payment at the same time' and 'payment without settlement', respectively. In Antonio and Plat (2014) these are also referred to as type 1, 2 and 3 events. As is done for estimation and simulation purposes in Section 4.3 in Antonio and Plat (2014), we introduce a finite partition $\mathcal{T} := \{t_0, t_1, \ldots, t_r\}$ of $\mathbb{R}_+$ on which the hazard rates of the three event types are piecewise constant, i.e.

$$h_e(t) = h_{e,l}, \quad t \in [t_{l-1}, t_l), \quad l = 1, \ldots, r,$$

for $e \in \{se, sep, p\}$. In Section 4, the moments of the outstanding claims payments will be computed given this discretisation scheme. Without the discretisation, the moments are expressed in terms of integrals that are generally not analytically tractable. To get analytically tractable expressions, one has to make assumptions on the rates, such as this particular discretisation. The usefulness of this discretisation scheme is of course not only in terms of calculating moments but also in terms of estimation. Estimating the rate function of a non-homogeneous Poisson process by assuming it to be piecewise constant on a set of intervals is an established heuristic that Henderson (2003) showed yields a consistent estimator if one shrinks the interval lengths at an appropriate rate when more data is available.

Based on the partition $\mathcal{T}$, it is natural to introduce the differences

$$\Delta t_l := t_l - t_{l-1}.$$

To ease the notational burden, as well as to increase interpretability, we will also define the two intensities

$$\bar{h}_{sl} := h_{se,l} + h_{sep,l},$$
$$\bar{h}_{pl} := h_{p,l} + h_{sep,l},$$

corresponding to the rates at which settlement events (with or without payment) and payment events (with or without settlement) occur, respectively.

We will denote the cumulative hazard rates by the corresponding capital letter, i.e.

$$H_e(t) := \int_0^t h_e(s)ds,$$

which, with the discretisation, is

$$H_e(t) = (t - t_{\kappa(t)-1})h_{e,\kappa(t)} + \sum_{l=0}^{\kappa(t)-1} \Delta t_l h_{e,l},$$

where $\kappa(t) := \inf_l\{l : t_l \geq t, t_l \in \mathcal{T}\}$. We define the corresponding quantities $\bar{H}_s(t)$ and $\bar{H}_p(t)$ in the same way.

Remark 1. *For our purposes, the partition $\mathcal{T}$ is finite. The results in this paper are straightforward to generalise to an infinite partition. However, the practical usefulness of this is questionable and therefore left out.*

4 Moments of the outstanding claims payments

In this section, we calculate the first two conditional moments of the outstanding claims payments split on IBNR and RBNS claims. We begin in Section 4.1 by calculating the mean and variance of one single RBNS claim. Given these moments, together with assuming that the claims are all (conditionally) independent, the corresponding quantities of the total outstanding claims payments from all RBNS claims will be a sum of the individual contributions. After computing the moments of the RBNS part, we move on in Section 4.2 to the moments of the IBNR part. The results in that section are in large part based on the results of the RBNS claims, since one can note that an IBNR claim will behave as an RBNS claim once reported. There will, however, be a random number of them and the developments of the claims will have random future starting times, adding a bit of complexity. Once we have both the RBNS and the IBNR moments, we have the moments of the total outstanding payments, since the RBNS and IBNR claims constitute two independent marked Poisson processes (as noted in, for instance, Norberg 1993). Therefore, both the mean and the variance of the total outstanding claims payments will be a sum of the RBNS and the IBNR part, i.e.

$$\mathbb{E}[R|\mathcal{F}_\tau] = \mathbb{E}[R^{\mathcal{I}}|\mathcal{F}_\tau] + \mathbb{E}[R^{\mathcal{R}}|\mathcal{F}_\tau], \tag{1}$$

$$\mathrm{Var}(R|\mathcal{F}_\tau) = \mathrm{Var}(R^{\mathcal{I}}|\mathcal{F}_\tau) + \mathrm{Var}(R^{\mathcal{R}}|\mathcal{F}_\tau), \tag{2}$$

where $R^{\mathcal{I}}$ and $R^{\mathcal{R}}$ are the outstanding claims payments from IBNR and RBNS claims, respectively, making up the total outstanding claims payments R. Moreover, $\mathcal{F}_\tau$ denotes the information available at the present time, τ.

The moments of the total outstanding claims payments are the main results, but we also touch upon computing moments of cash flows in specific time intervals. For the RBNS claims, it is possible to calculate such moments, and we do so in Section 4.1.1. For the IBNR claims, however, this is not possible. In Section 4.2.1 we discuss why this is the case.

4.1 Moments of the total outstanding payments from RBNS claims

In this section, we compute the conditional expectation and variance of the total outstanding payments from RBNS claims. By the assumption of independent claims, the expectation and variance of the total outstanding payments from all RBNS claims is the sum of all the expectations and variances of the individual contributions of each separate RBNS claim, i.e.

$$\mathbb{E}[R^{\mathcal{R}}|\mathcal{F}_\tau] = \sum_{i=1}^{N^{\mathcal{R}}} \mathbb{E}[R_i^{\mathcal{R}}|\mathcal{F}_\tau], \tag{3}$$

$$\mathrm{Var}(R^{\mathcal{R}}|\mathcal{F}_\tau) = \sum_{i=1}^{N^{\mathcal{R}}} \mathrm{Var}(R_i^{\mathcal{R}}|\mathcal{F}_\tau), \tag{4}$$

where $R_i^{\mathcal{R}}$ is the outstanding payments from the ith RBNS claim, and $N^{\mathcal{R}}$ is the number of RBNS claims. Therefore, we begin by computing the moments of one arbitrary RBNS claim, the result of which we summarise in Proposition 1 at the end of the present section.

For a given accident year, at the present time τ, we consider one arbitrary RBNS claim, say claim i, where $i = 1, \dots, N^{\mathcal{R}}$. Let τ_i denote the present time measured since reporting for this claim. By definition, settlement of this (RBNS) claim has not yet occurred, i.e. $V_i > \tau_i$.

Let $B_i(s,t)$ denote the payments made in the interval $[s,t)$, where time is measured from reporting and let $\kappa_i := \kappa(\tau_i)$. We will be particularly interested in the intervals $[t_{l-1}, t_l)$ for $l = \kappa_i + 1, \dots, r$ as well as the interval $[\tau_i, t_{\kappa_i})$ since the rates are constant on these intervals. For notational brevity, we let

$$B_{il} := \begin{cases} B_i(t_{l-1}, t_l) & \text{for } \kappa_i < l \le r, \\ B_i(\tau_i, t_{\kappa_i}) & \text{for } l = \kappa_i, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Given this definition, the total outstanding payments, $R_i^{\mathcal{R}}$, can be written as

$$R_i^{\mathcal{R}} = B_i(\tau_i, t_{\kappa_i}) + \sum_{l=\kappa_i+1}^{r} B_i(t_{l-1}, t_l) = \sum_{l=\kappa_i}^{r} B_{il}. \tag{6}$$

This relation can be further broken down into

$$B_i(s,t) = \sum_{j=1}^{N_i(s,t)} Y_{ij}(s,t), \tag{7}$$

where $N_i(s,t)$ is the number of payments made in $[s,t)$ and $Y_{ij}(s,t)$ is the size of the jth of these payments. For notational brevity, again, we define N_{il} in the same way as B_{il} in (5).

As mentioned in Section 2, the quantities on the right-hand side of (7) are all assumed to be mutually independent. Moreover, we assume that the payment size distribution only depends on covariates that are $\mathcal{F}_\tau$-measurable (i.e. covariates known at reporting), as is the case in Antonio and Plat (2014). Given this assumption, we can write

$$\mathbb{E}[Y_{ij}(s,t)|\mathcal{F}_\tau] = \mu_i,$$
$$\mathrm{Var}(Y_{ij}(s,t)|\mathcal{F}_\tau) = \sigma_i^2.$$

In Antonio and Plat (2014), μ_i and σ_i depend on the development year and the initial reserve category, both of which are known at reporting, as required. Given this, together with the independence of $Y_{ij}(s,t)$ and $N_i(s,t)$, it follows from (7) together with an application of iterated expectations that

$$\mathbb{E}[B_i(s,t)|\mathcal{F}_\tau] = \mu_i \mathbb{E}[N_i(s,t)|\mathcal{F}_\tau], \tag{8}$$

and with a variance decomposition that

$$\operatorname{Var}(B_i(s,t)|\mathcal{F}_\tau) = \sigma_i^2 \mathbb{E}[N_i(s,t)|\mathcal{F}_\tau] + \mu_i^2 \operatorname{Var}(N_i(s,t)|\mathcal{F}_\tau). \tag{9}$$

Moreover, for two disjoint intervals $[s,t)$ and $[s',t')$, a covariance decomposition yields

$$\operatorname{Cov}(B_i(s,t), B_i(s',t')|\mathcal{F}_\tau) = \mu_i^2 \operatorname{Cov}(N_i(s,t), N_i(s',t')|\mathcal{F}_\tau). \tag{10}$$

Taking (6) together with (8), (9) and (10) it is clear that

$$\mathbb{E}[R_i^{\mathcal{R}}|\mathcal{F}_\tau] = \sum_{l=\kappa_i}^{r} \mathbb{E}[B_{il}|\mathcal{F}_\tau] = \mu_i \sum_{l=\kappa_i}^{r} \mathbb{E}[N_{il}|\mathcal{F}_\tau], \tag{11}$$

and

$$\begin{aligned}\operatorname{Var}(R_i^{\mathcal{R}}|\mathcal{F}_\tau) &= \sum_{l=\kappa_i}^{r} \operatorname{Var}(B_{il}|\mathcal{F}_\tau) + 2\sum_{l=\kappa_i}^{r-1}\sum_{k=l+1}^{r} \operatorname{Cov}(B_{il}, B_{ik}|\mathcal{F}_\tau) \\ &= \sigma_i^2 \sum_{l=\kappa_i}^{r} \mathbb{E}[N_{il}|\mathcal{F}_\tau] + \mu_i^2 \sum_{l=\kappa_i}^{r} \operatorname{Var}(N_{il}|\mathcal{F}_\tau) \\ &\quad + 2\mu_i^2 \sum_{l=\kappa_i}^{r-1}\sum_{k=l+1}^{r} \operatorname{Cov}\left(N_{il}, N_{ik}|\mathcal{F}_\tau\right). \end{aligned} \tag{12}$$

We therefore see that to calculate the expectation and variance of the outstanding payments, we must compute $\mathbb{E}[N_{il}|\mathcal{F}_\tau]$ and $\operatorname{Var}(N_{il}|\mathcal{F}_\tau)$ for all $l \geq \kappa_i$ and $\operatorname{Cov}(N_{il}, N_{ik}|\mathcal{F}_\tau)$ for all $k > l \geq \kappa_i$. The computations of these quantities are left for the appendix, and we now conclude this section by stating a proposition giving the ingredients needed to compute the conditional expectation and variance of the total outstanding RBNS claims payments. The proof of the proposition can be found in Appendix A together with the computations of the above quantities.

Proposition 1. *The conditional expectation of the total outstanding claims payments from the ith RBNS claim is*

$$\mathbb{E}[R_i^{\mathcal{R}}|\mathcal{F}_\tau] = \mu_i \sum_{l=\kappa_i}^{r} \frac{\bar{h}_{pl}}{\bar{h}_{sl}} \left(1 - e^{-\bar{h}_{sl}\Delta t_{il}}\right) e^{-(\bar{H}_s(t_{l-1}) - \bar{H}_s(\tau_i))}, \tag{13}$$

and the variance is

$$\operatorname{Var}(R_i^{\mathcal{R}}|\mathcal{F}_\tau) = \sum_{l=\kappa_i}^{r} (\sigma_i^2 a_{il} + \mu_i^2 a_{il}(a_{il} + b_{il})) + 2\mu_i^2 \sum_{l=\kappa_i}^{r-1}\sum_{k=l+1}^{r} a_{ik}(c_{il} - a_{il}),$$

where

$$a_{il} := \frac{\bar{h}_{pl}}{\bar{h}_{sl}} \left(1 - e^{-\bar{h}_{sl}\Delta t_{il}}\right) e^{-(\bar{H}_s(t_{l-1}) - \bar{H}_s(\tau_i))} \tag{14}$$

$$b_{il} := 1 + 2\frac{\bar{h}_{pl}}{\bar{h}_{sl}} - 2h_{p,l}\Delta t_{i,l} \frac{e^{-\Delta t_{i,l}\bar{h}_{sl}}}{1 - e^{-\Delta t_{i,l}\bar{h}_{sl}}} \tag{15}$$

$$c_{il} := \Delta t_{il} h_{p,l}, \tag{16}$$

and

$$\Delta t_{ij} := \begin{cases} t_j - t_{j-1} & \text{if } j > \kappa_i, \\ t_{\kappa_i} - \tau_i & \text{if } j = \kappa_i. \end{cases}$$

Moreover, given the discretisation, we have

$$\bar{H}_s(t_{l-1}) - \bar{H}_s(\tau_i) = \sum_{j=\kappa_i}^{l-1} \bar{h}_{sj} \Delta t_{ij}.$$

A thing to note from the proof of Proposition 1 is that without the discretisation the expressions for the moments contain integrals. To be more specific, to compute the expected value for a general set of hazard rates we need to swap the terms

$$\frac{\bar{h}_{pl}}{\bar{h}_{sl}} \left(1 - e^{-\bar{h}_{sl}\Delta t_{il}}\right)$$

in (13) with an expression acquired by adding (43) and (46) of the proof together. The integrals in these equations are not generally solvable unless we specify the functional form of the hazard rates. This problem illustrates part of the convenience of performing the discretisation scheme introduced in Antonio and Plat (2014), since, as we see from the proposition, we then acquire explicit expressions while keeping a quite general hazard rate (given a fine enough partition).

4.1.1 Cash flows of RBNS claims payments

For the RBNS claims, it is possible to calculate moments of cash flows in specific time intervals $[s, t)$ for some time points $s < t$ measured since reporting, which we denote by $R_i^{\mathcal{R}}(s, t)$. To compute these, take the time points we want to consider and add them to the partition $\mathcal{T}$. Denote the resulting set by $\widetilde{\mathcal{T}}$, then the calculations of the previous section are still valid for this (finer) partition since the hazard rates are constant on the intervals it produces. This fact is what we will make use of in this section. It is important to note, however, that this is for prediction and not to be done for estimation.

For a specific time interval $[s, t)$, with $\tau_i \le s < t$ and $s, t \in \widetilde{\mathcal{T}}$, it holds that

$$R_i^{\mathcal{R}}(s, t) = \sum_{l=\tilde{\kappa}(s)}^{\tilde{\kappa}(t)} B_{il},$$

where $\tilde{\kappa}(t) := \inf_l\{l : t_l \ge t, t_l \in \widetilde{\mathcal{T}}\}$, and therefore we get, as before, that

$$\begin{aligned}
\mathbb{E}[R_i^{\mathcal{R}}(s, t)|\mathcal{F}_\tau] &= \mu_i \sum_{l=\tilde{\kappa}(s)}^{\tilde{\kappa}(t)} \mathbb{E}[N_{il}|\mathcal{F}_\tau], \\
\mathrm{Var}(R_i^{\mathcal{R}}(s, t)|\mathcal{F}_\tau) &= \sigma_i^2 \sum_{l=\tilde{\kappa}(s)}^{\tilde{\kappa}(t)} \mathbb{E}[N_{il}|\mathcal{F}_\tau] + \mu_i^2 \sum_{l=\tilde{\kappa}(s)}^{\tilde{\kappa}(t)} \mathrm{Var}(N_{il}|\mathcal{F}_\tau) \\
&\quad + 2\mu_i^2 \sum_{l=\tilde{\kappa}(s)}^{\tilde{\kappa}(t)-1} \sum_{k=l+1}^{\tilde{\kappa}(t)} \mathrm{Cov}\left(N_{il}, N_{ik}|\mathcal{F}_\tau\right).
\end{aligned}$$

Based on these equations, we get the following proposition in the same way as Proposition 1:

Proposition 2. *The conditional expectation of the total outstanding claims payments from the ith RBNS claim in the interval $[s, t)$ measured since reporting, with $s, t \in \widetilde{\mathcal{T}}$, is*

$$\mathbb{E}[R_i^{\mathcal{R}}(s, t)|\mathcal{F}_\tau] = \mu_i \sum_{l=\tilde{\kappa}(s)}^{\tilde{\kappa}(t)} \frac{\bar{h}_{pl}}{\bar{h}_{sl}} \left(1 - e^{-\bar{h}_{sl}\Delta t_{il}}\right) e^{-(\bar{H}_s(t_{l-1}) - \bar{H}_s(\tau_i))},$$

and the variance is

$$\mathrm{Var}(R_i^{\mathcal{R}}(s,t)|\mathcal{F}_\tau) = \sum_{l=\tilde{\kappa}(s)}^{\tilde{\kappa}(t)} (\sigma_i^2 a_{il} + \mu_i^2 a_{il}(a_{il}+b_{il})) + 2\mu_i^2 \sum_{l=\tilde{\kappa}(s)}^{\tilde{\kappa}(t)-1} \sum_{k=l+1}^{\tilde{\kappa}(t)} a_{ik}(c_{il}-a_{il}),$$

where a_{il}, b_{il} and c_{il} are given by (14)–(16), and Δt_{il} is understood to be the differences produced by $\widetilde{\mathcal{T}}$. Furthermore, for $s,t,s',t' \in \widetilde{\mathcal{T}}$ with $s < t < s' < t'$, the covariance terms are

$$\mathrm{Cov}(R_i^{\mathcal{R}}(s,t), R_i^{\mathcal{R}}(s',t')|\mathcal{F}_\tau) = \mu_i^2 \sum_{l=\tilde{\kappa}(s)}^{\tilde{\kappa}(t)} \sum_{k=\tilde{\kappa}(s')}^{\tilde{\kappa}(t')} a_{ik}(c_{il}-a_{il}).$$

If on the other hand $s < s' < t < t'$, then

$$\begin{aligned}\mathrm{Cov}(R_i^{\mathcal{R}}(s,t), R_i^{\mathcal{R}}(s',t')|\mathcal{F}_\tau) = {} & \mathrm{Var}(R_i^{\mathcal{R}}(s',t)|\mathcal{F}_\tau) \\ & + \mathrm{Cov}(R_i^{\mathcal{R}}(s,s'), R_i^{\mathcal{R}}(s',t)|\mathcal{F}_\tau) \\ & + \mathrm{Cov}(R_i^{\mathcal{R}}(s',t), R_i^{\mathcal{R}}(t,t')|\mathcal{F}_\tau).\end{aligned}$$

Note that, technically, Proposition 1 is a corollary to Proposition 2, but to not obfuscate the main result for RBNS claims, which is the moments of the total outstanding payments given in Proposition 1, it is presented here as a separate result.

Proposition 2 gives the first two moments for one RBNS claim. The moments of the outstanding payments from all RBNS claims, made in the time interval $[s,t)$ *measured since the beginning of the accident year*, is then

$$\mathbb{E}[R^{\mathcal{R}}(s,t)|\mathcal{F}_\tau] = \sum_{i=1}^{N^{\mathcal{R}}} \mathbb{E}[R_i^{\mathcal{R}}(s-\tau_i, t-\tau_i)|\mathcal{F}_\tau],$$
$$\mathrm{Var}(R^{\mathcal{R}}(s,t)|\mathcal{F}_\tau) = \sum_{i=1}^{N^{\mathcal{R}}} \mathrm{Var}(R_i^{\mathcal{R}}(s-\tau_i, t-\tau_i)|\mathcal{F}_\tau).$$

Note here that for $R^{\mathcal{R}}$, time is measured since the start of the accident year while it is measured since reporting for $R_i^{\mathcal{R}}$.

4.2 Moments of the total outstanding payments from IBNR claims

In this section, we calculate the expectation and variance of the outstanding payments from IBNR claims. We compute these by noting that an IBNR claim behaves as an RBNS claim once reported. Therefore, we can make use of the results from Section 4.1. We may also note that the outstanding payments from IBNR claims are independent of $\mathcal{F}_\tau$ and any conditioning will therefore disappear. Moreover, in Section 4.1 the payments had a mean and variance depending on the index i. For the IBNR claims, we will consider a finite number of possible means and variances. Each claim will then be given an initial reserve category C_i, as in Antonio and Plat (2014), determining which mean and variance the payment severities from this claim will have. We assume that these categories are randomly generated with probabilities $q_c = \mathbb{P}(C_i = c)$. We denote the mean and variance of the payment severities from a claim of category c by $\mu_{(c)}$ and $\sigma_{(c)}^2$, respectively.

We begin by noting that the total number of IBNR claims is distributed according to

$$N^{\mathcal{I}} \sim \mathrm{Po}\left(\int_0^\tau w_t \lambda_t (1 - P_{U|t}(\tau - t))dt\right), \tag{17}$$

see, e.g. Norberg (1993) and Antonio and Plat (2014). For notational brevity, we introduce the notation:

$$\Lambda_\tau := \int_0^\tau w_t \lambda_t (1 - P_{U|t}(\tau - t)) dt.$$

Let $R^{\mathcal{I},c}$ denote the outstanding payments from an IBNR claim belonging to reserve category c. The expectation of the outstanding payments from IBNR claims is then

$$\begin{aligned}\mathbb{E}\left[R^{\mathcal{I}} \middle| \mathcal{F}_\tau\right] &= \mathbb{E}\left[\sum_{i=0}^{N^{\mathcal{I}}} R_i^{\mathcal{I}}\right] = \mathbb{E}\left[N^{\mathcal{I}}\right] \mathbb{E}\left[R_i^{\mathcal{I}}\right] = \mathbb{E}\left[N^{\mathcal{I}}\right] \sum_{c=1}^{m} q_c \mathbb{E}\left[R^{\mathcal{I},c}\right] \\ &= \Lambda_\tau \left(\sum_{c=1}^{m} q_c \mu_{(c)}\right) \left(\sum_{l \geq 0} \frac{\bar{h}_{pl}}{\bar{h}_{sl}} \left(1 - e^{-\bar{h}_{sl}\Delta t_l}\right) e^{-\bar{H}_s(t_{l-1})}\right),\end{aligned}$$

where we have made use of Proposition 1 together with the fact that $\tau_i = 0$ by definition for the IBNR claims once reported. Moreover, by a variance decomposition, the variance is

$$\begin{aligned}\operatorname{Var}\left(R^{\mathcal{I}} \middle| \mathcal{F}_\tau\right) &= \mathbb{E}\left[N^{\mathcal{I}}\right] \mathbb{E}\left[\left(R_i^{\mathcal{I}}\right)^2\right] = \mathbb{E}\left[N^{\mathcal{I}}\right] \sum_{c=1}^{m} q_c \mathbb{E}[(R^{\mathcal{I},c})^2] \\ &= \Lambda_\tau \sum_{c=1}^{m} q_c \left(\sum_{l=0}^{m} \left(\sigma_{(c)}^2 a_l + \mu_{(c)}^2 a_l (2a_l + b_l)\right) + 2\mu_{(c)}^2 \sum_{l=0}^{r-1} \sum_{k=l+1}^{r} a_k c_l\right),\end{aligned}$$

with

$$a_l := \frac{\bar{h}_{pl}}{\bar{h}_{sl}} \left(1 - e^{-\bar{h}_{sl}\Delta t_l}\right) e^{-\bar{H}_s(t_{l-1})}, \tag{18}$$

$$b_l := 1 + 2\frac{\bar{h}_{pl}}{\bar{h}_{sl}} - 2h_{p,l}\Delta t_l \frac{e^{-\Delta t_l \bar{h}_{sl}}}{1 - e^{-\Delta t_l \bar{h}_{sl}}}, \tag{19}$$

$$c_l := \Delta t_l h_{p,l}. \tag{20}$$

We summarise the above in the following proposition:

Proposition 3. *The conditional expectation of the total outstanding claims payments from IBNR claims is*

$$\mathbb{E}\left[R^{\mathcal{I}} \middle| \mathcal{F}_\tau\right] = \Lambda_\tau \left(\sum_{c=1}^{m} q_c \mu^{(c)}\right) \sum_{l \geq 0} \frac{\bar{h}_{pl}}{\bar{h}_{sl}} \left(1 - e^{-\bar{h}_{sl}\Delta t_l}\right) e^{-\bar{H}_s(t_{l-1})},$$

and the variance is

$$\operatorname{Var}\left(R^{\mathcal{I}} \middle| \mathcal{F}_\tau\right) = \Lambda_\tau \left(\sum_{l=0}^{m} \left(\tilde{\sigma}^2 a_l + \tilde{\mu}^2 a_l (2a_l + b_l)\right) + 2\tilde{\mu}^2 \sum_{l=0}^{r-1} \sum_{k=l+1}^{r} a_k c_l\right),$$

with a_l, b_l and c_l as in (18)–(20)*, and*

$$\tilde{\mu}^2 := \sum_{c=1}^{m} q_c \mu_{(c)}^2,$$

$$\tilde{\sigma}^2 := \sum_{c=1}^{m} q_c \sigma_{(c)}^2.$$

4.2.1 Cash flows of IBNR claims payments

The cash flows in specific time intervals is a lot less straightforward for the IBNR claims. To see why this is the case, let $T_i^{\mathcal{I}}$ denote the time of reporting for the ith of the (unordered) IBNR claims. A well-known result, which is straightforward to prove, is that

$$\mathbb{P}(T_i^{\mathcal{I}} \leq s | N^{\mathcal{I}} = n) = \frac{\Lambda_s}{\Lambda_\tau}. \tag{21}$$

The specific form of this probability is not essential for our purposes. Instead, we merely note that it is indeed possible to derive the density of $T_i^{\mathcal{I}} | N^{\mathcal{I}} = n$, which we denote by g. This density can then be used to compute

$$\mathbb{E}[R^{\mathcal{I}}(s,t)] = \mathbb{E}[N^{\mathcal{I}}]\mathbb{E}[\mathbb{E}[R_i^{\mathcal{I}}(s - T_i^{\mathcal{I}}, t - T_i^{\mathcal{I}})|N^{\mathcal{I}}]],$$

where

$$\begin{aligned}
\mathbb{E}[R_i^{\mathcal{I}}(s - T_i^{\mathcal{I}}, t - T_i^{\mathcal{I}})|N^{\mathcal{I}}] &= \mathbb{E}[\mathbb{E}[R_i^{\mathcal{I}}(s - T_i^{\mathcal{I}}, t - T_i^{\mathcal{I}})|T_i^{\mathcal{I}}]|N^{\mathcal{I}}] \\
&= \int_0^\tau \mathbb{E}[R_i^{\mathcal{I}}(s - T_i^{\mathcal{I}}, t - T_i^{\mathcal{I}})|T_i^{\mathcal{I}} = u] g(u) du \\
&= \int_0^\tau \mathbb{E}[R_i^{\mathcal{I}}(s - u, t - u)] g(u) du.
\end{aligned}$$

The RBNS-result in Proposition 2 gives the expectation in the integral above

$$\mathbb{E}[R_i^{\mathcal{I}}(s-u, t-u)] = \tilde{\mu} \sum_{l=\kappa(s-u)}^{\kappa(t-u)} \frac{\bar{h}_{pl}}{\bar{h}_{sl}} \left(1 - e^{-\bar{h}_{sl}\Delta t_{il}}\right) e^{-\bar{H}_s(t_{l-1})},$$

where Δt_{il} is taken over the partition $\widetilde{\mathcal{T}}$ which includes the points $s-u$ and $t-u$, from which we see that the integral, while technically computable, is (likely) not analytically tractable.

5 Likelihood Estimation and properties of the parameter estimators

In this section, we take a closer look at the maximum likelihood estimation of the parameters in the model. We focus mainly on the parameters of the development process, showing asymptotic properties of the parameter estimators, such as consistency and asymptotic normality. Additionally, we show finite sample properties such as whether the estimators are biased or not. For the occurrence process and the reporting delays, we are not able to get as precise results unless we make certain assumptions about $\mathbb{P}_{U|t}$. One such assumption that we discuss in this section is that it forms an exponential family. We do not discuss the parameters, and their parameter estimators, of the payment severities since these entirely depend on the choice of the distribution $\mathbb{P}_p$, and not on any other parts of the model. Using the results of this section, we are also able to motivate the method used in Antonio and Plat (2014) of taking parameter uncertainty into account when quantifying prediction uncertainty, i.e. approximating the parameter uncertainty through the asymptotic normality of the parameter estimators.

Let (T_i^o, U_i^o) denote the ith pair of the observed occurrence times and reporting delays and let N^o denote the total number of observed/reported claims. For RBNS claims, τ_i denoted the current time in previous sections. In agreement with this, for all reported claims, settled or not settled, τ_i now denotes $(\tau - T_i - U_i) \wedge V_i$, which is consistent with the previous definition, and with Antonio and Plat (2014), since it corresponds to the current time for the RBNS claims.

The complete likelihood of the observed claims is, see for instance (5) in Antonio and Plat (2014),

$$
\begin{aligned}
L(\lambda, \boldsymbol{\alpha}, \boldsymbol{h}, \boldsymbol{\gamma}; \mathcal{F}_\tau) = &\left\{ \prod_{i=1}^{N^o} w(T_i^o)\lambda(T_i^o)\mathbb{P}_{U|t}(\tau - T_i^o; \boldsymbol{\alpha}) \right\} \\
&\quad \cdot \exp\left(- \int_0^\tau w(t)\lambda(t)\mathbb{P}_{U|t}(\tau - t; \boldsymbol{\alpha}) dt \right) \\
&\cdot \left\{ \prod_{i=1}^{N^o} \frac{\mathbb{P}_{U|t}(dU_i^o; \boldsymbol{\alpha})}{\mathbb{P}_{U|t}(\tau - T_i^o; \boldsymbol{\alpha})} \right\} \\
&\cdot \prod_{i=1}^{N^o} \Bigg\{ \prod_j \prod_e h_e^{\delta_{ije}}(V_{ij}) \\
&\quad \cdot \exp\left(- \int_0^{\tau_i} \left(h_{se}(u) + h_{sep}(u) + h_p(u)\right) du \right) \Bigg\} \\
&\cdot \prod_{i=1}^{N^o} \prod_j \mathbb{P}_p(dP_{ij'}; \boldsymbol{\gamma}),
\end{aligned} \tag{22}
$$

where δ_{ije} is a Kronecker delta in the last two indexes, equal to 1 if the jth event in the development of claim i is of type $e \in \{se, sep, p\}$, and 0 otherwise. Here $\boldsymbol{\alpha}$ denotes the parameter vector of the distribution of the reporting delay U_i, $\boldsymbol{\gamma}$ the one for the payment severity distribution, and $\boldsymbol{h}$ the vector of the hazard rate functions of the development process. The first three rows of (22) correspond to the occurrence times and the reporting delay, the fourth and fifth rows correspond to the development process, and the sixth row corresponds to the payment severities. These three parts of the likelihood may be optimised separately since there is no overlap in the dependence of the parameters. However, the first and second row cannot be optimised separately. This fact seems to be overlooked in Antonio and Plat (2014) where, based on Section 4.1, it appears as if the (truncated) distribution of the reporting delays are directly fitted to the data $\{U_i^o\}_{i=1}^{N^o}$. This fact is also commented upon in Section 4 of Sigmundsdóttir and Lindholm (2017).

Based on (22), it is straightforward to conclude that the log-likelihood is proportional to

$$
\begin{aligned}
l(\lambda, \boldsymbol{\alpha}, \boldsymbol{h}, \boldsymbol{\gamma}; \mathcal{F}_\tau) \propto &\sum_{i=1}^{N^o} \log \lambda(T_i^o) - \int_0^\tau w(t)\lambda(t)\mathbb{P}_{U|t}(\tau - t; \boldsymbol{\alpha}) dt \\
&+ \sum_{i=1}^{N^o} \log \mathbb{P}_{U|t}(dU_i^o; \boldsymbol{\alpha}) \\
&+ \sum_{i=1}^{N^o} \sum_j \sum_e \delta_{ije} \log h_e(V_{ij}) \\
&- \sum_{i=1}^{N^o} \int_0^{\tau_i} \left(h_{se}(u) + h_{sep}(u) + h_p(u)\right) du \\
&+ \sum_{i=1}^{N^o} \sum_{j'} \log \mathbb{P}_p(dP_{ij'}; \boldsymbol{\gamma}).
\end{aligned} \tag{23}
$$

In addition to the discretisation scheme of the rates of the development process, Antonio and Plat (2014) discretise $\lambda(t)$ and $w(t)$, something we also consider here. Let $\mathcal{D} := \{d_0, d_1, \ldots, d_m\}$ be a finite partition of $\mathbb{R}_+$. On the intervals made up by this partition, let $\lambda(t)$ and $w(t)$ be piecewise constant according to

$$\begin{aligned}\lambda(t) &= \lambda_i, \quad t \in [d_{i-1}, d_i), \quad i = 1, \ldots, m,\\ w(t) &= w_i, \quad t \in [d_{i-1}, d_i), \quad i = 1, \ldots, m.\end{aligned} \tag{24}$$

Given this and the previously introduced discretisation scheme, the log-likelihood in (23) becomes

$$\begin{aligned}l(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{h}, \boldsymbol{\gamma}; \mathcal{F}_\tau) \propto &\sum_{i=1}^{m} N_i^o \log \lambda_i - \sum_{i=1}^{m} w_i \lambda_i \int_{d_{i-1}}^{d_i} \mathbb{P}_{U|t}(\tau - t; \boldsymbol{\alpha}) dt\\ &+ \sum_{i=1}^{N^o} \log \mathbb{P}_{U|t}(dU_i^o; \boldsymbol{\alpha})\\ &+ \sum_{i=1}^{m} \left(N_{sep,i}^o \log h_{se,i} + N_{se,i}^o \log h_{sep,i} + N_{p,i}^o \log h_{p,i} \right)\\ &- \sum_{i=1}^{N^o} \sum_{j=1}^{m} (h_{se,j} + h_{sep,j} + h_{p,j}) \int_{t_{j-1}}^{t_j} 1_{\{u \le \tau_i\}} du\\ &+ \sum_{i=1}^{N^o} \sum_{j'} \log \mathbb{P}_p(dP_{ij'}; \boldsymbol{\gamma}),\end{aligned} \tag{25}$$

where N_i^o is the observed number of claims that occurred in $[d_{i-1}, d_i)$ and $N_{e,i}^o$ is the number of observed events of type e in $[t_{i-1}, t_i)$. Moreover, $\boldsymbol{\lambda}$ is a vector containing the piecewise constant rates λ_i and $\boldsymbol{h}$ now denotes the vector containing the rates $h_{e,k}$.

Based on this log-likelihood, we can conclude that the maximum likelihood estimators of the piecewise constant hazard rates are

$$\hat{h}_{e,k} = \frac{N_{e,k}^o}{\sum_{i=1}^{N^o} \int_{t_{k-1}}^{t_k} 1_{\{u \le \tau_i\}} du}, \tag{26}$$

for $e \in \{p, se, sep\}$ and $k = 1, \ldots, r$. Moreover, the Hessian of the log-likelihood is block-diagonal

$$J = \begin{bmatrix} J_{\lambda,\boldsymbol{\alpha}} & 0 & 0 \\ 0 & J_h & 0 \\ 0 & 0 & J_\gamma \end{bmatrix},$$

where, firstly, J_γ is the Hessian of the log-likelihood corresponding to the payment severities, and therefore wholly depends on the choice of $\mathbb{P}_p$. Secondly,

$$J_{\lambda,\boldsymbol{\alpha}} = \begin{bmatrix} \frac{\partial^2 l}{\partial \lambda_1^2} & 0 & \ldots & 0 & [\nabla_{\boldsymbol{\alpha}} \frac{\partial l}{\partial \lambda_1}]' \\ 0 & \frac{\partial^2 l}{\partial \lambda_2^2} & \ldots & 0 & [\nabla_{\boldsymbol{\alpha}} \frac{\partial l}{\partial \lambda_2}]' \\ 0 & 0 & \ddots & 0 & \vdots \\ 0 & 0 & \ldots & \frac{\partial^2 l}{\partial \lambda_m^2} & [\nabla_{\boldsymbol{\alpha}} \frac{\partial l}{\partial \lambda_m}]' \\ \nabla_{\boldsymbol{\alpha}} \frac{\partial l}{\partial \lambda_1} & \nabla_{\boldsymbol{\alpha}} \frac{\partial l}{\partial \lambda_2} & \ldots & \nabla_{\boldsymbol{\alpha}} \frac{\partial l}{\partial \lambda_m} & J_{\boldsymbol{\alpha}} \end{bmatrix},$$

where $\nabla_{\boldsymbol{\alpha}}$ denotes the gradient taken with respect to $\boldsymbol{\alpha}$ and $J_{\boldsymbol{\alpha}}$ is the Hessian of the log-likelihood with respect to $\boldsymbol{\alpha}$. Thirdly and finally, J_h is the Hessian of the log-likelihood with respect to the components of $\boldsymbol{h}$, which is a diagonal matrix consisting of the diagonal elements

$$\frac{\partial^2}{\partial h_{e,l}^2} l(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{h}; \mathcal{F}_\tau) = -\frac{N_{e,l}^o}{h_{e,l}^2},$$

for $l = 1, \ldots, r$ and $e \in \{se, sep, p\}$. These are all negative, and therefore, since J_h is diagonal, the log-likelihood of the development process is concave, and thus the parameter estimators $\hat{h}_{e,l}$ are unique. Based on the above Hessian, the observed Fisher information of the likelihood function of the development processes is a diagonal matrix with diagonal elements

$$\frac{N_{e,k}^o}{h_{e,k}^2}. \tag{27}$$

The expected Fisher information matrix, $I(\boldsymbol{h})$, is then acquired by taking the expectation of these elements. These are however difficult to compute, although we can note by taking (35) together with (40), (50) and (52) from the proof of Proposition 1 in Appendix A, that

$$\mathbb{E}\left[\frac{N_{e,k}^o}{h_{e,k}^2}\middle| N^o, \{\gamma_i\}\right] = \sum_{i=1}^{N^o} \frac{1_{\{\gamma_i > t_{k-1}\}}}{h_{e,k}\bar{h}_{sk}} \left(1 - e^{-\bar{h}_{sl}(t_k \wedge \gamma_i - t_{k-1})}\right) e^{-\bar{H}_s(t_{k-1})}, \tag{28}$$

where $\gamma_i := \tau - T_i^o - U_i^o$. The matrix consisting of these as its elements, normalised by N^o, is a consistent estimator of the expected Fisher information matrix in the exposure measure, i.e. as $w_l \to \infty$. To see that this is the case we can use the WLLN together with, e.g. part (iii) of Theorem 8.2 on page 302 in Gut (2005) since $N^o \overset{a.s.}{\to} \infty$ as $w_l \to \infty$. Of course, the observed Fisher information matrix is also a consistent estimator of the expected Fisher information, however, in (28) we are conditioning on more of the structure than if we were to use the observed Fisher information. It is however not entirely clear whether one should prefer the observed or the expected Fisher information when, for instance, approximating the variance of an estimator by the asymptotic variance (expected Fisher information), see for instance Efron and Hinkley (1978) for a discussion and a frequentist justification for using the observed over the expected Fisher information. Using the conditional elements in (28) can be motivated heuristically by the so-called conditionality principle since the likelihood in (22) factorises into the components described below (22), without dependence between the parameters in the resulting parts. However, we are considering all variables in the likelihood, and therefore the conditionality principle is not entirely applicable. For a detailed presentation of the conditionality principle, see Birnbaum (1962).

For the estimators in (26), it is possible to state the following proposition dealing with bias:

Proposition 4. *For all $k = 1, \ldots, r$ it holds that*

$$\mathbb{E}[\hat{h}_{p,k}] = h_{p,k},$$

and, for $e \in \{se, sep\}$, that

$$\mathbb{E}[\hat{h}_{e,k}] > h_{e,k},$$

i.e. the estimators of the hazard rates of pure payment events are unbiased, while there is an upwards bias of the estimators of the hazard rates of the settlement events.

The proof of Proposition 4 is given in Appendix A. While there is an upwards bias of the estimators, it is possible to state the following proposition dealing with consistency and asymptotic normality:

Proposition 5. *Assume either that $w_t = w\tilde{w}_t$ or that the discretisation of* (24) *holds, i.e. if the discretisation does not hold that the exposure is proportional to a scale factor w. Then, for $e \in \{p, se, sep\}$, it holds that*

$$\hat{h}_{e,k} \xrightarrow{p} h_{e,k}$$

as either $w \to \infty$ or $w_l \to \infty$ for at least one $l = 1, \dots, m$, i.e. the parameter estimators $\hat{h}_{e,k}$ are consistent in terms of the exposure measure. Moreover,

$$\sqrt{N^o}\left(\hat{\boldsymbol{h}} - \boldsymbol{h}\right) \xrightarrow{d} N\left(0, [I(\boldsymbol{h})]^{-1}\right),$$

as either $w \to \infty$ or $w_l \to \infty$ for at least one $l = 1, \dots, m$, i.e. $\sqrt{N^o}\left(\hat{\boldsymbol{h}} - \boldsymbol{h}\right)$ is asymptotically normal with mean vector zero and covariance matrix $[I(\boldsymbol{h})]^{-1}$, where $I(\boldsymbol{h})$ is the expected Fisher information of the part of the log-likelihood in (25) *corresponding to the development process.*

The proof of Proposition 5 is given in Appendix A. We can conclude from Proposition 5 that while the estimators are biased, they are at least asymptotically unbiased, something that will be useful in the next section. A thing to remark on is that it is possible to generalise the assumption on the non-discretised exposure in Proposition 5. This, however, obscures the result and is therefore left out. For other assumptions, it is straightforward to change the first part of the proof to see if the results still hold. Moreover, the assumption in the proposition is thought to capture most real-world scenarios.

What now remains to consider are the estimators of the λ_is and $\boldsymbol{\alpha}$. These do not have closed form solutions. However, we can make some assumptions on $\mathbb{P}_{U|t}$ such that $\hat{\lambda}_i$ and $\hat{\boldsymbol{\alpha}}$ have attractive properties. For an example of a case when these maximum likelihood estimators are unique, consider the case when the distribution of the reporting delay belongs to an exponential family. Some examples of standard distributions that are exponential families are the exponential, gamma and log-normal distributions. A family of distributions is said to form an exponential family if the distributions have densities of the form

$$f_U(u; \boldsymbol{\alpha}) = g(u) \exp\left\{\boldsymbol{\eta}(\boldsymbol{\alpha})' \boldsymbol{T}(u) - B(\boldsymbol{\alpha})\right\}, \tag{29}$$

with respect to some common measure. The vector $\boldsymbol{\eta}(\boldsymbol{\alpha})$ is usually referred to as the canonical parameter vector and $\boldsymbol{T}(u)$ the sufficient statistic. For more on exponential families, see Chapter 1.5 of Lehmann and Casella (1998).

Taking the logarithm of (29) and inserting it into the log-likelihood of the occurrence times and reporting delays yields a quantity proportional to

$$\begin{aligned}\sum_{i=1}^{m} N_i^o \log \lambda_i - \sum_{i=1}^{m} w_i \lambda_i \int_{d_{i-1}}^{d_i} \mathbb{P}_{U|t}(\tau - t; \boldsymbol{\alpha}) dt \\ + \sum_{i=1}^{N^o} \log g(U_i^o) + \boldsymbol{\eta}(\boldsymbol{\alpha})' \sum_{i=1}^{N^o} \boldsymbol{T}(U_i^o) - \sum_{i=1}^{m} N_i^o B(\boldsymbol{\alpha}).\end{aligned}$$

The second term depends only on the parameters, the third term only on data, while the other terms depend on both data and parameters. From this, we see that this belongs to an exponential family with canonical parameter vector

$$(\log \lambda_1 - B(\boldsymbol{\alpha}), \dots, \log \lambda_m - B(\boldsymbol{\alpha}), \boldsymbol{\eta}(\boldsymbol{\alpha})),$$

and sufficient statistic

$$\left(N_1^o, \dots, N_m^o, \sum_{i=1}^{N^o} \boldsymbol{T}(U_i^o)\right).$$

Therefore, the MLE in this parametrisation, if it exists, is unique, see, e.g. Example 6.3 of Section 6.6 in Lehmann and Casella (1998). Moreover, by Theorem 5.1 in Lehmann and Casella (1998), the conditions of which are shown to hold in the example mentioned above, the MLE is consistent, and, when multiplied by the square root of N^o, asymptotically normal. These convergences are, as in Proposition 5, in terms of the exposure measure. However, the theorem is in terms of the sample size, which in our situation is random. Nevertheless, we may use the same method as at the beginning of the proof of Proposition 5 to show that the convergences hold in terms of the exposure measure. Now, from the theorem, we know that the asymptotic variance-covariance matrix of the estimator of the parameter vector is the inverse of the expected Fisher information. In the canonical parametrisation, the expected Fisher information is the same as the observed Fisher information. This equality is a well-known result that can be seen to hold by taking the logarithm of (29) and computing the Hessian with respect to the canonical parameter vector $\boldsymbol{\eta}$, noting that the resulting matrix does not depend on data. Calculating this quantity by hand is tedious and not very illuminating, resulting in an expression involving integrals of $\mathbb{P}_{U|t}(\tau - t; \boldsymbol{\eta})$. Since the observed and expected Fisher information coincide, the most straightforward way of acquiring them is through standard methods when numerically maximising the log-likelihood.

Above we have shown that there are situations when it is motivated to assume that the asymptotic distribution of the parameter estimators is normal, as is assumed in Antonio and Plat (2014). We will comment more on this, and make use of the results in the present section, in Section 6.

6 Assessment of prediction uncertainty

An essential part of reserving is the assessment of prediction uncertainty taking estimation error into account. The standard method of doing this that is always applicable is the bootstrap, which can be either parametric or non-parametric. Antonio and Plat (2014) describe how to perform a parametric bootstrap for the model we consider in the present paper. In Section 5.3 they describe how to take parameter uncertainty into account by simulating from the asymptotic (normal) distribution of the parameter estimators, something we motivate in Section 5 of the present paper. Since there is not much more to add to this simulation/bootstrap approach, this section will focus on another way of assessing the prediction uncertainty (taking estimation error into account).

The method we consider here is that of the semi-analytical approximation of the conditional mean squared error of prediction (MSEP) introduced in Lindholm et al. (2018). This method is not necessarily more accurate than a bootstrap since it rests on an approximation. However, it has other redeeming qualities such as it requiring no simulations, and it is straightforward to implement relying only on standard output from a regular numerical maximum likelihood estimation. Since it requires no simulations, it has no Monte-Carlo error, there is no question of convergence of estimates (as would be the case in a bootstrap approach), and it is quick to compute. For another paper using this method, see Wahl et al. (2019) where it is applied to the models considered in Verrall et al. (2010), Miranda et al. (2011, 2012) and to the model introduced in the paper itself. For similar results to Lindholm et al. (2018) applied particularly to the distribution-free chain ladder model, see Diers et al. (2016), Röhr (2016) and Buchwalder et al. (2006).

To define the conditional MSEP, we introduce a random variable X, a σ-algebra $\mathcal{F}$, and a

$\mathcal{F}$-measurable predictor $\hat{X}$. The conditional MSEP is then defined as

$$\begin{aligned}\mathrm{MSEP}(X, \hat{X}) &= \mathbb{E}[(X - \hat{X})^2 | \mathcal{F}] \\ &= \mathrm{Var}(X|\mathcal{F}) + (E[X|\mathcal{F}] - \hat{X})^2.\end{aligned}$$

The approach in Lindholm et al. (2018) uses a re-sampling/bootstrap argument which results in the following approximation of the conditional MSEP:

$$\widehat{\mathrm{MSEP}}(X, \hat{X}) = \mathrm{Var}(X \mid \mathcal{F})(\hat{\boldsymbol{\theta}}) + \nabla\mathbb{E}[X|\mathcal{F}](\hat{\boldsymbol{\theta}})' \widehat{\mathrm{Cov}}(\hat{\boldsymbol{\theta}}) \nabla\mathbb{E}[X|\mathcal{F}](\hat{\boldsymbol{\theta}}). \tag{30}$$

The result relies on three assumptions: (i) the parameter estimator is unbiased; (ii) the process variance is computable; (iii) the covariance of the parameter estimator is computable. Based on Proposition 1 and 3, it is clear that (ii) holds. By Proposition 4, (i) does not hold. In the case of the estimators being biased, a correction term

$$\mathrm{Bias}(\hat{\boldsymbol{\theta}})' \nabla\mathbb{E}[R|\mathcal{F}_\tau](\hat{\boldsymbol{\theta}}) \nabla\mathbb{E}[R|\mathcal{F}_\tau](\hat{\boldsymbol{\theta}})' \mathrm{Bias}(\hat{\boldsymbol{\theta}}) \tag{31}$$

must be added to the approximation. The bias is not quantifiable in our setting and computing this expression is therefore not possible analytically. However, by Proposition 5 together with assuming that the reporting delay distribution follows an appropriate distribution, e.g. one that is part of an exponential family, we know that the estimators are consistent (and therefore asymptotically unbiased). In Section 8 we investigate this approximate unbiasedness through a numerical example for the parameter estimators of the discretised rates of the development process. There we see that the bias given by Proposition 4 is too small to be of any practical significance.

Consequently, we may approximate the MSEP by disregarding this (likely) small contribution. Assumption (iii) does not technically hold, especially not for the parameter estimators of the parameters in the occurrence part of the model, but neither is the variance of $\hat{h}_{e,k}$ in (26) computable. We may, however, approximate the covariance matrix by the asymptotic variance, which is the inverse Fisher information matrix. Since it is not computable for all parts of the likelihood, we may use the observed Fisher information, which we can obtain from a standard numerical maximum likelihood estimation through the negative of the Hessian of the log-likelihood evaluated at the maximum likelihood estimator. However, as mentioned before in Section 5, it is not entirely clear whether one should prefer the expected or the observed Fisher information, see Efron and Hinkley (1978).

In the setting of the present paper, when considering the total outstanding payments, the approximation in (30) is

$$\widehat{\mathrm{MSEP}}(R, \hat{R}) = \mathrm{Var}(R \mid \mathcal{F}_\tau)(\hat{\boldsymbol{\theta}}) + \nabla\mathbb{E}[R|\mathcal{F}_\tau](\hat{\boldsymbol{\theta}})' \widehat{\mathrm{Cov}}(\hat{\boldsymbol{\theta}}) \nabla\mathbb{E}[R|\mathcal{F}_\tau](\hat{\boldsymbol{\theta}}), \tag{32}$$

where $\hat{\boldsymbol{\theta}}$ here denotes the vector containing all parameter estimators considered in Section 5, and the moments of the outstanding payments R are given in Section 4, see (1) and (2) together with (3), (4), and Proposition 1 and 3. Note that we could consider any other range of quantities using this approximation. For instance, we could exchange R for the number of RBNS payments made in the future and $\hat{R}$ for its prediction.

Instead of considering the rates $\hat{h}_{e,k}$, we consider the transformed rates $\hat{\bar{h}}_{ek}$ since this simplifies computations of the gradients a bit. This transformation is a simple linear transformation, and therefore the asymptotic variance is $L[I(\boldsymbol{h})]^{-1}L'$ where L is the matrix defining the linear transformation from the rates $h_{e,k}$ to the rates $\bar{h}_{ek}$.

We end this section by including a proposition giving all the necessary gradients needed to calculate (32). The proof is a simple exercise in taking derivatives and keeping track of indices and is therefore left out.

Proposition 6. *The gradients needed to calculate* (32) *for the RBNS part are*

$$\frac{\partial}{\partial \bar{h}_{pk}} \mathbb{E}[R^{\mathcal{R}}|\mathcal{F}_\tau] = \sum_{i=1}^{N^{\mathcal{R}}} \frac{\mu_i a_{ik}}{\bar{h}_{pk}},$$
$$\frac{\partial}{\partial \bar{h}_{sk}} \mathbb{E}[R^{\mathcal{R}}|\mathcal{F}_\tau] = \sum_{i=1}^{N^{\mathcal{R}}} \mu_i \left(a_{ik}\Delta t_{ik} \frac{e^{-\Delta t_{ik}\bar{h}_{sk}}}{1-e^{-\Delta t_{ik}\bar{h}_{sk}}} - \frac{a_{ik}}{\bar{h}_{sk}} - \Delta t_{ik} \sum_{l=k+1}^{r} a_{il} \right),$$
$$\frac{\partial}{\partial \mu_i} \mathbb{E}[R^{\mathcal{R}}|\mathcal{F}_\tau] = \sum_{i=1}^{N^{\mathcal{R}}} \frac{1}{\mu_i} \mathbb{E}[R_i^{\mathcal{R}}|\mathcal{F}_\tau],$$

where the a_{ik}s are defined by (14)*. The gradients needed to calculate* (32) *for the IBNR part are*

$$\frac{\partial}{\partial \mu_{(c)}} \mathbb{E}[R^{\mathcal{I}}|\mathcal{F}_\tau] = \frac{q_c}{\sum_{c=1}^m q_c \mu_{(c)}} \mathbb{E}[R^{\mathcal{I}}|\mathcal{F}_\tau]$$
$$\frac{\partial}{\partial q_c} \mathbb{E}[R^{\mathcal{I}}|\mathcal{F}_\tau] = \frac{\mu_{(c)}}{\sum_{c=1}^m q_c \mu_{(c)}} \mathbb{E}[R^{\mathcal{I}}|\mathcal{F}_\tau]$$
$$\frac{\partial}{\partial \bar{h}_{sk}} \mathbb{E}[R^{\mathcal{I}}|\mathcal{F}_\tau] = \Lambda_\tau \left(\sum_{c=1}^m q_c \mu_{(c)} \right) \left(a_k \Delta t_k \frac{e^{-\Delta t_k \bar{h}_{sk}}}{1-e^{-\Delta t_k \bar{h}_{sk}}} - \frac{a_k}{\bar{h}_{sk}} - \Delta t_k \sum_{l=k+1}^{r} a_l \right)$$
$$\frac{\partial}{\partial \bar{h}_{pk}} \mathbb{E}[R^{\mathcal{I}}|\mathcal{F}_\tau] = \Lambda_\tau \left(\sum_{c=1}^m q_c \mu_{(c)} \right) \frac{a_k}{\bar{h}_{sk}}$$
$$\frac{\partial}{\partial \lambda_k} \mathbb{E}[R^{\mathcal{I}}|\mathcal{F}_\tau] = \frac{\mathbb{E}[R^{\mathcal{I}}|\mathcal{F}_\tau]}{\Lambda_\tau} \frac{\partial}{\partial \lambda_k} \Lambda_\tau$$
$$\nabla_{\boldsymbol{\alpha}} \mathbb{E}[R^{\mathcal{I}}|\mathcal{F}_\tau] = \frac{\mathbb{E}[R^{\mathcal{I}}|\mathcal{F}_\tau]}{\Lambda_\tau} \nabla_{\boldsymbol{\alpha}} \Lambda_\tau,$$

where a_k is given by (18) *and $\nabla_{\boldsymbol{\alpha}}$ denotes the gradient taken w.r.t. $\boldsymbol{\alpha}$.*

7 One-Year Reserve Risk and a comment on simulation

This section is closely related to Section 6.1 of Sigmundsdóttir and Lindholm (2017) where it is described how to simulate to get the Solvency Capital Requirement (SCR). Their method requires nested simulations and is quite computationally expensive. Here we use the formulas acquired in the present paper for the conditional moments of outstanding payments to simplify this to require only one layer of simulations, which should increase the computation speed drastically.

Under the Solvency II-directive, insurers must compute the SCR, which is supposed to *"ensure all quantifiable risks to which an insurance or reinsurance undertaking is exposed are taken into account"*. It corresponds to the 99.5% Value-at-Risk of the CDR over a one-year period, see Article 101(3) of European Commission (2015) and, for more on the one-year risk perspective, see Ohlsson and Lauzeningks (2009). The one-year CDR is the realised difference between the prediction of the ultimate claims payments made today and in a year, i.e. it is

$$\text{CDR} := \mathbb{E}[U|\mathcal{F}_\tau](\hat{\boldsymbol{\theta}}_{(\tau)}; \mathcal{F}_\tau) - \mathbb{E}[U|\mathcal{F}_{\tau+1}](\hat{\boldsymbol{\theta}}_{(\tau+1)}; \mathcal{F}_{\tau+1})$$

where U is the ultimate claims payments, and $\hat{\boldsymbol{\theta}}_{(t)}$ is the $\mathcal{F}_t$-measurable estimator of the parameter vector $\boldsymbol{\theta}$ based on the observations available at time t. Here we see that one has to take re-estimation into account. Based on this quantity, the one-year SCR according to the Solvency II-directive is

$$\mathrm{SCR} := \mathrm{VaR}_{0.995}(\mathrm{CDR}),$$

where VaR_α is the $100 \times \alpha\%$ Value-at-Risk.

Given the moments acquired in Section 4, computing the SCR through simulations is a simple matter. Proceed as follows:

1. Compute the expected value of the outstanding payments based on the data available today, i.e. calculate
$$\mathbb{E}[R|\mathcal{F}_\tau](\hat{\boldsymbol{\theta}}_{(\tau)}; \mathcal{F}_\tau).$$
2. Simulate one-year ahead n times to acquire a sample of datasets
$$\{\mathcal{F}_{\tau+1}^{(i)}\}_{i=1}^n.$$
3. For each $i = 1, \ldots, n$, estimate $\hat{\boldsymbol{\theta}}_{(\tau+1)}^{(i)}$ and compute
$$\mathbb{E}[R|\mathcal{F}_{\tau+1}](\hat{\boldsymbol{\theta}}_{(\tau+1)}^{(i)}; \mathcal{F}_{\tau+1}^{(i)}).$$
4. The sample of CDRs is then
$$\mathrm{CDR}_{(i)} := \mathbb{E}[R|\mathcal{F}_\tau](\hat{\boldsymbol{\theta}}_{(\tau)}; \mathcal{F}_\tau) - P_{(i)}(\tau, \tau+1) - \mathbb{E}[R|\mathcal{F}_{\tau+1}](\hat{\boldsymbol{\theta}}_{(\tau+1)}^{(i)}; \mathcal{F}_{\tau+1}^{(i)}),$$
where $P_{(i)}(\tau, \tau+1)$ is the total payments made in the interval $[\tau, \tau+1)$ based on the ith simulation.
5. An estimator of the SCR is then
$$\widehat{\mathrm{SCR}} := \widehat{\mathrm{VaR}}_{0.995}(\{\mathrm{CDR}_{(i)}\}_{i=1}^n),$$
where $\widehat{\mathrm{VaR}}_\alpha(\cdot)$ is the $100 \times \alpha\%$ empirical Value-at-Risk function (i.e. the empirical quantile function applied to the negative amounts).

Given the above procedure, it is straightforward to extend the computations to any other risk measure.

Simulation of the new datasets can be done in the same way as is described in Antonio and Plat (2014). It should, however, be noted that there is a slight error in the way Antonio and Plat (2014) simulates IBNR claims in steps (a) and (b) of Section 5.1. In step (a), they simulate the total number of IBNR claims in an interval $[d_{l-1}, d_l)$, denoted here $N^{\mathcal{I}}(l)$, according to

$$N^{\mathcal{I}}(l) \sim \mathrm{Poisson}\left(\lambda_l w_l \int_{d_{l-1}}^{d_l} (1 - \mathbb{P}_{U|t}(\tau - t))dt\right). \tag{33}$$

Given this number they say that *"the occurrence times of the claims are uniformly distributed in* $[d_{l-1}, d_l)$*"*. This statement is not true in general, although for a narrow enough interval it might hold approximately. Instead, it is the case that the reporting times (i.e. the occurrence times plus the reporting delay) are distributed according to a slight modification of (21) given $N^{\mathcal{I}}(l)$. More precisely, for $s \in [d_{l-1}, d_l)$ it is distributed according to

$$\frac{\int_{d_{l-1}}^{s} (1 - \mathbb{P}_{U|t}(\tau - t))dt}{\int_{d_{l-1}}^{d_l} (1 - \mathbb{P}_{U|t}(\tau - t))dt} \tag{34}$$

on $[d_{l-1}, d_l)$. If we assume that the payment severities do not depend on the occurrence times and the reporting delays, we can note that given a set of simulated reporting times there is no need to invert to find the separate occurrence times and reporting delays since these are then of no particular interest, i.e. step (b) would be redundant.

Instead of simulating IBNR claims using (33) and (34), a perhaps more straightforward (but less efficient) method is to simulate the whole process from time 0 up until at least one year from now (i.e. time $\tau + 1$) and then only keep those claims that were IBNR at the current time τ. This way we only need to simulate from a (piecewise) homogeneous Poisson process and then for each occurrence simulate a reporting delay according to $\mathbb{P}_{U|t}$. However, if there are relatively few IBNR claims compared to RBNS claims at time τ, this will be highly inefficient.

8 Numerical Illustration

In this section, we simulate the development process using parameter values inspired by Antonio and Plat (2014). The purpose is to investigate the size of the biases of the parameter estimators in the development process part of the model to see if it is appropriate to disregard the bias correction in the semi-analytical approximation of the MSEP given in Section 6. We do not illustrate the MSEP approximation itself, nor do we illustrate other results of this paper such as the CDR computations in Section 7. Illustrating these methods would require us to simulate from the complete model, requiring us to make assumptions on all the different parts of the model. To make the illustration convincing we would need to make a quite extensive simulation study, preferably calibrated to realistic parameter values, which is outside the scope for this paper. We therefore limit ourselves to the development process and the results in Section 5 regarding the parameter estimators of the development process parameters. For the interested reader, see Wahl et al. (2019) for a paper implementing the MSEP approximation of Lindholm et al. (2018), and Sigmundsdóttir and Lindholm (2017) for an implementation of the CDR computations in Section 7.

Instead of simulating the whole development process, we simulate one single interval. Any subsequent interval will behave similarly, but with a (random) smaller sample size since some claims will be settled. We therefore also investigate some different sample sizes to see the effect of having a small number of active claims in a particular interval of the development process. Even though there will be a smaller sample size available for intervals further from reporting, the importance of correctly estimating the rates in these intervals will become smaller the further out one gets since, out there, there are relatively fewer claims.

We use the case study of Antonio and Plat (2014) in order to pick realistic parameter values and sample sizes. In Section 2.3 they state that there are 491,912 claims in total. These claims are split into two types, material damage ('material') and bodily injury ('injury'). To keep this illustration conservative sample size wise, based on Figure 3 in Antonio and Plat (2014), we assume that there are 40,000 material claims and 900 injury claims since these are the numbers available the last accident year (2008). Apart from being conservative, it might be the case that claims reported in 2008 (the last year in their data), compared to the claims reported in 1998 (the first year in their data), are more representative of the claims that are not yet settled. To investigate the behaviour in intervals further out from reporting we use 10% of these numbers, i.e. 4000 material claims and 90 injury claims.

For the rates, we use Figure 10 of Antonio and Plat (2014). We pick values that are close to the first piecewise constant estimate of the three types of events, both for material and injury claims. This choice is somewhat arbitrary, but most of the rates are, in any case, of the same magnitude. The rates we use are summarised in Table 1.

Lastly, we have to specify the length of the intervals to simulate over. These we choose according to the length of the interval for the first piecewise constant hazard given in Antonio and Plat (2014), see Section 4.3, i.e. four months for material claims and six months for injury claims.

Table 1: Parameter values used for simulation.

Claim type	h_{se}	h_{sep}	h_p
Injury	0.08	0.02	0.3
Material	0.45	0.35	0.15

Based on the above we simulate one million times for both the injury and material claims. The result of this is visualised in Figure 1 where kernel density estimators (KDE) of the simulated estimators of the rates given in Table 1 are shown for both the full and the conservative sample sizes, together with vertical lines indicating the parameter values used when simulating the data for comparison. We omit drawing vertical lines corresponding to the sample means of the simulated estimators since these coincide with the true parameter values to such an extent as to be indistinguishable, see Table 2 for the per mille differences. It is clear that the biases that the parameter estimators of the rates of the settlement events have according to Proposition 4 quickly becomes negligible, which motivates approximating the MSEP in Section 6 by neglecting the bias correction given by (31).

A final thing to note here is that even though the sample size is quite tiny for the conservative case for material claims (90 claims), the widths of the distributions of the estimators are not strikingly large.

Table 2: Per mille differences of the sample mean of the estimators compared to the (true) parameter values used for the simulations.

Sample size	Claim type	h_{se}	h_{sep}	h_p
Full	Injury	0.8	0.9	0.04
Conservative	Injury	7.0	6.6	0.009
Full	Material	0.009	0.02	0.0006
Conservative	Material	0.23	0.25	0.05

9 Conclusion

In this paper, we have analysed the model considered in Antonio and Plat (2014), which is based on the model class of Norberg (1993) and Haastrup and Arjas (1996). We have computed moments of the outstanding payments, split on IBNR and RBNS claims. Based on this, we have shown how it is straightforward to simulate to acquire the moments of the CDR. This method requires only one layer of simulations and is therefore quick to run, and convergence is convenient to assess since there is no need for nested simulations. We had a closer look at the maximum likelihood estimation, showing, for instance, the asymptotic normality of the parameter estimators (under weak assumptions), which motivates the use of the normal distribution when assessing parameter uncertainty, as is done in Antonio and Plat (2014). Based on Lindholm et al. (2018) and the asymptotic results for the parameter estimators, we have given a semi-analytical approximation of the conditional MSEP as a convenient method of assessing prediction uncertainty taking parameter uncertainty into account. This method requires no simulations and only uses standard output from numerical maximum likelihood estimation. Finally, we have shown that the estimators of the rates in

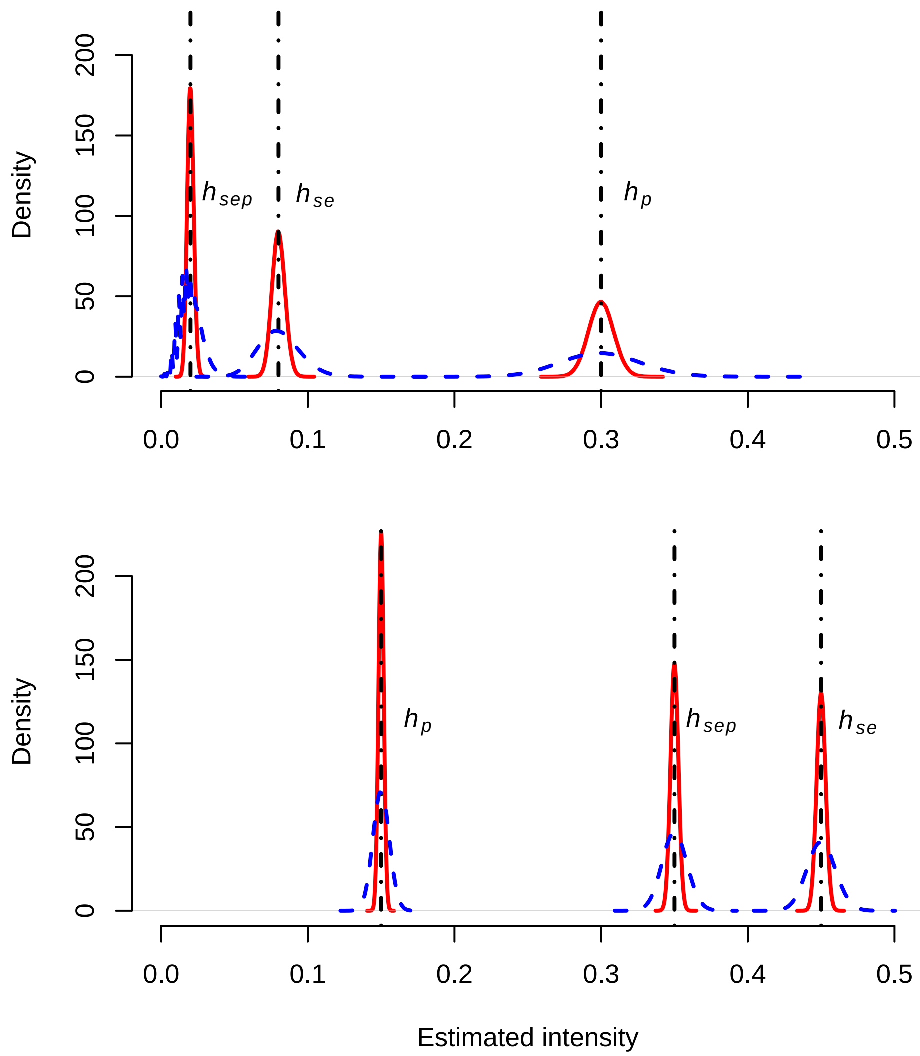

Figure 1: Kernel density estimators (KDE) of the estimators of the rates of the development process given in Table 1 based on 10^6 simulations. The top figure corresponds to the injury claims while the bottom figure to the material claims. The red solid KDE corresponds to using the full sample size (40,000 for material and 900 for injury claims), the blue dashed KDE corresponds to using the convervative sample size (10% of the full sample size), and the black dashed-dotted vertical lines are the true parameter values. The sample means of the parameter estimators are not shown in the figures since these cannot be visually distinguished from the true parameter values, see Table 2 for the per mille differences between the sample means of the estimators and the true parameter values.

the development process are well behaved given some realistic parameter values and sample sizes based on the work in Antonio and Plat (2014).

Acknowledgement

The author wants to thank Mathias Lindholm for insightful discussions, for formulating the original idea, and for giving helpful comments on preliminary versions of this paper.

References

Antonio, K. and 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.

Birnbaum, A. (1962). On the foundations of statistical inference. Journal of the American Statistical Association, 57(298):269–306.

Buchwalder, M., Bühlmann, H., Merz, M., and Wüthrich, M. V. (2006). The mean square error of prediction in the chain ladder reserving method (Mack and Murphy revisited). ASTIN Bulletin: The Journal of the IAA, 36(2):521–542.

Bühlmann, H., Schnieper, R., and Straub, E. (1980). Claims reserves in casualty insurance based on a probabilistic model. Bulletin of the Swiss Association of Actuaries, 80:21–45.

Diers, D., Linde, M., and Hahn, L. (2016). Addendum to 'the multi-year non-life insurance risk in the additive reserving model'[insurance math. econom. 52 (3)(2013) 590–598]: Quantification of multi-year non-life insurance risk in chain ladder reserving models. Insurance: Mathematics and Economics, 67:187–199.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. Biometrika, 65(3):457–483.

European Commission (2015). Commission Delegated Regulation (EU) 2015/35 of 10 October 2014. Official Journal of the European Union.

Gut, A. (2005). Probability: a graduate course. Springer-Verlag, New York.

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

Henderson, S. G. (2003). Estimation for nonhomogeneous Poisson processes from aggregated data. Operations Research Letters, 31(5):375–382.

Lehmann, E. L. and Casella, G. (1998). Theory of point estimation. Springer-Verlag, New York, second edition.

Lindholm, M., Lindskog, F., and Wahl, F. (2017). Valuation of non-life liabilities from claims triangles. Risks, 5(3):39.

Lindholm, M., Lindskog, F., and Wahl, F. (2018). Estimation of conditional mean squared error of prediction for claims reserving. Available at SSRN 3302576. Accepted for publication in Annals of Actuarial Science.

Mikosch, T. (2009). Non-life insurance mathematics: an introduction with the Poisson process. Springer Science & Business Media.

Miranda, M. D. M., Nielsen, B., Nielsen, J. P., and Verrall, R. (2011). Cash flow simulation for a model of outstanding liabilities based on claim amounts and claim numbers. ASTIN Bulletin: The Journal of the IAA, 41(1):107–129.

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

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. (1999). Prediction of outstanding liabilities II. Model variations and extensions. ASTIN Bulletin: The Journal of the IAA, 29(1):5–25.

Ohlsson, E. and Lauzeningks, J. (2009). The one-year non-life insurance risk. Insurance: Mathematics and Economics, 45(2):203–208.

Röhr, A. (2016). Chain ladder and error propagation. ASTIN Bulletin: The Journal of the IAA, 46(2):293–330.

Sigmundsdóttir, J. and Lindholm, M. (2017). One-year non-life Solvency II risk calculations for point process micro models: Methods and applications. Stockholm university research reports in Mathematical statistics, 2017:2.

Verrall, R., Nielsen, J. P., and Jessen, A. H. (2010). Prediction of RBNS and IBNR claims using claim amounts and claim counts. ASTIN Bulletin: The Journal of the IAA, 40(2):871–887.

Wahl, F., Lindholm, M., and Verrall, R. (2019). The collective reserving model. Insurance: Mathematics and Economics, 87:34–50.

A Proofs

Proof of Proposition 1. To calculate the expectation and variance of the outstanding payments, we must compute $\mathbb{E}[N_{il}|\mathcal{F}_\tau]$ and $\mathrm{Var}(N_{il}|\mathcal{F}_\tau)$ for all $l \geq \kappa_i$ and $\mathrm{Cov}(N_{il}, N_{ik}|\mathcal{F}_\tau)$ for all $k > l \geq \kappa_i$. The following will be useful for this purpose: For $\tau_i \leq s < t$ and all $k \in \mathbb{N}_+$ it holds that

$$\mathbb{E}[N_i^k(s,t)|\mathcal{F}_\tau] = \mathbb{E}[N_i^k(s,t)|V_i > \tau_i] = \mathbb{E}[N_i^k(s,t)|V_i > s]\mathbb{P}(V_i > s|V_i > \tau_i), \quad (35)$$

from which it follows that

$$\begin{aligned}\mathrm{Var}(N_i(s,t)|\mathcal{F}_\tau) &= \mathbb{E}[N_i^2(s,t)|V_i > s]\mathbb{P}(V_i > s|V_i > \tau_i) \\ &\quad - \mathbb{E}[N_i(s,t)|V_i > s]^2\mathbb{P}(V_i > s|V_i > \tau_i)^2. \end{aligned} \quad (36)$$

Moreover, for $\tau_i < s < t < s' < t'$, the covariance terms are given by

$$\begin{aligned}\mathbb{E}[N_i(s,t)N(s',t')|\mathcal{F}_\tau] &= \mathbb{E}[N_i(s,t)N(s',t')|V_i > \tau_i] \\ &= \mathbb{E}[N_i(s,t)|V_i > s']\mathbb{E}[N_i(s',t')|V_i > s']\mathbb{P}(V_i > s'|V_i > \tau_i), \end{aligned} \quad (37)$$

where we can conclude that, for $[s,t] \subset [t_{l-1}, t_l)$ for some $l \geq \kappa_i$,

$$\mathbb{E}[N_i(s,t)|V_i > s'] = (t-s)h_{p,l}, \quad (38)$$

since the conditioning implies no settlement in the interval $[s,t)$ and therefore $N_i(s,t)$ is equivalent to the Poisson process counting only the payments without settlement, not stopped by a settlement event, which has the given expectation. In (35)–(37) we have made use of the fact that the only information in $\mathcal{F}_\tau$ that is relevant for the event processes is whether settlement has occurred or not since the three event generating processes are mutually independent (non-homogeneous) Poisson processes that stop if a settlement event occurs. Therefore, conditioning on $\mathcal{F}_\tau$ is equivalent to conditioning on the event $V_i > \tau_i$.

The probabilities in (35) and (36) are, for $s < t$, given by

$$\begin{aligned}\mathbb{P}(V_i > t|V_i > s) &= \frac{\mathbb{P}(V_i > t, V_i > s)}{\mathbb{P}(V_i > s)} \\ &= \frac{\mathbb{P}(V_i > t)}{\mathbb{P}(V_i > s)}.\end{aligned}$$

Since events occur according to mutually independent Poisson processes and settlement occurs if any of the two settlement processes has an event, these probabilities are given by

$$\begin{aligned}\mathbb{P}(V_i > s) &= \mathbb{P}(\text{No event with settlement in } [0,s]) \\ &= \exp\left\{-\int_0^s (h_{se}(t) + h_{sep}(t))\, dt\right\}.\end{aligned} \quad (39)$$

Therefore, for $s < t$ it holds that

$$\mathbb{P}(V_i > t|V_i > s) = \exp\left\{-\int_s^t (h_{se}(t) + h_{sep}(t))\, dt\right\}. \quad (40)$$

What now remains to compute in (35) is $\mathbb{E}[N_i^k(s,t)|V_i > s]$ for $k = 1, 2$, and $s, t \in [t_{l-1}, t_l)$. To do this, we first note using (39), that the probability density function of V_i conditioned on $V_i > s$, is

$$f_{V_i|V_i>s}(t) = \frac{f_{V_i}(t)}{1 - F_{V_i}(s)} = \bar{h}_s(t)e^{-(\bar{H}_s(t) - \bar{H}_s(s))}. \quad (41)$$

Define $N_{i,e}(s,t)$ to be the number of events of type $e \in \{p, se, sep\}$ for the ith claim in the interval $[s,t)$ when settlement stops the processes. We are interested in the moments of the total number of payments, i.e. of

$$N_i(s,t) := N_{i,p}(s,t) + N_{i,sep}(s,t). \tag{42}$$

For this, define $M_{i,e}(s,t)$ to be the Poisson process generating events of type e, not stopped by settlement events. To acquire the moments of the number of payment (without settlement) events we note that, conditional on V_i, $M_{i,p}$ is still a Poisson process on the interval $[0, V_i)$ since $M_{i,p}$ is independent of $M_{i,e}$ for $e \in \{se, sep\}$. Therefore

$$\begin{aligned}
\mathbb{E}[N_{i,p}(s,t)|V_i > s] &= \mathbb{E}[M_{i,p}(s, t \wedge V_i)|V_i > s] \\
&= \mathbb{E}[H_p(t \wedge V_i) - H_p(s)|V_i > s] \\
&= \int_s^\infty (H_p(t \wedge u) - H_p(s))\, f_{V_i|V_i>s}(u)du \\
&= \int_s^t (H_p(u) - H_p(s))\, f_{V_i|V_i>s}(u)du \\
&\quad + (H_p(t) - H_p(s))\, \mathbb{P}(V_i > t|V_i > s),
\end{aligned} \tag{43}$$

and

$$\begin{aligned}
\mathbb{E}[N_{i,p}^2(s,t)|V_i > s] &= \mathbb{E}[M_{i,p}^2(s, t \wedge V_i)|V_i > s] \\
&= \mathbb{E}[N_i(s,t)|V_i > s] + \mathbb{E}[(H_p(t \wedge V_i) - H_p(s))^2\, |V_i > s],
\end{aligned} \tag{44}$$

where

$$\begin{aligned}
\mathbb{E}[(H_p(t \wedge V_i) - H_p(s))^2\, |V_i > s] &= (H_p(t) - H_p(s))^2\, \mathbb{P}(V_i > t|V_i > s) \\
&\quad + \int_s^t (H_p(u) - H_p(s))^2\, f_{V_i|V_i>s}(u)du.
\end{aligned} \tag{45}$$

For the moments of the settlement events, let V_i^e denote the first time the Poisson process generating events of type e has an event. For $e \in \{se, sep\}$ we note that $N_{i,e}(s,t)|V_i > s$ is a Bernoulli random variable with probability parameter

$$\begin{aligned}
\mathbb{P}(N_{i,e}(s,t) = 1|V_i > s) &= \mathbb{E}[1_{\{V_i^e = V_i < t\}}|V_i > s] \\
&= \mathbb{E}\left[\frac{h_e(V_i)}{\bar{h}_s(V_i)} 1_{\{V_i < t\}} \middle| V_i > s\right] \\
&= \int_s^t \frac{h_e(v)}{\bar{h}_s(v)} f_{V_i|V_i>s}(v)dv,
\end{aligned} \tag{46}$$

where we have made use of iterated expectations with

$$\mathbb{E}[1_{\{V_i^e = V_i < t\}}|V_i] = \frac{h_e(V_i)}{\bar{h}_s(V_i)} 1_{\{V_i < t\}}. \tag{47}$$

This last equality can be seen to be the case by the following standard result: Take $dt > 0$ and note that

$$\begin{aligned}
\mathbb{P}(N_{i,e}(t, t+dt) = 1|V_i \in (t, t+dt)) &= \frac{\mathbb{P}(N_{i,e}(t, t+dt) = 1)}{\mathbb{P}(V_i \in (t, t+dt))} \\
&= \frac{h_e(t)dt + o(dt)}{\bar{h}_s(t)dt + o(dt)},
\end{aligned}$$

for $e \in \{se, sep\}$. Since $N_{i,e}(s,t)|V_i > s$ is a Bernoulli random variable, the probability in (46) is the sought after first two moments of $N_{ie}(s,t)|V_i > s$.

Finally, we acquire the first moment of (42) by adding together (43) and (46), while the second moment is acquired by computing

$$\begin{aligned}\mathbb{E}[N_i^2(s,t)|V_i > s] = \mathbb{E}[N_{i,p}^2(s,t)|V_i > s] + \mathbb{E}[N_{i,sep}^2(s,t)|V_i > s] \\ + 2\mathbb{E}[N_{i,p}(s,t)N_{i,sep}(s,t)|V_i > s],\end{aligned} \tag{48}$$

where

$$\begin{aligned}\mathbb{E}[N_{i,p}(s,t)N_{i,sep}(s,t)|V_i > s] &= \mathbb{E}[M_{i,p}(s,V_i)1_{\{V_i^e = V_i < t\}}|V_i > s] \\ &= \mathbb{E}\left[(H_p(V_i) - H_p(s))\frac{h_e(V_i)}{\bar{h}_s(V_i)}1_{\{V_i<t\}}\middle| V_i > s\right] \\ &= \int_s^t (H_p(v) - H_p(s))\frac{h_e(v)}{\bar{h}_s(v)}f_{V_i|V_i>s}(v)dv,\end{aligned} \tag{49}$$

where we have made use of the independence between the processes.

Now we move on to computing the moments given the discretisation scheme introduced in Section 3. Taking $s, t \in [t_{l-1}, t_l)$ for some $l \in \{1, \dots, r\}$, we get from (43) that

$$\begin{aligned}\mathbb{E}[N_{i,p}(s,t)|V_i > s] &= \int_s^t h_{p,l}\,(u-s)\,\bar{h}_{sl}e^{-\bar{h}_{sl}(u-s)}du + (t-s)\,h_{p,l}e^{-\bar{h}_{sl}(t-s)} \\ &= \frac{h_{p,l}}{\bar{h}_{sl}}\left(1 - e^{-\bar{h}_{sl}(t-s)}\right),\end{aligned} \tag{50}$$

and, together with (45), that

$$\begin{aligned}\mathbb{E}[N_{i,p}^2(s,t)|V_i > s] =& \frac{h_{p,l}}{\bar{h}_{sl}}\left(1 - e^{-\bar{h}_{sl}(t-s)}\right)\left(1 + 2\frac{h_{p,l}}{\bar{h}_{sl}}\right) \\ &- 2\frac{h_{p,l}^2}{\bar{h}_{sl}}(t-s)e^{-\bar{h}_{sl}(t-s)}.\end{aligned} \tag{51}$$

Moreover, from (46) it holds for $e \in \{se, sep\}$ that

$$\mathbb{E}[N_{i,e}^2(s,t)|V_i > s] = \mathbb{E}[N_{i,e}(s,t)|V_i > s] = \frac{h_{e,l}}{\bar{h}_{sl}}\left(1 - e^{-\bar{h}_{sl}(t-s)}\right). \tag{52}$$

Finally, from (49) we get

$$\mathbb{E}[N_{i,p}(s,t)N_{i,sep}(s,t)|V_i > s] = \frac{h_{p,l}h_{sep,l}}{\bar{h}_{sl}}\left(1 - (1 + \bar{h}_{sl}(t-s))e^{-\bar{h}_{sl}(t-s)}\right). \tag{53}$$

Therefore, from (50) and (52) it holds that

$$\mathbb{E}[N_i(s,t)|V_i > s] = \frac{\bar{h}_{pl}}{\bar{h}_{sl}}\left(1 - e^{-\bar{h}_{sl}(t-s)}\right), \tag{54}$$

and from (51)–(53) together with (48) that

$$\begin{aligned}\mathbb{E}[N_i^2(s,t)|V_i > s] =& \frac{\bar{h}_{pl}}{\bar{h}_{sl}}\left(1 + 2\frac{\bar{h}_{pl}}{\bar{h}_{sl}}\right)\left(1 - e^{-\bar{h}_{sl}(t-s)}\right) \\ &- 2\frac{\bar{h}_{pl}h_{p,l}}{\bar{h}_{sl}}(t-s)e^{-\bar{h}_{sl}(t-s)}.\end{aligned} \tag{55}$$

The proof is finished by taking the probability in (40) together with the moments in (54) and (55) and inserting these into (35) and (36). Taking these together with (37) and inserting them into (11) and (12) yields the result. □

Proof of Proposition 4. Let $\gamma_i := \tau - T_i^o - U_i^o$ so that $\tau_i = \gamma_i \wedge V_i$. All expectations in this proof are conditional on $\sigma((T_i^o, U_i^o)_{i=1}^{N^o})$, but we suppress this to keep expressions less cumbersome. Therefore, we treat γ_i as a constant.

Beginning with the estimators of the hazard rates of pure payment events, we have

$$\begin{aligned}
\mathbb{E}[\hat{h}_{p,k}] &= \mathbb{E}\left[\frac{\sum_i N_{i,p}(s \wedge \tau_i, t \wedge \tau_i)}{\sum_i(\tau_i \wedge t - \tau_i \wedge s)}\right] \\
&= \mathbb{E}\left[\frac{\sum_i M_{i,p}(s \wedge \tau_i, t \wedge \tau_i)}{\sum_i(\tau_i \wedge t - \tau_i \wedge s)}\right] \\
&= \mathbb{E}\left[\frac{\sum_i H_p(t \wedge \tau_i) - H_p(s \wedge \tau_i)}{\sum_i(t \wedge \tau_i - s \wedge \tau_i)}\right] \\
&= h_{p,l}\mathbb{E}\left[\frac{\sum_i(t \wedge \tau_i - s \wedge \tau_i)}{\sum_i(t \wedge \tau_i - s \wedge \tau_i)}\right] \\
&= h_{p,l},
\end{aligned}$$

showing that the estimators are unbiased.

For the corresponding quantities for settlement events, let $\mathcal{I} = \{i = 1, \ldots, N_{e,k}^o : \gamma_i > s\}$ and $\mathcal{V} := \sigma(\{V_i\}_{i \in \mathcal{I}})$. Moreover, for notational brevity, let $s := t_{k-1}$ and $t := t_k$. Finally, for $i \in \mathcal{I}$, let A_i be the event that $V_i \in (s, t \wedge \gamma_i)$ (i.e. A_i^c is the event $V_i > t \wedge \gamma_i$). We begin by the following computation for $e \in \{se, sep\}$ where we make use of (47):

$$\begin{aligned}
\mathbb{E}[\hat{h}_{e,k}] &= \mathbb{E}\left[\mathbb{E}\left[\frac{\sum_{i \in \mathcal{I}} M_{i,e}(s \wedge V_i, t \wedge V_i \wedge \gamma_i)}{\sum_{i \in \mathcal{I}}(V_i \wedge \gamma_i \wedge t - V_i \wedge s)}\middle| \mathcal{V}\right]\right] \\
&= \frac{h_{e,k}}{\bar{h}_{sk}}\mathbb{E}\left[\frac{\sum_{i \in \mathcal{I}} 1_{A_i}}{\sum_{i \in \mathcal{I}}(V_i \wedge \gamma_i \wedge t - V_i \wedge s)}\right].
\end{aligned}$$

Now define $\mathcal{A} := \sigma((1_{A_1}, \ldots, 1_{A_{N_{e,k}^o}}))$, then

$$\begin{aligned}
\mathbb{E}[\hat{h}_{e,k}] &= \frac{h_{e,k}}{\bar{h}_{sk}} \sum_{i \in \mathcal{I}} \mathbb{E}\left[\frac{1_{A_i}}{\sum_{j \in \mathcal{I}}(t \wedge \gamma_j - s)1_{A_j^c} + \sum_{l \in \mathcal{I}}(V_l - s)1_{A_l}}\right] \\
&= \frac{h_{e,k}}{\bar{h}_{sk}} \sum_{i \in \mathcal{I}} \mathbb{E}\left[\mathbb{E}\left[\frac{1}{\sum_{j \in \mathcal{I}}(t \wedge \gamma_j - s)1_{A_j^c} + \sum_{l \in \mathcal{I}}(V_l - s)1_{A_l}}\middle| \mathcal{A}\right]\middle| A_i\right] \mathbb{P}(A_i) \\
&\geq \frac{h_{e,k}}{\bar{h}_{sk}} \sum_{i \in \mathcal{I}} \mathbb{E}\left[\frac{1}{\sum_{j \in \mathcal{I}}(t \wedge \gamma_j - s)1_{A_j^c} + \sum_{l \in \mathcal{I}} 1_{A_l}\mathbb{E}[V_l - s | A_l]}\middle| A_i\right] \mathbb{P}(A_i) \\
&\geq \frac{h_{e,k}}{\bar{h}_{sk}} \sum_{i \in \mathcal{I}} \frac{\mathbb{P}(A_i)}{\mathbb{E}\left[\sum_{j \in \mathcal{I}\setminus\{i\}}(t \wedge \gamma_j - s)1_{A_j^c} + \sum_{l \in \mathcal{I}} 1_{A_l}\mathbb{E}[V_l - s | A_l]\middle| A_i\right]},
\end{aligned}$$

where, using (39) together with

$$f_{V_i | A_i}(v) = \frac{f_{V_i}(v)}{F_{V_i}(t) - F_{V_i}(s)},$$

we have

$$\begin{aligned}
\mathbb{E}[V_i - s | A_i] &= \int_s^{t \wedge \gamma_i} (v - s) \frac{\bar{h}_{sk} e^{-\bar{h}_{sk}(v-s)}}{1 - e^{-\bar{h}_{sk}(t \wedge \gamma_i - s)}} dv \\
&= \frac{1}{\bar{h}_{sk}} - \frac{t \wedge \gamma_i - s}{1 - e^{-\bar{h}_{sk}(t \wedge \gamma_i - s)}} e^{-\bar{h}_{sk}(t \wedge \gamma_i - s)}.
\end{aligned}$$

Next, note that

$$1_{A_i}|\gamma_i > s \sim \text{Bernoulli}(p_i)$$

with probability parameter

$$p_i := \mathbb{P}(A_i) = \mathbb{P}(V_i \in (s, t \wedge \gamma_i)) = 1 - e^{-\bar{h}_{sk}(t\wedge\gamma_i - s)}.$$

Using this together with the fact that

$$-\log(1-p_i) = \bar{h}_{sk}(t \wedge \gamma_i - s),$$

we have that

$$\begin{aligned}
\frac{\mathbb{E}[\hat{h}_{e,k}]}{h_{e,k}} &\geq \sum_{i\in\mathcal{I}} \frac{p_i}{\mathbb{E}\left[\sum_{l\in\mathcal{I}} 1_{A_l}\left(1 + \frac{1-p_l}{p_l}\log(1-p_l)\right) - \sum_{j\in\mathcal{I}\setminus\{i\}} 1_{A_j^c}\log(1-p_j)\Big|A_i\right]} \\
&= \sum_{i\in\mathcal{I}} \frac{p_i}{\sum_{l\in\mathcal{I}\setminus\{i\}} p_l\left(1+\frac{\log(1-p_l)}{p_l}\right) + 1 + \frac{\log(1-p_i)}{p_i} - \sum_{j\in\mathcal{I}}\log(1-p_j)} \\
&= \sum_{i\in\mathcal{I}} \frac{p_i}{\sum_{l\in\mathcal{I}\setminus\{i\}} p_l + 1 + \frac{\log(1-p_i)}{p_i} - \log(1-p_i)} \\
&= \sum_{i\in\mathcal{I}} \frac{p_i}{1 + \frac{1-p_i}{p_i}\log(1-p_i) + \sum_{l\in\mathcal{I}\setminus\{i\}} p_l} \\
&> \sum_{i\in\mathcal{I}} \frac{p_i}{1-(1-p_i) + \sum_{l\in\mathcal{I}\setminus\{i\}} p_l} \\
&= 1,
\end{aligned}$$

where the last (strict) inequality holds since

$$\frac{1-x}{x}\log(1-x) < -(1-x),$$

for all $x \in (0,1)$. □

Proof of Proposition 5. To prove the results of the proposition, we make use of Theorem 5.1 on page 463 in Lehmann and Casella (1998). To use this theorem, we must verify that some regularity conditions hold. However, before that, we must verify that it is OK that the log-likelihood in (25) has a random number of terms, N^o. For the observed claims we know, in the same way as for the IBNR claims in (17), that the number of them is distributed according to

$$N^o \sim \text{Po}\left(\int_0^\tau w_t\lambda_t P_{U|t}(\tau - t)dt\right),$$

see, e.g. Norberg (1993) and Antonio and Plat (2014). For notational brevity, let

$$\tilde{\Lambda}_\tau := \int_0^\tau w_t\lambda_t P_{U|t}(\tau-t)dt.$$

We know by the same well-known result giving us (21), that

$$\mathbb{P}(T^o \leq t|N^o = n) = \frac{\tilde{\Lambda}_t}{\tilde{\Lambda}_\tau}$$

Therefore, it is the case that, conditional on N^o, the terms of the log-likelihood are all i.i.d., and, moreover, the distribution of the terms does not depend on the value of N^o.

Furthermore, by assuming that $w_t = w\tilde{w}_t$, it holds that

$$\frac{N^o}{\tilde{w}_t} \xrightarrow{p} \int_0^\tau \tilde{w}_t \lambda_t \mathbb{P}_{U|t}(\tau - t)dt,$$

as $w \to \infty$. We may, therefore, apply (a special case of) Anscombe's Theorem, see e.g. Theorem 3.2 on page 346 in Gut (2005), to the proof of Theorem 5.1 in Lehmann and Casella (1998) to see that part (b) and (c) of the theorem applies to our situation of a random number of terms in the log-likelihood. Additionally, since $N^o \overset{a.s.}{\to} \infty$ as $w \to \infty$, it holds by, e.g. part (iii) of Theorem 8.2 on page 302 in Gut (2005), that part (a) of the theorem also applies in our situation.

We now move on to prove the proposition by verifying the assumptions of the theorem. The assumptions (A0)-(A2) are trivially satisfied. Therefore, we move on to check assumptions (A)-(D).

The log-density corresponding to the development process of the ith claim is

$$\log f(\boldsymbol{h}) = \sum_{j=1}^{m} \sum_{e} \left(N_{i,e}(t_{j-1}, t_j) \log h_{e,j} - h_{e,j} \int_{t_{j-1}}^{t_j} 1_{\{u<\tau_i\}} du \right),$$

where we have suppressed the dependence on the data for notational convenience. The third partial derivatives of this log-density are

$$\frac{\partial^3}{\partial h_{e,j} \partial h_{e',k} \partial h_{e'',l}} \log f(\boldsymbol{h}) = 2 \frac{N_{i,e}(t_{j-1}, t_j)}{h_{e,j}^3} 1_{\{e=e'=e''\}} 1_{\{j=k=l\}},$$

and thus $f(\boldsymbol{h})$ satisfies Assumption (A). Moreover, Assumption (D) is fulfilled since

$$\mathbb{E}\left[\left|\frac{\partial^3}{\partial h_{e,j} \partial h_{e',k} \partial h_{e'',l}} \log f(\boldsymbol{h})\right|\right] = 2 \frac{\mathbb{E}[N_{i,e}(t_{j-1}, t_j)]}{h_{e,j}^3} 1_{\{e=e'=e''\}} 1_{\{j=k=l\}} < \infty.$$

To show that Assumption (B) holds, we first note that

$$\frac{\partial}{\partial h_{e,j}} \log f(\boldsymbol{h}) = \frac{N_{i,e}(t_{j-1}, t_j)}{h_{e,j}} - \int_{t_{j-1}}^{t_j} 1_{\{u<\tau_i\}} du. \tag{56}$$

Since

$$\begin{aligned}
\mathbb{E}\left[\int_{t_{j-1}}^{t_j} 1_{\{u<\tau_i\}} \middle| \gamma_i\right] &= \mathbb{E}\left[V_i \wedge \gamma_i \wedge t_j - t_{j-1} | V_i > t_{j-1}, \gamma_i\right] \mathbb{P}(V_i > t_{j-1}) 1_{\{\gamma_i > t_{j-1}\}} \\
&= \frac{1_{\{\gamma_i > t_{j-1}\}}}{\bar{h}_{sj}} \left(1 - e^{-\bar{h}_{sj}(t_j \wedge \gamma_i - t_{j-1})}\right) \mathbb{P}(V_i > t_{j-1}) \\
&= \mathbb{E}\left[\frac{N_{i,e}(t_{j-1}, t_j)}{h_{e,j}} \middle| \gamma_i\right],
\end{aligned}$$

where the second equality follows from computing the expectation using the density in (41), and the last equality follows from (50) and (52), it holds that

$$\mathbb{E}\left[\frac{\partial}{\partial h_{e,j}} \log f(\boldsymbol{h})\right] = 0. \tag{57}$$

What remains to show for Assumption (B) is that

$$\mathbb{E}\left[-\frac{\partial^2}{\partial h_{e,j}\partial h_{e',k}}\log f(\boldsymbol{h})\right]=\mathbb{E}\left[\frac{\partial}{\partial h_{e,j}}\log f(\boldsymbol{h})\frac{\partial}{\partial h_{e',k}}\log f(\boldsymbol{h})\right]. \tag{58}$$

For this, we take the partial derivatives of (56) and note that

$$\mathbb{E}\left[-\frac{\partial^2}{\partial h_{e,j}\partial h_{e',k}}\log f(\boldsymbol{h})\middle|\gamma_i\right]=\mathbb{E}\left[\frac{N_{i,e}(t_{j-1},t_j)}{h_{e,j}^2}\middle|\gamma_i\right]1_{\{e=e'\}}1_{\{j=k\}},$$

We must, therefore, show that the right-hand side of (58) is equal to this expression. By using (56), we see that it holds for $j<k$ that

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial}{\partial h_{e,j}}\log f(\boldsymbol{h})\frac{\partial}{\partial h_{e',k}}\log f(\boldsymbol{h})\right]\\
&=\mathbb{E}\left[\frac{\partial}{\partial h_{e,j}}\log f(\boldsymbol{h})\middle|V_i>t_{k-1}\right]\mathbb{E}\left[\frac{\partial}{\partial h_{e',k}}\log f(\boldsymbol{h})\middle|V_i>t_{k-1}\right]\mathbb{P}(V_i>t_{k-1})\\
&=0
\end{aligned}$$

since the second expectation is 0 by (57), showing that the expectation of the right-hand side of (58) is zero if $j\neq k$. To show that the expectation of the product is zero if $e\neq e'$, we take $e\in\{se,sep\}$ and then first compute the following

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial}{\partial h_{p,j}}\log f(\boldsymbol{h})\frac{\partial}{\partial h_{e,j}}\log f(\boldsymbol{h})\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\frac{\partial}{\partial h_{p,j}}\log f(\boldsymbol{h})\frac{\partial}{\partial h_{e,j}}\log f(\boldsymbol{h})\middle|\mathcal{V}\right]\right]\\
&=\mathbb{E}\left[\left(\frac{h_{p,j}\int_{t_{j-1}}^{t_j}1_{\{u<\tau_i\}}du}{h_{p,j}}-\int_{t_{j-1}}^{t_j}1_{\{u<\tau_i\}}du\right)\frac{\partial}{\partial h_{e,j}}\log f(\boldsymbol{h})\right]\\
&=0,
\end{aligned}$$

where the second equality comes from the fact that

$$\mathbb{E}[N_{i,p}(t_{j-1},t_j)|\mathcal{V}]=h_{p,j}\int_{t_{j-1}}^{t_j}1_{\{u<\tau_i\}}du, \tag{59}$$

which follows by the same reasoning as the motivation for (38). Secondly, similar to in the proof of Proposition 4, define A_i^e to be the event that $V_i^e\in(s,t\wedge\gamma_i)$, then

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial}{\partial h_{se,j}}\log f(\boldsymbol{h})\frac{\partial}{\partial h_{sep,j}}\log f(\boldsymbol{h})\right]\\
&=\mathbb{E}\left[\left(\frac{1_{A_j^{se}}}{h_{se,j}}-\int_{t_{j-1}}^{t_j}1_{\{u<\tau_i\}}du\right)\left(\frac{1_{A_j^{sep}}}{h_{sep,j}}-\int_{t_{j-1}}^{t_j}1_{\{u<\tau_i\}}du\right)\right]\\
&=\mathbb{E}\left[-\left(\frac{1_{A_j^{se}}}{h_{sep,j}}+\frac{1_{A_j^{sep}}}{h_{se,j}}\right)\int_{t_{j-1}}^{t_j}1_{\{u<\tau_i\}}du+\left(\int_{t_{j-1}}^{t_j}1_{\{u<\tau_i\}}du\right)^2\right]\\
&=\mathbb{E}\left[-2\frac{1_{\{V_i\in(t_{j-1},t_j)\}}}{\bar{h}_{sj}}\int_{t_{j-1}}^{t_j}1_{\{u<\tau_i\}}du+\left(\int_{t_{j-1}}^{t_j}1_{\{u<\tau_i\}}du\right)^2\right]\\
&=0,
\end{aligned} \tag{60}$$

where the third equality is acquired by (47) and the last equality with zero from computing the expectation using the density in (41). What remains are the squared terms where both indexes are the same in the product. These are:

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial}{\partial h_{p,j}} \log f(\boldsymbol{h})\right)^2\right] &= \mathbb{E}\left[\left(\frac{N_{i,p}(t_{j-1}, t_j)}{h_{p,j}} - \int_{t_{j-1}}^{t_j} 1_{\{u<\tau_i\}} du\right)^2\right] \\
&= \mathbb{E}\left[\frac{N_{i,p}^2(t_{j-1}, t_j)}{h_{p,j}^2} - \left(\int_{t_{j-1}}^{t_j} 1_{\{u<\tau_i\}} du\right)^2\right] \\
&= \mathbb{E}\left[\frac{N_{i,p}(t_{j-1}, t_j)}{h_{p,j}^2}\right],
\end{aligned}
$$

where the second equality comes from expanding the square and using (59) and the last equality follows from (44); and for $e \in \{se, sep\}$

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial}{\partial h_{e,j}} \log f(\boldsymbol{h})\right)^2\right] &= \mathbb{E}\left[\left(\frac{N_{i,e}^2(t_{j-1}, t_j)}{h_{e,j}^2} - \int_{t_{j-1}}^{t_j} 1_{\{u<\tau_i\}} du\right)^2\right] \\
&= \mathbb{E}\left[\frac{N_{i,e}(t_{j-1}, t_j)}{h_{e,j}^2}\right],
\end{aligned}
$$

where the last step follows by expanding the square and then using the last step in (60), and thus it is clear that (58) holds.

Finally, we must show Assumption (C). Doing this is straightforward since the expected Fisher information is a diagonal matrix with diagonal elements given by the expectation of the elements in (27), which are all non-negative, implying it is positive semi-definite.

Therefore, all assumptions of Theorem 5.1 on page 463 in Lehmann and Casella (1998) are fulfilled, and the proof is thus complete. □