A COLLECTIVE RESERVING MODEL WITH CLAIM OPENNESS

By

Mathias Lindholm and Henning Zakrisson

Abstract

The present paper introduces a simple aggregated reserving model based on claim count and payment dynamics, which allows for claim closings and re-openings. The modelling starts off from individual Poisson process claim dynamics in discrete time, keeping track of accident year, reporting year and payment delay. This modelling approach is closely related to the one underpinning the so-called double chain-ladder model, and it allows for producing separate reported but not settled and incurred but not reported reserves. Even though the introduction of claim closings and re-openings will produce new types of dependencies, it is possible to use flexible parametrisations in terms of, for example, generalised linear models (GLM) whose parameters can be estimated based on aggregated data using quasi-likelihood theory. Moreover, it is possible to obtain interpretable and explicit moment calculations, as well as having consistency of normalised reserves when the number of contracts tend to infinity. Further, by having access to simple analytic expressions for moments, it is computationally cheap to bootstrap the mean squared error of prediction for reserves. The performance of the model is illustrated using a flexible GLM parametrisation evaluated on non-trivial simulated claims data. This numerical illustration indicates a clear improvement compared with models not taking claim closings and re-openings into account. The results are also seen to be of comparable quality with machine learning models for aggregated data not taking claim openness into account.

Keywords

Aggregated reserving, open claim dynamics, RBNS and IBNR reserves.

1 INTRODUCTION

In the present paper, focus is on aggregated reserving making use of claim count and claim payment data, which allows for handling claim closings and re-openings. The modelling approach follows the same underlying idea as in Wahl *et al*. (2019), which introduces the so-called collective reserving model (CRM), where individual claims are modelled using discrete Poisson processes keeping track of accident year, reporting year, and payment delay in years, and which allows for multiple payments per claim. Due to this, the introduced model will be referred to as the CRMO – a CRM with openness dynamics. Further, this modelling approach is similar to the one used in, for example, Verrall *et al*. (2010) which underpins the double chain-ladder model introduced in Martinez-Miranda *et al*. (2012). Further, by using this slightly more granular representation than the one used with the standard chain-ladder technique makes it possible to obtain separate reported, but not settled (RBNS) and incurred, but not reported (IBNR) reserves. Moreover, reserve moments can be calculated explicitly and are possible to give constructive interpretations. This will also be the case when allowing for claim closings and re-openings, although this will introduce additional dependencies. That is, starting from individual claims, it is natural that you need to keep track of claims staying open as well as those that have been closed, but could become re-opened. Examples of individual claim models with claim closings are, for example, Antonio and Plat (2014), Crevecoeur and Antonio (2019), Bettonville *et al*. (2020), and Delong *et al*. (2020), where the latter two also allow for claim re-openings. Another type of computer intensive approach is to use self-exciting processes as in Maciak *et al*. (2021). Their approach allows for re-openings, and self-exciting processes are able to capture complex dynamics and dependencies, but with the potential drawback of being more complex than standard marked point process reserving models.

Further, a discrete time setup will allow us to estimate parameters in, for example, flexible generalised linear models (GLM) using aggregated data and quasi-likelihood theory, but now having to keep track of the number of claims staying open together with the number of claims becoming re-opened. This is natural, but perhaps not ideal, since the main driver behind future payments is, reasonably, the number of open claims, regardless of whether they have been re-opened or not. With this in mind, a more restrictive version of the model is considered that allows for estimating parameters based on only knowing the number of open claims.

Moreover, the current model is an extension of the CRM from Wahl *et al*. (2019) and by re-using arguments for the CRM, it is possible to obtain consistency for normalised reserves when letting the number of contracts tend to infinity.

The modelling approach used in the current paper is a constructive one, starting from detailed claim dynamics. Another approach is to instead consider less explicit modelling, but instead allowing for flexible parametrisations

using machine learning techniques. Successful examples of aggregated reserving models using machine learning techniques can be found in, for example, Gabrielli *et al.* (2020) where neural networks (NN) are combined with the over-dispersed Poisson (ODP) chain-ladder model and Lindholm *et al.* (2020), which considers ODP chain-ladder type models using gradient boosting machines (GBM), as well as GBM and NN versions of CRM type models. By construction, Gabrielli *et al.* (2020), Lindholm *et al.* (2020) use aggregated data, but individual claim reserving models can be found in, for example, Delong *et al.* (2020), which uses a similar basic process approach as the CRM using NN:s, or Lopez *et al.* (2019) using trees. The present paper introduces a model that also could be parametrised using machine learning techniques. As will be seen in Section 6, however, already using a rather simple GLM formulation of the CRMO will produce predictions whose performance is comparable with aggregated machine learning reserving models based on simulated data from Gabrielli and Wüthrich (2018). This illustrates the potential benefits of modelling claim openness dynamics explicitly.

The remainder of the paper is organised as follows: Section 2 introduces data and notation, followed by Section 3, which introduces the CRMO. Section 4 discusses moment calculations, which are needed for the quasi-likelihood approach treated in Section 5, consistency of normalised reserve estimates, together with a more detailed comparison with the CRM from Wahl *et al.* (2019). Section 5 also briefly discusses how the bootstrap procedure for calculating MSEP for the CRM can be adapted to the CRMO, before moving to a number of numerical examples in Section 6 based on data generated by the individual claims history simulation machine presented in Gabrielli and Wüthrich (2018). The paper ends with a number of concluding remarks given in Section 7.

Since many of the calculations closely follow those for the CRM in Wahl *et al.* (2019), those calculations not considered essential for the exposition are omitted.

2 DATA AND NOTATION

Following the notation in Wahl *et al.* (2019), let N_{ij} denote the number of claims that occurred in accident year i, $i = 1, ..., m$ and were reported j, $j = 1, \ldots, d$ years after occurring, and let X_{ijk} denote the sum of all incremental payments from claims that incurred in year i, were reported j years after that and paid k, $k = 1, ..., \kappa$ years after being reported, where κ is the maximum payment delay. While it is possible to use different maximum delays for reporting claims and payments, respectively, only $d = \kappa = m - 1$ will be considered in this paper. For more on this, see, for example, Verrall *et al.* (2010), Wahl *et al.* (2019).

Further, let N_{ijk}^{open} denote the number of claims that incurred in accident year i, were reported j years thereafter that are *open* at the end of the k:th year

thereafter, that is open at the end of year $i + j + k$. Thus, $N_{ijk}^{\text{open}} \leq N_{ij}$ for all $k = 0, ..., m - 1$. Let $N_{ijk}^{\text{stay-open}}$ and $N_{ijk}^{\text{re-open}}$ denote the number of open claims that were and were not open in the previous year, respectively. Note that

$$N_{ijk}^{\text{open}} = N_{ijk}^{\text{re-open}} + N_{ijk}^{\text{stay-open}}.$$

Consider a setting where one has access to data from the perspective of the first day of the $m + 1$:th year. That means that all data N_{ij} such that $i + j \leq m$ and all data $X_{ijk}, N_{ijk}^{\text{open}}, N_{ijk}^{\text{re-open}}, N_{ijk}^{\text{stay-open}}$ such that $i + j + k \leq m$ is available, which can be formalised using the index set

$$\mathcal{A}_0 = \{(i, j, k): i + j + k \leq m\},$$

as the data sets

$$\mathcal{X}_0 = \left\{ X_{ijk}: (i, j, k) \in \mathcal{A}_0 \right\},$$
$$\mathcal{N}_0^+ = \left\{ N_{ij}, N_{ijk}^{\text{re-open}}, N_{ijk}^{\text{stay-open}}: (i, j, k) \in \mathcal{A}_0 \right\}.$$

In certain cases, we might not have access to how many of the open claims stem from re-openings or not, and in those cases $\mathcal{N}_0^+$ will have to be replaced by the less granular set

$$\mathcal{N}_0 = \left\{ N_{ij}, N_{ijk}^{\text{open}}: (i, j, k) \in \mathcal{A}_0 \right\}.$$

3 EXPLICIT MODELLING OF CLOSINGS AND RE-OPENINGS OF CLAIMS – THE CRMO MODEL

The aim of the current paper is to introduce an extension of the CRM that was introduced in Wahl *et al.* (2019), which allows for closing and re-opening of claims, including multiple re-openings and closings per claim. This new model will be abbreviated CRMO, where "O" stands for "openness dynamics".

Following Wahl *et al.* (2019), we let N_{ij} follow an over-dispersed Poisson distribution with intensity parameter ν_{ij} and over-dispersion parameter ϕ, that is

$$\mathbb{E}\left[N_{ij}\right] = \nu_{ij},$$
$$\text{Var}\left(N_{ij}\right) = \phi \mathbb{E}\left[N_{ij}\right]. \tag{3.1}$$

Further, note that N_{ijk}^{open} can be expressed as a sum of boolean variables,

$$N_{ijk}^{\text{open}} = \sum_{l=1}^{N_{ij}} N_{ijkl}^{\text{open}},$$

where $N_{ijkl}^{\text{open}} = 1$ if claim l of claims incurred in year i and reported j years after that is open at the end of the k:th year thereafter. Conditioned on $N_{i,j,k-1,l}^{\text{open}}$, this

can be viewed as a Bernoulli variable, with staying-open probability p_{ijk} and re-opening probability q_{ijk} such that for $k = 1, ..., m$,

$$N_{ijkl}^{\text{open}} \mid N_{i,j,k-1,l}^{\text{open}} = \begin{cases} 1, & \text{with probability } p_{ijk}, & \text{if } N_{i,j,k-1,l}^{\text{open}} = 1 \\ 0, & \text{with probability } 1 - p_{ijk}, & \text{if } N_{i,j,k-1,l}^{\text{open}} = 1 \\ 1, & \text{with probability } q_{ijk}, & \text{if } N_{i,j,k-1,l}^{\text{open}} = 0 \\ 0, & \text{with probability } 1 - q_{ijk}, & \text{if } N_{i,j,k-1,l}^{\text{open}} = 0 \end{cases},$$

where $p_{ijk}, q_{ijk} \in [0, 1]$. This means that N_{ijk}^{open} will be the sum of two sums of i.i.d. Bernoulli variables

$$N_{ijk}^{\text{open}} = N_{ijk}^{\text{stay-open}} + N_{ijk}^{\text{re-open}},$$

where

$$N_{ijk}^{\text{stay-open}} = \sum_{l : N_{i,j,k-1,l}^{\text{open}} = 1} N_{i,j,k,l}^{\text{open}}$$

and

$$N_{ijk}^{\text{re-open}} = \sum_{l : N_{i,j,k-1,l}^{\text{open}} = 0} N_{i,j,k,l}^{\text{open}}.$$

For $k = 0$, there can be no re-openings, and the distribution of the openness-booleans can instead be expressed as

$$N_{i,j,0,l}^{\text{open}} = \begin{cases} 1, & \text{with probability } p_{ij0} \\ 0, & \text{with probability } 1 - p_{ij0} \end{cases},$$

where $p_{ij0} \in [0, 1]$ and

$$N_{ij0}^{\text{stay-open}} = N_{ij0}^{\text{open}} = \sum_{l=1}^{N_{ij}} N_{i,j,0,l}^{\text{open}}.$$

Thus, this construction yields the following conditional distributions

$$N_{ijk}^{\text{stay-open}} \mid N_{i,j,k-1}^{\text{open}} \sim \text{Bin}\left(N_{i,j,k-1}^{\text{open}}, p_{ijk}\right),$$

$$N_{ijk}^{\text{re-open}} \mid N_{ij}, N_{i,j,k-1}^{\text{open}} \sim \text{Bin}\left(N_{ij} - N_{i,j,k-1}^{\text{open}}, q_{ijk}\right).$$

and

$$N_{ij0}^{\text{open}} \mid N_{ij} \sim \text{Bin}\left(N_{ij}, p_{ij0}\right).$$

Continuing, every open claim is assumed to produce a number of payments following a conditional Poisson distribution. That is, let N_{ijk}^{paid} denote the number

of claims that incurred in accident year i, were reported j years thereafter, and that are paid during the k:th year thereafter, whose conditional distribution is given by

$$N_{ijk}^{\text{paid}} \mid N_{i,j,k-1}^{\text{open}} \sim \text{Po}\left(\lambda_{ijk} N_{i,j,k-1}^{\text{open}}\right) \tag{3.2}$$

for $k \geq 1$ where $\lambda_{ijk} \geq 0$. Hence, the λ_{ijk}:s denote the average number of payments during year k stemming from a claim that incurred in accident year i, was reported j years after that and was open at the end of year $i + j + k - 1$. For $k = 0$, the number of payments depends on the number of reported claims as

$$N_{ij0}^{\text{paid}} \mid N_{ij} \sim \text{Po}\left(\lambda_{ij0} N_{ij}\right).$$

This means that the interpretation of λ_{ij0} differs slightly from that of other λ_{ijk}, as it is the expected number of payments made for a claim during the remaining calendar year after being reported.

Concerning paid amounts, let Y_{ijkl} be the size of the l:th payment of the N_{ijk}^{paid} payments and assume that all individual payments are independent over l with moments

$$\mathbb{E}\left[Y_{ijkl}\right] = \mu_{ijk} \geq 0,$$

$$\text{Var}\left(Y_{ijkl}\right) = \sigma_{ijk}^2 \geq 0.$$

Thus, finally, incremental payments X_{ijk} can be written as

$$X_{ijk} = \sum_{l=1}^{N_{ijk}^{\text{paid}}} Y_{ijkl}.$$

Remark 1

(a) *In order to simplify the exposition, we introduce the column vector*

$$\boldsymbol{\theta} = (\boldsymbol{v}, \phi, \boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\sigma})',$$

where $\boldsymbol{v} = \{v_{ij} : 1 \leq i \leq m, 0 \leq j < m\}$, $\boldsymbol{p} = \{p_{ijk} : (i,j,k) \in \mathcal{I}\}$, $\boldsymbol{q} = \{q_{ijk} : (i,j,k) \in \mathcal{I}\}$, $\boldsymbol{\lambda} = \{\lambda_{ijk} : (i,j,k) \in \mathcal{I}\}$, $\boldsymbol{\mu} = \{\mu_{ijk} : (i,j,k) \in \mathcal{I}\}$, $\boldsymbol{\sigma} = \{\sigma_{ijk} : (i,j,k) \in \mathcal{I}\}$, *where all* $v_{ij}, \lambda_{ijk}, \mu_{ijk}, \sigma_{ijk} \geq 0$, $p_{ijk}, q_{ijk} \in [0,1]$, *and where*

$$\mathcal{I} = \{(i,j,k) : 1 \leq i \leq m, 0 \leq j < m, 0 \leq k < m\}.$$

For practical purposes, we will, however, assume that the parametrisation of $\boldsymbol{\theta}$ *is estimable based on the information contained in* $\mathcal{X}_0, \mathcal{N}_0$ *(or* $\mathcal{N}_0^+$ *), but usually without specifying this in detail. That is, for example,* $v_{ij} := \exp\{\alpha_i + \beta_j\}$ *and similarly, which in practice will make the number of unique parameters in* $\boldsymbol{\theta}$ *considerably less than* $O(m^3)$*. A concrete example of this is given in Section 6.*

(b) *The simple sequential structure defined by (3.2) can, of course, be replaced with a longer historical dependence in terms of delay. For ease of exposition and to keep calculations manageable, we will only focus on definition (3.2). Further, note that the construction above, notwithstanding the particular choice of dependence structure, is in line with how individual claim dynamics would look in a micro model. These dynamics will naturally lead to a distinction between claims staying open and becoming re-opened, even though it is the total number of open claims that is the driver behind future payments. One can, however, note that if $p_{ijk} = q_{ijk} = \pi_{ijk}$ it follows that*

$$N_{ijk}^{open} \mid N_{ij}, N_{i,j,k-1}^{open} \sim Bin(N_{ij}, \pi_{ijk}), \qquad (3.3)$$

which means that π_{ijk} corresponds to the probability that a claim is open in (i, j, k), regardless of whether it has been closed prior to k or not. Thus, it is possible to have that $N_{i,j,k+s}^{open} \geq N_{ijk}^{open}$ for $s \geq 1$, that is, net re-openings are allowed.

(c) *The motivation of including dynamics for closing and re-opening claims is a generalisation of the Collective Reserving Model (CRM) introduced in Wahl et al. (2019), and it is seen that the CRM is retrieved by setting all p_{ijk}:s equal to one. A more detailed comparison with the CRM is found in Section 5.2. Further, there are other models in the literature that only capture closing of claims, without treating re-openings, see, for example, Bettonville et al. (2020) and Crevecoeur and Antonio (2019).*

(d) *As with the models discussed in Verrall et al. (2010), Martinez-Miranda et al. (2011), and Wahl et al. (2019), the current model is defined based on individual claim dynamics. Still, due to the current Poisson assumption, under suitable conditionings, the aggregated model can be described in terms of an over-dispersed Poisson model. From this perspective, the sufficient statistics for the aggregated model tell us that no information is lost while aggregating. For more on aggregation of "micro" level information, see, for example, Huang et al. (2015, 2016), Charpentier and Pigeon (2016), Lindholm and Verrall (2020).*

(e) *Note that the above model can be expressed in terms of the hierarchical model framework presented in Crevecoeur and Antonio (2019), using number of open claims as one of the layers. For the purpose of this paper, this observation does not lead to additional insights and will not be discussed further.*

4 MOMENTS

In order to be able to obtain a reserve predictor, we need to first obtain an expression for the expected value of the outstanding payments. Further, since the model structure is defined in terms of (i, j, k)-tuples, it is possible to obtain separate IBNR and RBNS reserves. Moreover, the simple model structure will

allow us to obtain recursive expressions for (co)variances, which can be used in (semi-)analytical approximations of the reserve MSEP. Furthermore, due to the sequential structure defining the dynamics of the number of claim payments as a function of time, given by (3.2), it is clear that it is not meaningful to condition on more granular information than that contained in $\mathcal{N}_0$. That is, the least granular information that does not loose information is given by (i, j, k)-level knowledge of the N_{ijk}^{open}:s, and when calculating first and second moments, the final expressions will not require (explicit) knowledge of $N_{ijk}^{\text{stay-open}}$ or $N_{ijk}^{\text{re-open}}$.

The derivation of moments will make use of the standard sum representation of outstanding payments, with RBNS and IBNR reserves defined by

$$R = R^{\mathcal{R}} + R^{\mathcal{I}},$$

where

$$R^{\mathcal{R}} = \sum_{i=1}^{m} R_i^{\mathcal{R}}, \quad R^{\mathcal{I}} = \sum_{i=1}^{m} R_i^{\mathcal{I}} \tag{4.1}$$

and

$$R_i^{\mathcal{R}} = \sum_{j=1}^{m-i} R_{ij}, \quad R_i^{\mathcal{I}} = \sum_{j=I-i+1}^{m-1} R_{ij}, \tag{4.2}$$

where R_{ij} is the sum of all outstanding payments from claims stemming from accident year i with a reporting delay of j, that is,

$$R_{ij} = \sum_{k=m-i-j+1}^{m-1} X_{ijk}. \tag{4.3}$$

Here one can note that the R_{ij}:s are independent. This is, however, not in general the case for the X_{ijk}:s. Further, let $\boldsymbol{\theta}$ be defined as in Remark 1, and let the theoretical reserve predictor be denoted by

$$h\left(\boldsymbol{\theta}; \mathcal{N}_0\right) := \mathbb{E}\left[R \mid \mathcal{N}_0\right](\boldsymbol{\theta}),$$

while the corresponding RBNS and IBNR estimators are denoted by

$$h^{\mathcal{R}}\left(\boldsymbol{\theta}; \mathcal{N}_0\right) := \mathbb{E}\left[R^{\mathcal{R}} \mid \mathcal{N}_0\right](\boldsymbol{\theta}), \quad h^{\mathcal{I}}\left(\boldsymbol{\theta}; \mathcal{N}_0\right) := \mathbb{E}\left[R^{\mathcal{I}} \mid \mathcal{N}_0\right] = \mathbb{E}\left[R^{\mathcal{I}}\right](\boldsymbol{\theta}).$$

The standard computable (plug-in) reserve predictors are, hence, given by

$$\widehat{R} := h\left(\widehat{\boldsymbol{\theta}}; \mathcal{N}_0\right)$$

and analogously for the RBNS and IBNR reserves. Similar estimators can be constructed for individual R_i and R_{ij}. The derivation of moments for outstanding payments closely follows those given in Wahl *et al.* (2019) and are based on suitable repeated conditionings that are either omitted or merely outlined.

4.1 Expected values

The expected value of R_{ij} is

$$\mathbb{E}\left[R_{ij} \mid \mathcal{N}_0\right] = \sum_{k=m-i-j+1}^{m-1} \mathbb{E}\left[X_{ijk} \mid \mathcal{N}_0\right],$$

where the expected value of the payments is obtained by repeated conditioning, and where the split into RBNS and IBNR follows from the indexation given by (4.2) and (4.3). In more detail, it follows that

$$\mathbb{E}\left[X_{ijk} \mid \mathcal{N}_0\right] = \mu_{ijk}\lambda_{ijk}\mathbb{E}\left[N_{i,j,k-1}^{\text{open}} \mid \mathcal{N}_0\right] \tag{4.4}$$

for $k = 1, ..., m-1$, and

$$\mathbb{E}\left[X_{ij0} \mid \mathcal{N}_0\right] = \mu_{ij0}\lambda_{ij0}\mathbb{E}\left[N_{ij} \mid \mathcal{N}_0\right],$$

where the expected value of the number of open claims is most conveniently expressed recursively as

$$\mathbb{E}\left[N_{ijk}^{\text{open}} \mid \mathcal{N}_0\right]$$
$$= \begin{cases} N_{ijk}^{\text{open}}, & (i,j,k) \in \mathcal{A}_0 \\ \left(p_{ijk} - q_{ijk}\right)\mathbb{E}\left[N_{i,j,k-1}^{\text{open}} \mid \mathcal{N}_0\right] + q_{ijk}\mathbb{E}\left[N_{ij} \mid \mathcal{N}_0\right], & (i,j,k) \notin \mathcal{A}_0 \end{cases}, \tag{4.5}$$

for $k = 1, ..., m-1$, and

$$\mathbb{E}\left[N_{ij0}^{\text{open}} \mid \mathcal{N}_0\right] = p_{ij0}\mathbb{E}\left[N_{ij} \mid \mathcal{N}_0\right],$$

where

$$\mathbb{E}\left[N_{ij} \mid \mathcal{N}_0\right] = \begin{cases} N_{ij}, & (i,j,0) \in \mathcal{A}_0 \\ v_{ij}, & (i,j,0) \notin \mathcal{A}_0 \end{cases}.$$

Remark 2

(a) *In agreement with Verrall et al. (2010) and Wahl et al. (2019), it is seen that (4.4) can be re-written in terms of $\psi_{ijk} = \mu_{ijk}\lambda_{ijk}$, which is a parametrisation that will be used when discussing estimation in Section 5.*

(b) *By using the reduced model where $p_{ijk} = q_{ijk}$ discussed in Remark 1(b), it is seen that the expected values not in $\mathcal{A}_0$ reduce to only depending on N_{ij}:s, as it should.*

4.2 (Co)variances

Compared with the situation in Verrall *et al.* (2010) and Wahl *et al.* (2019), the payments X_{ijk} are no longer independent for equal i and j (unless all $p_{ijk} \equiv 1$, when the model coincides with the CRM), and the variance of outstanding payments must be calculated in full according to

$$\text{Var}\left(R_{ij} \mid \mathcal{N}_0\right) = \sum_{k=m-i-j+1}^{m-1} \left\{ \text{Var}\left(X_{ijk} \mid \mathcal{N}_0\right) + 2 \sum_{l=k+1}^{m-1} \text{Cov}\left(X_{ijk}, X_{ijl} \mid \mathcal{N}_0\right) \right\},$$

(4.6)

where the split into RBNS and IBNR follows from the indexation given by (4.2) and (4.3). Still, the components of (4.6) can be represented recursively according to

$$\text{Var}\left(X_{ijk} \mid \mathcal{N}_0\right) = \mu_{ijk}^2 \lambda_{ijk}^2 \text{Var}\left(N_{i,j,k-1}^{\text{open}} \mid \mathcal{N}_0\right)$$
$$+ \left(\mu_{ijk}^2 + \sigma_{ijk}^2\right) \lambda_{ijk} \mathbb{E}\left[N_{i,j,k-1}^{\text{open}} \mid \mathcal{N}_0\right]$$

(4.7)

for $k = 1, ..., m-1$, and

$$\text{Var}\left(X_{ij0} \mid \mathcal{N}_0\right) = \mu_{ij0}^2 \lambda_{ij0}^2 \text{Var}\left(N_{ij} \mid \mathcal{N}_0\right)$$
$$+ \left(\mu_{ij0}^2 + \sigma_{ij0}^2\right) \lambda_{ij0} \mathbb{E}\left[N_{ij} \mid \mathcal{N}_0\right].$$

For $(i, j, k) \in \mathcal{A}_0$,

$$\text{Var}\left(N_{ijk}^{\text{open}} \mid \mathcal{N}_0\right) = 0,$$

and for $(i, j, k) \notin \mathcal{A}_0$,

$$\text{Var}\left(N_{ijk}^{\text{open}} \mid \mathcal{N}_0\right) = \left(p_{ijk} - q_{ijk}\right)^2 \text{Var}\left(N_{i,j,k-1}^{\text{open}} \mid \mathcal{N}_0\right)$$
$$+ \left(p_{ijk}\left(1 - p_{ijk}\right) - q_{ijk}\left(1 - q_{ijk}\right)\right) \mathbb{E}\left[N_{i,j,k-1}^{\text{open}} \mid \mathcal{N}_0\right]$$
$$+ q_{ijk}\left(1 - q_{ijk}\right) \mathbb{E}\left[N_{ij} \mid \mathcal{N}_0\right]$$
$$+ q_{ijk}^2 \text{Var}\left(N_{ij} \mid \mathcal{N}_0\right)$$
$$+ 2\left(p_{ijk} - q_{ijk}\right) q_{ijk} \text{Cov}\left(N_{i,j,k-1}^{\text{open}}, N_{ij} \mid \mathcal{N}_0\right)$$

(4.8)

for $k = 1, ..., m-1$ and

$$\text{Var}\left(N_{ij0}^{\text{open}} \mid \mathcal{N}_0\right) = p_{ij0}\left(1 - p_{ij0}\right) \mathbb{E}\left[N_{ij} \mid \mathcal{N}_0\right]$$
$$+ p_{ij0}^2 \text{Var}\left(N_{ij} \mid \mathcal{N}_0\right).$$

where

$$\text{Cov}\left(N_{ijk}^{\text{open}}, N_{ij} \mid \mathcal{N}_0\right) = (p_{ijk} - q_{ijk}) \, \text{Cov}\left(N_{ij}, N_{i,j,k-1}^{\text{open}} \mid \mathcal{N}_0\right)$$
$$+ q_{ijk} \text{Var}\left(N_{ij} \mid \mathcal{N}_0\right)$$

for $k = 1, ..., m-1$ and

$$\text{Cov}\left(N_{ij0}^{\text{open}}, N_{ij} \mid \mathcal{N}_0\right) = p_{ij0} \text{Var}\left(N_{ij} \mid \mathcal{N}_0\right).$$

Also,

$$\text{Var}\left(N_{ij} \mid \mathcal{N}_0\right) = \begin{cases} 0, & (i,j,0) \in \mathcal{A}_0 \\ \phi v_{ij}, & (i,j,0) \notin \mathcal{A}_0 \end{cases}.$$

As for the covariance terms of the payments, it holds that

$$\text{Cov}\left(X_{ijk}, X_{ijl} \mid \mathcal{N}_0\right) = \mu_{ijk} \mu_{ijl} \lambda_{ijk} \lambda_{ijl} \text{Cov}\left(N_{i,j,k-1}^{\text{open}}, N_{i,j,l-1}^{\text{open}} \mid \mathcal{N}_0\right),$$

when $k = 1, ...l-1$ and

$$\text{Cov}\left(X_{ij0}, X_{ijl} \mid \mathcal{N}_0\right) = \mu_{ij0} \mu_{ijl} \lambda_{ij0} \lambda_{ijl} \text{Cov}\left(N_{ij}, N_{i,j,l-1}^{\text{open}} \mid \mathcal{N}_0\right).$$

Here,

$$\text{Cov}\left(N_{ijk}^{\text{open}}, N_{ijl}^{\text{open}} \mid \mathcal{N}_0\right) = (p_{ijl} - q_{ijl}) \, \text{Cov}\left(N_{ijk}^{\text{open}}, N_{i,j,l-1}^{\text{open}} \mid \mathcal{N}_0\right)$$
$$+ q_{ijl} \text{Cov}\left(N_{ijk}^{\text{open}}, N_{ij} \mid \mathcal{N}_0\right)$$

for $k = 0, ..., l-1$. The derivation of the above (co)variances is again given by straightforward repeated conditioning.

Remark 3

(a) *By inspecting the expressions for the (co)variances, it is clear that the corresponding quantities for the CRM are retrieved by setting all $p_{ijk} := 1$. That is, $N_{ij} \equiv N_{ij,k}^{open}$, for all k, and those covariance terms that are not immediately cancelled will be turned into variance terms.*

(b) *For the model discussed in Remark 1(b) where $p_{ijk} = q_{ijk}$, the moments are simplified considerably, and all explicit dependencies between N_{ijk}^{open}:s are replaced with variances only depending on N_{ij}:s.*

(c) *Note that $\text{Var}\left(X_{ijk} \mid \mathcal{N}_0\right) = \varphi_{ijk} \mathbb{E}[X_{ijk} \mid \mathcal{N}_0]$ for all $(i,j,k) \in \mathcal{A}_0$, where φ_{ijk} is a function of $\boldsymbol{\theta}$, hence satisfying the ODP moment assumption. Moreover, $\text{Cov}(X_{ijk}, X_{i,j,l} \mid \mathcal{N}_0) = 0$ for all $k \neq l$ and $(i,j,k), (i,j,l) \in \mathcal{A}_0$, and, in particular, it holds that all $X_{ijk}, X_{i,j,l}$ where $k \neq l$ and $(i,j,k), (i,j,l) \in \mathcal{A}_0$ are conditionally independent, given $\mathcal{N}_0$.*

(d) *Analogously to the discussion in Remark 2 it is seen that (co)variance-expressions can be simplified by introducing $\psi_{ijk} := \mu_{ijk} \lambda_{ijk}$.*

5 ESTIMATION

The CRMO as introduced in Section 3 is very flexible and in practice we need to make assumptions concerning the parametrisation of this model. Concerning the claim occurrence part of the model, a natural way to proceed is to use the over-dispersed chain-ladder parametrisation, that is,

$$v_{ij} = e^{\alpha_i + \beta_j}, \tag{5.1}$$

which corresponds to a standard over-dispersed Poisson GLM. Following this approach, ϕ can be estimated, as in, for example, Martinez-Miranda *et al.* (2011), Eq. (11), according to

$$\widehat{\phi} = \frac{1}{\frac{m(m+1)}{2} - (1 + 2(m-1))} \sum_{(i,j) \in \mathcal{A}_0} \frac{(N_{ij} - \widehat{v}_{ij})^2}{\widehat{v}_{ij}}.$$

Furthermore, recall that the motivation for using a quasi-likelihood approach for modelling the parameters defining the distribution of N_{ij}, that is, v_{ij} and ϕ, relies on that all N_{ij} are independent and that

$$\mathrm{Var}\,(N_{ij}) = \phi\mathbb{E}[N_{ij}] = \phi v_{ij},$$

and that by using the parametrisation (5.1), it is sufficient to only have observations on $\mathcal{A}_0$ and still be able to predict future claim occurrences. In Wahl *et al.* (2019), this was exploited for the payment part of the process as well, by observing that the CRM results in that

$$\mathrm{Var}\,(X_{ijk}^{\mathrm{CRM}} \mid \mathcal{N}_0) = \widetilde{\varphi}\mathbb{E}[X_{ijk}^{\mathrm{CRM}} \mid \mathcal{N}_0] = \widetilde{\varphi}\widetilde{\psi}_{ijk}N_{ij}, \tag{5.2}$$

where $\widetilde{\psi}_{i,jk}, \widetilde{\varphi} > 0$, and where all X_{ijk}^{CRM} independent.

Remark 4

(a) *For the numerical illustrations in Wahl et al. (2019), the parametrisation $\widetilde{\psi}_{ijk} = \mu_{ijk}\widetilde{\lambda}_{ijk} = \widetilde{\psi}_k$ is used, based on that $\widetilde{\lambda}_{ijk} = \widetilde{\lambda}_k$ and $\mu_{ijk} = \mu$. This combined with that $\sigma_{ijk}^2 = \sigma^2$ implies a constant over-dispersion parameter $\widetilde{\varphi}$. (Note that we do not have to introduce $\widetilde{\mu}_{ijk}$ and $\widetilde{\sigma}_{ijk}$, since these parameters will always have the same meaning for the CRMO and the CRM.) This particular choice is partly motivated by the fact that the CRM can be estimated using more aggregated X_{ij}-data, which is not an alternative in the current paper, since we are interested in capturing the dynamics of claim closings and re-openings.*

(b) *In Lindholm et al. (2020), it is assumed that*

$$\widetilde{\varphi}_{ijk} := \frac{\mu_{ijk}^2 + \sigma_{ijk}^2}{\mu_{ijk}} \equiv \widetilde{\varphi}, \tag{5.3}$$

which is an artificial, and in many situations unreasonable, assumption. Still, it allows us to estimate the $\widetilde{\psi}_{ijk}$:s flexibly and reduces the number

of parameters. Note, however, that $\widetilde{\varphi}_{ijk}$ is only a function of payment size parameters, and we could just as well assume that these are independent of (i, j, k). The practical consequences of assuming (5.3) are primarily that the model interpretation becomes less clear. This will be discussed further in Sections 6 and 7.

In the current setting, for the CRMO, this relationship holds true as well, but only conditionally independent for all $(i, j, k) \in \mathcal{A}_0$. That is,

$$\mathrm{Var}\left(X_{ijk} \mid \mathcal{N}_0\right) = \varphi \mathbb{E}[X_{ijk} \mid \mathcal{N}_0] = \varphi \psi_{ijk} N_{i,j,k-1}^{\mathrm{open}}, \quad (i, j, k) \in \mathcal{A}_0, \qquad (5.4)$$

where $\psi_{ijk} = \mu_{ijk} \lambda_{ijk}$, and where

$$\varphi = \frac{\mu_{ijk}^2 + \sigma_{ijk}^2}{\mu_{ijk}}$$

which follows directly from (4.4) and (4.7) for all X_{ijk} where $(i, j, k) \in \mathcal{A}_0$, and recall from Remark 3 that these X_{ijk} are conditionally independent. Moreover, as for the CRM, the functional form describing the ψ_{ijk}:s should be possible to estimate based on data in $\mathcal{A}_0$ alone and that the functional form should still allow for predictions into the future. Furthermore, concerning the estimation of φ, this can be done analogously as for the CRM based on X_{ijk}-level data by using

$$\widehat{\varphi} = \frac{1}{\frac{m(m+1)(m+2)}{6} - (1 + 3(m-1))} \sum_{(i,j,k) \in \mathcal{A}_0} \frac{\left(X_{ijk} - \widehat{\psi}_{ijk} N_{i,j,k-1}^{\mathrm{open}}\right)^2}{\widehat{\psi}_{ijk} N_{i,j,k-1}^{\mathrm{open}}}. \qquad (5.5)$$

Thus, the ψ_{ijk}:s and the over-dispersion parameter φ can be estimated using a quasi-Poisson likelihood. Consistency of these parameter estimators follows the arguments from Wahl *et al.* (2019), Prop. 2 for the CRM.

What remains now is to estimate the probabilities defining the dynamics for closing and re-opening of claims. In the ideal situation, one has access to the additional information $N_{ijk}^{\mathrm{stay\text{-}open}}$ and $N_{ijk}^{\mathrm{re\text{-}open}}$, and given this information, the probabilities p_{ijk} and q_{ijk} are estimated using standard binomial likelihoods. This approach guarantees consistency of the parameter estimators. In Section 6, we use GLM:s with canonical link functions. Although this modelling approach is simple, it requires a lot of data, data which are not needed for reserve prediction, see also the discussion in the first paragraph of Section 4. Another way to estimate the p_{ijk}:s and q_{ijk}:s is to only use N_{ijk}^{open}-data and make the (artificial) assumption $p_{ijk} = q_{ijk} = \pi_{ijk}$ for all (i, j, k). While it is unlikely that the probability of a claim re-opening is equal to that of a claim staying open, the parameter π_{ijk} can instead be interpreted as the probability of a claim *being open* k years after being reported in (i, j). Moreover, the π_{ijk}:s can then be estimated using, for example, standard GLM:s with canonical

link functions, since the number of open claims will then be (conditionally) likelihood equivalent with

$$N_{ijk}^{\text{open}} \mid N_{ij} \sim \text{Bin}\left(N_{ij}, \pi_{ijk}\right). \tag{5.6}$$

As commented on in Section 4, the moment expressions for N_{ijk}^{open} will also be vastly simplified for this more restrictive model.

An alternative to estimate model parameters only based on N_{ijk}^{open} data is to treat the $N_{ijk}^{\text{stay-open}}$:s and $N_{ijk}^{\text{re-open}}$:s as missing values. This situation is straight-forward to approach using the EM-algorithm, even though this approach tends to be sensitive to initial values, since the sum-of-Binomial structure will, when using more flexible parametrisations, likely lead to identification problems for the probabilities, resulting in many local maxima for the joint likelihood. Even if this approach may produce reasonable reserve predictions, this will have implications on the variance structure.

5.1 Comments on consistency of normalised reserves

Based on the constructive arguments that underpin both the CRM and the CRMO in terms of individual claim dynamics, it is natural to explicitly intro-duce accident year specific exposures w_i that corresponds to the number of contracts for accident year i. That is, assume that $N_{ij} \sim \text{ODP}(\nu_{ij}, \phi)$ where

$$\nu_{ij} := w_i \alpha_i \beta_j = w_i \overline{n}_{ij}(\boldsymbol{\theta}),$$

where $\overline{n}_{ij}(\boldsymbol{\theta}) := \alpha_i \beta_j$, and note that N_{ij} can be expressed as

$$N_{ij} \overset{d}{=} \sum_{l=1}^{w_i} \widetilde{N}_{ijl}, \tag{5.7}$$

where all $\widetilde{N}_{ijl}$ are i.i.d. $\text{ODP}(\alpha_i \beta_j, \phi)$. Hence, since the $\widehat{R}_{ij}$s will be proportional to the corresponding N_{ij}s (or expectations thereof), $\widehat{R}_{ij}/w_i$ can be thought of as the approximate reserve from a single contract. Thus, consistency of $\widehat{R}_{ij}/w_i$ has a natural interpretation and is also treated in, for example, Huang *et al.* (2015, 2016).

Further, note that when letting $w_i \to \infty$ the dimension of $\widehat{\boldsymbol{\theta}}$ is fixed. Moreover, if we assume that the parametrisation of $\boldsymbol{\theta}$ is estimable based on $\mathcal{X}$ and $\mathcal{N}_0^+$ data (or assuming that $p_{ijk} = q_{ijk}$, see also Remark 1), by the arguments in Wahl *et al.* (2019), Proposition 2(ii), it follows that $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}$ in prob-ability as $w_i \to \infty$. Next, given that the reserves can be split into an IBNR and a RBNS part, the consistency of $\widehat{\boldsymbol{\theta}}$ immediately yields that $\widehat{R}_{ij}^{\mathcal{I}}/w_i$ as $w_i \to \infty$. Continuing with the RBNS reserves, from Section 4 for (i, j, k) such that X_{ijk} corresponds to an RBNS payment, it holds that N_{ij} and $N_{i,j,m-i-j}^{\text{open}}$ are $\mathcal{N}_0$

measurable, which gives us that

$$
\begin{aligned}
h_{ijk}(\boldsymbol{\theta}; \mathcal{N}_0) &:= \mathbb{E}[X_{ijk} \mid \mathcal{N}_0](\boldsymbol{\theta}) \\
&:= \mathbb{E}[N_{i,j,k-1}^{\mathrm{open}} \mid \mathcal{N}_0](\boldsymbol{\theta}) \psi_{ijk} \\
&= \psi_{ijk} N_{i,j,m-i-j}^{\mathrm{open}} \prod_{l=m-i-j+1}^{k-1} \left(p_{ijl} - q_{ijl} \right) \\
&\quad + \psi_{ijk} N_{ij} \sum_{l=m-i-j+1}^{k-1} q_{ijl} \prod_{n=l+1}^{k-1} \left(p_{ijn} - q_{ijn} \right)
\end{aligned}
\tag{5.8}
$$

by repeated conditioning using (4.5), and analogous calculations yields

$$
\begin{aligned}
\mathbb{E}[h_{ijk}(\boldsymbol{\theta}; \mathcal{N}_0)] &= \mathbb{E}[N_{ij}] \left(\sum_{l=1}^{k-1} q_{ijl} \prod_{n=l+1}^{k-1} \left(p_{ijn} - q_{ijn} \right) + p_{ij0} \prod_{l=1}^{k-1} \left(p_{ijl} - q_{ijl} \right) \right) \psi_{ijk} \\
&= w_i \bar{n}_{ij}(\boldsymbol{\theta}) \left(\sum_{l=1}^{k-1} q_{ijl} \prod_{n=l+1}^{k-1} \left(p_{ijn} - q_{ijn} \right) + p_{ij0} \prod_{l=1}^{k-1} \left(p_{ijl} - q_{ijl} \right) \right) \psi_{ijk} \\
&= w_i \bar{x}_{ijk}(\boldsymbol{\theta}),
\end{aligned}
\tag{5.9}
$$

where

$$
\bar{x}_{ijk} := \bar{n}_{ij}(\boldsymbol{\theta}) \left(\sum_{l=1}^{k-1} q_{ijl} \prod_{n=l+1}^{k-1} \left(p_{ijn} - q_{ijn} \right) + p_{ij0} \prod_{l=1}^{k-1} \left(p_{ijl} - q_{ijl} \right) \right) \psi_{ijk}.
$$

Further, by also noting that both the expected value and variance of (5.8) conditional on N_{ij} are proportional to N_{ij} gives us that

$$
\mathrm{Var}\left(\mathbb{E}[h_{ijk}(\boldsymbol{\theta}; \mathcal{N}_0)] \right) \propto w_i,
$$

where "$\propto$" corresponds to equality up to scaling factors (being functions only of $\boldsymbol{\theta}$) in front of leading terms, that is

$$
\mathrm{Var}\left(\mathbb{E}[h_{ijk}(\boldsymbol{\theta}; \mathcal{N}_0)] \right) = w_i v_{ijk}(\boldsymbol{\theta}) < \infty, \text{ for } w_i < \infty,
$$

where $v_{ijk}(\boldsymbol{\theta})$ is a function expressed in terms of the parameter vector $\boldsymbol{\theta}$. Consequently, it holds that

$$
\mathbb{P}(| \frac{1}{w_i} h_{ijk}(\boldsymbol{\theta}; \mathcal{N}_0) - \bar{x}_{ijk}| > \epsilon) \leq \frac{v_{ijk}(\boldsymbol{\theta})}{\epsilon^2 w_i} \to 0 \text{ as } w_i \to \infty,
$$

that is $h_{ijk}(\boldsymbol{\theta}; \mathcal{N}_0)/w_i \to \bar{x}_{ijk}(\boldsymbol{\theta})$ in probability as $w_i \to \infty$.

Further, by estimating the parameters $\boldsymbol{\theta}$ using $\mathcal{N}_0^+$ and combining with the above it follows that

$$
\frac{1}{w_i} h_{ijk}(\widehat{\boldsymbol{\theta}}; \mathcal{N}_0) \xrightarrow{p} \bar{x}_{ijk} \text{ as } w_i \to \infty,
$$

due to the continuous mapping theorem, and, hence,

$$\frac{1}{w_i}\widehat{R}_{ij}^{\mathcal{R}} := \frac{1}{w_i} \sum_{k=m-i-j+1}^{m-1} h_{ijk}(\widehat{\boldsymbol{\theta}}; \mathcal{N}_0) \xrightarrow{p} \sum_{k=m-i-j+1}^{m-1} \overline{x}_{ijk}(\boldsymbol{\theta}) \text{ as } w_i \to \infty,$$

which follows from elementary operations of convergence in probability (Cramér-Slutsky). For more details and alternative proofs, see, for example, Wahl *et al*. (2019).

Remark 5 *Note that the overall consistency of $\widehat{R}_{ij}/w_i$ as $w_i \to \infty$ is a mixture of QMLE consistency of $\widehat{\boldsymbol{\theta}}$ and the trivial N_{ij}/w_i consistency. One can, however, note that N_{ij}/w_i in light of (5.7) can be thought of as the QMLE of $\overline{n}_{ij}(\boldsymbol{\theta})$ based on the i.i.d. individual claim counts $\widetilde{N}_{ijl}, l = 1, \ldots, w_i$. Further, the interpretation of the above is that by increasing the number of underlying contracts in the portfolio, we expect that the relative reserve in each cell will converge, without letting the dimension of the observed data tend to infinity. Consequently, for sufficiently large portfolios, the influence of conditioning on actually observed claim counts instead of using conditional expected counts will diminish, which partly motivates the reasonability of using chain-ladder technique models, see Martinez-Miranda et al. (2012, Thm. 1).*

5.2 Comparison with the CRM

As already discussed in Remarks 1, 2, and 3, the CRM, as presented in Wahl *et al*. (2019), is a special case of the model introduced in the current paper, corresponding to $p_{ijk} = 1$ for all combinations of i, j, k. Further, the expected value of a payment conditioned on the number of reported claims for the CRM can be expressed as

$$\mathbb{E}\left[X_{ijk} \mid N_{ij}\right] = \widetilde{\psi}_{i,j,k} N_{ij},$$

for some parametrisation $\widetilde{\psi}_{i,j,k}$, if we restrict our attention to $(i, j, k) \in \mathcal{A}_0$ (otherwise replace N_{ij} with v_{ij}). By assuming that the true claim dynamics are given by the CRMO, one obtains

$$\mathbb{E}\left[X_{ijk} \mid N_{ij}\right]$$
$$= \mu_{ijk}\lambda_{ijk} \left(\sum_{l=1}^{k-1} q_{ijl} \prod_{m=l+1}^{k-1} \left(p_{ijm} - q_{ijm}\right) + p_{ij0} \prod_{l=1}^{k-1} \left(p_{ijl} - q_{ijl}\right) \right) N_{ij}, \quad (5.10)$$

which follows from repeated conditioning, see (5.9). Further, in Wahl *et al*. (2019), Lindholm *et al*. (2020) GLM:s of the form

$$\log \widetilde{\psi}_{ijk} := \alpha_i + \beta_j + \gamma_k, \quad (5.11)$$

are used whereas in Lindholm *et al*. (2020), machine learning algorithms are used in order to learn non-linear representations of $\widetilde{\psi}_{ijk}$ using flexible

parametrisations in terms of functional forms $\widetilde{\psi}(i, j, k; \boldsymbol{\theta})$. Consequently, given that the CRMO dynamics are the correct ones, the CRM parametrisation (5.11) will likely not produce sufficient flexibility compared with (5.10). This could possibly be achieved by using machine learning techniques as those discussed in Lindholm *et al.* (2020).

To conclude this far, theoretically it is possible to equate the conditional expected values of X_{ijk} between the CRM and the CRMO, when conditioning on N_{ij}:s, but the computable predictors of these moments will in general not agree.

Concerning higher moments, the variance of X_{ijk}, $(i, j, k) \in \mathcal{A}_0$ satisfies

$$
\begin{aligned}
\mathrm{Var}\,(X_{ijk} \mid N_{ij}) &= \varphi \psi_{ijk} \mathbb{E}[N_{i,j,k-1}^{\mathrm{open}} \mid N_{ij}] + \psi_{ijk}^2 \,\mathrm{Var}\,(N_{i,j,k-1}^{\mathrm{open}} \mid N_{ij}) \\
&= \varphi \mathbb{E}[X_{ijk} \mid N_{ij}] + \psi_{ijk}^2 \,\mathrm{Var}\,(N_{i,j,k-1}^{\mathrm{open}} \mid N_{ij}) \\
&= \widetilde{\varphi} \mathbb{E}[X_{ijk} \mid N_{ij}].
\end{aligned}
\tag{5.12}
$$

Thus, if we assume that the true underlying data generating process follows the CRMO, (5.12) tells us that for $(i, j, k) \in \mathcal{A}_0$, the ODP moment assumption will hold for the CRM as well. But, the over-dispersion parameter in the CRM, here denoted $\widetilde{\varphi}$, will not correspond to the correct parametric relationship, since from (5.12) it is clear that $\widetilde{\varphi}$ also depends on p_{ijk}:s and q_{ijk}:s, hence not being of the form (5.3). That is, when estimating $\widetilde{\varphi}$ directly using the CRM based on data generated by the CRMO, unless there are no claim closings or re-openings, the CRM will not be consistent with the data generating process. Moreover, (5.12) also tells us that the variance of the CRMO, when only conditioning on N_{ij} information, will be greater than the corresponding variance of the CRM, when using the true, unobservable, parameters. This is expected, since an additional source of variation has been introduced.

Remark 6 *In Wahl et al. (2019), the relationship between the CRM and the VNJ model is discussed w.r.t. computable reserve predictors and their corresponding variances. This comparison was possible due to that the parameter estimates in the VNJ model can be made identical to those for the CRM. This is not the case for the CRMO, and this type of comparison is, hence, omitted.*

5.3 Comments on the computation of the mean squared error of prediction of reserves

The CRMO is an extension of the CRM, and consequently the basic steps in the computation of the mean squared error of prediction closely follow those from the CRM, which is described in Wahl *et al.* (2019) and Lindholm *et al.* (2020). This procedure can be thought of as a parametric bootstrap making use of explicit moment calculations, including the process variance, and the only qualitative difference with the CRM is the introduction of claim closings and re-openings, which is only an addition of a binomial structure to the modelling.

This added complexity, however, exactly follows the same procedures concerning the split in terms of RBNS and IBNR reserves and is, hence, omitted.

Alternatively, as in, for example, Buchwalder *et al.* (2006), Diers *et al.* (2016), Lindholm *et al.* (2020), one could Taylor approximate the estimation error part of the MSEP calculation directly, making use of the Hessian from the QMLE-fitting. This results in a semi-analytical MSEP approximation, but since bootstrapping is straightforward (and quick), we will use the bootstrap MSEP-procedure in Section 6.

6 NUMERICAL ILLUSTRATIONS

In order to illustrate the usefulness of the model in a reserving setting, the following section tests the models' reserve prediction accuracy on simulated individual claims data from the algorithm described in Gabrielli and Wüthrich (2018), which allows us to compare reserve predictions with true simulated outcomes, that are not simulated in accordance with the assumed model structure. In particular, we will use the same simulated data set as used in Gabrielli *et al.* (2020) and Lindholm *et al.* (2020), which consists of 6 individual portfolios, or Lines of Business (LoBs), each having roughly 100,000–250,000 individual claims equipped with reporting, payment and claim openness history. By using simulated data on the individual claim level, it is straightforward to aggregate data to obtain $\mathcal{X}_0$, $\mathcal{N}_0$ and $\mathcal{N}_0^+$.

Concerning properties of the simulated data, apart from the data exploration already described for this particular simulated data set that can be found in Gabrielli and Wüthrich (2018), Gabrielli *et al.* (2020), Lindholm *et al.* (2020), we briefly want to highlight additional aspects that are of interest for the CRMO.

The CRMO allows for the amount paid in a period to depend not only on accident year and reporting delay but also on payment delay. For the simulated data used in the numerical illustration, it is seen that this type of dependence is present in every LoB, see Figure 1. Further, the CRMO allows for dynamic closings and re-openings of claims. In Figure 2, the average ratio of open claims per reported claim for every year after being reported can be seen for different reporting delays for each respective LoB. As can be seen, the majority of claims close within a few years since reporting. But, interestingly, the average number of open claims are occasionally higher than in the previous development year, which is especially noticeable at $k = 2$ years since reporting. This would prove a problem for a model using $q_{ijk} = 0$, that is, assuming no re-openings.

Moreover, a closer look into re-openings is provided by Figure 3, which shows the average ratio of open claims that were open at the end of the previous year, $N_{ijk}^{\text{stay-open}}/N_{ijk}^{\text{open}}$. From Figure 3, it is seen that re-openings are primarily occurring during the first few years after reporting, but can make up a non-negligible fraction of the open claims and should hence be modelled.
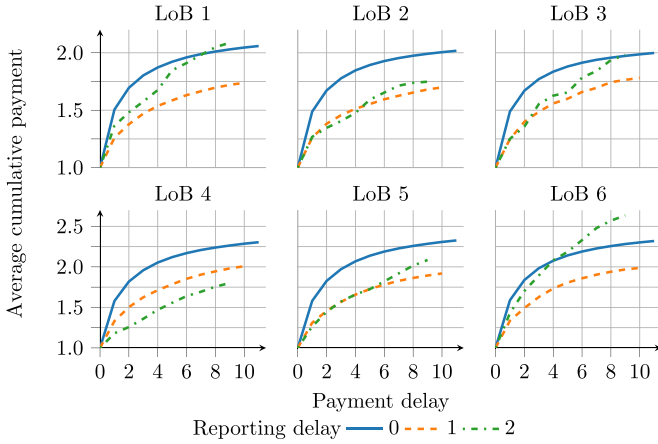
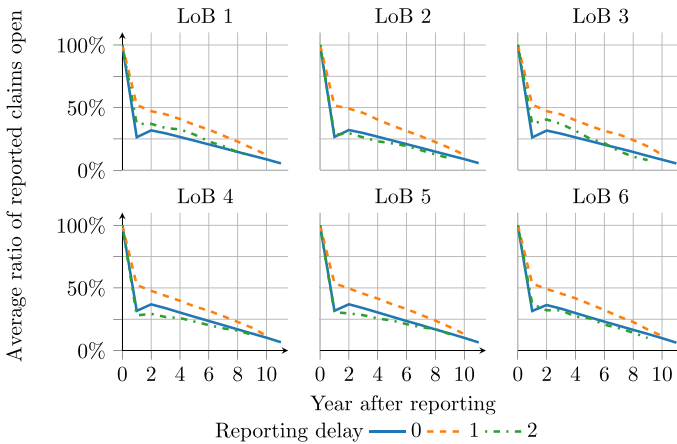FIGURE 1: Average cumulative payment per payment delay for different reporting delays in all LoBs.

FIGURE 2: Average ratio of open claims to reported claims for different reporting delays in all LoBs.

Continuing, the definition of the CRMO is a theoretical construction, and for estimation purposes, we need to make assumptions concerning the form of the regression functions. In this short numerical illustration, the CRMO is fitted to the 6 LoBs (respectively) using the following (canonical) GLM formulations

$$\log \nu_{ij} = \alpha_i^{\nu} + \beta_j^{\nu}, \tag{6.1}$$

$$\log \psi_{ijk} = \alpha_i^{\psi} + \beta_j^{\psi} + \gamma_k^{\psi}, \tag{6.2}$$

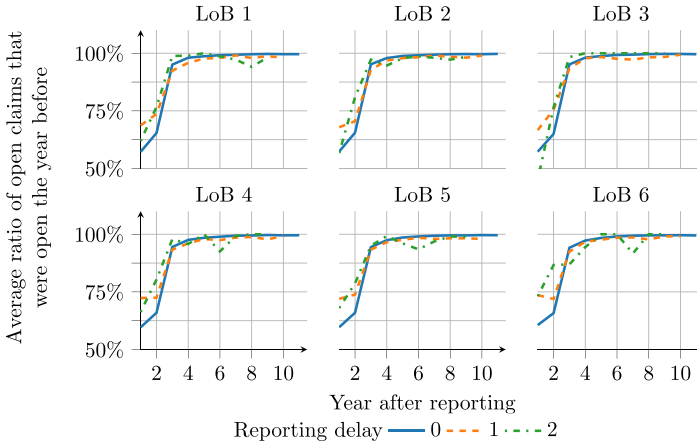$$\mathrm{sigm}^{-1}\left(p_{ijk}\right) = \alpha_i^{p} + \beta_j^{p} + \gamma_k^{p}, \tag{6.3}$$

FIGURE 3: Average ratio of open claims that were also open in the end of the previous year, $N_{ijk}^{\text{stay-open}}/N_{ijk}^{\text{open}}$.

and

$$\text{sigm}^{-1}\left(q_{ijk}\right) = \alpha_i^q + \beta_j^q + \gamma_k^q, \tag{6.4}$$

where

$$\text{sigm}\left(x\right) = \frac{1}{1 + e^{-x}}.$$

Moreover, the ν_{ij}:s and the ψ_{ijk}:s will be estimated using quasi-likelihood theory, and it will be assumed that $\varphi_{ijk} \equiv \varphi$, an assumption that will be discussed further below, also recall Remark 4 in Section 5.

As pointed out in Section 5, for the purpose of reserve prediction, only data in $\mathcal{N}_0$ is needed. Still, $\mathcal{N}_0^+$ is needed for direct estimation when assuming non-zero re-opening probabilities and $p_{ijk} \neq q_{ijk}$. The model fitted using this data is denoted CRMO$^{(1)}$ in the numerical illustration. The alternative model, only requiring $\mathcal{N}_0$ for parameter estimation by assuming $p_{ijk} = q_{ijk}$, is denoted CRMO$^{(2)}$. Both CRMO$^{(1)}$ and CRMO$^{(2)}$ can handle net re-openings, that is, situations where $N_{ijk}^{\text{open}} > N_{i,j,k-1}^{\text{open}}$, which is the case with the data being studied here, see Figure 2.

The different CRMO models are also benchmarked against the chain-ladder technique and the CRM using (6.1) and (6.2). In Table 1, the relative reserve residuals, $R/\widehat{R} - 1$, for the models are presented. The best performing model for every individual LoB is marked with a box.

From Table 1, it is seen that the CRMO model outperforms the CL and CRM models on all the data sets. Moreover, the model assuming $p_{ijk} = q_{ijk}$, CRMO$^{(2)}$, outperforms the model fitted to the granular data sets, CRMO$^{(1)}$,

RESERVE PREDICTIONS TOGETHER WITH RELATIVE MODEL PERFORMANCE DEFINED AS $R/\widehat{R} - 1$, WHERE $\text{CRMO}^{(1)}$ IS FITTED USING $\mathcal{N}_0^+$, THAT IS USING DETAILED OPENNESS INFORMATION, AND $\text{CRMO}^{(2)}$ ASSUMES $p_{ijk} = q_{ijk}$ FOR ALL (i, j, k), AND ESTIMATION IS BASED ON $\mathcal{N}_0$. THE SMALLEST ABSOLUTE RESIDUAL IS MARKED WITH A BOX.

	LoB 1	LoB 2	LoB 3	LoB 4	LoB 5	LoB 6
True	39,689	37,037	16,878	71,630	72,548	31,117
CL	38,569	35,460	15,692	67,574	70,166	29,409
	−2.82%	−4.26%	−7.02%	−5.66%	−3.28%	−5.49%
CRM	38,321	35,144	15,456	67,051	69,480	29,174
	−3.45%	−5.11%	−8.43%	−6.39%	−4.23%	−6.24%
$\text{CRMO}^{(1)}$	39,464	36,124	15,902	69,468	71,146	30,729
	−0.57%	−2.47%	−5.79%	−3.02%	−0.55%	−1.25%
$\text{CRMO}^{(2)}$	39,392	36,163	15,944	69,207	72,123	30,391
	−0.75%	−2.36%	−5.53%	−3.38%	−0.59%	−2.33%

on two LoB's, which could be a result of the full model overfitting to in-sample data.

When taking a closer look into the predicted dynamics of the different CRMO models, it becomes clear that the $\text{CRMO}^{(2)}$ on average tends to produce better predictions of the total number of open claims than the granular fit, at least for low k. An example of this is given in Figure 4, which illustrates predictions of the ratio of open to reported claims for the last accident year and reporting delay 1 in LoB 1. Since it is the number of open claims that set the basis for the prediction of future payments, it does not matter that the $\text{CRMO}^{(2)}$ model formulation does not produce more granular claim openness predictions, as its prediction of number of open claims still seems to be sufficiently accurate.

In order to assess the model accuracy for different combinations of i, j, k, the relative residual, $R/\widehat{R} - 1$, for the $\text{CRMO}^{(1)}$-model for aggregated payments stemming from accident years $i = 1, ..., m$ and reporting plus payment delay $j + k = 0, ..., m - 1$ for LoB 1 is shown in Figure 5. Here one can note a slight tendency towards underestimating the reserves in the lower right triangle (out-of-sample), even though the overall pattern looks very good. Still, the tendency of diagonal effects could be a result of inflation, that is that payments or the number of open claims increases with calendar year $i + j + k$ – something considered in Lindholm *et al*. (2020), where the results indicated that taking such effects into account could improve accuracy.

Concerning the assumption of $\varphi_{ijk} \equiv \varphi$, this does not seem to be a too troublesome assumption for the data being analysed: Let

$$\widehat{\varphi}_{ijk} = \frac{\left(X_{ijk} - \widehat{\psi}_{ijk} N_{i,j,k-1}^{\text{open}}\right)^2}{\widehat{\psi}_{ijk} N_{i,j,k-1}^{\text{open}}}, \tag{6.5}$$
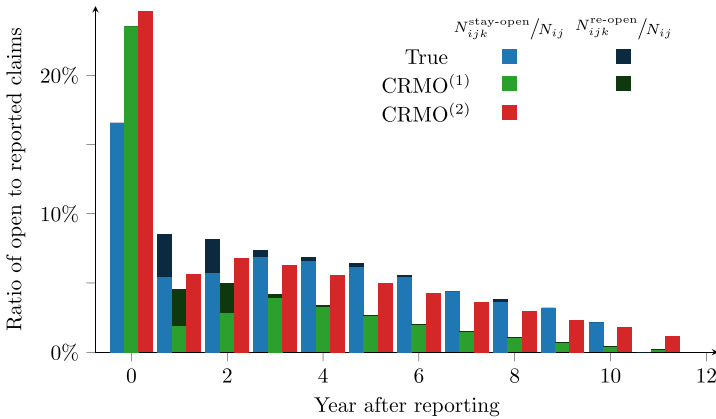
FIGURE 4: Ratio of open to reported claims for accident year 12 and reporting delay 1, outcome versus predictions from the models fitted using granular data (CRMO$^{(1)}$) and the model assuming $p_{ijk} = q_{ijk}$ (CRMO$^{(2)}$), respectively, in LoB 1. (Note that CRMO$^{(2)}$ cannot distinguish between claims that have been re-opened or that have stayed open.)
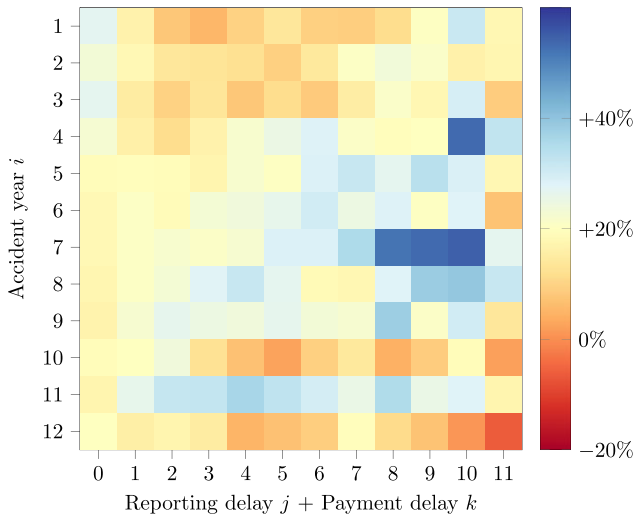

FIGURE 5: Average relative model performance defined as $\frac{X_{ijk}}{\hat{X}_{ijk}} - 1$ for the CRMO$^{(1)}$ for different lags in LoB 1, where model parameters are fitted using $\mathcal{N}_0^+$.

which can be compared with the overall estimate calculated using (5.5). In Figure 6 (6.5) is compared with (5.5) for LoB 1 for different combinations of i and $j + k$ for the first 5 development years. The reason for only using the first 5 development years is due to the scarcity in payments leading to extreme values for later years, and the fact that 90% of payments stem from the first 5 development years in all LoBs. It is obvious that while the assumption is not completely fulfilled, it here serves as a viable approximation.

TABLE 2.

ROOT MEAN SQUARED ERROR OF PREDICTION.

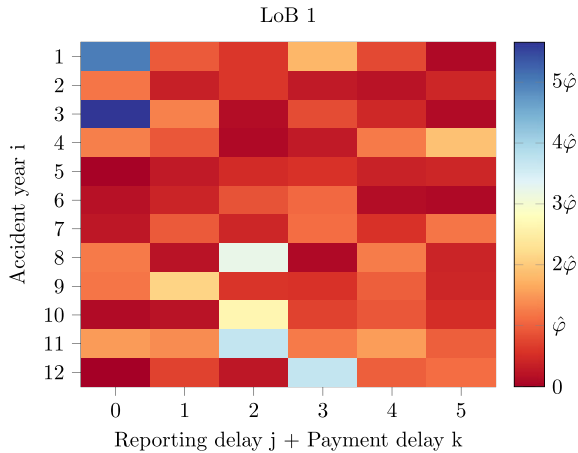|	LoB 1	LoB 2	LoB 3	LoB 4	LoB 5	LoB 6
CL	1112	1321	476	2200	1957	980
CRM	1070	1188	623	1747	2374	1193
CRMO[(1)]	1156	1374	636	2615	2255	1880
CRMO[(2)]	1168	1351	620	2586	2235	1866

FIGURE 6:   Relative overdispersion $\widehat{\varphi}/\widehat{\varphi}_{ijk} - 1$ calculated using (5.5) and (6.5) for LoB 1, using the CRMO[(1)].

Following the algorithm for calculating the conditional mean squared error of prediction as presented in Lindholm *et al.* (2020), using 10,000 bootstrap simulations for all LoBs, the (root) MSEP for the model is presented in Table 2. Note that even though the GLM formulations of the CRMO introduce additional parameters compared with the simpler CRM, the (rooted) MSEP values are of the same order as the CRM and often not too far off from the far simpler ODP CL model.

6.1 Comparisons to machine learning techniques and individual claims modelling

As mentioned in the beginning of Section 6, the simulated data used here are also used in Gabrielli *et al.* (2020), Lindholm *et al.* (2020) where machine learning techniques are used to learn a general representation of the ψ_{ijk}:s and ν_{ij}:s (or corresponding quantities) using feed-forward neural networks (NNs) and gradient boosting machines (GBMs). In Table 4, the CRMO[(1)] is compared with an NN version and a GBM version of the CRM, see Lindholm *et al.* (2020) for details concerning the NN and GBM models. From Table 4 it is seen

TABLE 3

RESERVE PREDICTIONS TOGETHER WITH RELATIVE MODEL PERFORMANCE FOR
DATA AND MODEL USED IN BETTONVILLE *et al.* (2020). CRMO$^{(1)}$ IS FITTED USING $\mathcal{N}_0$
AND $\mathcal{X}_0$ DATA, WHILE CRMO$^{(*)}$ USES OUT-OF-SAMPLE PAYMENT DATA FOR
PARAMETER ESTIMATION, BUT USES OBSERVED COUNTS FROM $\mathcal{N}_0$ FOR PREDICTION.

True	CL	CRMO$^{(1)}$	CRMO$^{(*)}$	Bettonville *et al.* (2020)
27,223	33,842	34,857	27,386	24,707
–	+24.31%	+28.04%	+0.60%	−9.24%

TABLE 4.

RESERVE PREDICTIONS TOGETHER WITH RELATIVE MODEL PERFORMANCE DEFINED AS
$\frac{R}{R} - 1$ FOR THE MODEL FITTED TO $\mathcal{N}_0^+$ (CRMO$^{(1)}$), AS COMPARED TO CRM MODELS FITTED
USING GRADIENT BOOSTING MACHINES (GBMS) AND NEURAL NETWORKS (NNS) FROM
LINDHOLM *et al.* (2020). THE SMALLEST ABSOLUTE RESIDUAL IS MARKED WITH A BOX. THE
GBM AND NN RESULTS ARE FOUND IN 4 IN LINDHOLM *et al.* (2020), TOGETHER WITH
DETAILED MODEL DESCRIPTIONS.

	LoB 1	LoB 2	LoB 3	LoB 4	LoB 5	LoB 6
True	39,689	37,037	16,878	71,630	72,548	31,117
GBM	38,324	37,053	16,327	73,386	70,486	32,100
	−3.44%	+0.04%	−3.26%	+2.45%	−2.84%	+3.16%
NN	41,587	37,587	15,680	71,155	71,309	28,984
	+4.78%	+1.48%	−7.10%	−0.66%	−1.71%	−6.86%
CRMO$^{(1)}$	39,464	36,124	15,902	69,468	71,146	30,729
	−0.57%	−2.47%	−5.79%	−3.02%	−0.55%	−1.25%

that by using the CRMO with canonical link-functions and the parametrisations (6.1)–(6.4), not using any interactions, the CRMO results are not too far off from those obtained using more flexible machine learning techniques.

Further, the present modelling approach builds on that from, for example, Verrall *et al.* (2010), Wahl *et al.* (2019), which makes use of that the underlying stochastic processes allows for preserving various desirable properties when aggregating claim payments and claim counts. Another approach is to instead consider individual claims modelling directly, which naturally allows for closing of claims. One such model that has been evaluated on publicly available data generated using the procedure from Gabrielli and Wüthrich (2018) is the semi-Markov model introduced in Bettonville *et al.* (2020), which allows for claim closings, but no re-openings. The data used in Bettonville *et al.* (2020) differ from the previously used six LoBs, here a single LoB with approximately 50,000 claims is used, but we refer to Bettonville *et al.* (2020) for more details. The results for the model from Bettonville *et al.* (2020) are given in Table 3 together with the CRMO$^{(1)}$ model.

It is here seen that both models perform rather poorly (including standard chain-ladder), but as commented on in Bettonville *et al.* (2020), a closer

inspection of the data reveals a number of large outliers. Without going into details concerning outliers, by instead of estimating $\boldsymbol{\theta}$ based on future data, but making predictions based on the observed number of open claims, the CRMO performance is greatly improved. This model is denoted CRMO[(*)] in Table 3, and the results for CRMO[(*)] indicate that the underlying model structure of the CRMO is sufficiently flexible and accurate in this situation as well. It should, however, be stressed that the result for CRMO[(*)] is not to be compared to the Bettonville *et al*. (2020) model results, since the former is fitted to out-of-sample data and the latter is not. This model is merely included to illustrate that the underlying model structure is reasonable, even though the parameter estimates in this application are poor. A final comment concerns the fact that we use a rather large number of parameters in the above GLM formulation of the CRMO. In practice, it would be reasonable to try to reduce the number of parameters by using regularisation techniques.

7 CONCLUDING REMARKS

The CRMO model is a generalisation of the CRM presented in Wahl *et al*. (2019), allowing for individual claim openness dynamics, but which can be estimated using aggregated payment and count data. Further, the constructive nature of the modelling allows for explicit and interpretable RBNS and IBNR reserve calculations. In the present paper, estimation of different GLM formulations of the model based on data availability and re-opening assumptions is presented and tested.

The model performs well in the numerical illustrations, with reserve residuals that outperform both the standard CL model and the generalised CRM model. It is also seen that for several simulated LoBs the CRMO performance is on par with machine learning versions of the CRM presented in Lindholm *et al*. (2020), see Table 4. A benefit with the CRMO in this comparison is that the CRMO is transparent, that is the structure is completely known, while the machine learning techniques will always contain "black box" elements. Still, similarly as in Lindholm *et al*. (2020), it is straightforward to combine the CRMO with NNs or GBMs directly if using $\mathcal{N}_0^+$ data or by using $\mathcal{N}_0$ data together with $p_{ijk} = q_{ijk}$. For the situation where $p_{ijk} \neq q_{ijk}$ and only $\mathcal{N}_0$ data are available bespoke loss functions are needed, which makes a machine learning technique implementation somewhat more involved.

In order to assess the CRMO model strength compared to other models considering claim openness, a small numerical comparison was made using the data from Bettonville *et al*. (2020). However, the CRMO model fell short to the model presented in Bettonville *et al*. (2020), due to that the realised payment and openness structure in the out-of-sample data was too far off from the data used for parameter estimation. Still, by estimating parameters based on out-of-sample data, something not possible to do in practice, but producing RBNS and IBNR predictions using the actual observed counts,

the reserve predictions became close to perfect, see CRMO(*) in Table 3. This artificial construction can be seen as an indication of that the underlying model structure for claim openness is reasonable even for this data.

ACKNOWLEDGEMENTS

CONFLICTS OF INTEREST

The author(s) declare none.

REFERENCES

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

BETTONVILLE, C., D'OULTREMONT, L., DENUIT, M., TRUFIN, J. and VAN OIRBEEK, R. (2020) Matrix calculation for ultimate and 1-year risk in the semi-markov individual loss reserving model. Scandinavian Actuarial Journal, 1–28.

BETTONVILLE, C., D'OULTREMONT, L., DENUIT, M., TRUFIN, J. and OIRBEEK, R.V. (2021) Matrix calculation for ultimate and 1-year risk in the Semi-Markov individual loss reserving model, *Scandinavian Actuarial Journal*, **2021**(5), 380–407.

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*, **36**(2), 521–542.

CHARPENTIER, A. and PIGEON, M. (2016) Macro vs. micro methods in non-life claims reserving (an econometric perspective). *Risks*, **4**(2), 12.

CREVECOEUR, J. and ANTONIO, K. (2019) A hierarchical reserving model for reported non-life insurance claims. arXiv preprint arXiv:1910.12692.

DELONG, L., LINDHOLM, M. and WÜTHRICH, M.V. (2020) Collective reserving using individual claims data. Available at SSRN.

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.

GABRIELLI, A., RICHMAN, R. and WÜTHRICH, M.V. (2020) Neural network embedding of the over-dispersed poisson reserving model. *Scandinavian Actuarial Journal*, **2020**(1), 1–29.

GABRIELLI, A. and WÜTHRICH, M.V. (2018) An individual claims history simulation machine. *Risks*, **6**(2), 29.

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

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

LINDHOLM, M., LINDSKOG, F. and WAHL, F. (2020) Estimation of conditional mean squared error of prediction for claims reserving. *Annals of Actuarial Science*, **14**(1), 93–128.

LINDHOLM, M. and VERRALL, R. (2020) Regression based reserving models and partial information. *Insurance: Mathematics and Economics*, **94**, 109–124.

LINDHOLM, M., VERRALL, R., WAHL, F. and ZAKRISSON, H. (2020) Machine learning, regression models, and prediction of claims reserves.

LINDHOLM, M., VERRALL, R., WAHL, F. and ZAKRISSON, H. (2020) Machine learning, regression models, and prediction of claims reserves. *Casualty Actuarial Society E-Forum*, Summer 2020, 1–47.

LOPEZ, O., MILHAUD, X. and THÉROND, P.-E. (2019) A tree-based algorithm adapted to microlevel reserving and long development claims. *Astin Bulletin*, **49**(3), 741–762.

MACIAK, M., OKHRIN, O. and PEŠTA, M. (2021) Infinitely stochastic micro reserving. *Insurance: Mathematics and Economics*, **100**, 30–58.

MARTINEZ-MIRANDA, M.D., 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*, **41**(1), 107–129.

MARTINEZ-MIRANDA, M.D., NIELSEN, J.P. and VERRALL, R.J. (2012) Double chain ladder. *Astin Bulletin*, **42**(1), 59–76.

VERRALL, R.J., NIELSEN, J.P. and JESSEN, A.H. (2010) Prediction of rbns and ibnr claims using claim amounts and claim counts. *Astin Bulletin*, **40**(2), 871–887.

WAHL, F., LINDHOLM, M. and VERRALL, R. (2019) The collective reserving model. *Insurance: Mathematics and Economics*, **87**, 34–50.

MATHIAS LINDHOLM (CORRESPONDING AUTHOR)
Division of Mathematical Statistics, Department of Mathematics,
Stockholm University
Stockholm, Sweden
e-mail: lindholm@math.su.se

HENNING ZAKRISSON
Division of Mathematical Statistics, Department of Mathematics,
Stockholm University
Stockholm, Sweden
e-mail: zakrisson@math.su.se