Mathematical Statistics
Stockholm University

# On distribution-free reserving and partial information

Mathias Lindholm and Richard Verrall

**Research Report 2019:14**

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 **2019:14**,
http://www.math.su.se

# On distribution-free reserving and partial information

Mathias Lindholm[†] and Richard Verrall[‡]

October 2019

**Abstract**

A constructive distribution-free discrete-time micro model is introduced which shares the key properties of the models introduced in Verrall et al. (2010) and Wahl et al. (2019), including multiple payments per claim and separate RBNS and IBNR reserves. The introduced model class may be thought of as a sequence of two-stage conditional linear models with general variance structure. The models are distribution-free in the sense that they only rely on assumptions made on the dependence structure and the first two moments. For this model class it is possible to explicitly compute theoretical reserve predictors as well as their process variances. Moreover, the introduced model class is possible to define in terms of differently detailed information, and depending on the information available at hand, model parameters may be fitted using least square (LS) techniques or the generalised method of moments (GMM) making all computable reserve predictors, i.e. the theoretical predictor combined with actual parameter estimators, unbiased. In particular, if one assumes that detailed data include count and payment data where one keeps track of both time of reporting and time of payments for individual claims it is possible to analytically show that the variation of the computable reserve predictors always will be larger when using parameter estimators based on reduced information. This relation can also be shown to hold for the analytical mean squared error approximation estimate introduced in Lindholm et al. (2018). Furthermore, due to the linear structure of the introduced model class it is easy to construct bootstrap algorithms based on classical Pearson residuals.

The results are illustrated in a simulation study, which, in particular, highlights that it is the covariance structure and dispersion parameter estimates when using reduced information that is the likely cause of increase of the prediction error. It is also seen that the analytical MSEP-estimates are of comparable size as to the corresponding bootstrapped ones.

*Keywords:* Distribution-free linear models; Generalised least squares; Estimation error; Prediction error

†Department of Mathematics, Stockholm University, Sweden; E-mail: lindholm@math.su.se

‡Cass Business School, City, University of London, U.K.; E-mail: R.J.Verrall@city.ac.uk

# 1 Introduction

This paper develops a distribution-free analog of the "collective reserving model" (CRM) introduced in Wahl et al. (2019). This model is closely related to (conditional) distribution-free general linear models. In particular the computable reserve predictor for both the reported but not settled (RBNS) and incurred but not reported (IBNR) reserve can be expressed as linear functions in terms of what corresponds to regression parameters. By using this representation it is possible to (essentially) translate many standard statistical results into a reserving context, which allows us to make statements concerning optimality of reserve estimators and the effects on the quality of predictions when using less detailed or aggregated information.

In Verrall et al. (2010) a micro-structure model, henceforth referred to as the "VNJ" model, which allows for aggregated data estimation was introduced. This model relies on a convenient one-payment-per-claim assumption which simplifies various calculations. In particular, the modelling approach allows for the reserve predictor to be split into an RBNS and IBNR part, and the predictors themselves allow for constructive interpretations. Moreover, the VNJ model in itself, without further (ad hoc) adjustments, allows for prediction outside of the classical claims triangle. It is, however, believed that the one-payment-per-claim assumption may deter practitioners from using the method. Still, the VNJ model served as the starting point of the much studied double chain-ladder (DCL) model, see e.g. Miranda et al. (2012).

In order to overcome the weakness with the single payment-per-claim assumption underlying the VNJ model, the collective reserving model was introduced in Wahl et al. (2019). This model is based on a discrete time Poisson process approach. The CRM has an identical (plug-in) reserve predictor as the VNJ model, but with a different variance structure.

By either using the single payment-per-claim assumption from the VNJ model or the discrete time Poisson process approach from the CRM it is possible to define detailed dynamics at the individual level, which allows for controlled aggregation to the population level, e.g. a sum of Poisson variables is still Poisson. What this means in practice is that we may define detailed dynamics, motivating aggregate level predictors based only on aggregated information, whose parameters as well may be estimated solely based on aggregated data. More specifically, estimation may be carried out using quasi-maximum likelihood (QML) theory.

In the present paper we will take a slightly different route forward starting from partial sums of payments and claims, which, as it turns out, will result in a distribution-free analog of the CRM model. For this distribution-free class of models it is possible to explicitly compute the theoretical reserve predictors and process variances with more flexibility than the VNJ and CRM models. Moreover, this class of models may be fitted using least squares (LS) techniques and/or the generalised method of moments (GMM). Consequently, most results rely on using properties concerning random vectors, which tends to be tedious, but straightforward, see e.g. (Seber and Lee, 2012, Ch. 1 – 3) for a comprehensive introduction to the most relevant results.

The approach adopted will be distribution-free in the same sense as Mack (1993), i.e. assumptions are only made w.r.t. the first two moments together with a specification of relevant dependence structures. It is, however, important to note that the *theoretical* reserve predictors will coincide with the VNJ/CRM models, but the actually used computable, so-called "plug-in", reserve predictors will *not* coincide – the VNJ/CRM fit is based on QML theory, which will produce different estimates than those obtained using LS theory. An advantage with using LS estimators is that these are explicit and may be obtained without the use of numerical procedures such as those used for (Q)ML estimation. Moreover, LS estimators are typically robust and their estimation error distributions are well studied, even for finite sample sizes, allowing for explicit expressions for their covariances.

Further, as will be shown, the expected payment and count dynamics will only be functions in terms of parameters which in a standard linear model setting corresponds to regression coefficients. As a consequence of this, and that these parameters are estimated unbiasedly using LS techniques, the expected payment and count predictors, and, more importantly, the reserve predictors will be unbiased as well. Another natural question is to consider the prediction error of these estimators, i.e. the (conditional) mean squared error of prediction (MSEP). As pointed out above, the covariance structure for LS estimators is well-known, but in order for these to be computable, there is need to estimate dispersion parameters. Given that this is possible, one can use the analytical (non-simulation based) conditional MSEP approximations discussed in Lindholm et al. (2018). Moreover, due to the introduced model class' inherent linear structure it is easy to bootstrap models of this type using the techniques discussed in e.g. England and Verrall (2002).

Another question which has received attention is the effect on reserve prediction when using reserving methods based on differently detailed information spanning from micro models such as e.g. Norberg (1993); Antonio and Plat (2014) to more classical aggregated macro models such as the classical chain-ladder type models, see e.g. Mack (1993); Renshaw and Verrall (1998). The definition of "detailed" information in the present context will correspond to having access to count and payment data where both (discrete) time points of individual reporting and payments are known, following the setup of Bühlmann et al. (1980). A nice feature with the approach is that it is possible to define models using differently detailed information, and in particular it is possible to express the reserve predictors of less detailed models as explicit linear transformations of the, in the current context, most detailed one. This makes it possible to analytically show that computable reserve predictors estimated using reduced data will produce larger variation than the corresponding un-reduced data predictor. It is even possible to show that this relation holds for the analytical MSEP-approximations from Lindholm et al. (2018).

These results are illustrated in a simulation study based on the data from Verrall et al. (2010). This study indicates that it is the change in covariance structure and the dispersion parameter estimates which is the likely cause of the increase in the prediction error when using less detailed data for estimation. Moreover, it is seen that the analytical MSEP-approximation from Lindholm et al. (2018) is of comparable size as the corresponding bootstrapped estimates.

Due to the already mentioned connections between the introduced model class and distribution-free linear models, the majority of the proofs are based on slight generalisations of well-known results, resulting in tedious, but elementary calculations. As a consequence of this, many proofs are omitted or only outlined with references to relevant literature.

The remainder of the paper is structured as follows: Section 2 defines the structure of data and introduces relevant notation. Section 3 introduces the distribution-free class of reserving models and the theoretical reserve predictors and process variances are derived in Section 4. In Section 5 estimation and properties of the estimators are discussed when varying the level of granularity in the available data. Section 6 discusses properties of the computable reserve predictor including MSEP and bootstrapping, in particular w.r.t. when data with different levels of granularity is used for estimation. The paper ends with a numerical illustration given in Section 7.

## 2 Data and notation

This section introduces the notation and relevant data used throughout this paper. The notation is in accordance with that of Wahl et al. (2019), where $m$ denotes the total number of accident years and where $d$ denotes the maximal delay with which claims payments are made after being reported. Further, let $X_{ij}$ denote the incremental aggregated claims payments during development year $j \in \{0, \dots, m + d - 1\}$ for accident year $i \in \{1, \dots, m\}$. Similarly, let $N_{ij}$ denote the incremental number of claims incurred in development year $j \in \{0, \dots, m-1\}$ from accident year $i \in \{1, \dots, m\}$, but note that this implies that no claims may be reported later than $m$ years from the date of the accident. By construction, the fully observed data can be represented in terms of the following rectangles:

$$\begin{array}{cccc} N_{10} & \dots & N_{1,m-1} \\ N_{20} & \dots & N_{2,m-1} \\ \vdots & & \vdots \\ N_{m0} & \dots & N_{m,m-1} \end{array} \qquad \begin{array}{cccc} X_{10} & \dots & X_{1,m+d-1} \\ X_{20} & \dots & X_{2,m+d-1} \\ \vdots & & \vdots \\ X_{m0} & \dots & X_{m,m+d-1} \end{array}$$

In accordance with Wahl et al. (2019), the data known "today" and the data used for prediction consists of the $X_{ij}$ and $N_{ij}$ where $(i, j) \in \mathcal{A}_0$ where

$$\mathcal{A}_0 := \{(i, j) \in \mathbb{N} \times \mathbb{N}_0 : i + j \leq m\},$$

which corresponds to the following standard, partially observed, data triangles:

$$\begin{array}{ccccc} N_{10} & N_{11} & \dots & N_{1,m-2} & N_{1,m-1} \\ N_{20} & N_{21} & \dots & N_{2,m-2} & \\ \vdots & \ddots & & & \\ N_{m0} & & & & \end{array} \qquad \begin{array}{ccccc} X_{10} & X_{11} & \dots & X_{1,m-2} & X_{1,m-1} \\ X_{20} & X_{21} & \dots & X_{2,m-2} & \\ \vdots & \ddots & & & \\ X_{m0} & & & & \end{array}$$

Another quantity which will turn out to be useful for e.g. estimation purposes, is $N_{ij}^{\text{paid}}$, which corresponds to the total number of payments in $(i, j)$, where $j \in \{0, \dots, m + d - 1\}$.

Further, as in Wahl et al. (2019), the ultimate goal with the prediction, is to fill in the lower right trapezoid of $X_{ij}$s with indices in the set

$$\mathcal{A}_0^* := \{(i,j) \in \mathbb{N} \times \mathbb{N}_0 : i \leq m,\ j \leq m-1+d,\ i+j \geq m+1\}.$$

Note that $\mathcal{A}_0^*$ allows for predictions further into the future than possible for e.g. the standard chain-ladder technique, which only allows for predictions restricted to the following set of indices

$$\tilde{\mathcal{A}}_0^* := \{(i,j) \in \mathcal{A}_0^* : j \leq m-1\}.$$

It will also be convenient to introduce notation for the filtration

$$\mathcal{N}_0 := \sigma\{N_{i,j} : (i,j) \in \mathcal{A}_0\},$$

corresponding to the information generated by the number of claims observed up until "today", together with the filtration

$$\overline{\mathcal{N}}_{i,j} := \sigma\{N_{i,k} : k \leq j\},$$

i.e. the information concerning the number of claims observed up until development period $j$ for accident year $i$. It will also be convenient to introduce

$$\overline{\mathcal{N}}_{\bullet,j}^0 := \sigma\{N_{i,k} : k \leq j, (i,k) \in \mathcal{A}_0\}.$$

Note that the above quantities do not keep track of payments w.r.t. when the corresponding claim was made. Due to this, let $X_{i,j,k}$ denote payments stemming from claims that have been reported in $(i,j)$ that are paid $k$ years later, and let $\mathcal{F}_0$ denote the filtration which summarises all available information from the $X_{i,j,k}$s and the $N_{i,j}$s known up until today. It will also be useful to introduce the following auxiliary notation: Let

$N_{i,j.k}^{\text{paid}}$ denote the number of payments from claims that have been reported at $(i,j)$ that are paid $k$ periods later,

$Z_{i,j,k,l}$ correspond to the number of payments that stem from the $l$th claim that was reported in $(i,j)$ and which were paid $k$ periods later,

$Y_{i,j,k,l}$ denote the $l$th of the $N_{i,j,k}^{\text{paid}}$ payments.

## 3 Distribution-free regression models

In Verrall et al. (2010); Wahl et al. (2019) the dynamics of individual claims is described in a discrete time framework, which allows for estimation and prediction using aggregated data consisting of claim counts and claim payments. Both of these models rely on certain distributional assumptions. We will now take a slightly different approach, still using constructive arguments, but were we only make assumptions on the first two moments of the claims dynamics, in this way mimicking

the distribution-free chain-ladder model, see Mack (1993). Similarly to Mack (1993), we will obtain a (conditional) distribution-free general linear model structure whose parameters may be fitted using least squares (LS) and/or (generalized) method of moments (MM) techniques, see e.g. Seber and Lee (2012) and Mátyás et al. (1999), respectively. Before going into details about estimation, we will derive two main conditional general linear models: One model for claim payments, conditional on observed claim counts, and one model for how claims are observed.

### 3.1 A distribution-free general linear model for payments conditional on claim counts

We will start by assuming that there exists a model for claim counts and analyse the implied payment dynamics. The following assumptions will be used:

(A1) all claims are i.i.d.

(A2) all claims are reported no later than development year $m-1$,

(A3) the payment pattern for claims is independent of time of reporting, but no payments are made later than $d$ periods since reporting, and the number of payments made with different delays for each claim is independent,

(A4) all individual payments are i.i.d. and, hence, independent of time of reporting and time of payment.

By using the notation from Section 2 allows us to introduce the following constructive representation of the main quantities of interest

$$N_{i,j,k}^{\text{paid}} = \sum_{l=0}^{N_{i,j}} Z_{i,j,k,l}, \tag{1}$$

and that

$$X_{i,j,k} = \sum_{l=0}^{N_{i,j,k}^{\text{paid}}} Y_{i,j,k,l}, \tag{2}$$

and finally

$$X_{i,j} = \sum_{k=0}^{j \wedge d} X_{i,j-k,k}, \tag{3}$$

which form a natural basis for calculation of various moments.

Before proceeding further, since claims reserving amounts to making predictions of future payments stemming from already incurred claims based on current information, it is natural to consider moment calculations w.r.t. suitable conditionings. The constructive representation of the payment dynamics given by (1) – (3) is closely connected to the models introduced in Verrall et al. (2010); Miranda et al. (2012);

Wahl et al. (2019) which suggests that conditioning ought to be carried out w.r.t. $\mathcal{N}_0$. The intuition behind using this conditioning is that we may treat already observed claim counts as measures of exposure for future payments originating from already incurred claims. Another alternative is to condition on observed claim payments, but we argue that it is primarily the *number of observed claims* that drives the future payment dynamics, rather than the *amounts previously paid* – although the latter, of course, may carry information. We may now state the following result on moments conditional on $\mathcal{N}_0$:

**Proposition 1.** *Assume that Assumption (A1) – (A4) holds, and let* $\mathbb{E}[Z_{i,j,k,l}] = \lambda_k$, *and* $\operatorname{Var}(Z_{i,j,k,l}) = \theta_k^2$, *and let* $\mathbb{E}[Y_{i,j,k,l}] = \mu$, *and* $\operatorname{Var}(Y_{i,j,k,l}) = \sigma^2$. *It then holds that*

$$\begin{aligned}
\mathbb{E}[N_{i,j,k}^{\text{paid}} \mid \mathcal{N}_0] &= \lambda_k \mathbb{E}[N_{i,j} \mid \mathcal{N}_0], \\
\mathbb{E}[X_{i,j,k} \mid \mathcal{N}_0] &= \psi_k \mathbb{E}[N_{i,j} \mid \mathcal{N}_0], \\
\mathbb{E}[X_{i,j} \mid \mathcal{N}_0] &= \sum_{k=0}^{j \wedge d} \psi_k \mathbb{E}[N_{i,j-k} \mid \mathcal{N}_0],
\end{aligned}$$

*where* $\psi_k := \mu \lambda_k$, *and that*

$$\begin{aligned}
\operatorname{Var}(N_{i,j,k}^{\text{paid}} \mid \mathcal{N}_0) &= \theta_k^2 \mathbb{E}[N_{i,j} \mid \mathcal{N}_0] + \lambda_k^2 \operatorname{Var}(N_{i,j} \mid \mathcal{N}_0), \\
\operatorname{Var}(X_{i,j,k} \mid \mathcal{N}_0) &= \nu_k^2 \mathbb{E}[N_{i,j} \mid \mathcal{N}_0] + \psi_k^2 \operatorname{Var}(N_{i,j} \mid \mathcal{N}_0), \\
\operatorname{Var}(X_{i,j} \mid \mathcal{N}_0) &= \sum_{k=0}^{j \wedge d} \nu_k^2 \mathbb{E}[N_{i,j-k} \mid \mathcal{N}_0] + \sum_{k=0}^{j \wedge d} \psi_k^2 \operatorname{Var}(N_{i,j-k} \mid \mathcal{N}_0) \\
&\quad + 2 \sum_{k<l} \psi_k \psi_l \operatorname{Cov}(N_{i,j-k}, N_{i,j-l} \mid \mathcal{N}_0),
\end{aligned}$$

*where* $\nu_k^2 := \lambda_k \sigma^2 + \mu^2 \theta_k^2$

The proof of Proposition 1 is based on standard variance decomposition and tower property arguments, faithfully following the steps of the proofs in Wahl et al. (2019), and is hence omitted.

It is clear that we need to specify a model for the $N_{i,j}$s in order for Proposition 1 to become useful in practice. Moreover, note that

(i) by replacing $\mathcal{N}_0$ in Proposition 1 with $\overline{\mathcal{N}}_{i,j}^0$ (or $\overline{\mathcal{N}}_{\bullet,j}^0$) it holds that $\mathbb{E}[N_{i,j} \mid \overline{\mathcal{N}}_{i,j}^0] = N_{i,j}$, and that $\operatorname{Var}(N_{i,j} \mid \overline{\mathcal{N}}_{i,j}^0) = 0$, which yields

$$\begin{cases}
\mathbb{E}[X_{i,j,k} \mid \overline{\mathcal{N}}_{i,j}^0] & = \psi_k N_{i,j}, \\
\operatorname{Var}(X_{i,j,k} \mid \overline{\mathcal{N}}_{i,j}^0) & = \nu_k^2 N_{i,j},
\end{cases}$$

and analogously for $X_{i,j}$, by noting that also $\operatorname{Cov}(N_{i,j-k}, N_{i,j-l} \mid \overline{\mathcal{N}}_{i,j}^0) = 0$,

(ii) the relations from (i) holds for all $(i,j) \in \mathcal{A}_0$.

Hence, the above observations allows us to introduce the following linear models, defined *conditionally* on $\overline{\mathcal{N}}_{i,j}$, which have the same first two moments as those in (i)

$$X_{i,j,k} = \psi_k N_{i,j} + \nu_k \sqrt{N_{i,j}} \delta_{i,j,k}, \tag{LM 1}$$

$$X_{i,j} = \sum_{k=0}^{j \wedge d} (\psi_k N_{i,j-k} + \nu_k \sqrt{N_{i,j-k}} \delta_{i,j-k,k}), \tag{LM 2}$$

where the $\delta_{i,j,k}$s are random variables with mean 0 and variance 1 that are independent of everything else. Thus, the (conditional) general linear models defined by (LM 1) and (LM 2) are motivated constructively based on Assumptions (A1) – (A4), together with a certain parametrisation of low order moments. Also note that, if $\theta_k^2 = \lambda_k$, then (LM 2) will have the same variance structure as the (quasi/over-dispersed) Poisson models from Verrall et al. (2010) and Wahl et al. (2019). Further, note that we introduced parameters $\psi_k$ and $\nu_k$. From now on we will focus on this more parsimonious parametrisation. This choice is not only based on reducing the number of parameters to be fitted, but we also believe that by neglecting more detailed parametric dependencies the model becomes less vulnerable to model misspecification. Moreover, note that an equivalent representation of model (LM 2) w.r.t. the first two moments is the following one:

$$X_{i,j} = \sum_{k=0}^{j \wedge d} \psi_k N_{i,j-k} + \sqrt{\sum_{k=0}^{j \wedge d} \nu_k^2 N_{i,j-k}} \delta_{i,j}, \tag{LM 2$'$}$$

where the $\delta_{i,j}$s are independent random variables with mean 0 and variance 1 that are independent of everything else. Thus, by comparing model (LM 1) and (LM 2$'$), it is clear that having access to less granular data will make it harder to estimate the individual dispersion parameters $\nu_k^2$, since we can no longer separate out the individual payment delay effects directly. That is, model (LM 2) (and model (LM 2$'$)) will not only have fewer observations, making inference less precise, but the observations themselves will be less informative, since they are sums of more granular observations. These complications will be discussed further in Section 5 and 6, also w.r.t. other types of reductions of data.

Continuing, Proposition 1 is essentially of little practical value unless we specify a model for claim counts. Further, above we gave constructive arguments based on Assumptions (A1) – (A4) for how to arrive at the (distribution-free) general linear models (LM 1) and (LM 2). In the next section we will make further assumptions on the dynamics of reporting of claims which allows us to provide a similar motivation of the distribution-free chain-ladder model for counts, which, as already commented upon above, is another general linear model.

### 3.2 A distribution-free general linear model for claim counts

When introducing the general linear models (LM 1) and (LM 2) above, we made a point of that these are *conditional* models – conditional on observed claim counts. In the current section we will provide a similarly motivated model for claim counts. In

order to do so, recall that in the standard distribution-free chain-ladder model there is no model for the first (or zeroth) development period. Moreover, based on Assumption (A2) it follows that all claims shall be reported no later than development period $m-1$, which suggests that we should include parameters $\pi_j, j = 0, \ldots, m-1$ which governs the distribution of how claims are reported, and we may assume that $\sum \pi_j = 1$. Furthermore, in order for the first development period to carry any information when conditioned upon, it is clear that there can not be independence between claims reported in different development periods. For ease of exposition, we will refer to a claim to be reported in period $j$ as being of type $j$. We propose the following conditional dynamics for the claims generating process for each accident year, given that a total of $\bar{n}_j$ claims have been reported up until development period $j$, $j = 1, \ldots, m-1$:

(A5) a claim of type $j$ is generated independently of everything else with probability $\pi_j$, $j = 0, \ldots, m-1$,

(A6) claims of *all types* are generated until the total number of claims of type $l \in \{0, \ldots, j\}$ equals $\bar{n}_j$.

Based on Assumption (A5) – (A6) it is clear that the total number of claims per accident year is random, *conditional* on what has been observed. This mechanism can be described in terms of certain Negative Multinomial distributions, see e.g. the construction described in (Sibuya et al., 1964, Sec. 3.a). That is, the vector $(U_1, \ldots, U_r) \mid U_0 = k \sim \text{NegMult}(k, \theta_1, \ldots, \theta_r)$ has probability mass function given by

$$\mathbb{P}(U_1 = u_1, \ldots, U_r = u_r \mid U_0 = k) = \binom{k-1+\sum_{i=1}^r u_i}{k-1, u_1, \ldots, u_r} (1 - \sum_{i=1}^r \theta_i)^k \prod_{i=1}^r \theta_i^{u_i}, \quad (4)$$

where $\sum_{i=0}^r \theta_i = 1$ and $\theta_i, u_i \geq 0$ and $k \geq 1$, see e.g. Sibuya et al. (1964) or (Bishop et al., 1975, Ch. 13.8) for the mean-value parametrisation. This allows us to the state the following result:

**Proposition 2.** *Let $\overline{N}_{i,j}$ denote the total number of claims reported up to and including development period $j$, and let $\overline{\pi}_j = \sum_{k=0}^j \pi_k$, with $\overline{\pi}_{m-1} = 1$, and assume that Assumption (A5) – (A6) are fulfilled. For $j = 1, \ldots, m-1$, it then holds that*

*(i)* $(N_{i,j}, \ldots, N_{i,m-1}) \mid \overline{N}_{i,j-1} \sim \text{NegMult}(\overline{N}_{i,j-1}, \pi_j, \ldots, \pi_{m-1})$,

*(ii)* $N_{i,j} \mid \overline{N}_{i,j-1} \sim \text{NegBin}(\overline{N}_{i,j-1}, \pi_j/(\overline{\pi}_j))$,

*(iii)* $\mathbb{E}[\overline{N}_{i,j} \mid \overline{N}_{i,j-1}] = \overline{N}_{i,j-1} \frac{\overline{\pi}_j}{\overline{\pi}_{j-1}}$,
$\text{Var}(\overline{N}_{i,j} \mid \overline{N}_{i,j-1}) = \frac{\pi_j \overline{\pi}_j}{\overline{\pi}_{j-1}^2} \overline{N}_{i,j-1}$.

The proof of Proposition 2 follows directly from (Sibuya et al., 1964, Eq. (2.5) – (2.7)) by noting that $\pi_j + \overline{\pi}_{j-1} = \overline{\pi}_j$. Note that Proposition 2 does *not* have a model for the first (or zeroth) development period, and also note that the negative binomial model of Verrall (2000) is retrieved.

A direct consequence of Proposition 2 is the following general linear model, conditional on $\overline{N}_{i,j-1}$

$$\overline{N}_{i,j} = \alpha_j \overline{N}_{i,j-1} + \beta_j \sqrt{\overline{N}_{i,j-1}} \epsilon_{i,j}, \tag{LM 3}$$

where $\alpha_j := \overline{\pi}_j / \overline{\pi}_{j-1}$ and $\beta_j^2 := \pi_j \overline{\pi}_j / \overline{\pi}_{j-1}^2$, and where all $\epsilon_{i,j}$ are random variables with mean 0 and variance 1 that are independent of everything else. The most noteworthy feature of the model defined by (LM 3), apart from fulfilling the moment conditions from Proposition 2(iii), is that it is a claim count analog of the distribution-free chain-ladder model, see e.g. Mack (1993). Thus, by combining models (LM 1) – (LM 3) we obtain a fully distribution-free general linear model analog of the claim and payment dynamics defined in Verrall et al. (2010); Wahl et al. (2019) – a model which can be fitted using least squares techniques, see Section 5. It is, however, clear that the model (LM 3) will produce non-integer valued $\overline{N}_{i,j}$s, but this is no problem w.r.t. estimation.

Further, from Renshaw and Verrall (1998) we know that the linear ("plug-in") predictor from the over-dispersed chain-ladder model and the distribution-free chain-ladder model coincides, which, as pointed out in Verrall et al. (2010), implies that the linear predictor of the IBNR reserve for the models in Verrall et al. (2010); Wahl et al. (2019) are the same regardless of whether the distribution-free chain-ladder model or the over-dispersed chain-ladder model is used. This even further strengthens the connection between the current general linear model approach and the models in Verrall et al. (2010); Wahl et al. (2019).

**Remark 1.** *Note that by using the general linear model* (LM 3) *we have discarded the parametric relationship between the $\pi_j$s hidden in the $\alpha_j$s and $\beta_j$s. It is, however, straightforward to deduce that e.g.*

$$\mathbb{E}[\overline{N}_{i,m-i+k} \mid \mathcal{N}_0] = \overline{N}_{i,m-i} \prod_{j=m-i+1}^{m-i+k} \alpha_j = \overline{N}_{i,m-i} \frac{\overline{\pi}_{m-i+k}}{\overline{\pi}_{m-i}},$$

*and*

$$\frac{\mathbb{E}[\overline{N}_{i,m-i+k} \mid \mathcal{N}_0]}{\mathbb{E}[\overline{N}_{i,m-1} \mid \mathcal{N}_0]} = \overline{\pi}_{m-i+k},$$

*which shows that the standard interpretation of $\alpha_j$s in* (LM 3) *actually corresponds to the correct counterpart in terms of $\pi_j$s. Moreover, one can also note that by assuming that $\mathbb{E}[N_{i,0}] := e_i \pi_0$, where $e_i$ corresponds to a suitable exposure measure, it follows that*

$$\mathbb{E}[N_{i,j}] = \mathbb{E}[\overline{N}_{i,j} - \overline{N}_{i,j-1}] = e_i \pi_j,$$

*which is equivalent to the expected value in the cross-classified over-dispersed Poisson chain-ladder model.*

**Remark 2.** *Assumptions (A5) and (A6) may feel odd – it is often assumed that reporting of claims follows a standard multinomial distribution. This assumption, however, is in the above setting only reasonable* conditional *on the total number of claims. Concerning the negative multinomial distribution given by* (4)*, it holds that*

$(U_1, \ldots, U_r) \mid U_0 = k, \sum_{i=1}^r U_i = u \sim \text{Mult}(u, \frac{\theta_1}{\sum_{i=1}^r \theta_i}, \ldots, \frac{\theta_r}{\sum_{i=1}^r \theta_i})$, *see e.g. (Sibuya et al., 1964, Eq. (2.7)), which in some sense is the best we can achieve, given that we do not have a model for* $U_0$. *One can also note that the resulting multinomial distribution is independent of* $U_0 = k$.

**Remark 3.** *Note that model* (LM 3) *does* not *take into account that* $\overline{N}_{i,j} \in \mathbb{Z}_+$. *Still, given that all* $\overline{N}_{i,j} \gg 1$, *it is clear that the difference between using* $[\overline{N}_{i,j}]$ *(i.e. the integer part of* $\overline{N}_{i,j}$*) and* $\overline{N}_{i,j}$ *is negligible. On the other hand, if it is of great importance to ascertain that the* $\overline{N}_{i,j}$ *are integer valued one can use the estimates of the* $\alpha_j$*s to obtain estimates of the* $\pi_j$*s and e.g. use the negative binomial structure from Proposition 2. From the perspective of reserve estimation this will not be a problems, since these estimates will (of course) only be expressed in terms of parameter estimates, historical observations and moments.*

In the next sub-section we continue to explore these connections by analysing the IBNR and RBNS reserve predictors, together with their variances.

# 4 Reserve predictors and process variances

When motivating the conditional linear models (LM 1) and (LM 2) the moments from Proposition 1 were used. These moments are all calculated conditional on $\mathcal{N}_0$ and essentially contains all components needed in order to fully specify the reserve predictor. Further, based on the dynamics of (LM 1) – (LM 3) it follows, in analogy with Verrall et al. (2010); Wahl et al. (2019), that

(i) the reserve predictor may be split into a reported but not settled (RBNS) and an incurred but not reported (IBNR) part,

(ii) the model provides predictions for a total of $m - 1 + d$ development periods ahead.

Rather than directly re-using Proposition 1 we will use the following representation of the payments related to the RBNS reserve for accident year $i$, denoted $R_i^{\mathcal{R}}$, and the the payments related to the IBNR reserve for accident year $i$, denoted $R_i^{\mathcal{I}}$:

$$\begin{cases} R_i^{\mathcal{R}} := & \sum_{j=0}^{m-i} \sum_{k=(m-i+1-j)\wedge d}^{d} X_{i,j,k}, \\ R_i^{\mathcal{I}} := & \sum_{j=m-i+1}^{m-1} \sum_{k=0}^{d} X_{i,j,k}. \end{cases} \tag{5}$$

That is, $R_i^{\mathcal{R}}$ only sums over column indices $j$ belonging to $\mathcal{A}_0$, and the summation over $k$ ascertains that the delay forces the payments to be paid made beyond development period $m - i$, which is the last observed period. Similarly for $R_i^{\mathcal{I}}$, where the summation over column indices $j$ belongs to $\tilde{\mathcal{A}}_0^*$, i.e. the lower right triangle, thus making sure that the payments are based on claims observed beyond development period $m - i$, and here all $k = 0, \ldots, j \wedge d$, are feasible. Also note that (5) allows for payments up until development period $m - 1 + d$. Note that the representation of the payments relating to RBNS and IBNR reserves given by (5) differs from the

one given in e.g. (Wahl et al., 2019, Eq. (6) & (7)), although both representations are equivalent. The reason for using (5) is to facilitate the derivation of the process variances stated in Proposition 3.

Before stating the results on the moments corresponding to Proposition 1 we introduce the following results on the distribution-free chain-ladder model expressed in terms of claim counts:

**Lemma 1** (Mack (1993); England and Verrall (2002))**.** *The model for claim counts given by* (LM 3) *have the following moments*

$$\begin{aligned}
\mathbb{E}[\overline{N}_{i,j} \mid \mathcal{N}_0] &= \overline{N}_{i,m-i} \prod_{l=m-i+1}^{j} \alpha_l, \quad j = m-i+1, \dots, m-1, \\
\mathbb{E}[N_{i,j} \mid \mathcal{N}_0] &= \mathbb{E}[\overline{N}_{i,j} \mid \mathcal{N}_0] - \mathbb{E}[\overline{N}_{i,j-1} \mid \mathcal{N}_0] \\
&= \overline{N}_{i,m-i}(\alpha_j - 1) \prod_{l=m-i+1}^{j-1} \alpha_l, \quad j = m-i+1, \dots, m-1, \\
\operatorname{Var}(\overline{N}_{i,m-1} \mid \mathcal{N}_0) &= \operatorname{Var}(N_{i,m-1} \mid \mathcal{N}_0) \\
&= \alpha_{m-1}^2 \operatorname{Var}(\overline{N}_{i,m-2} \mid \mathcal{N}_0) + \beta_{m-1}^2 \mathbb{E}[\overline{N}_{i,m-2} \mid \mathcal{N}_0] \\
&= \mathbb{E}[\overline{N}_{i,m-1} \mid \mathcal{N}_0]^2 \sum_{j=m-i+1}^{m-1} \frac{\beta_j^2/\alpha_j^2}{\mathbb{E}[\overline{N}_{i,j-1} \mid \mathcal{N}_0]}.
\end{aligned}$$

The proof of Lemma 1 is found in the above mentioned references.

By using the representation (5) together with Proposition 1 and Lemma 1 we may derive the following result:

**Proposition 3.** *Assume that the claim count and payment dynamics are defined by* (LM 1) – (LM 3)*. The first two moments of $R_i^{\mathcal{R}}$ and $R_i^{\mathcal{I}}$ defined by* (5) *are then given by*

$$\begin{aligned}
\mathbb{E}[R_i^{\mathcal{R}} \mid \mathcal{N}_0] &= \sum_{j=0}^{m-i} \overline{\psi}_{i,j} N_{i,j}, \\
\mathbb{E}[R_i^{\mathcal{I}} \mid \mathcal{N}_0] &= \sum_{j=m-i+1}^{m-1} \overline{\psi}_{\bullet} \mathbb{E}[N_{i,j} \mid \mathcal{N}_0],
\end{aligned}$$

*where*

$$\overline{\psi}_{i,j} := \sum_{k=(m-i+1-j)\wedge d}^{d} \psi_k, \qquad \overline{\psi}_{\bullet} := \sum_{k=0}^{d} \psi_k,$$

*and*

$$\begin{aligned}
\operatorname{Var}(R_i^{\mathcal{R}} \mid \mathcal{N}_0) &= \sum_{j=0}^{m-i} \overline{\nu}_{i,j}^2 N_{i,j}, \\
\operatorname{Var}(R_i^{\mathcal{I}} \mid \mathcal{N}_0) &= \overline{\nu}_{\bullet}^2 \mathbb{E}[\overline{N}_{i,m-1} - \overline{N}_{i,m-i} \mid \mathcal{N}_0] + \overline{\psi}_{\bullet}^2 \operatorname{Var}(\overline{N}_{i,m-1} \mid \mathcal{N}_0),
\end{aligned}$$

*where*

$$\overline{\nu}^2_{i,j} := \sum_{k=(m-i+1-j)\wedge d}^{d} \nu_k, \qquad \overline{\nu}^2_{\bullet} := \sum_{k=0}^{d} \nu_k^2,$$

*and where* $\mathbb{E}[N_{i,j} \mid \mathcal{N}_0]$ *and* $\mathrm{Var}(\overline{N}_{i,m-1} \mid \mathcal{N}_0)$ *are given in Lemma 1.*

The proof of Proposition 3 is given in the appendix.

Note that we have not stated the total process variances, but due to independence between RBNS and IBNR reserve payments, as well as independence between accident years, this is obtained by summing over all relevant components. Also note that, as mentioned above, the representation (5) is different, but equivalent, to the one in e.g. (Wahl et al., 2019, Eq. (6) & (7)). Hence, by rearranging the sums in Proposition 3 it is clear that the theoretical reserve predictors in the two models are equivalent, and the same consequently holds for the model in Verrall et al. (2010). Further, Proposition 3 is a purely theoretical result, which cannot be used in practice without replacing the unknown parameters with their corresponding estimates. That is, if we introduce

$$h_i^{\mathcal{R}}(\boldsymbol{\theta}; \mathcal{N}_0) := \mathbb{E}[R_i^{\mathcal{R}} \mid \mathcal{N}_0](\boldsymbol{\theta}), \tag{6}$$

$$h_i^{\mathcal{I}}(\boldsymbol{\theta}; \mathcal{N}_0) := \mathbb{E}[R_i^{\mathcal{I}} \mid \mathcal{N}_0](\boldsymbol{\theta}), \tag{7}$$

where $\boldsymbol{\theta} := (\boldsymbol{\alpha}, \boldsymbol{\beta}^2, \boldsymbol{\psi}, \boldsymbol{\nu}^2)$, the reserve predictors that are suggested to be used in practice are the following ones:

$$\widehat{R}_i^{\mathcal{R}} := h_i^{\mathcal{R}}(\widehat{\boldsymbol{\theta}}; \mathcal{N}_0), \tag{8}$$

$$\widehat{R}_i^{\mathcal{I}} := h_i^{\mathcal{I}}(\widehat{\boldsymbol{\theta}}; \mathcal{N}_0). \tag{9}$$

The computable reserve predictors given by (8) and (9), will be referred to as "plug-in" estimators. Further, note that (8) and (9) may be expressed according to

$$\widehat{R}_i^{\mathcal{R}} := \underline{\boldsymbol{N}}_i^{\mathcal{R}} \widehat{\boldsymbol{\psi}}, \tag{10}$$

$$\widehat{R}_i^{\mathcal{I}} := \widehat{\underline{\boldsymbol{N}}}_i^{\mathcal{I}} \widehat{\boldsymbol{\psi}}, \tag{11}$$

where

$$(\underline{\boldsymbol{N}}_i^{\mathcal{R}})_k := \sum_{j=(m-i+1-k)\vee 0}^{m-i} N_{i,j}, \tag{12}$$

$$(\widehat{\underline{\boldsymbol{N}}}_i^{\mathcal{I}})_k := \sum_{j=m-i+1}^{m-1} \mathbb{E}[N_{i,j} \mid \mathcal{N}_0](\widehat{\boldsymbol{\alpha}}), \tag{13}$$

and we may, hence, e.g. write

$$\widehat{R}_i := \widehat{R}_i^{\mathcal{R}} + \widehat{R}_i^{\mathcal{I}} = (\underline{\boldsymbol{N}}_i^{\mathcal{R}} + \widehat{\underline{\boldsymbol{N}}}_i^{\mathcal{I}})\widehat{\boldsymbol{\psi}}, \tag{14}$$

which is a representation which will prove useful in Section 6 below. Further, note that the moments from Proposition 3 are only functions of $\boldsymbol{\theta}$ and $N_{i,j}$s. Thus, assuming knowledge of the $N_{i,j}$s needed for computing the moments in Proposition 3, the only effect on the reserve predictions will be w.r.t. the level of detail in the data used for estimation of $\boldsymbol{\theta}$.

Before turning to the question of estimation, we end this section with the following remark:

**Remark 4.** *The derivation of the expressions for the expected value of* (5) *is based on standard tower property arguments. Thus, by instead considering the claim count and payment dynamics defined in terms of*

$$X_{i,j,k} = \psi_k N_{i,j-k} + \nu_k w^X_{i,j} \delta_{i,j,k}, \tag{LM 1$'$}$$

*and*

$$\overline{N}_{i,j} = \alpha_j \overline{N}_{i,j-1} + \beta_j w^N_{i,j} \epsilon_{i,j}, \tag{LM 3$'$}$$

*where $w^X_{i,j} := w^X_{i,j}(\mathcal{N}_0)$ and $w^N_{i,j} := w^N_{i,j}(\mathcal{N}_0)$ corresponds to arbitrary non-negative weight functions, only defined in terms of $N_{i,j}$s, it follows that the expressions for the expected values given in Proposition 3 still hold. This is, however, not the case with the process variances – in fact, unless the weight functions are chosen with some care the process variances are not even analytically tractable. With respect to the constructive argumentation taken in the present paper, it is also less clear how to interpret other choices of weights than the currently used "$\sqrt{N_{i,j}}$"-weights. For examples and further discussion of models of the form* (LM 3$'$) *using other choices of weights, see e.g. Portugal et al. (2018) and the references therein.*

# 5 Estimation — the effect of using data on different levels of granularity

To start off, note that all (conditional) models of payment dynamics discussed in the previous sections can be expressed according to

$$\boldsymbol{X} = \boldsymbol{N}\boldsymbol{\psi} + \boldsymbol{\delta}, \tag{LM-$\boldsymbol{X}$}$$

where $\boldsymbol{X}$ is an $n \times 1$ vector, $\boldsymbol{N}$ is an $n \times (d+1)$ matrix with known covariates, $\boldsymbol{\psi}$ is a $(d+1) \times 1$ vector with unknown parameters, and $\boldsymbol{\delta}$ is a random $n \times 1$ vector with with mean $\mathbf{0}$ and a covariance matrix depending on both unknown parameters $\boldsymbol{\nu}^2$ and (functions of) $\boldsymbol{N}$. Without specifying the covariance structure any further, it is still possible to use the ordinary least squares (OLS) technique, see e.g. (Seber and Lee, 2012, Ch. 3.10), to obtain the estimator

$$\widehat{\boldsymbol{\psi}} = (\boldsymbol{N}'\boldsymbol{N})^{-1}\boldsymbol{N}'\boldsymbol{X} =: \boldsymbol{N}^{+}\boldsymbol{X}, \tag{15}$$

which satisfies

$$\mathbb{E}[\widehat{\boldsymbol{\psi}} \mid \boldsymbol{N}] = \mathbb{E}[\widehat{\boldsymbol{\psi}} \mid \mathcal{N}_0] = \boldsymbol{\psi}. \tag{16}$$

Here one can note, that the corresponding estimators of $\boldsymbol{\psi}$ in Verrall et al. (2010); Wahl et al. (2019) are *not* unbiased, since the parameter estimators are obtained using the QML technique. Also note that the unbiasedness (16) holds when using general weight functions, although other choices of weight functions tend to lack constructive interpretation. In Section 5.1 we continue the discussion for weighted least squares estimators, which also provide unbiased estimators.

If we turn to the count models, similarly, these can be expressed as a sequence of conditional linear models

$$\overline{\boldsymbol{N}}_j = \overline{\boldsymbol{N}}_{j-1}\alpha_j + \boldsymbol{\delta}_j, \tag{LM-$\overline{\boldsymbol{N}}$}$$

where $j = 1, \ldots, m-1$, and where $\overline{\boldsymbol{N}}_j$ and $\overline{\boldsymbol{N}}_{j-1}$ are $n_j \times 1$ vectors, and where $\boldsymbol{\delta}_j$ is a random vector with mean $\mathbf{0}$ and a covariance matrix depending on $\beta_j^2$ and (functions of) $\overline{\boldsymbol{N}}_{j-1}$. Note that the sequential structure of model (LM-$\overline{\boldsymbol{N}}$) directly implies that the regression coefficient estimators are (conditionally) uncorrelated:

**Lemma 2.** *Let $\widehat{\alpha}_j$ denote the OLS estimator of $\alpha_j, j = 1, \ldots, m-1$, from model* (LM-$\overline{\boldsymbol{N}}$)*. It then holds that*

$$\mathbb{E}[\widehat{\alpha}_k\widehat{\alpha}_j \mid \overline{\mathcal{N}}^0_{\bullet,k-1}] = \alpha_k\widehat{\alpha}_j, \quad j \leq k,$$

*and that*

$$\mathbb{E}[\widehat{\alpha}_k\widehat{\alpha}_j] = \alpha_k\alpha_j.$$

The proof of Lemma 2 follows directly from the tower property, see e.g. Mack (1993) for additional details.

We will now continue to analyse the effect of estimation when using data on different levels of detail, including estimation of dispersion parameters.

### 5.1 Estimation using $X_{i,j,k}$s

Similarly to the count model situation, when we have access to $X_{i,j,k}$-level data we can use the following representation of model (LM-$\boldsymbol{X}$): Let $\boldsymbol{X}_k$ and $\boldsymbol{N}_k$ be defined as follows

$$\boldsymbol{X}_k = \begin{pmatrix} X_{1,0,k} \\ X_{1,1,k} \\ \vdots \\ X_{1,m-1,k} \\ X_{2,0,k} \\ \vdots \\ X_{k,0,m-k} \end{pmatrix}, \quad \boldsymbol{N}_k = \begin{pmatrix} N_{1,0} \\ N_{1,1} \\ \vdots \\ N_{1,m-1} \\ N_{2,0} \\ \vdots \\ N_{k,0} \end{pmatrix},$$

and let $\boldsymbol{w}_k$ denote a weight vector consisting of functions of $\boldsymbol{N}_k$ and let $\boldsymbol{\delta}_k$ be a random vector with independent elements with mean 0 and variance 1 analogously indexed as $\boldsymbol{N}_k$ and $\boldsymbol{X}_k$. Thus, by letting

$$\boldsymbol{W}_k = \text{diag}(\boldsymbol{w}_k),$$

the joint regression model for the $X_{i,j,k}$s in $\boldsymbol{X}_k$ may be expressed as

$$\boldsymbol{X}_k = \boldsymbol{N}_k\psi_k + \nu_k\boldsymbol{W}_k\boldsymbol{\delta}_k, \tag{17}$$

which results in the following weighted least squares (WLS) estimators, see e.g. (Seber and Lee, 2012, Ch. 3.10):

$$\begin{cases} \widehat{\psi}_k &= (\boldsymbol{N}_k'\boldsymbol{\Sigma}_k^{-1}\boldsymbol{N}_k)^{-1}\boldsymbol{N}_k'\boldsymbol{\Sigma}_k^{-1}\boldsymbol{X}_k, \\ \widehat{\nu}_k^2 &= \frac{1}{|\boldsymbol{X}_k|-1}\widehat{\boldsymbol{e}}_k'\boldsymbol{\Sigma}_k^{-1}\widehat{\boldsymbol{e}}_k, \end{cases} \tag{18}$$

where

$$\boldsymbol{\Sigma}_k = \boldsymbol{W}_k'\boldsymbol{W}_k,$$
$$\widehat{\boldsymbol{e}}_k = \boldsymbol{X}_k - \boldsymbol{N}_k\widehat{\psi}_k.$$

For WLS estimators it is well-known, see e.g. Fahrmeir et al. (2007), that

(i) $\mathbb{E}[\widehat{\psi}_k \mid \mathcal{N}_0] = \psi_k$,

(ii) $\mathbb{E}[\widehat{\nu}_k^2 \mid \mathcal{N}_0] = \nu_k^2$,

(iii) $\text{Var}(\widehat{\psi}_k \mid \mathcal{N}_0) = \nu_k^2(\boldsymbol{N}_k'\boldsymbol{\Sigma}_k^{-1}\boldsymbol{N}_k)^{-1}$.

In particular, when using weights $\boldsymbol{w}_k$ corresponding to the elementwise square-root of $\boldsymbol{N}_k$, the estimators from (18) simplifies according to

$$\begin{cases} \widehat{\psi}_k &= \frac{\sum_{i=1}^{k}\sum_{j=0}^{m-i-k} X_{i,j,k}}{\sum_{i=1}^{k}\sum_{j=0}^{m-i-k} N_{i,j}}, \\ \widehat{\nu}_k^2 &= \frac{1}{|\boldsymbol{X}_k|-1}\sum_{i=1}^{k}\sum_{j=0}^{m-i-k} N_{i,j}(\frac{X_{i,j,k}}{N_{i,j}} - \widehat{\psi}_k)^2, \end{cases} \tag{19}$$

and that

$$\text{Var}(\widehat{\psi}_k \mid \mathcal{N}_0) = \nu_k^2 / \sum_{i=1}^{k}\sum_{j=0}^{m-i-k} N_{i,j},$$

which are easily implemented in a spreadsheet. Also not the close resemblance between the WLS-estimators and the classical chain-ladder technique estimators, see e.g. Mack (1993), but where $\widehat{\psi}_k$ now corresponds to the average payment per claim. From (19) it is also clear that if all $N_{i,j}, X_{i,j,k} \geq 0$ implies that all $\psi_k \geq 0$.

Furthermore, note that, since

$$\mathbb{E}[X_{i,j,k}X_{u,v,l} \mid \mathcal{N}_0] = \mathbb{E}[X_{i,j,k} \mid \mathcal{N}_0]\mathbb{E}[X_{u,v,l} \mid \mathcal{N}_0]$$

it follows that

$$\mathbb{E}[\widehat{\psi}_k\widehat{\psi}_j \mid \mathcal{N}_0] = \psi_k\psi_j, \quad j \leq k,$$

i.e. $\widehat{\psi}_k$ and $\widehat{\psi}_j$ are conditionally uncorrelated.

**Remark 5.** *Note that the count model defined by model* (LM 3) *may be written on the form given by* (17)*. Thus, all results on estimation for* (17) *translates directly to the model* (LM 3)*.*

When turning to the situation with estimation based on less detailed data, the following vector representation of the detailed data model will turn out to be useful:

$$\boldsymbol{X} = \boldsymbol{N}\boldsymbol{\psi} + \boldsymbol{W}\boldsymbol{\Xi}^{1/2}\boldsymbol{\delta}, \tag{20}$$

where

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_0 \\ \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_d \end{pmatrix}, \ \boldsymbol{N} = \begin{pmatrix} \boldsymbol{N}_0 & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{N}_1 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots & \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{N}_d \end{pmatrix},$$

and

$$\boldsymbol{\Xi} = \begin{pmatrix} \nu_0^2 \operatorname{diag}(\boldsymbol{1}_0) & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \nu_1^2 \operatorname{diag}(\boldsymbol{1}_1) & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots & \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \nu_d^2 \operatorname{diag}(\boldsymbol{1}_d) \end{pmatrix},$$

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{W}_0 & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W}_1 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots & \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{W}_d \end{pmatrix},$$

where $\boldsymbol{1}_k$ is a vector of ones with $\dim(\boldsymbol{1}_k) = \dim(\boldsymbol{N}_k)$. Moreover, by using the representation (20) it follows that

$$\widehat{\boldsymbol{\psi}} = (\boldsymbol{N}'\boldsymbol{\Sigma}^{-1}\boldsymbol{N})^{-1}\boldsymbol{N}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}, \tag{21}$$

where $\boldsymbol{\Sigma} = \boldsymbol{W}'\boldsymbol{W}$, with

$$\begin{aligned} \operatorname{Cov}(\widehat{\boldsymbol{\psi}} \mid \mathcal{N}_0) &= (\boldsymbol{N}'\boldsymbol{\Sigma}^{-1}\boldsymbol{N})^{-1}\boldsymbol{N}'\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\boldsymbol{\Xi}\boldsymbol{W}'\boldsymbol{\Sigma}^{-1}\boldsymbol{N}(\boldsymbol{N}'\boldsymbol{\Sigma}^{-1}\boldsymbol{N})^{-1} \\ &= (\boldsymbol{N}'\boldsymbol{\Sigma}^{-1}\boldsymbol{N})^{-1}\boldsymbol{N}'\boldsymbol{\Xi}\boldsymbol{\Sigma}^{-1}\boldsymbol{N}(\boldsymbol{N}'\boldsymbol{\Sigma}^{-1}\boldsymbol{N})^{-1}, \end{aligned} \tag{22}$$

where the last equality follows due to that both $\boldsymbol{W}$ and $\boldsymbol{\Xi}$ are diagonal, i.e. $\boldsymbol{W}' = \boldsymbol{W}$ and $\boldsymbol{\Xi}\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Xi}$. Furthermore, due to construction, $\operatorname{Cov}(\widehat{\boldsymbol{\psi}} \mid \mathcal{N}_0)$ from (22) is diagonal, and it holds that

$$(\boldsymbol{N}'\boldsymbol{\Xi}\boldsymbol{\Sigma}^{-1}\boldsymbol{N})_{kk} = \nu_k^2(\boldsymbol{N}'\boldsymbol{\Sigma}^{-1}\boldsymbol{N})_{kk},$$

which gives us that $\operatorname{Cov}(\widehat{\boldsymbol{\psi}} \mid \mathcal{N}_0)_{kk} = \operatorname{Var}(\widehat{\psi}_k \mid \mathcal{N}_0) = \nu_k^2(\boldsymbol{N}_k'\boldsymbol{\Sigma}_k^{-1}\boldsymbol{N}_k)^{-1}$ as it should.

### 5.2 Estimation using $X_{i,j}$s

The natural starting point for discussing reduced data estimation is to consider the model formulation given by (LM 2′) and let $\boldsymbol{X}^{\text{det}}$ denote the vector of $X_{i,j,k}$s from (20) and let $\boldsymbol{X}^{\text{agg}}$ denote the vector of stacked $X_{i,j}$s, and analogously for all other quantities referred to as "det" or "agg". It then holds that there exists a $|\boldsymbol{X}^{\text{agg}}| \times |\boldsymbol{X}^{\text{det}}|$ "selection" matrix $\boldsymbol{S}$ only consisting of zeros and ones such that

$$\boldsymbol{X}^{\text{agg}} = \boldsymbol{S}\boldsymbol{X}^{\text{det}}, \tag{23}$$

which corresponds to that

$$\boldsymbol{X}^{\text{agg}} = \boldsymbol{S}\boldsymbol{N}^{\text{det}} + \boldsymbol{S}\boldsymbol{W}\boldsymbol{\Xi}^{1/2}\boldsymbol{\delta}^{\text{det}}.$$

Thus, w.r.t. to equivalence of the first two moments, (LM 2′) (and (LM 2)) may be represented as

$$\boldsymbol{X}^{\text{agg}} = \boldsymbol{N}^{\text{agg}}\boldsymbol{\psi} + \widetilde{\boldsymbol{W}}\boldsymbol{\delta}^{\text{agg}}, \tag{24}$$

where $\boldsymbol{N}^{\text{agg}} := \boldsymbol{S}\boldsymbol{N}^{\text{det}}$ and $\widetilde{\boldsymbol{W}} := (\boldsymbol{S}\boldsymbol{\Xi}(\boldsymbol{W})^2\boldsymbol{S}')^{1/2} = \text{Cov}(\boldsymbol{X}^{\text{agg}} \mid \mathcal{N}_0)^{1/2}$, which is well-defined since it corresponds to the element-wise square-root of a diagonal matrix, and where $\boldsymbol{\delta}^{\text{agg}}$ is a $|\boldsymbol{X}^{\text{det}}| \times 1$ random vector with independent components having mean 0 and variance 1. From this definition it is clear that if one assumes $\nu_k = \nu$ for all $k$ the WLS technique may be used since (24) reduces to

$$\boldsymbol{X}^{\text{agg}} = \boldsymbol{N}^{\text{agg}}\boldsymbol{\psi} + \nu(\boldsymbol{S}(\boldsymbol{W})^2\boldsymbol{S}')^{1/2}\boldsymbol{\delta}^{\text{agg}},$$

and the $\boldsymbol{\psi}$ estimator for this model is again given by (18). In this situation, however, there is nothing that guarantees that the individual regression coefficients are uncorrelated.

Further, in the situation where we have dispersion parameters $\nu_k$ that are specific to different payment delay periods, it is clear that the model (24) is only a more detailed description of the model (LM-$\boldsymbol{X}$), and it follows that $\boldsymbol{\psi}$ may be estimated using the OLS estimator given by (15). This estimator is equivalent to a (generalized) method of moments, (G)MM, estimator, see e.g. (Mátyás et al., 1999, Ch. 1), and by using the GMM-technique it is also possible to estimate the dispersion parameters. By letting $\boldsymbol{N}^{\text{agg}}_{i,\bullet}$ denote the $i$th row in $\boldsymbol{N}^{\text{agg}}$, it is possible to state the following proposition:

**Proposition 4.** *The parameters in the linear model* (24) *may be estimated using*

$$\widehat{\boldsymbol{\psi}} = (\boldsymbol{N}^{\text{agg}})^{+}\boldsymbol{X}^{\text{agg}},$$

*and*

$$\widehat{\boldsymbol{\nu}}^2 = (\boldsymbol{N}^{\text{agg}})^{+}\widehat{\boldsymbol{u}}$$

*where*

$$\widehat{\boldsymbol{u}}_i := (\boldsymbol{N}^{\text{agg}}_{i,\bullet}\widehat{\boldsymbol{\psi}} - \boldsymbol{X}^{\text{agg}}_i)^2.$$

*Further, it also holds that*

$$\begin{aligned}\mathbb{E}[\widehat{\boldsymbol{\psi}} \mid \mathcal{N}_0] &= \boldsymbol{\psi},\\ \mathbb{E}[\widehat{\boldsymbol{\nu}}^2 \mid \mathcal{N}_0] &\neq \boldsymbol{\nu}^2,\\ \mathrm{Cov}(\widehat{\boldsymbol{\psi}} \mid \mathcal{N}_0) &= (\boldsymbol{N}^{\mathrm{agg}})^{+}\widetilde{\boldsymbol{W}}^2((\boldsymbol{N}^{\mathrm{agg}})^{+})'.\end{aligned}$$

The proof of Proposition 4 is given in the appendix.

Note that Proposition 4 may be seen as a two-stage regression, where one in a first step regresses $\boldsymbol{X}$ on $\boldsymbol{N}$, and then, in a second step, regresses the squared residuals $\boldsymbol{U}$ on $\boldsymbol{N}$. Moreover, from Proposition 4 we see that the suggested estimator for the vector with dispersion parameters is biased. It is, however, not possible to make statements about the asymptotic efficiency of the estimator, without making further specifications of higher moments. We will not go into more details about this, but refer the reader to the discussion of asymptotic efficiency of (G)MM estimators found in e.g. (Mátyás et al., 1999, Ch. 1), but see also White (1980). Still, as will be seen in Section 6, Proposition 4 contains all relevant building blocks in order to compute an analytical approximation of the conditional MSEP for the claims reserve using the techniques from Lindholm et al. (2018). We end this section with the following remarks.

**Remark 6.** *Compared with the results on estimation of $\boldsymbol{\psi}$ based on $X_{i,j,k}$-data, from Proposition 4 it follows that there is no guarantee that $\widehat{\psi}_k \geq 0$, even though all $X_{i,j}, N_{i,j} \geq 0$, or that the regression coefficients are uncorrelated. Moreover, the estimator of $\boldsymbol{\nu}^2$ is also of a least squares type, hence being unconstrained, meaning that individual $\widehat{\nu}_k^2$ could become negative.*

**Remark 7.** *One can also note that the cause of complication when estimating the dispersion parameters $\boldsymbol{\nu}^2$ from* (24)*, leading up to Proposition 4, is that the diagonal elements of $\widetilde{\boldsymbol{W}}$ are* mixtures *of $\nu_k^2$s and $N_{i,j}$s. If we assume that all diagonal elements of $\widetilde{\boldsymbol{W}}$ are distinct, which is a reasonable assumption, given the accident year specific $N_{i,j}$ weights, an alternative would be to first directly estimate the covariance of $\widehat{\boldsymbol{\psi}}$ without making any parametric assumptions, and based on this estimate, in a second step, estimate $\boldsymbol{\nu}^2$. This corresponds to solving the following equation w.r.t. $\widehat{\boldsymbol{\nu}}^2$*

$$\begin{aligned}\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\psi}} \mid \mathcal{N}_0) &= (\boldsymbol{N}^{\mathrm{agg}})^{+}\widehat{\widetilde{\boldsymbol{W}}}^2((\boldsymbol{N}^{\mathrm{agg}})^{+})'\\ &= (\boldsymbol{N}^{\mathrm{agg}})^{+}\,\mathrm{diag}((\boldsymbol{N}^{\mathrm{agg}})\widehat{\boldsymbol{\nu}}^2)((\boldsymbol{N}^{\mathrm{agg}})^{+})',\end{aligned}$$

*which is equivalent to solving*

$$\mathrm{diag}((\boldsymbol{N}^{\mathrm{agg}})\widehat{\boldsymbol{\nu}}^2) = \widehat{\widetilde{\boldsymbol{W}}}^2.$$

*Further, a famous consistent estimator of* $\mathrm{Cov}(\widehat{\boldsymbol{\psi}} \mid \mathcal{N}_0)$ *is the White-estimator, see White (1980):*

$$\widehat{\mathrm{Cov}}_{White}(\widehat{\boldsymbol{\psi}} \mid \mathcal{N}_0) = (\boldsymbol{N}^{\mathrm{agg}})^{+}\,\mathrm{diag}(\widehat{\boldsymbol{u}})((\boldsymbol{N}^{\mathrm{agg}})^{+})',$$

*which immediately gives us that $\widehat{\boldsymbol{\nu}}^2_{White} = (\boldsymbol{N}^{\mathrm{agg}})^{+}\widehat{\boldsymbol{u}}$, which coincides with the estimator from Proposition 4.*

**Remark 8.** *Note that the selection matrix* $\boldsymbol{S}$ *from* (23)*, which only consists of zeros and ones, may be replaced by an arbitrary transformation matrix* $\widetilde{\boldsymbol{S}}$*, hence corresponding to an arbitrary data reduction. In the present paper we primarily focus on* $\boldsymbol{S}$ *which is a selection matrix corresponding to the mapping from* $X_{i,j,k}$*-level data to* $X_{i,j}$*-level data because of the natural interpretation of this particular reduction of data. In Section 6 results are derived which holds for arbitrary transformations* $\widetilde{\boldsymbol{S}}$ *as long as the transformation preserves unbiasedness of the resulting estimator of* $\boldsymbol{\psi}$*. Concerning estimation using arbitrary* $\widetilde{\boldsymbol{S}}$ *it is clear from the discussion above that a standard OLS-approach is always feasible.*

# 6 Properties of the computable reserve predictors and the mean squared error of prediction

One of the main purposes with the current paper is to illustrate the effect of using data on different levels of granularity. In Section 5 some aspects of this has already been covered w.r.t. estimation. We will now continue to illustrate how the suggested estimators introduced in Section 5 will affect the corresponding reserve predictors.

To start off, note that the *theoretical* reserve predictors, e.g. $h_i^{\mathcal{R}}(\boldsymbol{\theta};\mathcal{N}_0)$ and $h_i^{\mathcal{I}}(\boldsymbol{\theta};\mathcal{N}_0)$ from (6) and (7), respectively, are, by definition, $\mathcal{N}_0$-measurable functions, whereas the corresponding *computable* reserve predictors $\widehat{R}_i^{\mathcal{R}}$ and $\widehat{R}_i^{\mathcal{I}}$ from (8) and (9), respectively, are $\mathcal{F}_0$-measurable, where $\mathcal{N}_0 \subset \mathcal{F}_0$. Further, by using the computable reserve representation from (10) and (11), i.e.

$$\widehat{R}_i^{\mathcal{R}} := \underline{\boldsymbol{N}}_i^{\mathcal{R}}\widehat{\boldsymbol{\psi}},$$
$$\widehat{R}_i^{\mathcal{I}} := \widehat{\underline{\boldsymbol{N}}}_i^{\mathcal{I}}\widehat{\boldsymbol{\psi}},$$

it is possible to formulate the following general result:

**Proposition 5.** *Assume that* $\widehat{\underline{\boldsymbol{N}}}_i^{\mathcal{I}}$ *is* $\mathcal{N}_0$*-measurable, and* $\widehat{\boldsymbol{\psi}}$ *is not* $\mathcal{N}_0$*-measurable, and that*

$$\mathbb{E}[\widehat{\underline{\boldsymbol{N}}}_i^{\mathcal{I}}] = \mathbb{E}[\underline{\boldsymbol{N}}_i^{\mathcal{I}}], \text{ and } \mathbb{E}[\widehat{\boldsymbol{\psi}} \mid \mathcal{N}_0] = \boldsymbol{\psi},$$

*it holds that*

$$\mathbb{E}[\widehat{R}_i \mid \mathcal{N}_0] = \mathbb{E}[\widehat{R}_i^{\mathcal{R}} + \widehat{R}_i^{\mathcal{I}} \mid \mathcal{N}_0] = (\underline{\boldsymbol{N}}_i^{\mathcal{R}} + \widehat{\underline{\boldsymbol{N}}}_i^{\mathcal{I}})\boldsymbol{\psi},$$

*and that*

$$\mathbb{E}[\widehat{R}_i] = (\underline{\boldsymbol{N}}_i^{\mathcal{R}} + \mathbb{E}[\underline{\boldsymbol{N}}_i^{\mathcal{I}}])\boldsymbol{\psi}.$$

Note that Proposition 5 is a result concerning random vectors. The proof is omitted, but can be found in e.g. (Seber and Lee, 2012, Ch. 3). Moreover, note that Proposition 5 implies that $\widehat{R}_i$ is an unbiased estimator of $\mathbb{E}[R_i \mid \mathcal{N}_0]$, and that $\widehat{\underline{\boldsymbol{N}}}_i^{\mathcal{I}}$ is an unbiased estimator of $\mathbb{E}[\underline{\boldsymbol{N}}_i^{\mathcal{I}} \mid \mathcal{N}_0]$. Concerning the particular estimators from

Section 5, it is known from Mack (1993) that $\mathbb{E}[\widehat{\underline{\boldsymbol{N}}}_i^{\mathcal{I}}] = \mathbb{E}[\underline{\boldsymbol{N}}_i^{\mathcal{I}}]$, and all LS-type estimators of $\boldsymbol{\psi}$ discussed in Section 5 are unbiased, consequently Proposition 5 applies. Another question of interest is to consider the actual values of the computable reserve predictors when using different estimators of $\boldsymbol{\psi}$. In this situation it is not possible to order the estimators, since if $X_{i,j,k}$-level data is used it holds that all $\psi_k \geq 0$ if all $X_{i,j,k}, N_{i,j} \geq 0$, something which cannot be ascertained in the corresponding situation when using $X_{i,j}$-level data, see Remark 6.

When turning to the variation in the computable reserve estimators, one may note that all unbiased estimators of $\boldsymbol{\psi}$ need to be projections of one another, see e.g. (Seber and Lee, 2012, Ch. 3). In particular, this allows us to formulate the following result

**Proposition 6.** *Amongst all estimators of $\boldsymbol{\psi}$ from the model* (20) *the WLS-estimators from* (21) *minimises the variance of*

$$\widehat{R} := \sum_{i=1}^{m} \widehat{R}_i,$$

*conditional on $\mathcal{N}_0$, whose variance is given by*

$$\operatorname{Var}(\widehat{R} \mid \mathcal{N}_0) = \left( \sum_{i=1}^{m} (\underline{\boldsymbol{N}}_i^{\mathcal{R}} + \widehat{\underline{\boldsymbol{N}}}_i^{\mathcal{I}}) \right)' \operatorname{Cov}(\widehat{\boldsymbol{\psi}} \mid \mathcal{N}_0) \left( \sum_{i=1}^{m} (\underline{\boldsymbol{N}}_i^{\mathcal{R}} + \widehat{\underline{\boldsymbol{N}}}_i^{\mathcal{I}}) \right),$$

*with $\operatorname{Cov}(\widehat{\boldsymbol{\psi}} \mid \mathcal{N}_0)$ from* (22)*. Further, let $\widehat{\widetilde{\boldsymbol{\psi}}}$ denote any other unbiased estimator of $\boldsymbol{\psi}$, and let $\widehat{\widetilde{R}}$ denote the corresponding computable reserve predictor. It then holds that*

$$\begin{aligned}
&\operatorname{Var}(\widehat{\widetilde{R}}) - \operatorname{Var}(\widehat{R}) \\
&\quad = \mathbb{E}\left[ \left( \sum_{i=1}^{m} (\underline{\boldsymbol{N}}_i^{\mathcal{R}} + \widehat{\underline{\boldsymbol{N}}}_i^{\mathcal{I}}) \right)' (\operatorname{Cov}(\widehat{\widetilde{\boldsymbol{\psi}}} \mid \mathcal{N}_0) - \operatorname{Cov}(\widehat{\boldsymbol{\psi}} \mid \mathcal{N}_0)) \left( \sum_{i=1}^{m} (\underline{\boldsymbol{N}}_i^{\mathcal{R}} + \widehat{\underline{\boldsymbol{N}}}_i^{\mathcal{I}}) \right) \right] \\
&\quad \geq 0.
\end{aligned}$$

*These relations also hold for the RBNS and IBNR reserves separately.*

Note that Proposition 6 merely states that the detailed level estimator is "BLUE" (best linear unbiased estimator), and the proof follows the usual steps for linear models: The part of the proof concerning $\operatorname{Var}(\widehat{R} \mid \mathcal{N}_0)$ follows from a slight generalisation of Theorem 3.2 in Seber and Lee (2012) re-using the arguments from (Seber and Lee, 2012, Ch. 3.10), by noting that $\Xi$ from (20) is diagonal with constant $\nu_k^2$ for all rows of $\boldsymbol{N}$ corresponding to $\boldsymbol{N}_k$, which is the same type of argument used for the motivation of (22). You may also consult the proof of Proposition 4 for more details on the linear algebra being used. The unconditional variance result then follows from that $\operatorname{Var}(\widehat{R} \mid \mathcal{N}_0) \leq \operatorname{Var}(\widehat{\widetilde{R}} \mid \mathcal{N}_0)$ a.s., together with that $\operatorname{Var}(\mathbb{E}[\widehat{\widetilde{R}} \mid \mathcal{N}_0]) = \operatorname{Var}(\mathbb{E}[\widehat{R} \mid \mathcal{N}_0])$.

Note that Proposition 6 tells us that all unbiased estimators of $\boldsymbol{\psi}$ other than (21) will have a higher variance in the reserve predictor, hence covering arbitrary reduced

data estimators as discussed in Remark 8. This result, hence, does not make any distinction w.r.t. *how* a candidate estimator to (21) is obtained; it only tells us that every other is worse in this respect. This may perhaps seem somewhat disappointing, since it does not give any guidance to whether e.g. a standard OLS-estimator of $\boldsymbol{\psi}$ defined in terms of $X_{i,j,k}$-level data will outperform an aggregated dito or not. Still, it is not surprising that it is hard to say more, since one possible detailed estimator is to e.g. neglect all observations but one for each $\psi_k$, which still is an unbiased estimator defined on a detailed level, but likely with a very large variance. Moreover, note that the unconditional result only tells us that the variance of any other unbiased estimator will be greater, but in general it is not possible to analytically compute what the actual variances are. Furthermore, when using arbitrary forms of data reductions it is neither clear how to estimate the corresponding dispersion parameters.

Continuing, Proposition 6 tells us about the variation in the computable predictors in terms of relations which again are functions of the unknown parameters $\boldsymbol{\theta}$. Another question of interest is to try to quantify the variation in the computable reserve predictors w.r.t. the true reserve, which can be done using the (conditional) mean squared error of prediction (MSEP). The conditional MSEP w.r.t. the computable reserve predictor is defined according to

$$\mathrm{MSEP}_{\mathcal{F}_0}(R, \widehat{R}) := \mathbb{E}[(R - \widehat{R})^2 \mid \mathcal{F}_0], \tag{25}$$

see e.g. Mack (1993), which in the current context, with parameters $\boldsymbol{\theta} := (\boldsymbol{\alpha}, \boldsymbol{\beta}^2, \boldsymbol{\psi}, \boldsymbol{\nu}^2)$, is equivalent to

$$\begin{aligned} \mathrm{MSEP}_{\mathcal{F}_0}(R, \widehat{R}) &:= \mathbb{E}[(R - h(\widehat{\boldsymbol{\theta}}; \mathcal{N}_0))^2 \mid \mathcal{F}_0] \\ &= \mathbb{E}[(R - h((\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\psi}}); \mathcal{N}_0))^2 \mid \mathcal{F}_0], \end{aligned} \tag{26}$$

where $\mathcal{N}_0$ ($\subset \mathcal{F}_0$) corresponds to the so-called basis of prediction, see Lindholm et al. (2018), and where $\widehat{\boldsymbol{\theta}}$ is some $\mathcal{F}_0$-measurable estimator. By using the analytical MSEP-approximation suggested in (Lindholm et al., 2018, Eq. (6)) it follows that (26) may be approximated according to

$$\mathrm{MSEP}^*_{\mathcal{F}_0}(R, \widehat{R}) := \mathrm{Var}(R \mid \mathcal{N}_0)(\boldsymbol{\theta}) + \mathrm{err}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)(\boldsymbol{\theta}), \tag{27}$$

where

$$\mathrm{err}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)(\boldsymbol{\theta}) := \nabla h((\boldsymbol{\alpha}, \boldsymbol{\psi}); \mathcal{N}_0)' \Lambda((\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\psi}}); \mathcal{F}_0)(\boldsymbol{\theta}) \nabla h((\boldsymbol{\alpha}, \boldsymbol{\psi}); \mathcal{N}_0) \tag{28}$$

which is based on a certain independent copy/re-sampling argument for the $\mathcal{F}_0$-measurable estimator $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\psi}})$, together with a Taylor expansion of $h(\boldsymbol{z}; \mathcal{N}_0)$ around $\boldsymbol{z} = \boldsymbol{\theta}$, resulting in $\nabla h((\boldsymbol{\alpha}, \boldsymbol{\psi}); \mathcal{F}_0)$, which corresponds to the column vector with partial derivatives of $h(\boldsymbol{z}; \mathcal{N}_0)$ evaluated at $\boldsymbol{z} = (\boldsymbol{\alpha}, \boldsymbol{\psi})$, and where $\Lambda((\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\psi}}); \mathcal{F}_0)$ is a function (to be specified) relating to the covariance of $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\psi}})$. The form of (27) relies on that $\mathbb{E}[\widehat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$, which holds for the estimators studied in the present paper. For more on when this is not fulfilled, see e.g. Lindholm et al. (2018); Wahl et al. (2019). Further, the MSEP-approximation given by (27) is a theoretical quantity,

and the natural computable approximation is given by

$$\begin{aligned}\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(R,\widehat{R}) &= \mathrm{MSEP}_{\mathcal{F}_0}(R,\widehat{R})(\widehat{\boldsymbol{\theta}}) \\ &:= \mathrm{Var}(R \mid \mathcal{N}_0)(\widehat{\boldsymbol{\theta}}) \\ &\quad + \nabla h((\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\psi}});\mathcal{N}_0)'\Lambda((\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\psi}});\mathcal{F}_0)(\widehat{\boldsymbol{\theta}})\nabla h((\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\psi}});\mathcal{N}_0), \qquad (29)\end{aligned}$$

where the process variances are given in Proposition 3, and the partial derivatives, $\nabla h(\widehat{\boldsymbol{\theta}};\mathcal{N}_0)$, are provided in the following proposition which is analogous to (Wahl et al., 2019, Prop. 10 ):

**Proposition 7.** *The gradients needed to calculate* (27) *based on the reserve predictors from* (8) *and* (9) *are:*

$$\begin{aligned}\frac{\partial}{\partial\alpha_l}h_i^{\mathcal{I}}((\boldsymbol{\alpha},\boldsymbol{\psi});\mathcal{N}_0) &= 1_{\{m-i+1\le l\}}\sum_{k=m-i+1+l}^{m-1}\frac{1}{\alpha_l}(1_{\{k>l\}}\mathbb{E}[N_{i,k}\mid\mathcal{N}_0]+1_{\{k=l\}}\mathbb{E}[\overline{N}_{i,l}\mid\mathcal{N}_0])\overline{\psi}_{\bullet}, \\ \frac{\partial}{\partial\psi_l}h_i^{\mathcal{I}}((\boldsymbol{\alpha},\boldsymbol{\psi});\mathcal{N}_0) &= \sum_{k=m-i+1}^{m-1}\mathbb{E}[N_{ik}\mid\mathcal{N}_0], \\ \frac{\partial}{\partial\psi_l}h_i^{\mathcal{R}}(\boldsymbol{\psi};\mathcal{N}_0) &= \sum_{k=(m-i+1-l)\vee 0}^{m-i}N_{ik}.\end{aligned}$$

What remains is to define $\Lambda((\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\psi}});\mathcal{F}_0)$. As already noted when discussing the reserve predictors' variances in Proposition 6 is that the unconditional covariance of $\widehat{\boldsymbol{\psi}}$ is not analytically tractable, and the same, hence, holds true for $\widehat{\boldsymbol{\alpha}}$. Still, certain *conditional* covariances of $(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\psi}})$ are computable. That is, one suggestion is to use

$$\Lambda((\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\psi}});\mathcal{F}_0) = \Lambda((\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\psi}});\mathcal{N}_0) := \begin{pmatrix}\mathrm{Cov}(\widehat{\boldsymbol{\alpha}}\mid\overline{\mathcal{N}}_0) & \mathbf{0}\\ \mathbf{0} & \mathrm{Cov}(\widehat{\boldsymbol{\psi}}\mid\mathcal{N}_0)\end{pmatrix}, \qquad (30)$$

which is motivated by

$$\mathrm{Cov}(\widehat{\alpha}_i,\widehat{\psi}_j\mid\overline{\mathcal{N}}^0_{\bullet,i-1},\mathcal{N}_0) = \mathrm{Cov}(\widehat{\alpha}_i,\widehat{\psi}_j\mid\mathcal{N}_0) = 0,$$

and where

$$\mathrm{Cov}(\widehat{\boldsymbol{\alpha}}\mid\overline{\mathcal{N}}_0) := \begin{pmatrix}\mathrm{Var}(\widehat{\alpha}_1\mid\overline{\mathcal{N}}^0_{\bullet,0}) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathrm{Var}(\widehat{\alpha}_2\mid\overline{\mathcal{N}}^0_{\bullet,1}) & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathrm{Var}(\widehat{\alpha}_{m-1}\mid\overline{\mathcal{N}}^0_{\bullet,m-2})\end{pmatrix}$$

which is motivated by that $\mathrm{Cov}(\widehat{\alpha}_i,\widehat{\alpha}_j\mid\overline{\mathcal{N}}^0_{\bullet,i-1},\overline{\mathcal{N}}^0_{\bullet,j-1}) = 0, i\neq j$, and which is in alignment with the arguments in Mack (1993). Also note that this implies that $\mathrm{err}(\widehat{\boldsymbol{\theta}};\mathcal{F}_0)(\widehat{\boldsymbol{\theta}}) = \mathrm{err}(\widehat{\boldsymbol{\theta}};\mathcal{N}_0)(\widehat{\boldsymbol{\theta}})$.

Further, for the estimators discussed in the present paper it holds that $\mathbb{E}[\widehat{\alpha}_k\mid\overline{\mathcal{N}}^0_{\bullet,k-1}] = \alpha_k$, $\mathbb{E}[\widehat{\alpha}_k\mid\mathcal{N}_0] = \widehat{\alpha}_k$, and $\mathbb{E}[\widehat{\boldsymbol{\psi}}\mid\mathcal{N}_0] = \boldsymbol{\psi}$, which gives us that

$$\mathbb{E}[\Lambda((\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\psi}});\mathcal{F}_0)] = \begin{pmatrix}\mathrm{Cov}(\widehat{\boldsymbol{\alpha}}) & \mathbf{0}\\ \mathbf{0} & \mathrm{Cov}(\widehat{\boldsymbol{\psi}})\end{pmatrix}.$$

That is, $\Lambda((\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\psi}}); \mathcal{F}_0)$ from (30) is an unbiased estimator of the (analytically intractable) unconditional dito. For a more detailed discussion on having a conditional or unconditional view on estimation error, see e.g. Lindholm et al. (2018) and the references therein. We may now state the following result

**Proposition 8.** *The theoretical MSEP-approximation given by* (27) *defined using* $\Lambda((\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\psi}}); \mathcal{F}_0)$ *from* (30) *is minimised when using the WLS-estimators of* $\boldsymbol{\psi}$ *from* (19)*.*

The proof follows by repeating the arguments from the proof of Proposition 6 verbatim, since it is the same type of quadratic form appearing in both propositions.

Again, as noted in the discussion of Proposition 6, Proposition 8 is only able to order the *theoretical* MSEP approximations, with no clear implications when it comes to the *computable* MSEP approximations given by (29) – which is reasonable w.r.t. the discussion following Proposition 6 concerning Remark 8.

### 6.1 Bootstrapping

As discussed in Section 4 when using the square-root weight functions it is possible to compute the process variances analytically, see Proposition 3. These results are one part of the analytical MSEP-approximation given in (27), where the other part corresponds to an estimate of the reserve estimation error – corresponding to the quadratic form in (27) which involves the partial derivatives $\nabla h(\boldsymbol{\theta}; \mathcal{N}_0)$ from Proposition 7.

An alternative approach to estimating the reserve estimation error is to use standard bootstrapping techniques as those discussed in e.g. England and Verrall (2002), which is a natural approach due to the linear model structure of both the count model and the payment model, given counts. This structure suggests the use of an iterative bootstrap starting with the bootstrapping of $N^*_{i,j}$s, and given the $N^*_{i,j}$s, in a second step bootstrap $X^*_{i,j,k}$s or $X^*_{i,j}$, depending on the available level of detail in data. Moreover, as the underlying model structure is (conditionally) linear it is natural to consider the classical (conditional) standardised Pearson residuals:

$$\widehat{\delta}^N_{i,j} := \frac{\overline{N}_{i,j} - \widehat{\alpha}_j \overline{N}_{i,j-1}}{\widehat{\beta}_j \sqrt{\overline{N}_{i,j-1}}}, \tag{31}$$

$$\widehat{\delta}^X_{i,j,k} := \frac{X_{i,j,k} - \widehat{\psi}_k}{\widehat{\nu}_k \sqrt{N_{i,j}}}. \tag{32}$$

Note that (32) corresponds to detailed data residuals, which may be replaced with the obvious analogous $X_{i,j}$-residuals. Thus, by first drawing $\widehat{\delta}^N$-residuals with replacement, denoted $\widehat{\delta}^{N,*}$, generating

$$\overline{N}^*_{i,j} = \widehat{\alpha}_j \overline{N}^*_{i,j-1} + \widehat{\beta}_j \sqrt{\overline{N}^*_{i,j-1}} \widehat{\delta}^{N,*}_{i,j},$$

and in a second step draw $\widehat{\delta}^X$-residuals with replacement, denoted $\widehat{\delta}^{X,*}$, producing

$$X^*_{i,j,k} = \widehat{\psi}_k N^*_{i,j} + \widehat{\nu}_k \sqrt{N^*_{i,j}} \widehat{\delta}^{X,*}_{i,j,k},$$

one generates new $\mathcal{N}_0^*$ and $\mathcal{F}_0^*$ which are used to obtain $\widehat{\boldsymbol{\theta}}^*$. Consequently, by denoting the $b$th bootstrapped estimate of $\boldsymbol{\theta}$ with $\widehat{\boldsymbol{\theta}}^{*,b}$, it follows that

$$\operatorname{err}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}^{\text{boot}}) := \frac{1}{n_b} \sum_{b=1}^{n_b} (h(\widehat{\boldsymbol{\theta}}^{*,b}; \mathcal{N}_0) - \overline{h}^*)^2 \tag{33}$$

where

$$\overline{h}^* := \frac{1}{n_b} \sum_{b=1}^{n_b} h(\widehat{\boldsymbol{\theta}}^{*,b}; \mathcal{N}_0).$$

Also note that the conditional reserve estimation error is used, since $\mathcal{N}_0$ is kept fixed when evaluating the $h$-functions, which is reasonable due to that the process variance is calculated conditional on $\mathcal{N}_0$. That is, $\operatorname{err}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}^{\text{boot}})$ from (33) is another approximation which should be comparable to $\operatorname{err}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)(\widehat{\boldsymbol{\theta}})$ from (28). Thus, a bootstrapped estimate of MSEP is obtained by using the plug-in estimates of the process variances from Proposition 3 combined with $\operatorname{err}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}^{\text{boot}})$ from (33), i.e.

$$\operatorname{MSEP}_{\mathcal{N}_0}(R, \widehat{R})(\widehat{\boldsymbol{\theta}}^{\text{boot}}) := \operatorname{Var}(R \mid \mathcal{N}_0)(\widehat{\boldsymbol{\theta}}) + \operatorname{err}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}^{\text{boot}}). \tag{34}$$

**Remark 9.** *Note that* (31) *neglects the fact that the $N_{i,j}$s are non-negative integers, see also Remark 3. In practice this is only a potential problem if $N_{i,j} \approx 1$, which typically only occurs when little remains to be paid and, hence, only having a negligible effect on the results. The numerical illustration discussed in Section 7 uses simple rounding of the produced $N_{i,j}^*$s. For more on bootstrap techniques, see England and Verrall (2002) and the references therein.*

# 7 Numerical illustration

We will now illustrate the main results of the paper using simulation. The starting point will be the data from Verrall et al. (2010), although this data is not available at the $X_{i,j,k}$-level. Still, by using the count data together with the parameter estimates from Verrall et al. (2010) it is easy to simulate data which is in agreement with the models introduced in the present paper. As discussed in Section 5.2 (see in particular Remark 8), we may express all reduced data estimators of $\boldsymbol{\psi}$ in terms of transformations of the detailed $X_{i,j,k}$-data. Thus, we will only simulate detailed $X_{i,j,k}$-data and transform data into the $X_{i,j}$-format, and let $\widehat{\boldsymbol{\theta}}^{\text{agg}}$ denote the corresponding $\boldsymbol{\theta}$-estimator from Proposition 4. Similarly, let $\widehat{\boldsymbol{\theta}}^{\text{det}}$ denote the corresponding detailed estimator from (18). In order to stress this we will occasionally use the notation $(\widehat{\boldsymbol{\theta}}^{\text{agg}}, \mathcal{N}_0^{\text{det}})$ and $(\widehat{\boldsymbol{\theta}}^{\text{det}}, \mathcal{N}_0^{\text{det}})$.

The following procedure will be used to generate data:

- Given a column of $N_{i,0}$-values, it is possible to use the negative binomial recursion from Proposition 2 to generate count data. The parametrisation used is based on the $\widehat{\alpha}_j$s from Verrall et al. (2010), which are transformed into the corresponding $\widehat{\pi}_j$s of the negative binomial model from Proposition 2. In

order to include a model for the first column of $N_{i,0}$-values, we let each $N_{i,0}$ be Poisson-distributed with mean corresponding to the observed count from Verrall et al. (2010).

- The number of payments per claim is modelled as negative binomial random variable with mean 1.5. In Verrall et al. (2010) the total mean and variance of a claim is estimated to 203 and 3 496 125, respectively. These values are scaled with the expected number of payments to produce the corresponding values for a single claim payment, and where each claim payment is modelled as a log-normal random variable.

- The distribution of claim payments per claim is modelled as a multinomial distribution parametrised in terms of $p_0, \ldots, p_7$ given in (Verrall et al., 2010, Table 5).

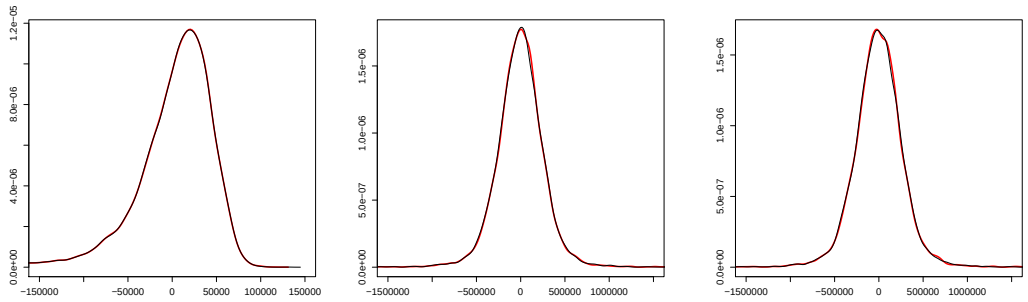
By using this procedure we generate $N_{i,j}$ and $X_{i,j,k}$ data corresponding to 10 accident years that have been observed for at most 10 development years, implying that we will try to fit $\alpha_j, j = 1, \ldots, 9$, and $\psi_k, k = 0, \ldots, 9$. Note, however, that a simulated data set may contain a lot of zero observations, in particular for later development years. This will make it impossible to invert the necessary matrices needed in the estimation of e.g. $\psi$. One can, hence, manually remove zero-patterns which will give rise to this problem, or one can use a pseudo-inverse based on a standard singular value decomposition. We have chosen to use the latter approach due to that we want to use automatic re-estimation for each simulated and bootstrapped dataset.

In Figure 1(A) – 1(C) kernel density estimates of the difference between the predicted IBNR, RBNS and total reserve and the true simulated future payments are shown, when we use both detailed $X_{i,j,k}$-level data (red) and aggregated $X_{i,j}$-level data (black) based on 10 000 simulated complete data sets. From this figure it is clear that both methods perform satisfactory, e.g. centred around zero, and are very close, hence, implying that $\widehat{\boldsymbol{\psi}}^{\text{agg}} \approx \widehat{\boldsymbol{\psi}}^{\text{det}}$.

In Table 1 the predicted (plug-in) total IBNR and RBNS reserves are shown for one simulated dataset, together with the corresponding (plug-in) process variances from Proposition 3. From Table 1 it is seen that the both the IBNR and RBNS reserves are close regardless of whether detailed or aggregated data is used for parameter estimation (but note that the detailed reserve is larger), which is in line with Figure 1(A) – 1(C). This is, however, not the case for the estimated process variances, hence illustrating that using reduced data estimation may have a large effect on the precision of your variance parameter estimates and the parameters' covariance structure. It is also worth mentioning that $\widehat{\psi}_9^{\text{agg}} < 0$ and that there exists negative estimates of $(\widehat{\nu}_k^{\text{agg}})^2$, see Remark 6.

Concerning prediction errors, in Table 2 the reserve estimation errors for the IBNR and RBNS reserves are summarised when calculated using detailed, accumulated and bootstrapped data, in accordance with (28) and (33), together with the corresponding MSEP-values. The bootstrapped estimates are based 10 000 samples following the procedure described in Section 6.1, and all generated $N_{i,j}$-values are made integer valued using simple rounding, see Remark 9. From the table it is clear

that the detailed data estimation produces lower variation regardless of the type of approximation being used and that the analytical approximations are of the same order of magnitude as the corresponding bootstrapped ones. It is also worth recalling that, at least, the analytical ordering may change, depending on the parameter estimates actually used – recall the comments following Proposition 8.



(A) (B) (C)

Figure 1: Fig. 1(A)–1(C): From left to right, kernel density estimates of IBNR, RBNS, and total reserve predictions compared with true simulated future payments based on 10 000 simulations. Red lines corresponds to estimation of $\boldsymbol{\psi}$ based on $X_{i,j,k}$-data using (19), black lines corresponds to estimation of $\boldsymbol{\psi}$ using $X_{i,j}$-data using the estimator from Proposition 4.

Table 1: Predicted (plug-in) total IBNR and RBNS reserves together with the corresponding estimated process variances, based on Proposition 3.

| | $\widehat{R}^{\mathcal{I}}$ | $\widehat{R}^{\mathcal{R}}$ | $\sqrt{\mathrm{Var}(R^{\mathcal{I}} \mid \mathcal{N}_0)(\widehat{\boldsymbol{\theta}})}$ | $\sqrt{\mathrm{Var}(R^{\mathcal{R}} \mid \mathcal{N}_0)(\widehat{\boldsymbol{\theta}})}$ |
|---|---|---|---|---|
| $(\widehat{\boldsymbol{\theta}}^{\mathrm{det}}, \mathcal{N}_0^{\mathrm{det}})$ | 202 906 | 2 742 867 | 31 709 | 134 064 |
| $(\widehat{\boldsymbol{\theta}}^{\mathrm{agg}}, \mathcal{N}_0^{\mathrm{det}})$ | 202 868 | 2 738 703 | 38 714 | 158 953 |

## References


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

Bishop, Y. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis MIT Press.*

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.

England, P. D. and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3):443–518.

Fahrmeir, L., Kneib, T., Lang, S., et al. (2007). *Regression.* Springer.

Lindholm, M., Lindskog, F., and Wahl, F. (2018). Estimation of conditional mean squared error of prediction for claims reserving. *Annals of Actuarial Science*, pages 1–36.

Table 2: Estimates of the reserve estimation errors for total IBNR and RBNS reserves, computed according to (28) and (33), together with the corresponding MSEP estimates calculated according to (29) and (34). The analytical MSEP approximation is based on $\Lambda((\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\psi}}); \mathcal{N}_0)$ from (30), and the bootstrapped estimates are based on 10 000 samples.

| | $\text{err}^{\mathcal{I}}(\widehat{\boldsymbol{\theta}}; \mathcal{N}_0)(\widehat{\boldsymbol{\theta}})$ | $\text{err}^{\mathcal{R}}(\widehat{\boldsymbol{\theta}}; \mathcal{N}_0)(\widehat{\boldsymbol{\theta}})$ | $\sqrt{\text{MSEP}_{\mathcal{N}_0}(R^{\mathcal{I}}, \widehat{R}^{\mathcal{I}})(\widehat{\boldsymbol{\theta}})}$ | $\sqrt{\text{MSEP}_{\mathcal{N}_0}(R^{\mathcal{R}}, \widehat{R}^{\mathcal{R}})(\widehat{\boldsymbol{\theta}})}$ |
|---|---|---|---|---|
| $(\widehat{\boldsymbol{\theta}}^{\text{det}}, \mathcal{N}_0^{\text{det}})$ | 4 714 | 131 795 | 32 058 | 187 998 |
| $(\widehat{\boldsymbol{\theta}}^{\text{agg}}, \mathcal{N}_0^{\text{det}})$ | 5 645 | 151 661 | 39 123 | 219 698 |
| $(\widehat{\boldsymbol{\theta}}^{\text{det,boot}}, \mathcal{N}_0^{\text{det,boot}})$ | 5 629 | 164 887 | 32 205 | 212 511 |
| $(\widehat{\boldsymbol{\theta}}^{\text{agg,boot}}, \mathcal{N}_0^{\text{det,boot}})$ | 5 687 | 168 229 | 39 129 | 231 446 |


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

Mátyás, L., Phillips, P. C., Gourieroux, C., et al. (1999). *Generalized method of moments estimation*, volume 5. Cambridge University Press.

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

Portugal, L., Pantelous, A. A., and Assa, H. (2018). Claims reserving with a stochastic vector projection. *North American Actuarial Journal*, 22(1):22–39.

Renshaw, A. E. and Verrall, R. J. (1998). A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, 4(4):903–923.

Seber, G. A. and Lee, A. J. (2012). *Linear regression analysis, 2nd ed.* John Wiley & Sons.

Sibuya, M., Yoshimura, I., and Shimizu, R. (1964). Negative multinomial distribution. *Annals of the Institute of Statistical Mathematics*, 16(1):409–426.

Verrall, R. (2000). An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: Mathematics and Economics*, 26(3):91–99.

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.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838.

# A Proofs

*Proof of Proposition 3.* We start by considering the RBNS reserve. By combining (5) and (LM 1) it follows that

$$\begin{aligned} R_i^{\mathcal{R}} &= \sum_{j=0}^{m-i} \sum_{k=(m-i+1-j)\wedge d}^{d} X_{i,j,k} \\ &= \sum_{j=0}^{m-i} \sum_{k=(m-i+1-j)\wedge d}^{d} (\psi_k N_{i,j} + \nu_k \sqrt{N_{i,j}} \delta_{i,j,k}), \end{aligned}$$

from which it follows that

$$\begin{aligned} \mathbb{E}[R_i^{\mathcal{R}} \mid \mathcal{N}_0] &= \sum_{j=0}^{m-i} \sum_{k=(m-i+1-j)\wedge d}^{d} \psi_k N_{i,j} \\ &= \sum_{j=0}^{m-i} N_{i,j} \sum_{k=(m-i+1-j)\wedge d}^{d} \psi_k \\ &= \sum_{j=0}^{m-i} N_{i,j} \overline{\psi}_{i,j}. \end{aligned}$$

By using the same representation of $R_i^{\mathcal{R}}$, the process variance of the RBNS reserve becomes

$$\begin{aligned} \mathrm{Var}(R_i^{\mathcal{R}} \mid \mathcal{N}_0) &= \mathrm{Var}\left( \sum_{j=0}^{m-i} \sum_{k=(m-i+1-j)\wedge d}^{d} (\psi_k N_{i,j} + \nu_k \sqrt{N_{i,j}} \delta_{i,j,k}) \mid \mathcal{N}_0 \right) \\ &= \mathrm{Var}\left( \sum_{j=0}^{m-i} \sum_{k=(m-i+1-j)\wedge d}^{d} \nu_k \sqrt{N_{i,j}} \delta_{i,j,k} \mid \mathcal{N}_0 \right) \\ &= \sum_{j=0}^{m-i} \sum_{k=(m-i+1-j)\wedge d}^{d} \nu_k^2 N_{i,j} \\ &= \sum_{j=0}^{m-i} \sum_{k=(m-i+1-j)\wedge d}^{d} \overline{\nu}_{i,j}^2 N_{i,j}, \end{aligned}$$

where the second to last equality follows due to independence.

Similarly, for the IBNR reserve, (5) and (LM 1) yields

$$R_i^{\mathcal{I}} = \sum_{j=m-i+1}^{m-1} \sum_{k=0}^{d} (\psi_k N_{i,j} + \nu_k \sqrt{N_{i,j}} \delta_{i,j,k}),$$

which by conditioning on $\overline{\mathcal{N}}_{i,m-1}$ gives us that

$$\begin{aligned}
\mathbb{E}[R_i^{\mathcal{I}} \mid \mathcal{N}_0] &= \mathbb{E}[\mathbb{E}[R_i^{\mathcal{I}} \mid \overline{\mathcal{N}}_{i,m-1}], \mathcal{N}_0] \\
&= \mathbb{E}\left[\mathbb{E}\left[\sum_{j=m-i+1}^{m-1}\sum_{k=0}^{d}(\psi_k N_{i,j} + \nu_k\sqrt{N_{i,j}}\delta_{i,j,k}) \mid \overline{\mathcal{N}}_{i,m-1}\right] \mid \mathcal{N}_0\right] \\
&= \mathbb{E}\left[\sum_{j=m-i+1}^{m-1}\sum_{k=0}^{d}\psi_k N_{i,j} \mid \mathcal{N}_0\right] \\
&= \sum_{j=m-i+1}^{m-1}\mathbb{E}\left[N_{i,j} \mid \mathcal{N}_0\right]\sum_{k=0}^{d}\psi_k \\
&= \overline{\psi}_{\bullet}\sum_{j=m-i+1}^{m-1}\mathbb{E}\left[N_{i,j} \mid \mathcal{N}_0\right].
\end{aligned}$$

By re-using the same repeated conditioning, the process variance of the IBNR reserve becomes

$$\begin{aligned}
\mathrm{Var}(R_i^{\mathcal{I}} \mid \mathcal{N}_0) &= \mathbb{E}[\mathrm{Var}(R_i^{\mathcal{I}} \mid \mathcal{N}_{i,m-1}) \mid \mathcal{N}_0] + \mathrm{Var}(\mathbb{E}[R_i^{\mathcal{I}} \mid \mathcal{N}_{i,m-1}] \mid \mathcal{N}_0) \\
&= \mathbb{E}\left[\sum_{j=m-i+1}^{m-1}\sum_{k=0}^{d}\nu_k^2 N_{i,j} \mid \mathcal{N}_0\right] + \mathrm{Var}\left(\sum_{j=m-i+1}^{m-1}\sum_{k=0}^{d}\psi_k N_{i,j} \mid \mathcal{N}_0\right) \\
&= \overline{\nu}_{\bullet}^2\mathbb{E}\left[\overline{N}_{i,m-1} - \overline{N}_{i,m-i} \mid \mathcal{N}_0\right] + \overline{\psi}_{\bullet}^2\,\mathrm{Var}(\overline{N}_{i,m-1} \mid \mathcal{N}_0),
\end{aligned}$$

as stated in the proposition. □

*Proof of Proposition 4.* The estimator of $\boldsymbol{\psi}$ is a standard OLS-estimator, and we will not discuss it any further.

Concerning the estimator of $\boldsymbol{\nu}^2$, recall that by the definition of $\boldsymbol{W}$ and $\boldsymbol{\Xi}$, it follows that

$$\boldsymbol{\Xi}\boldsymbol{W}^2 = \mathrm{diag}(\boldsymbol{N}^{\mathrm{det}}\boldsymbol{\nu}^2),$$

which in turn gives us that

$$\mathrm{Cov}(\boldsymbol{X}^{\mathrm{agg}} \mid \mathcal{N}_0) = \boldsymbol{S}\,\mathrm{diag}(\boldsymbol{N}^{\mathrm{det}}\boldsymbol{\nu}^2)\boldsymbol{S}'.$$

Thus, by defining

$$\begin{aligned}
\boldsymbol{e}_i &:= \boldsymbol{X}_i^{\mathrm{agg}} - \boldsymbol{N}_{i,\bullet}^{\mathrm{agg}}\boldsymbol{\psi} \\
&= \mathbf{1}_i'(\boldsymbol{X}^{\mathrm{agg}} - \boldsymbol{N}^{\mathrm{agg}}\boldsymbol{\psi}) \\
&= \mathbf{1}_i'(\boldsymbol{S}\,\mathrm{diag}(\boldsymbol{N}^{\mathrm{det}}\boldsymbol{\nu}^2)\boldsymbol{S}')^{1/2}\boldsymbol{\delta}^{\mathrm{agg}},
\end{aligned}$$

where $\mathbf{1}_i$ is the $|\boldsymbol{X}^{\mathrm{agg}}| \times 1$ vector consisting of only zeros except for position $i$ where

there is a 1, it follows that

$$\begin{aligned}
\mathbb{E}[\boldsymbol{e}_i^2 \mid \mathcal{N}_0] &= \mathbf{1}_i'(\boldsymbol{S}\,\mathrm{diag}(\boldsymbol{N}^{\mathrm{det}}\boldsymbol{\nu}^2)\boldsymbol{S}')\mathbf{1}_i \\
&= \boldsymbol{S}_{i,\bullet}\,\mathrm{diag}(\boldsymbol{N}^{\mathrm{det}}\boldsymbol{\nu}^2)\boldsymbol{S}_{i,\bullet}' \\
&= \mathrm{tr}(\boldsymbol{S}_{i,\bullet}\,\mathrm{diag}(\boldsymbol{N}^{\mathrm{det}}\boldsymbol{\nu}^2)\boldsymbol{S}_{i,\bullet}') \\
&= \mathrm{tr}(\boldsymbol{S}_{i,\bullet}'\boldsymbol{S}_{i,\bullet}\,\mathrm{diag}(\boldsymbol{N}^{\mathrm{det}}\boldsymbol{\nu}^2)) \\
&= \mathbf{1}_i\boldsymbol{N}^{\mathrm{agg}}\boldsymbol{\nu}^2 \\
&= \boldsymbol{N}_{i,\bullet}^{\mathrm{agg}}\boldsymbol{\nu}^2,
\end{aligned}$$

where $\mathrm{tr}(\cdot)$ is the trace operator, and where we have used that $\boldsymbol{S}_{i,\bullet}'\boldsymbol{S}_{i,\bullet}$ is a square matrix whose only non-zero diagonal elements corresponds to the selection pattern $\boldsymbol{S}_{i,\bullet}$. That is, a moment estimator of $\boldsymbol{\nu}^2$, corresponds to the $\widehat{\boldsymbol{\nu}}^2$ which solves the equation system

$$\boldsymbol{N}^{\mathrm{agg}}\widehat{\boldsymbol{\nu}}^2 = \widehat{\boldsymbol{u}},$$

whose solution is $\widehat{\boldsymbol{\nu}}^2 = (\boldsymbol{N}^{\mathrm{agg}})^+\widehat{\boldsymbol{u}}$.

Further, in order for $\widehat{\boldsymbol{\nu}}^2$ to be unbiased, it is clear that

$$\mathbb{E}[\widehat{\boldsymbol{u}} \mid \mathcal{N}_0] = \boldsymbol{N}^{\mathrm{agg}}\boldsymbol{\nu}^2$$

must hold. We will now show that this in general is not the case, and we will do this by considering a single element $\widehat{\boldsymbol{u}}_i$ re-using the arguments from the derivation of the estimator $\widehat{\boldsymbol{\nu}}^2$:

$$\begin{aligned}
\mathbb{E}[\widehat{\boldsymbol{u}}_i^2 \mid \mathcal{N}_0] &= \mathbb{E}[(\mathbf{1}_i'(\boldsymbol{X}^{\mathrm{agg}} - \boldsymbol{N}^{\mathrm{agg}}\widehat{\boldsymbol{\psi}}))^2 \mid \mathcal{N}_0] \\
&= \mathbb{E}[(\mathbf{1}_i'(\boldsymbol{I} - \boldsymbol{N}^{\mathrm{agg}}(\boldsymbol{N}^{\mathrm{agg}})^+)\boldsymbol{X}^{\mathrm{agg}})^2 \mid \mathcal{N}_0] \\
&= \mathbf{1}_i'(\boldsymbol{I} - \boldsymbol{N}^{\mathrm{agg}}(\boldsymbol{N}^{\mathrm{agg}})^+)(\boldsymbol{S}\,\mathrm{diag}(\boldsymbol{N}^{\mathrm{det}}\boldsymbol{\nu}^2)\boldsymbol{S}')(\boldsymbol{I} - \boldsymbol{N}^{\mathrm{agg}}(\boldsymbol{N}^{\mathrm{agg}})^+)'\mathbf{1}_i \\
&= \mathrm{tr}(\mathbf{1}_i'(\boldsymbol{I} - \boldsymbol{N}^{\mathrm{agg}}(\boldsymbol{N}^{\mathrm{agg}})^+)(\boldsymbol{S}\,\mathrm{diag}(\boldsymbol{N}^{\mathrm{det}}\boldsymbol{\nu}^2)\boldsymbol{S}')(\boldsymbol{I} - \boldsymbol{N}^{\mathrm{agg}}(\boldsymbol{N}^{\mathrm{agg}})^+)'\mathbf{1}_i) \\
&= \mathrm{tr}(\mathbf{1}_i'\boldsymbol{S}\,\mathrm{diag}(\boldsymbol{N}^{\mathrm{det}}\boldsymbol{\nu}^2)\boldsymbol{S}'\mathbf{1}_i) \\
&\quad + \mathrm{tr}(\mathbf{1}_i'\boldsymbol{N}^{\mathrm{agg}}(\boldsymbol{N}^{\mathrm{agg}})^+\boldsymbol{S}\,\mathrm{diag}(\boldsymbol{N}^{\mathrm{det}}\boldsymbol{\nu}^2)\boldsymbol{S}'(\boldsymbol{N}^{\mathrm{agg}}(\boldsymbol{N}^{\mathrm{agg}})^+)'\mathbf{1}_i) \\
&\quad - \mathrm{tr}(\mathbf{1}_i'\boldsymbol{N}^{\mathrm{agg}}(\boldsymbol{N}^{\mathrm{agg}})^+(\boldsymbol{S}\,\mathrm{diag}(\boldsymbol{N}^{\mathrm{det}}\boldsymbol{\nu}^2)\boldsymbol{S}'\mathbf{1}_i) \\
&\quad - \mathrm{tr}(\mathbf{1}_i'(\boldsymbol{S}\,\mathrm{diag}(\boldsymbol{N}^{\mathrm{det}}\boldsymbol{\nu}^2)\boldsymbol{S}'(\boldsymbol{N}^{\mathrm{agg}}(\boldsymbol{N}^{\mathrm{agg}})^+)'\mathbf{1}_i) \\
&= \boldsymbol{N}_{i,\bullet}^{\mathrm{agg}}\boldsymbol{\nu}^2 + \mathrm{tr}(\cdot) - \mathrm{tr}(\cdot) - \mathrm{tr}(\cdot) \\
&\neq \boldsymbol{N}_{i,\bullet}^{\mathrm{agg}}\boldsymbol{\nu}^2,
\end{aligned}$$

which shows that $\widehat{\boldsymbol{\nu}}^2$ is biased. □