Mathematical Statistics
Stockholm University

Estimation of conditional mean squared error of prediction for claims reserving

Mathias Lindholm, Filip Lindskog and Felix Wahl

Research Report 2018:6

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

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

Estimation of conditional mean squared error of prediction for claims reserving

Mathias Lindholm[*], Filip Lindskog[†] and Felix Wahl[‡]

December 2018

Abstract

This paper studies estimation of conditional mean squared error of prediction, conditional on what is known at the time of prediction. The particular problem considered is the assessment of actuarial reserving methods given data in the form of runoff triangles (trapezoids), where the use of prediction assessment based on out-of-sample performance is not an option. The prediction assessment principle advocated here can be viewed as a generalization of Akaike's final prediction error. A direct application of this simple principle in the setting of a data generating process given in terms of a sequence of general linear models yields an estimator of conditional mean square error of prediction that can be computed explicitly for a wide range of models within this model class. Mack's distribution-free chain ladder model and the corresponding estimator of the prediction error for the ultimate claim amount is shown to be a special case. It is demonstrated that the prediction assessment principle easily applies to quite different data generating processes and results in estimators that have been studied in the literature.

Keywords: Mean squared error of prediction; Reserving methods; Prediction error; Estimation error; Ultimate claim amount; Claims development result; Chain ladder method.

1 Introduction

Actuarial reserving amounts to forecasting future claim costs from incurred claims that the insurer is unaware of and from claims known to the insurer that may lead to future claim costs. The predictor commonly used is an expectation of future claim costs computed with respect to a parametric model, conditional on the currently observed data, where the unknown parameter vector is replaced by a parameter estimator. A natural question is how to calculate an estimate of the conditional mean squared error of prediction, MSEP, given the observed data, so that this estimate is a fair assessment of the accuracy of the predictor. The main question is how the variability of the predictor due to estimation error should be accounted for and quantified.

Mack's seminal paper Mack (1993) addressed this question for the chain ladder reserving method. Given a set of model assumptions, referred to as Mack's distribution-free chain ladder model, Mack justified the use of the chain ladder reserve predictor and, more importantly, provided an estimator of the conditional MSEP for the chain ladder predictor. Another significant contribution to measuring variability in reserve estimation is the paper England and Verrall (1999) which introduced bootstrap techniques to actuarial science. For more on other approaches to assess the effect of estimation error in claims reserving, see e.g. Buchwalder et al. (2006); Gisler (2006); Wüthrich and Merz (2008b); Röhr (2016); Diers et al. (2016) and the references therein.

Even though Mack (1993) provided an estimator of conditional MSEP for the chain ladder predictor of the ultimate claim amount, the motivation for the approximations in the derivation of the conditional MSEP estimator is somewhat opaque – something commented upon in e.g. Buchwalder et al. (2006). Moreover, by inspecting the above references it is clear that there is no general agreement on how estimation error should be accounted for when assessing prediction error.

Many of the models underlying commonly encountered reserving methods, such as Mack's distribution-free chain ladder model, have an inherent conditional or autoregressive structure. This conditional structure will make the observed data not only a basis for parameter estimation, but also a basis for prediction. More precisely, expected future claim amounts are functions, expressed in terms of observed claim amounts, of the unknown model parameters. These functions form the basis for prediction. Predictors are obtained by replacing the unknown model parameters by their estimators. In particular, the same data are used for the basis for prediction and param-

eter estimation. In order to estimate prediction error in terms of conditional MSEP it is necessary to account for the fact that the parameter estimates differ from the unknown parameter values. As demonstrated in Mack (1993), not doing so will make the effect of estimation error vanish in the conditional MSEP estimation.

We start by considering assessment of a prediction method without reference to a specific model. Given a random variable X to be predicted and a predictor $\widehat{X}$, the conditional MSEP, conditional on the available observations, is defined as

$$\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X}) := \mathbb{E}\big[(X - \widehat{X})^2 \mid \mathcal{F}_0\big] \tag{1}$$
$$= \mathrm{Var}(X \mid \mathcal{F}_0) + \mathbb{E}\big[(\widehat{X} - \mathbb{E}[X \mid \mathcal{F}_0])^2 \mid \mathcal{F}_0\big].$$

The variance term is usually referred to as the process variance and the expected value is referred to as the estimation error. Notice that $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$ is the optimal predictor of the squared prediction error $(X - \widehat{X})^2$ in the sense that it minimizes $\mathbb{E}[((X - \widehat{X})^2 - V)^2]$ over all $\mathcal{F}_0$-measurable random variables V having finite variance. However, $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$ typically depends on unknown parameters.

Typically, the predictor $\widehat{X}$ is taken as the plug-in estimator of the conditional expectation $\mathbb{E}[X \mid \mathcal{F}_0]$: if X has a probability distribution with a parameter vector $\boldsymbol{\theta}$, then we may write

$$h(\boldsymbol{\theta}; \mathcal{F}_0) := \mathbb{E}[X \mid \mathcal{F}_0], \quad \widehat{X} := h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0),$$

where $\boldsymbol{z} \mapsto h(\boldsymbol{z}; \mathcal{F}_0)$ is an $\mathcal{F}_0$-measurable function and $\widehat{\boldsymbol{\theta}}$ is an $\mathcal{F}_0$-measurable estimator of $\boldsymbol{\theta}$. (Note that this definition of a plug-in estimator, i.e. the estimator obtained by replacing an unknown parameter $\boldsymbol{\theta}$ with an estimator $\widehat{\boldsymbol{\theta}}$ of the parameter, is not to be confused with the so-called plug-in principle, see e.g. (Efron and Tibshirani, 1994, Ch. 4.3).) Since the plug-in estimator of

$$\mathbb{E}\big[(\widehat{X} - \mathbb{E}[X \mid \mathcal{F}_0])^2 \mid \mathcal{F}_0\big] = (h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) - h(\boldsymbol{\theta}; \mathcal{F}_0))^2 \geq 0, \tag{2}$$

is equal to 0, it is clear that the plug-in estimator of $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$ coincides with the plug-in estimator of $\mathrm{Var}(X \mid \mathcal{F}_0)$,

$$\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})(\widehat{\boldsymbol{\theta}}) = \mathrm{Var}(X \mid \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}),$$

which fails to account for estimation error and underestimates $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$.

In the present paper we suggest a simple general approach to estimate conditional mean squared error of prediction. The basis of this approach is as follows. Notice that (1) may be written as

$$\text{MSEP}_{\mathcal{F}_0}(X, \widehat{X}) := \mathbb{E}\big[(X - h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0))^2 \mid \mathcal{F}_0\big]$$

whose plug-in estimator, as demonstrated above, is flawed. Consider a random variable $\widehat{\boldsymbol{\theta}}^*$ such that $\widehat{\boldsymbol{\theta}}^*$ and X are conditionally independent, given $\mathcal{F}_0$. Let

$$\begin{aligned}
\text{MSEP}^*_{\mathcal{F}_0}(X, \widehat{X}) &:= \mathbb{E}\big[(X - h(\widehat{\boldsymbol{\theta}}^*; \mathcal{F}_0)^2 \mid \mathcal{F}_0\big] \\
&= \text{Var}(X \mid \mathcal{F}_0) + \mathbb{E}\big[(h(\widehat{\boldsymbol{\theta}}^*; \mathcal{F}_0) - h(\boldsymbol{\theta}; \mathcal{F}_0)^2 \mid \mathcal{F}_0\big].
\end{aligned}$$

The definition of $\text{MSEP}^*_{\mathcal{F}_0}(X, \widehat{X})$ is about disentangling the basis of prediction $\boldsymbol{z} \mapsto h(\boldsymbol{z}; \mathcal{F}_0)$ and the parameter estimator $\widehat{\boldsymbol{\theta}}$ that together form the predictor $\widehat{X}$. Both are expressions in terms of the available noisy data generating $\mathcal{F}_0$, the "statistical basis" in the terminology of Norberg (1986).

The purpose of this paper is to demonstrate that a straightforward estimator of $\text{MSEP}^*_{\mathcal{F}_0}(X, \widehat{X})$ is a good estimator of $\text{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$ that coincides with estimators that have been proposed in the literature for specific models and methods, with Mack's distribution-free chain ladder method as the canonical example. If $\widehat{\boldsymbol{\theta}}^*$ is chosen as $\widehat{\boldsymbol{\theta}}^\perp$, an independent copy of $\widehat{\boldsymbol{\theta}}$, independent of $\mathcal{F}_0$, then $\text{MSEP}^*_{\mathcal{F}_0}(X, \widehat{X})$ coincides with Akaike's final prediction error (FPE) in the conditional setting, see e.g. Remark 1 below for details. Akaike's FPE is a well-studied quantity used for model selection in time series analysis, see Akaike (1969), Akaike (1970), and further elaborations and analysis in Bhansali and Downham (1977) and Speed and Yu (1993). $\widehat{\boldsymbol{\theta}}^*$ should be chosen to reflect the variability of the parameter estimator $\widehat{\boldsymbol{\theta}}$. Different choices of $\widehat{\boldsymbol{\theta}}^*$ may be justified and we will in particular consider choices that make the quantity $\text{MSEP}^*_{\mathcal{F}_0}(X, \widehat{X})$ computationally tractable. In Diers et al. (2016), "pseudo-estimators" are introduced as a key step in the analysis of prediction error in the setting of the distribution-free chain ladder model. Upon identifying the vector of "pseudo-estimators" with $\widehat{\boldsymbol{\theta}}^*$, the approach in Diers et al. (2016) and the one presented in the present paper coincide in the setting of the distribution-free chain ladder model. Moreover, the approaches considered in Buchwalder et al. (2006) are compatible with the general approach of the present paper for the special case of the distribution-free chain ladder model when assessing the prediction error of the ultimate claim amount.

When considering so-called distribution-free models, i.e. models only defined in terms of a set of (conditional) moments, analytical calculation of

$\text{MSEP}^*_{\mathcal{F}_0}(X, \widehat{X})$ requires the first order approximation

$$h(\widehat{\boldsymbol{\theta}}^*; \mathcal{F}_0) \approx h(\boldsymbol{\theta}; \mathcal{F}_0) + \nabla h(\boldsymbol{\theta}; \mathcal{F}_0)'(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}),$$

where $\nabla h(\boldsymbol{\theta}; \mathcal{F}_0)$ denotes the gradient of $\boldsymbol{z} \mapsto h(\boldsymbol{z}; \mathcal{F}_0)$ evaluated at $\boldsymbol{\theta}$. However, this is the only approximation needed. The use of this kind of linear approximation is very common in the literature analyzing prediction error. For instance, it appears naturally in the error propagation argument used for assessing prediction error in the setting of the distribution-free chain ladder model in Röhr (2016), although the general approach taken in Röhr (2016) is different from the one presented here.

Before proceeding with the general exposition, one can note that, as pointed out above, Akaike's original motivation for introducing FPE was as a device for model selection in autoregressive time series modelling. In Section 4 a class of conditional, autoregressive, reserving models is introduced for which the question of model selection is relevant. This topic will not be pursued any further, but it is worth noting that the techniques and methods discussed in the present paper allow for "distribution-free" model selection.

In Section 2 we present in detail the general approach to estimation of conditional mean squared error of prediction briefly summarized above. Moreover, in Section 2 we illustrate how the approach applies to the situation with runoff triangle based reserving when we are interested in calculating conditional MSEP for the ultimate claim amount and the claims development result (CDR). We emphasize the fact that the conditional MSEP given by (1) is the standard (conditional) L^2 distance between a random variable and its predictor. The MSEP quantities considered in Wüthrich et al. (2009) in the setting of the distribution-free chain ladder model are not all conditional MSEP in the sense of (1).

In Section 3 we put the quantities introduced in the general setting in Section 2 in the specific setting where data emerging during a particular time period (calendar year) form a diagonal in a runoff triangle (trapezoid).

In Section 4, development-year dynamics for the claim amounts are given by a sequence of general linear models. Mack's distribution-free chain ladder model is a special case but the model structure is more general and include e.g. development-year dynamics given by sequences of autoregressive models. Given the close connection between our proposed estimator of conditional MSEP and Akaike's FPE, our approach naturally lends itself to model selection within a set of models.

In Section 5 we show that we retrieve Mack's famous conditional MSEP estimator for the ultimate claim amount and demonstrate that our approach

coincides with the approach in Diers et al. (2016) to estimation of conditional MSEP for the ultimate claim amount for Mack's distribution-free chain ladder model. We also argue that conditional MSEP for the CDR is simply a special case, choosing CDR as the random variable of interest instead of e.g. the ultimate claim amount. In Section 5 we show agreement with certain CDR-expressions obtained in Wüthrich et al. (2009) for the distribution-free chain ladder model, while noting that the estimation procedure is different from those used in e.g. Wüthrich et al. (2009); Diers et al. (2016).

Although Mack's distribution-free chain ladder model and the associated estimators/predictors provide canonical examples of the claim amount dynamics and estimators/predictors of the kind considered in Section 4, analysis of the chain ladder method is not the purpose of the present paper. In Section 6 we demonstrate that the general approach to estimation of conditional MSEP presented here applies naturally to non-sequential models such as the overdispersed Poisson chain ladder model. Moreover, for the overdispersed Poisson chain ladder model we derive a (semi-) analytical MSEP-approximation which turns out to coincide with the well-known estimator from Renshaw (1994).

2   Estimation of conditional MSEP in a general setting

We will now formalize the procedure briefly described in Section 1. All random objects are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{T} = \{\underline{t}, \underline{t} + 1, \ldots, \overline{t}\}$ be an increasing sequence of integer times with $\underline{t} < 0 < \overline{t}$ and $0 \in \mathcal{T}$ representing current time. Let $((S_t, S_t^{\perp}))_{t \in \mathcal{T}}$ be a stochastic process generating the relevant data. $(S_t)_{t \in \mathcal{T}}$ and $(S_t^{\perp})_{t \in \mathcal{T}}$ are independent and identically distributed stochastic processes, where the former represents outcomes over time in the real world and the latter represents outcomes in an imaginary parallel universe. Let $(\mathcal{F}_t)_{t \in \mathcal{T}}$ denote the filtration generated by $(S_t)_{t \in \mathcal{T}}$. It is assumed that the probability distribution of $(S_t)_{t \in \mathcal{T}}$ is parametrized by an unknown parameter vector $\boldsymbol{\theta}$. Consequently, the same applies to $(S_t^{\perp})_{t \in \mathcal{T}}$. The problem considered in this paper is the assessment of the accuracy of the prediction of a random variable X, that may be expressed as some functional applied to $(S_t)_{t \in \mathcal{T}}$, given the currently available information represented by $\mathcal{F}_0$. The natural object to consider as the basis for predicting X is

$$h(\boldsymbol{\theta}; \mathcal{F}_0) := \mathbb{E}[X \mid \mathcal{F}_0], \tag{3}$$

which is an $\mathcal{F}_0$-measurable function evaluated at $\boldsymbol{\theta}$. The corresponding predictor is then obtained as the plug-in estimator

$$\widehat{X} := h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0), \tag{4}$$

where $\widehat{\boldsymbol{\theta}}$ is an $\mathcal{F}_0$-measurable estimator of $\boldsymbol{\theta}$. We define

$$\begin{aligned} \mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X}) :&= \mathbb{E}\big[(X - \widehat{X})^2 \mid \mathcal{F}_0\big] \\ &= \mathrm{Var}(X \mid \mathcal{F}_0) + \mathbb{E}\big[(\widehat{X} - \mathbb{E}[X \mid \mathcal{F}_0])^2 \mid \mathcal{F}_0\big]. \end{aligned}$$

and notice that

$$\begin{aligned} \mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X}) &= \mathrm{Var}(X \mid \mathcal{F}_0) + \mathbb{E}\big[(h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) - h(\boldsymbol{\theta}; \mathcal{F}_0))^2 \mid \mathcal{F}_0\big] \quad (5) \\ &= \mathrm{Var}(X \mid \mathcal{F}_0)(\boldsymbol{\theta}) + (h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) - h(\boldsymbol{\theta}; \mathcal{F}_0))^2. \end{aligned}$$

We write

$$H(\boldsymbol{\theta}; \mathcal{F}_0) := \mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$$

to emphasize that $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$ can be seen as an $\mathcal{F}_0$-measurable function of $\boldsymbol{\theta}$. Consequently, the plug-in estimator of $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$ is given by

$$H(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) = \mathrm{Var}(X \mid \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}) + 0,$$

which coincides with the plug-in estimator of the process variance leading to a likely underestimation of $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$. This problem was highlighted already in Mack (1993) in the context of prediction/reserving using the distribution-free chain ladder model. The analytical MSEP approximation suggested for the chain ladder model in Mack (1993) is, in essence, based on replacing the second term on the right-hand side in (5), relating to estimation error, by another term based on certain conditional moments, conditioning on σ-fields strictly smaller than $\mathcal{F}_0$. These conditional moments are natural objects and straightforward to calculate due to the conditional structure of the distribution-free chain-ladder claim-amount dynamics. This approach to estimate conditional MSEP was motivated heuristically as "average over as little as possible", see (Mack, 1993, p. 219). In the present paper, we present a conceptually clear approach to quantifying the variability due to estimation error that is not model specific. The resulting conditional MSEP estimator for the ultimate claim amount is found to coincide with that found in Mack (1993) for the distribution-free chain ladder model, see Section 5. This is further illustrated by applying the same approach to non-sequential,

unconditional, models, see Section 6, where it is shown that the introduced method can provide an alternative motivation of the estimator from Renshaw (1994) for the overdispersed Poisson chain ladder model.

With the aim of finding a suitable estimator of $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$, notice that the predictor $\widehat{X} := h(\widehat{\boldsymbol{\theta}}, \mathcal{F}_0)$ is obtained by evaluating the $\mathcal{F}_0$-measurable function $\boldsymbol{z} \mapsto h(\boldsymbol{z}, \mathcal{F}_0)$ at $\widehat{\boldsymbol{\theta}}$. The chosen model and the stochastic quantity of interest, X, together form the function $\boldsymbol{z} \mapsto h(\boldsymbol{z}, \mathcal{F}_0)$ that is held fixed. This function may be referred to as the basis of prediction. However, the estimator $\widehat{\boldsymbol{\theta}}$ is a random variable whose observed outcome may differ substantially from the unknown true parameter value $\boldsymbol{\theta}$. In order to obtain a meaningful estimator of the $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$, the variability in $\widehat{\boldsymbol{\theta}}$ should be taken into account. Towards this end, consider the random variable $\widehat{\boldsymbol{\theta}}^*$ which is not $\mathcal{F}_0$-measurable, which is constructed to share key properties with $\widehat{\boldsymbol{\theta}}$. Based on $\widehat{X}^* := h(\widehat{\boldsymbol{\theta}}^*; \mathcal{F}_0)$ we will introduce versions of conditional MSEP from which estimators of conditional MSEP in (5) will follow naturally.

Assumption 2.1. $\widehat{\boldsymbol{\theta}}^*$ *and X are conditionally independent, given $\mathcal{F}_0$.*

Definition 2.1. *Define* $\mathrm{MSEP}_{\mathcal{F}_0}^*(X, \widehat{X})$ *by*

$$\mathrm{MSEP}_{\mathcal{F}_0}^*(X, \widehat{X}) := \mathbb{E}[(X - h(\widehat{\boldsymbol{\theta}}^*; \mathcal{F}_0))^2 \mid \mathcal{F}_0].$$

Definition 2.1 and Assumption 2.1 together immediately yield

$$\mathrm{MSEP}_{\mathcal{F}_0}^*(X, \widehat{X}) = \mathrm{Var}(X \mid \mathcal{F}_0)(\boldsymbol{\theta}) + \mathbb{E}[(h(\widehat{\boldsymbol{\theta}}^*; \mathcal{F}_0) - h(\boldsymbol{\theta}; \mathcal{F}_0))^2 \mid \mathcal{F}_0].$$

In general, evaluation of the second term on the right-hand side above requires full knowledge about the model. Typically, we want only to make weaker moment assumptions. The price paid is the necessity to consider the approximation

$$h(\widehat{\boldsymbol{\theta}}^*; \mathcal{F}_0) \approx h(\boldsymbol{\theta}; \mathcal{F}_0) + \nabla h(\boldsymbol{\theta}; \mathcal{F}_0)'(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}) =: h^{\nabla}(\widehat{\boldsymbol{\theta}}^*; \mathcal{F}_0),$$

where $\nabla h(\boldsymbol{\theta}; \mathcal{F}_0)$ denotes the gradient of $\boldsymbol{z} \mapsto h(\boldsymbol{z}; \mathcal{F}_0)$ evaluated at $\boldsymbol{\theta}$.

Notice that if $\mathbb{E}[\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0] = \boldsymbol{\theta}$ and $\mathrm{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0)$ exists finitely a.s., then

$$\begin{aligned}
\mathbb{E}[(h^{\nabla}&(\widehat{\boldsymbol{\theta}}^*; \mathcal{F}_0) - h(\boldsymbol{\theta}; \mathcal{F}_0))^2 \mid \mathcal{F}_0] \\
&= \nabla h(\boldsymbol{\theta}; \mathcal{F}_0)' \mathbb{E}[(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})' \mid \mathcal{F}_0] \nabla h(\boldsymbol{\theta}; \mathcal{F}_0) \\
&= \nabla h(\boldsymbol{\theta}; \mathcal{F}_0)' \mathrm{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0) \nabla h(\boldsymbol{\theta}; \mathcal{F}_0).
\end{aligned}$$

Assumption 2.2. $\mathbb{E}[\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0] = \boldsymbol{\theta}$, $\mathrm{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0)$ *exists finitely a.s.*

Definition 2.2. *Define* $\mathrm{MSEP}_{\mathcal{F}_0}^{*;\nabla}(X,\widehat{X})$ *by*

$$\mathrm{MSEP}_{\mathcal{F}_0}^{*;\nabla}(X,\widehat{X}) := \mathrm{Var}(X \mid \mathcal{F}_0)(\boldsymbol{\theta}) + \nabla h(\boldsymbol{\theta};\mathcal{F}_0)' \, \mathrm{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0) \nabla h(\boldsymbol{\theta};\mathcal{F}_0).$$

Notice that $\mathrm{MSEP}_{\mathcal{F}_0}^*(X,\widehat{X}) = \mathrm{MSEP}_{\mathcal{F}_0}^{*;\nabla}(X,\widehat{X})$ *if* $h^\nabla(\widehat{\boldsymbol{\theta}}^*;\mathcal{F}_0) = h(\widehat{\boldsymbol{\theta}}^*;\mathcal{F}_0)$.

Remark 1. *Akaike presented, in Akaike (1969, 1970), the quantity* FPE *(final prediction error) for assessment of the accuracy of a predictor, intended for model selection by rewarding models that give rise to small prediction errors. Akaike demonstrated the merits of* FPE *when used for order selection among autoregressive processes.*

Akaike's FPE *assumes a stochastic process* $(S_t)_{t\in\mathcal{T}}$ *of interest and an independent copy* $(S_t^\perp)_{t\in\mathcal{T}}$ *of that process. Let* $\mathcal{F}_0$ *be the σ-field generated by* $(S_t)_{t\in\mathcal{T},t\leq 0}$ *and let* X *be the result of applying some functional to* $(S_t)_{t\in\mathcal{T}}$ *such that* X *is not $\mathcal{F}_0$-measurable. If* $(S_t)_{t\in\mathcal{T}}$ *is 1-dimensional, then* $X = S_t$, *for some* $t > 0$, *is a natural example. Let* $h(\boldsymbol{\theta};\mathcal{F}_0) := \mathbb{E}[X \mid \mathcal{F}_0]$ *and let* $h(\widehat{\boldsymbol{\theta}};\mathcal{F}_0)$ *be the corresponding predictor of* X *based on the $\mathcal{F}_0$-measurable parameter estimator* $\widehat{\boldsymbol{\theta}}$. *Let* $\mathcal{F}_0^\perp$, $X^\perp$, $h(\boldsymbol{\theta};\mathcal{F}_0^\perp)$ *and* $\widehat{\boldsymbol{\theta}}^\perp$ *be the corresponding quantities based on* $(S_t^\perp)_{t\in\mathcal{T}}$. FPE *is defined as*

$$\mathrm{FPE}(X,\widehat{X}) := \mathbb{E}\big[(X^\perp - h(\widehat{\boldsymbol{\theta}};\mathcal{F}_0^\perp))^2\big]$$

and it is clear that the roles of $(S_t)_{t\in\mathcal{T}}$ *and* $(S_t^\perp)_{t\in\mathcal{T}}$ *may be interchanged to get*

$$\begin{aligned}
\mathrm{FPE}(X,\widehat{X}) &= \mathbb{E}\big[(X^\perp - h(\widehat{\boldsymbol{\theta}};\mathcal{F}_0^\perp))^2\big] \\
&= \mathbb{E}\big[(X - h(\widehat{\boldsymbol{\theta}}^\perp;\mathcal{F}_0))^2\big].
\end{aligned}$$

Naturally, we may consider the conditional version of FPE *which gives*

$$\begin{aligned}
\mathrm{FPE}_{\mathcal{F}_0}(X,\widehat{X}) &= \mathbb{E}\big[(X^\perp - h(\widehat{\boldsymbol{\theta}};\mathcal{F}_0^\perp))^2 \mid \mathcal{F}_0^\perp\big] \\
&= \mathbb{E}\big[(X - h(\widehat{\boldsymbol{\theta}}^\perp;\mathcal{F}_0))^2 \mid \mathcal{F}_0\big].
\end{aligned}$$

Clearly, $\widehat{\boldsymbol{\theta}}^* = \widehat{\boldsymbol{\theta}}^\perp$ *gives*

$$\mathrm{MSEP}_{\mathcal{F}_0}^*(X,\widehat{X}) := \mathbb{E}\big[(X - h(\widehat{\boldsymbol{\theta}}^*;\mathcal{F}_0))^2 \mid \mathcal{F}_0\big] = \mathrm{FPE}_{\mathcal{F}_0}(X,\widehat{X}).$$

If $h^\nabla(\widehat{\boldsymbol{\theta}}^*;\mathcal{F}_0) = h(\widehat{\boldsymbol{\theta}}^*;\mathcal{F}_0)$, *then choosing* $\widehat{\boldsymbol{\theta}}^* = \widehat{\boldsymbol{\theta}}^\perp$ *gives*

$$\begin{aligned}
\mathrm{MSEP}_{\mathcal{F}_0}^{*;\nabla}(X,\widehat{X}) &= \mathrm{FPE}_{\mathcal{F}_0}(X,\widehat{X}) \\
&= \mathrm{Var}(X \mid \mathcal{F}_0)(\boldsymbol{\theta}) + \nabla h(\boldsymbol{\theta};\mathcal{F}_0)' \, \mathrm{Cov}(\widehat{\boldsymbol{\theta}}) \nabla h(\boldsymbol{\theta};\mathcal{F}_0).
\end{aligned}$$

9

Since $\text{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0)$ is a $\mathcal{F}_0$-measurable function of $\boldsymbol{\theta}$ we may write

$$\text{MSEP}_{\mathcal{F}_0}^{*,\nabla}(X, \widehat{X}) = \text{Var}(X \mid \mathcal{F}_0) + \nabla h(\boldsymbol{\theta}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\boldsymbol{\theta}; \mathcal{F}_0) \nabla h(\boldsymbol{\theta}; \mathcal{F}_0), \quad (6)$$

where

$$\boldsymbol{\Lambda}(\boldsymbol{\theta}; \mathcal{F}_0) := \text{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0).$$

We write

$$H^*(\boldsymbol{\theta}; \mathcal{F}_0) := \text{MSEP}_{\mathcal{F}_0}^*(X, \widehat{X}), \quad H^{*,\nabla}(\boldsymbol{\theta}; \mathcal{F}_0) := \text{MSEP}_{\mathcal{F}_0}^{*,\nabla}(X, \widehat{X})$$

to emphasize that $\text{MSEP}_{\mathcal{F}_0}^*(X, \widehat{X})$ and $\text{MSEP}_{\mathcal{F}_0}^{*,\nabla}(X, \widehat{X})$ are $\mathcal{F}_0$-measurable functions of $\boldsymbol{\theta}$.

The plug-in estimator $H^*(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)$ of $\text{MSEP}_{\mathcal{F}_0}^*(X, \widehat{X})$ may appear to be a natural estimator of $\text{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$. However, in most situations there will not be sufficient statistical evidence to motivate specifying the full distribution of $\widehat{\boldsymbol{\theta}}^*$. Therefore, $H^*(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)$ is not likely to be an attractive estimator of $\text{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$. The plug-in estimator $H^{*,\nabla}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)$ of $\text{MSEP}_{\mathcal{F}_0}^{*,\nabla}(X, \widehat{X})$ is more likely to be a computable estimator of $\text{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$, requiring only the covariance matrix $\boldsymbol{\Lambda}(\boldsymbol{\theta}; \mathcal{F}_0) := \text{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0)$ as a matrix-valued function of the parameter $\boldsymbol{\theta}$ instead of the full distribution of $\widehat{\boldsymbol{\theta}}^*$. We will henceforth focus solely on the estimator $H^{*,\nabla}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)$.

Definition 2.3. *The estimator of the conditional mean squared error of prediction is given by*

$$\widehat{\text{MSEP}}_{\mathcal{F}_0}(X, \widehat{X}) := \text{Var}(X \mid \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}) + \nabla h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) \nabla h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0). \quad (7)$$

We emphasize that the estimator we suggest in Definition 2.3 relies on **one** approximation and **one** modeling choice. The approximation refers to

$$h(\boldsymbol{z}; \mathcal{F}_0) \approx h(\boldsymbol{\theta}; \mathcal{F}_0) + \nabla h(\boldsymbol{\theta}; \mathcal{F}_0)'(\boldsymbol{z} - \boldsymbol{\theta})$$

and **no other approximations will appear**. The modeling choice refers to deciding on how the estimation error should be accounted for in terms of the conditional covariance structure $\text{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0)$, where $\widehat{\boldsymbol{\theta}}^*$ satisfies the requirement $\mathbb{E}[\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0] = \boldsymbol{\theta}$.

Before proceeding further with the specification of $\widehat{\boldsymbol{\theta}}^*$, one can note that in many situations it will be natural to structure data according to e.g. accident year. In these situation it will be possible to express X as $X = \sum_{i \in \mathcal{I}} X_i$ and consequently also $h(\boldsymbol{\theta}; \mathcal{F}_0) = \sum_{i \in \mathcal{I}} h_i(\boldsymbol{\theta}; \mathcal{F}_0)$. This immediately implies that the estimator (7) of conditional MSEP can be expressed in a way that simplifies computations.

Lemma 2.1. *Given that $X = \sum_{i \in \mathcal{I}} X_i$ and $h(\boldsymbol{\theta}; \mathcal{F}_0) = \sum_{i \in \mathcal{I}} h_i(\boldsymbol{\theta}; \mathcal{F}_0)$, the estimator (7) takes the form*

$$\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(X, \widehat{X}) = \sum_{i \in \mathcal{I}} \widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(X_i, \widehat{X}_i)$$
$$+ 2 \sum_{i,j \in \mathcal{I}, i<j} \Big(\mathrm{Cov}(X_i, X_j \mid \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}) + Q_{i,j}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) \Big),$$

where

$$\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(X_i, \widehat{X}_i) := \mathrm{Var}(X_i \mid \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}) + Q_{i,i}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0),$$
$$Q_{i,j}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) := \nabla h_i(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) \nabla h_j(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0).$$

The proof of Lemma 2.1 follows from expanding the original quadratic form in the obvious way, see Appendix C. Even though Lemma 2.1 is trivial, it will be used repeatedly in later sections when the introduced methods are illustrated using e.g. different models for the data generating process.

Assumption 2.3. $\mathbb{E}[\widehat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$, $\mathrm{Cov}(\widehat{\boldsymbol{\theta}})$ *exists finitely.*

Given Assumption 2.3, one choice of $\widehat{\boldsymbol{\theta}}^*$ is to choose $\widehat{\boldsymbol{\theta}}^*$ as an independent copy $\widehat{\boldsymbol{\theta}}^\perp$, based entirely on $(S_t^\perp)_{t \in \tau}$, of $\widehat{\boldsymbol{\theta}}$, independent of $\mathcal{F}_0$. An immediate consequence of this choice is

$$\mathbb{E}[\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0] = \boldsymbol{\theta}, \quad \mathrm{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0) = \mathrm{Cov}(\widehat{\boldsymbol{\theta}}) =: \boldsymbol{\Lambda}(\boldsymbol{\theta}), \quad \widehat{\boldsymbol{\theta}}^* := \widehat{\boldsymbol{\theta}}^\perp.$$

Since the specification $\widehat{\boldsymbol{\theta}}^* := \widehat{\boldsymbol{\theta}}^\perp$ implies that $\mathrm{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0)$ does not depend on $\mathcal{F}_0$ we refer to $\widehat{\boldsymbol{\theta}}^\perp$ as the unconditional specification of $\widehat{\boldsymbol{\theta}}^*$. In this case, as described in Remark 1, $\mathrm{MSEP}^*_{\mathcal{F}_0}(X, \widehat{X})$ coincides with Akaike's FPE in the conditional setting. Moreover,

$$\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(X, \widehat{X}) = \mathrm{Var}(X \mid \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}) + \nabla h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}) \nabla h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0).$$

For some models for the data generating process $(S_t)_{t \in \mathcal{T}}$, such as the conditional linear models investigated in Section 4 below, computation of the unconditional covariance matrix $\mathrm{Cov}(\widehat{\boldsymbol{\theta}})$ is not feasible. Moreover, it may be argued that observed data should be considered also in the specification of $\widehat{\boldsymbol{\theta}}^*$ although there is no statistical principle justifying this argument. The models investigated in Section 4 are such that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p)$ and there exist nested σ-fields $\mathcal{G}_1 \subseteq \ldots \mathcal{G}_p \subseteq \mathcal{F}_0$ such that $\mathbb{E}[\widehat{\boldsymbol{\theta}}_k \mid \mathcal{G}_k] = \boldsymbol{\theta}_k$ for $k = 1, \ldots, p$ and $\widehat{\boldsymbol{\theta}}_k$ is $\mathcal{G}_{k+1}$-measurable for $k = 1, \ldots, p-1$. Consequently,

$\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_j, \widehat{\boldsymbol{\theta}}_k \mid \mathcal{G}_j, \mathcal{G}_k) = 0$ for $j \neq k$. If further the covariance matrices $\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_k \mid \mathcal{G}_k)$ can be computed explicitly, as demonstrated in Section 4, then we may choose $\widehat{\boldsymbol{\theta}}^* := \widehat{\boldsymbol{\theta}}^{*,c}$ such that $\mathbb{E}[\widehat{\boldsymbol{\theta}}_k^* \mid \mathcal{F}_0] = \mathbb{E}[\widehat{\boldsymbol{\theta}}_k \mid \mathcal{G}_k]$, $\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_k^* \mid \mathcal{F}_0) = \mathrm{Cov}(\widehat{\boldsymbol{\theta}}_k \mid \mathcal{G}_k)$ for $k = 1, \ldots, p$ and $\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_j^*, \widehat{\boldsymbol{\theta}}_k^* \mid \mathcal{F}_0) = 0$ for $j \neq k$. Since the specification $\widehat{\boldsymbol{\theta}}^* := \widehat{\boldsymbol{\theta}}^{*,c}$ implies that $\mathrm{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0)$ depends on $\mathcal{F}_0$ we refer to $\widehat{\boldsymbol{\theta}}^{*,c}$ as the conditional specification of $\widehat{\boldsymbol{\theta}}^*$. In this case,

$$\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(X, \widehat{X}) = \mathrm{Var}(X \mid \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}) + \nabla h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) \nabla h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0).$$

Notice that if $\widehat{\boldsymbol{\theta}}^{*,u} := \widehat{\boldsymbol{\theta}}^{\perp}$ and $\widehat{\boldsymbol{\theta}}^{*,c}$ denote the unconditional and conditional specifications of $\widehat{\boldsymbol{\theta}}^*$, respectively, then covariance decomposition yields

$$\begin{aligned}
\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_k^{*,u} \mid \mathcal{F}_0) &= \mathrm{Cov}(\widehat{\boldsymbol{\theta}}_k) \\
&= \mathbb{E}[\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_k \mid \mathcal{G}_k)] + \mathrm{Cov}(\mathbb{E}[\widehat{\boldsymbol{\theta}}_k \mid \mathcal{G}_k]) \\
&= \mathbb{E}[\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_k \mid \mathcal{G}_k)] \\
&= \mathbb{E}[\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_k^{*,c} \mid \mathcal{F}_0)]
\end{aligned}$$

and therefore

$$\mathrm{Cov}(\widehat{\boldsymbol{\theta}}^{*,u} \mid \mathcal{F}_0) = \mathbb{E}[\mathrm{Cov}(\widehat{\boldsymbol{\theta}}^{*,c} \mid \mathcal{F}_0)]. \tag{8}$$

$\mathrm{Cov}(\widehat{\boldsymbol{\theta}}^{*,c} \mid \mathcal{F}_0)$ is thus an unbiased estimator of $\mathrm{Cov}(\widehat{\boldsymbol{\theta}}^{*,u} \mid \mathcal{F}_0)$.

The estimators of conditional MSEP for the distribution-free chain-ladder model given in Buchwalder et al. (2006) and, more explicitly, in Diers et al. (2016) are essentially based on the conditional specification of $\widehat{\boldsymbol{\theta}}^*$. We refer to Section 5 below for details.

2.1   Selection of estimators of conditional MSEP

As noted in the introduction, $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$ is the optimal predictor of the squared prediction error $(X - \widehat{X})^2$ in the sense that it minimizes $\mathbb{E}[((X - \widehat{X})^2 - V)^2]$ over all $\mathcal{F}_0$-measurable random variables V having finite variance. Therefore, given a set of estimators $\widehat{V}$ of $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$, the best estimator is the one minimizing $\mathbb{E}[((X - \widehat{X})^2 - \widehat{V})^2]$. Write $V := \mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$ and

$\widehat{V} := V + \Delta V$ and notice that

$$
\begin{aligned}
\mathbb{E}&[((X-\widehat{X})^2 - \widehat{V})^2] \\
&= \mathrm{Var}((X - \widehat{X})^2) + \mathbb{E}[(X - \widehat{X})^2]^2 - 2\mathbb{E}[(X - \widehat{X})^2(V + \Delta V)] \\
&\quad + \mathbb{E}[(V + \Delta V)^2] \\
&= \mathrm{Var}((X - \widehat{X})^2) + \mathbb{E}[V]^2 - 2\Big(\mathbb{E}[V^2] + \mathbb{E}[(X - \widehat{X})^2 \Delta V]\Big) \\
&\quad + \mathbb{E}[V^2] + \mathbb{E}[\Delta V^2] + 2\mathbb{E}[V \Delta V] \\
&= \mathrm{Var}((X - \widehat{X})^2) - \mathrm{Var}(V) + \mathbb{E}[\Delta V^2]. \qquad (9)
\end{aligned}
$$

In our setting,

$$
V = \mathrm{Var}(X \mid \mathcal{F}_0) + \Big(h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) - h(\boldsymbol{\theta}; \mathcal{F}_0)\Big)^2
$$

and, from Definition 2.3,

$$
\widehat{V} = \mathrm{Var}(X \mid \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}) + \nabla h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) \nabla h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0).
$$

Recall that $\boldsymbol{\Lambda}(\boldsymbol{\theta}; \mathcal{F}_0) = \mathrm{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0)$ depends on the specification of $\widehat{\boldsymbol{\theta}}^*$. Therefore we may in principle search for the optimal specification of $\widehat{\boldsymbol{\theta}}^*$. However, it is unlikely that any specifications will enable explicit computation of $\mathbb{E}[\Delta V^2]$. Moreover, for so-called distribution-free models defined only in terms of certain (conditional) moments, the required moments appearing in the computation of $\mathbb{E}[\Delta V^2]$ may be unspecified.

We may consider the approximations

$$
V = \mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X}) \approx \mathrm{Var}(X \mid \mathcal{F}_0) + \nabla h(\boldsymbol{\theta}; \mathcal{F}_0)' (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \nabla h(\boldsymbol{\theta}; \mathcal{F}_0),
$$
$$
\widehat{V} \approx \mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})^{*, \nabla} = \mathrm{Var}(X \mid \mathcal{F}_0) + \nabla h(\boldsymbol{\theta}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\boldsymbol{\theta}; \mathcal{F}_0) \nabla h(\boldsymbol{\theta}; \mathcal{F}_0)
$$

which yield

$$
\mathbb{E}[\Delta V^2] \approx \mathbb{E}\Big[\Big(\nabla h(\boldsymbol{\theta}; \mathcal{F}_0)'\big((\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' - \boldsymbol{\Lambda}(\boldsymbol{\theta}; \mathcal{F}_0)\big)\nabla h(\boldsymbol{\theta}; \mathcal{F}_0)\Big)^2\Big].
$$

Therefore, the specification of $\widehat{\boldsymbol{\theta}}^*$ should be such that

- $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'$ and $\boldsymbol{\Lambda}(\boldsymbol{\theta}; \mathcal{F}_0)$ are close, and

- $\boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)$ is computable.

Appendix D compares the two estimators of conditional MSEP based on unconditional and conditional, respectively, specification of $\widehat{\boldsymbol{\theta}}^*$, in the setting of Mack's distribution-free chain ladder model. No significant difference between the two estimators can be found. However, in the setting of Mack's distribution-free chain ladder model, only the estimator based on the conditional specification of $\widehat{\boldsymbol{\theta}}^*$ is computable.

3 Data in the form of runoff triangles

One of the main objectives of this paper is the estimation of the precision of reserving methods when the data in the form of runoff triangles (trapezoids), explained below, have conditional development-year dynamics of a certain form. Mack's chain ladder model, see e.g. Mack (1993), will serve as the canonical example.

Let $I_{i,j}$ denote the incremental claims payments during development year $j \in \{1, \ldots, J\} =: \mathcal{J}$ and from accidents during accident year $i \in \{i_0, \ldots, J\} =: \mathcal{I}$, where $i_0 \leq 1$. This corresponds to the indexation used in Mack (1993), i.e. $j = 1$ corresponds to the payments that are made during a particular accident year. Clearly, the standard terminology accident- and development year used here could refer to any other appropriate time unit. The observed payments as of today, at time 0, is what is called a runoff triangle or runoff trapezoid:

$$D_0 := \{I_{i,j} : (i,j) \in \mathcal{I} \times \mathcal{J}, i + j \leq J + 1\},$$

and let $\mathcal{F}_0 := \sigma(D_0)$. Notice that accident years $i \leq 1$ are fully developed. Notice also that in the often considered special case $i_0 = 1$, the runoff trapezoid takes the form of a triangle. Instead of incremental payments $I_{i,j}$ we may of course equivalently consider cumulative payments $C_{i,j} := \sum_{k=1}^{j} I_{i,k}$, noticing that $\mathcal{F}_0 = \sigma(\{C_{i,j} : (i,j) \in \mathcal{I} \times \mathcal{J}, i + j \leq J + 1\})$.

The incremental payments that occur between (calendar) time $t - 1$ and t corresponds to the following diagonal in the runoff triangle of incremental payments:

$$S_t = \{I_{i,j} : (i,j) \in \mathcal{I} \times \mathcal{J}, i + j = J + 1 + t\}.$$

Consequently the filtration $(\mathcal{F}_t)_{t \in \mathcal{T}}$ is given by

$$\mathcal{F}_t = \sigma(D_t), \quad D_t := \{I_{i,j} : (i,j) \in \mathcal{I} \times \mathcal{J}, i + j \leq J + 1 + t\}.$$

Let

$$B_k := \{I_{i,j} : (i,j) \in \mathcal{I} \times \mathcal{J}, j \leq k, i + j \leq J + 1\},$$

i.e. the subset of D_0 corresponding to claim amounts up to and including development year k, and notice that $\mathcal{G}_k := \sigma(B_k) \subset \mathcal{F}_0$, $k = 1, \ldots, J$, form an increasing sequence of σ-fields. Conditional expectations and covariances with respect to these σ-fields appear naturally when estimating conditional MSEP in the distribution-free chain ladder model, see Mack (1993), and also in the more general setting considered here when $\widehat{\boldsymbol{\theta}}^*$ is chosen according to the conditional specification. We refer to Section 4 for details.

3.1 Conditional MSEP for the ultimate claim amount

The outstanding claims reserve R_i for accident year i that is not yet fully developed, i.e. the future payments stemming from claims incurred during accident year i, and the total outstanding claims reserve R are given by

$$R_i := \sum_{j=J-i+2}^{J} I_{i,j} = C_{i,J} - C_{i,J-i+1}, \quad R := \sum_{i=2}^{J} R_i.$$

The ultimate claim amount U_i for accident year i that is not yet fully developed, i.e. the future and past payments stemming from claims incurred during accident year i, and the ultimate claim amount U are given by

$$U_i := \sum_{j=1}^{J} I_{i,j} = C_{i,J}, \quad U := \sum_{i=2}^{J} U_i.$$

Similarly, the amount of paid claims P_i for accident year i that is not yet fully developed, i.e. the past payments stemming from claims incurred during accident year i, and the total amount of paid claims P are given by

$$P_i := \sum_{j=1}^{J-i+1} I_{i,j} = C_{i,J-i+1}, \quad P := \sum_{i=2}^{J} P_i.$$

Obviously, $U_i = P_i + R_i$ and $U = P + R$.

We are interested in calculating the conditional MSEP of U and we can start by noticing that if the $\mathcal{F}_0$-measurable random variable P is added to the random variable R to be predicted, then the same applies to its predictor: $\widehat{U} = P + \widehat{R}$. Therefore,

$$\mathrm{MSEP}_{\mathcal{F}_0}(U, \widehat{U}) = \mathrm{MSEP}_{\mathcal{F}_0}(R, \widehat{R}).$$

Further, in order to be able compute the conditional MSEP estimators from Definition 2.1, and in particular the final plug-in estimator given by (7), we need to specify the basis of prediction, i.e. $\boldsymbol{z} \mapsto h(\boldsymbol{z}, \mathcal{F}_0)$, which is given by

$$h(\boldsymbol{\theta}, \mathcal{F}_0) := \mathbb{E}[U \mid \mathcal{F}_0],$$

as well as specify the choice of $\widehat{\boldsymbol{\theta}}^*$. Neither of this is meaningful to pursue any further, without specifying the underlying model structure. In Sections 4, 5 and 6 we discuss how this can be done for specific models using Lemma 2.1.

3.2 Conditional MSEP for the claims development result

Above we have described an approach for calculating MSEP for the ultimate claim amount. Another quantity which has received considerable attention is MSEP for the claims development result, CDR, which is the difference between the ultimate claim amount predictor and its update based on one more year of observations. For the chain ladder method, an estimator of the variability of CDR is provided in Wüthrich and Merz (2008a). We will now describe how this may be done consistently in terms of MSEP*. As will be seen, there is no conceptual difference compared to the calculations for the ultimate claim amount – all steps will follow verbatim from Section 2. For more on the estimator in Wüthrich and Merz (2008a) for the distribution-free chain ladder model, see Section 5.

Let

$$\mathrm{CDR} := h^{(0)}(\widehat{\boldsymbol{\theta}}^{(0)}; \mathcal{F}_0) - h^{(1)}(\widehat{\boldsymbol{\theta}}^{(1)}; \mathcal{F}_1),$$

where

$$h^{(0)}(\boldsymbol{\theta}; \mathcal{F}_0) := \mathbb{E}[U \mid \mathcal{F}_0], \quad h^{(1)}(\boldsymbol{\theta}; \mathcal{F}_1) := \mathbb{E}[U \mid \mathcal{F}_1],$$

and $\widehat{\boldsymbol{\theta}}^{(0)}$ and $\widehat{\boldsymbol{\theta}}^{(1)}$ are $\mathcal{F}_0$- and $\mathcal{F}_1$-measurable estimators of $\boldsymbol{\theta}$, based on the observations at times 0 and 1, respectively. Hence, CDR is simply the difference between the predictor at time 0 of the ultimate claim amount and that at time 1. Thus, given the above it follows by choosing

$$h(\boldsymbol{\theta}; \mathcal{F}_0) := \mathbb{E}[\mathrm{CDR} \mid \mathcal{F}_0] = h^{(0)}(\widehat{\boldsymbol{\theta}}^{(0)}; \mathcal{F}_0) - \mathbb{E}[h^{(1)}(\widehat{\boldsymbol{\theta}}^{(1)}; \mathcal{F}_1) \mid \mathcal{F}_0]$$

that we may again estimate $\mathrm{MSEP}_{\mathcal{F}_0}(\mathrm{CDR}, \widehat{\mathrm{CDR}})$ using Definitions 2.1 and 2.2 – in particular we may calculate the plug-in estimator given by (7). Note, from the definition of CDR, regardless of the specification of $\widehat{\boldsymbol{\theta}}^*$, that it directly follows that

$$\begin{aligned}
\mathrm{MSEP}&_{\mathcal{F}_0}(\mathrm{CDR}, \widehat{\mathrm{CDR}}) \\
&= \mathrm{MSEP}_{\mathcal{F}_0}\left(\mathbb{E}[U \mid \mathcal{F}_1](\widehat{\boldsymbol{\theta}}^{(1)}), \mathbb{E}[\mathbb{E}[U \mid \mathcal{F}_1](\widehat{\boldsymbol{\theta}}^{(1)}) \mid \mathcal{F}_0](\widehat{\boldsymbol{\theta}}^{(0)})\right) \\
&= \mathrm{MSEP}_{\mathcal{F}_0}\left(h^{(1)}(\widehat{\boldsymbol{\theta}}^{(1)}; \mathcal{F}_1), \mathbb{E}[h^{(1)}(\widehat{\boldsymbol{\theta}}^{(1)}; \mathcal{F}_1) \mid \mathcal{F}_0](\widehat{\boldsymbol{\theta}}^{(0)})\right), \quad (10)
\end{aligned}$$

where the $\mathcal{F}_0$-measurable term $h^{(0)}(\widehat{\boldsymbol{\theta}}^{(0)}; \mathcal{F}_0)$ appearing both in CDR and $\widehat{\mathrm{CDR}}$ has canceled out. Thus, from the above definition of $h(\boldsymbol{\theta}; \mathcal{F}_0)$, together

with the definition of MSEP*, Definition 2.1, it is clear that the estimation error will only correspond to the effect of perturbing $\boldsymbol{\theta}$ in $\mathbb{E}[h^{(1)}(\widehat{\boldsymbol{\theta}}^{(1)}; \mathcal{F}_1) \mid \mathcal{F}_0](\boldsymbol{\theta})$. Moreover, the notion of conditional MSEP and the suggested estimation procedure for the CDR is in complete analogy with that for the ultimate claim amount. This estimation procedure is however different from the ones used in e.g. Wüthrich and Merz (2008a); Wüthrich et al. (2009); Röhr (2016); Diers et al. (2016) for the distribution-free chain ladder model. For Mack's distribution-free chain ladder model,

$$\mathbb{E}[h^{(1)}(\widehat{\boldsymbol{\theta}}^{(1)}; \mathcal{F}_1) \mid \mathcal{F}_0](\widehat{\boldsymbol{\theta}}^{(0)}) = h^{(0)}(\widehat{\boldsymbol{\theta}}^{(0)}; \mathcal{F}_0)$$

and therefore $\mathrm{MSEP}_{\mathcal{F}_0}(\mathrm{CDR}, \widehat{\mathrm{CDR}}) = \mathrm{MSEP}_{\mathcal{F}_0}(\mathrm{CDR}, 0)$. This is however not true in general for other models. More details on CDR-calculations for the distribution-free chain ladder model are found in Section 5.

Moreover, by introducing

$$h^{(k)}(\boldsymbol{\theta}; \mathcal{F}_k) := \mathbb{E}[U \mid \mathcal{F}_k],$$

we can, of course, repeat the above steps to obtain the conditional MSEP for the k-year CDR by using the following definition

$$\mathrm{CDR}(k) := h^{(0)}(\widehat{\boldsymbol{\theta}}^{(0)}; \mathcal{F}_0) - h^{(k)}(\widehat{\boldsymbol{\theta}}^{(k)}; \mathcal{F}_k),$$

together with the obvious changes.

We want to stress that we have no particular interest in these CDR-calculations, one-year or k-year, but merely want to illustrate the applicability and transparency of the suggested approach. As an illustration, in Section 5 calculations for the ultimate claim amount and one-year CDR are carried out in more detail for the distribution-free chain ladder model. This is, again, based on using Lemma 2.1.

4 Dynamics in the form of sequential conditional linear models

We will now describe how the theory introduced in Section 2 applies to specific models. We will first introduce a class of sequential conditional linear models to which the distribution-free chain ladder model is a special case, but also contains more general autoregressive reserving models investigated in e.g. Kremer (1984) and Lindholm et al. (2017). Since this class of models

has a natural conditional structure it is interesting to discuss the specification of $\widehat{\boldsymbol{\theta}}^*$ as being either conditional or unconditional.

As concluded in Section 2, the parameter estimator $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\Lambda}(\boldsymbol{\theta}; \mathcal{F}_0)$ are needed in order to obtain a computable estimator of $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$ following (7). In the present section we will present rather general development-year dynamics for claim amounts that immediately give the estimator $\widehat{\boldsymbol{\theta}}$ and we will discuss how $\widehat{\boldsymbol{\theta}}^*$ can be specified which gives us $\boldsymbol{\Lambda}(\boldsymbol{\theta}; \mathcal{F}_0)$.

For the remainder of the current section we will focus on the following development-year dynamics for claim amounts:

$$\boldsymbol{Y}_{j+1} = \boldsymbol{A}_j \boldsymbol{\beta}_j + \sigma_j \boldsymbol{D}_j \boldsymbol{e}_{j+1}, \quad j = 1, \ldots, J - 1. \tag{11}$$

Here $\boldsymbol{Y}_{j+1}$ is a $|\mathcal{I}| \times 1$ vector that may represent incremental or cumulative claim amounts, corresponding to either $\boldsymbol{Y}_{j+1} = (I_{i,j+1})_{i \in \mathcal{I}}$ or $\boldsymbol{Y}_{j+1} = (C_{i,j+1})_{i \in \mathcal{I}}$, respectively, $\boldsymbol{A}_j$ is a $|\mathcal{I}| \times p_j$ matrix, $\boldsymbol{\beta}_j$ is a $p_j \times 1$ parameter vector, σ_j is a positive scalar parameter, $\boldsymbol{D}_j$ is a diagonal $|\mathcal{I}| \times |\mathcal{I}|$ matrix with positive diagonal elements and $\boldsymbol{e}_{j+1}$ is a $|\mathcal{I}| \times 1$ vector. The random matrices $\boldsymbol{A}_j$ and $\boldsymbol{D}_j$ and the random vector $\boldsymbol{e}_{j+1}$ all have independent rows. This requirement ensures that claim amounts stemming from different accident years are independent. Moreover, the components of $\boldsymbol{e}_{j+1}$ all have, conditional on $\boldsymbol{A}_j$ and $\boldsymbol{D}_j$, mean zero and variance one. Therefore, the same holds for the unconditional first two moments:

$$\mathbb{E}[e_{j+1,k}] = \mathbb{E}[\mathbb{E}[e_{j+1,k} \mid \boldsymbol{A}_j, \boldsymbol{D}_j]] = 0,$$
$$\mathbb{E}[e_{j+1,k}^2] = \mathbb{E}[\mathbb{E}[e_{j+1,k}^2 \mid \boldsymbol{A}_j, \boldsymbol{D}_j]] = 1.$$

Notice, however, that variables $e_{2,k}, \ldots, e_{J,k}$ are not required to be independent. In fact if the variables $Y_{2,k}, \ldots, Y_{J,k}$ are required to be positive, then $e_{2,k}, \ldots, e_{J,k}$ cannot be independent. See Remark 2 in Section 5 for an example, and Mack et al. (2006) for further comments in the setting of Mack's distribution-free chain ladder model.

The development-year dynamics (11) with the above dimensions of $\boldsymbol{A}_j$, $\boldsymbol{D}_j$ and $\boldsymbol{e}_{j+1}$ do not correspond to the dynamics of data observed at time 0. For runoff triangle data, observations come in the form of a diagonal. In particular, at time 0 only the first $n_j := J - j - i_0 + 1$ components of $\boldsymbol{Y}_{j+1}$ are observed. The development-year dynamics of claim amounts that are observed at time 0 are therefore of the form

$$\widetilde{\boldsymbol{Y}}_{j+1} = \widetilde{\boldsymbol{A}}_j \boldsymbol{\beta}_j + \sigma_j \widetilde{\boldsymbol{D}}_j \widetilde{\boldsymbol{e}}_{j+1}, \quad j = 1, \ldots, J - 1, \tag{12}$$

where $\widetilde{\boldsymbol{Y}}_{j+1}$ is a $n_j \times 1$ vector, $\widetilde{\boldsymbol{A}}_j$ is a $n_j \times p_j$ matrix, $\widetilde{\boldsymbol{D}}_j$ is a diagonal $n_j \times n_j$ matrix and $\widetilde{\boldsymbol{e}}_{j+1}$ is a $n_j \times 1$. We will throughout assume that

18

$n_j \geq p_j$. Hence, we will in what follows consider a sequence of conditional linear models where the dimension of the parameters is fixed whereas the dimension of the random objects vary with the development year. Notice that $\widetilde{\boldsymbol{Y}}_{j+1}$, $\widetilde{\boldsymbol{A}}_j$, $\widetilde{\boldsymbol{D}}_j$ and $\widetilde{\boldsymbol{e}}_{j+1}$ are the sub-vectors/matrices of $\boldsymbol{Y}_{j+1}$, $\boldsymbol{A}_j$, $\boldsymbol{D}_j$ and $\boldsymbol{e}_{j+1}$ obtained by considering only the first n_j rows.

Recall the following notation introduced in Section 2

$$B_k := \{I_{i,j} : (i,j) \in \mathcal{I} \times \mathcal{J}, j \leq k, i+j \leq J+1\},$$

i.e. the subset of D_0 corresponding to claim amounts up to and including development year k, $\widetilde{\boldsymbol{A}}_j$ and $\widetilde{\boldsymbol{D}}_j$ are both $\sigma(B_j)$-measurable with independent rows. Moreover, by the independence between the rows in $\boldsymbol{e}_{j+1}$, the components of $\widetilde{\boldsymbol{e}}_{j+1}$ all have, conditional on $\widetilde{\boldsymbol{A}}_j$ and $\widetilde{\boldsymbol{D}}_j$, mean zero and variance one. These observations form the basis of parameter estimation since it allows $\boldsymbol{\beta}_j$ to be estimated by the standard weighted least squares estimator from the theory of general linear models:

$$\widehat{\boldsymbol{\beta}}_j = \left(\widetilde{\boldsymbol{A}}_j' \widetilde{\boldsymbol{\Sigma}}_j^{-1} \widetilde{\boldsymbol{A}}_j\right)^{-1} \widetilde{\boldsymbol{A}}_j' \widetilde{\boldsymbol{\Sigma}}_j^{-1} \widetilde{\boldsymbol{Y}}_{j+1}, \quad \widetilde{\boldsymbol{\Sigma}}_j := \widetilde{\boldsymbol{D}}_j^2, \tag{13}$$

which is independent of σ_j. Notice in particular that

$$\mathbb{E}\big[\widehat{\boldsymbol{\beta}}_j \mid \widetilde{\boldsymbol{A}}_j, \widetilde{\boldsymbol{\Sigma}}_j\big] = \mathbb{E}\big[\widehat{\boldsymbol{\beta}}_j \mid B_j\big] = \boldsymbol{\beta}_j. \tag{14}$$

Moreover,

$$\text{Cov}\,\big(\widehat{\boldsymbol{\beta}}_j \mid B_j\big) = \sigma_j^2 \big(\widetilde{\boldsymbol{A}}_j' \widetilde{\boldsymbol{\Sigma}}_j^{-1} \widetilde{\boldsymbol{A}}_j\big)^{-1}. \tag{15}$$

The estimator of the dispersion parameter σ_j^2 is, for $j = 1, \ldots, J-1$, given by

$$\widehat{\sigma}_j^2 = \frac{1}{n_j - p_j} (\widetilde{\boldsymbol{Y}}_{j+1} - \widetilde{\boldsymbol{A}}_j \widehat{\boldsymbol{\beta}}_j)' \widetilde{\boldsymbol{\Sigma}}_j^{-1} (\widetilde{\boldsymbol{Y}}_{j+1} - \widetilde{\boldsymbol{A}}_j \widehat{\boldsymbol{\beta}}_j), \tag{16}$$

given that $n_j - p_j > 0$, i.e. given that $i_0 \leq J - j - p_j$. If $i_0 = 1$, then $\widehat{\sigma}_{J-1}^2$ has to be defined by an ad hoc choice. The weighted least squares estimator in (13) is the best linear unbiased estimator of $\boldsymbol{\beta}_j$ in the sense that, for any $\boldsymbol{a} \in \mathbb{R}^{p_j}$, $\widehat{\boldsymbol{\beta}}_j$ is such that $\boldsymbol{a}' \widehat{\boldsymbol{\beta}}_j$ has minimum variance among all unbiased linear estimators. Similarly the estimator in (16) is the best unbiased estimator of σ_j^2. For further details on weighted (generalized) least squares see e.g. (Seber and Lee, 2003, Sec. 3.10).

Basic properties of the estimators are presented next. The essential properties are that, for each j, $\widehat{\boldsymbol{\beta}}_j$ is unbiased and, for $j \neq k$, $\widehat{\boldsymbol{\beta}}_j$ and $\widehat{\boldsymbol{\beta}}_k$ are uncorrelated.

Proposition 4.1. *For each j,*

(i) $\mathbb{E}[\widehat{\boldsymbol{\beta}}_j] = \boldsymbol{\beta}_j$ *and, for $j \neq k$,* $\mathrm{Cov}\,(\widehat{\boldsymbol{\beta}}_j, \widehat{\boldsymbol{\beta}}_k) = \mathbf{0}$.

(ii) $\mathbb{E}[\widehat{\sigma}_j^2 \mid B_j] = \sigma_j^2$ *given that $i_0 \leq J - j - p_j$.*

The proof of Proposition 4.1 is given in the appendix.

Recall that the overall aim is estimation of $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$, where X is a stochastic quantity of interest, e.g. the ultimate claim amount U or the claims development result CDR, whose distribution depends on an unknown parameter $\boldsymbol{\theta}$. Here,

$$\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma}), \quad \boldsymbol{\beta} := (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{J-1}), \quad \boldsymbol{\sigma} := (\sigma_1, \ldots, \sigma_{J-1}).$$

Considering the similarities of the model considered here and general linear models, it is clear that there are conditions ensuring that $h(\boldsymbol{\theta}; \mathcal{F}_0) = \mathbb{E}[X \mid \mathcal{F}_0]$ depends on $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma})$ only through $\boldsymbol{\beta}$ and not $\boldsymbol{\sigma}$, e.g. $h(\boldsymbol{\beta}; \mathcal{F}_0) = \mathbb{E}[U \mid \mathcal{F}_0]$. In what follows we hence make the following assumption:

Assumption 4.1. $h((\boldsymbol{\beta}, \boldsymbol{\sigma}); \mathcal{F}_0) = \mathbb{E}[X \mid \mathcal{F}_0]$ *is independent of $\boldsymbol{\sigma}$.*

Assumption 4.1 is fulfilled by e.g. the distribution-free chain ladder model, see Section 5, as well as the models stated in Appendix A, which cover e.g. Kremer (1984); Lindholm et al. (2017).

Given Assumption 4.1 we write $h(\boldsymbol{\beta}; \mathcal{F}_0)$ for $h((\boldsymbol{\beta}, \boldsymbol{z}); \mathcal{F}_0)$ for an arbitrary $\boldsymbol{z}$.

Recall from Section 2 that $\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X})$ is approximated by (6) which in turn has a computable estimator (7). Under Assumption 4.1,

$$\nabla h(\boldsymbol{\theta}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\boldsymbol{\theta}; \mathcal{F}_0) \nabla h(\boldsymbol{\theta}; \mathcal{F}_0) = \nabla_{\boldsymbol{\beta}} h(\boldsymbol{\beta}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\boldsymbol{\theta}; \mathcal{F}_0) \nabla_{\boldsymbol{\beta}} h(\boldsymbol{\beta}; \mathcal{F}_0)$$

and therefore (6) simplifies as follows:

$$\mathrm{MSEP}_{\mathcal{F}_0}^{*, \nabla}(X, \widehat{X}) = \mathrm{Var}(X \mid \mathcal{F}_0) + \nabla_{\boldsymbol{\beta}} h(\boldsymbol{\beta}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\boldsymbol{\theta}; \mathcal{F}_0) \nabla_{\boldsymbol{\beta}} h(\boldsymbol{\beta}; \mathcal{F}_0). \quad (17)$$

4.1 Specification of $\widehat{\boldsymbol{\theta}}^*$

Recall from Section 2 that we introduced the two independent and identically distributed stochastic processes $(S_t)_{t \in \mathcal{T}}$ and $(S_t^\perp)_{t \in \mathcal{T}}$, where the former is the one generating data that can be observed. In the current setting we have a parallell universe (another independent runoff triangle) with development year dynamics

$$\boldsymbol{Y}_{j+1}^\perp = \boldsymbol{A}_j^\perp \boldsymbol{\beta}_j + \sigma_j \boldsymbol{D}_j^\perp \boldsymbol{e}_{j+1}^\perp, \quad j = 1, \ldots, J - 1.$$

If the unconditional specification of $\widehat{\boldsymbol{\theta}}^*$ is chosen, i.e. $\widehat{\boldsymbol{\theta}}^{*,u} = \widehat{\boldsymbol{\theta}}^\perp$, then

$$\widehat{\boldsymbol{\beta}}_j^{*,u} = \left(\{\widetilde{\boldsymbol{A}}_j^\perp\}'\{\widetilde{\boldsymbol{\Sigma}}_j^\perp\}^{-1}\widetilde{\boldsymbol{A}}_j^\perp\right)^{-1}\{\widetilde{\boldsymbol{A}}_j^\perp\}'\{\widetilde{\boldsymbol{\Sigma}}_j^\perp\}^{-1}\widetilde{\boldsymbol{Y}}_{j+1}^\perp$$

$$= \boldsymbol{\beta}_j + \sigma_j\left(\{\widetilde{\boldsymbol{A}}_j^\perp\}'\{\widetilde{\boldsymbol{\Sigma}}_j^\perp\}^{-1}\widetilde{\boldsymbol{A}}_j^\perp\right)^{-1}\{\widetilde{\boldsymbol{A}}_j^\perp\}'\{\widetilde{\boldsymbol{D}}_j^\perp\}^{-1}\widetilde{\boldsymbol{e}}_{j+1}^\perp,$$

i.e. simply the weighted least squares estimator applied to the data in the independent triangle with identical features as the observable one. It follows directly from Proposition 4.1 that

$$\operatorname{Cov}(\widehat{\boldsymbol{\beta}}^* \mid \mathcal{F}_0) = \operatorname{Cov}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\Lambda}(\boldsymbol{\beta},\boldsymbol{\sigma}), \quad \widehat{\boldsymbol{\beta}}^* := \widehat{\boldsymbol{\beta}}^{*,u},$$

is a block-diagonal covariance matrix with blocks $\operatorname{Cov}(\widehat{\boldsymbol{\beta}}_j)$ of dimension $p_j \times p_j$. It is also clear that these unconditional covariances $\operatorname{Cov}(\widehat{\boldsymbol{\beta}}_j)$ are not possible to compute analytically.

On the other hand, if we specify $\widehat{\boldsymbol{\theta}}^*$ conditionally, then

$$\widehat{\boldsymbol{\beta}}_j^{*,c} := \left(\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{A}}_j\right)^{-1}\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{\Sigma}}_j^{-1}\left(\widetilde{\boldsymbol{A}}_j\boldsymbol{\beta}_j + \sigma_j\widetilde{\boldsymbol{D}}_j\widetilde{\boldsymbol{e}}_{j+1}^\perp\right)$$

$$= \boldsymbol{\beta}_j + \sigma_j\left(\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{A}}_j\right)^{-1}\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{D}}_j^{-1}\widetilde{\boldsymbol{e}}_{j+1}^\perp,$$

which is identical to $\widehat{\boldsymbol{\beta}}_j$ except that $\widetilde{\boldsymbol{e}}_{j+1}^\perp$ appears instead of $\widetilde{\boldsymbol{e}}_{j+1}$. Notice that this definition of $\widehat{\boldsymbol{\beta}}_j^*$ satisfies Assumption 2.1. Notice also that

$$\operatorname{Cov}(\widehat{\boldsymbol{\beta}}_j^{*,c} \mid \mathcal{F}_0) = \operatorname{Cov}(\widehat{\boldsymbol{\beta}}_j \mid B_j) = \sigma_j^2\left(\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{A}}_j\right)^{-1}.$$

Hence,

$$\operatorname{Cov}(\widehat{\boldsymbol{\beta}}^* \mid \mathcal{F}_0) = \boldsymbol{\Lambda}(\boldsymbol{\sigma};\mathcal{F}_0), \quad \widehat{\boldsymbol{\beta}}^* := \widehat{\boldsymbol{\beta}}^{*,c},$$

where

$$\boldsymbol{\Lambda}(\boldsymbol{\sigma};\mathcal{F}_0) = \begin{bmatrix} \operatorname{Cov}(\widehat{\boldsymbol{\beta}}_1 \mid B_1) & 0 & \ldots & & 0 \\ 0 & & \ddots & & \\ \vdots & & & & \\ 0 & & & & \operatorname{Cov}(\widehat{\boldsymbol{\beta}}_{J-1} \mid B_{J-1}) \end{bmatrix}.$$

Further, in Section 2 arguments were given for when the conditional specification of $\widehat{\boldsymbol{\theta}}^*$ resulting in $\boldsymbol{\Lambda}(\boldsymbol{\sigma};\mathcal{F}_0)$ may be seen as an unbiased estimator of $\boldsymbol{\Lambda}(\boldsymbol{\beta},\boldsymbol{\sigma})$, given by the corresponding unconditional $\widehat{\boldsymbol{\theta}}^*$, see (8). Within the class of models given by (11) this relation may be strengthened: Proposition 4.2 below tells us that $\boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}};\mathcal{F}_0)$ is an unbiased estimator of $\operatorname{Cov}(\widehat{\boldsymbol{\beta}})$ and an empirical estimator of $\operatorname{Cov}(\widehat{\boldsymbol{\beta}})$ based on a single claims trapezoid.

Proposition 4.2. $\mathbb{E}[\boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0)] = \mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}\right)$ *given that* $i_0 \leq J - j - p_j$ *for all* j.

The proof of Proposition 4.2 is given in the appendix.

Moreover, in Appendix B we have collected a number of asymptotic results where it is shown that, given suitable regularity conditions, $\mathrm{Cov}(\widehat{\boldsymbol{\beta}})$ and $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_j \mid B_j)$ will converge to the same limit as the number of accident years tends to infinity, see Proposition B.1. This implies that given a sufficient amount of data the two views on estimation error will result in conditional MSEP estimates that are close. In Section 5 this is shown to be the case in an illustration based on real data.

5 Mack's distribution-free chain ladder

The classical chain ladder reserving method is a prediction algorithm for predicting the ultimate claim amount. In order to justify the use of this method and in order to measure the prediction accuracy, Mack introduced, in Mack (1993), conditions that should be satisfied by the underlying model. The chain ladder method with Mack's conditions is referred to as Mack's distribution-free chain ladder model. We will see that this setting is compatible with the development-year dynamics (11) in Section 4 and we will show in Proposition 5.1 that the estimator of $\mathrm{MSEP}_{\mathcal{F}_0}(U, \widehat{U})$ from Section 3.1 calculated according to Definition 2.3 coincides with the celebrated estimator of $\mathrm{MSEP}_{\mathcal{F}_0}(U, \widehat{U})$ provided by Mack in Mack (1993).

In accordance with Mack's distribution-free chain ladder model, assume that, for $j = 1, \ldots, J-1$, there exist constants $f_j > 0$, called development factors, and constants $\sigma_j^2 \geq 0$ such that

$$\mathbb{E}[C_{i,j+1} \mid C_{i,j}, \ldots, C_{i,1}] = f_j C_{i,j}, \tag{18}$$

$$\mathrm{Var}(C_{i,j+1} \mid C_{i,j}, \ldots, C_{i,1}) = \sigma_j^2 C_{i,j}, \tag{19}$$

where $i = i_0, \ldots, J$. Moreover, assume that,

$$\{C_{i_0,1}, \ldots, C_{i_0,J}\}, \ldots, \{C_{J,1}, \ldots, C_{J,J}\} \quad \text{are independent.} \tag{20}$$

Notice that the claim amounts during the first development year $I_{i_0,1}, \ldots, I_{J,1}$ are independent but not necessarily identically distributed.

Mack's distribution-free chain ladder fits into the development-year dy-

namics (11) in Section 4 as follows: for $j = 1, \ldots, J-1$, set $p_j = 1$, $\boldsymbol{\beta}_j = f_j$,

$$
\widetilde{\boldsymbol{Y}}_{j+1} = \begin{bmatrix} C_{i_0,j+1} \\ C_{i_0+1,j+1} \\ \vdots \\ C_{J-j,j+1} \end{bmatrix}, \quad \widetilde{\boldsymbol{A}}_j = \begin{bmatrix} C_{i_0,j} \\ C_{i_0+1,j} \\ \vdots \\ C_{J-j,j} \end{bmatrix}, \quad \widetilde{\boldsymbol{\Sigma}}_j = \mathrm{diag}\begin{bmatrix} C_{i_0,j} \\ C_{i_0+1,j} \\ \vdots \\ C_{J-j,j} \end{bmatrix},
$$

where $\mathrm{diag}[\boldsymbol{a}]$ denotes a diagonal matrix with diagonal $[\boldsymbol{a}]$. Notice that this choice of $(\boldsymbol{Y}_{j+1}, \boldsymbol{A}_j, \boldsymbol{\Sigma}_j)$ corresponds to a special case of (38) of Assumption A.1. Therefore, the statement of Assumption 4.1 holds.

Remark 2. *For the elements of $\boldsymbol{\Sigma}_j$ to have positive diagonal elements we need the additional condition $\{\boldsymbol{e}_{j+1}\}_i > -f_j C_{i,j}^{1/2}/\sigma_j$. This somewhat odd requirement is easily satisfied. For instance, set $W_{i,j} := f_j C_{i,j}^{1/2}/\sigma_j$, let $Z_{i,j}$ be standard normal independent of $W_{i,j}$ and set*

$$
\{\boldsymbol{e}_{j+1}\}_i := \exp\left\{\mu(W_{i,j}) + \sigma(W_{i,j})Z_{i,j}\right\} - W_{i,j},
$$

$$
\sigma(W_{i,j}) := \sqrt{\log(1 + W_{i,j}^{-2})}, \quad \mu(W_{i,j}) := \log(W_{i,j}) - \sigma^2(W_{i,j})/2.
$$

In this case, conditional on $C_{i,j}$, $\{\boldsymbol{e}_{j+1}\}_i$ is simply a translated lognormal random variable, translated by $-f_j C_{i,j}^{1/2}/\sigma_j$, with zero mean and unit variance.

Notice that

$$
\widehat{\boldsymbol{\beta}}_j = \left(\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{A}}_j\right)^{-1}\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{Y}}_{j+1} = \frac{\sum_{i=i_0}^{J-j} C_{i,j+1}}{\sum_{i=i_0}^{J-j} C_{i,j}} = \widehat{f}_j,
$$

which coincides with the classical chain ladder development factor estimator, hence, being a standard weighted least-squares estimator for the model (11). Furthermore,

$$
\widehat{\sigma}_j^2 = \frac{1}{n_j - p_j}(\widetilde{\boldsymbol{Y}}_{j+1} - \widetilde{\boldsymbol{A}}_j\widehat{\boldsymbol{\beta}}_j)'\widetilde{\boldsymbol{\Sigma}}_j^{-1}(\widetilde{\boldsymbol{Y}}_{j+1} - \widetilde{\boldsymbol{A}}_j\widehat{\boldsymbol{\beta}}_j)
$$

$$
= \frac{1}{J - j - i_0}\sum_{i=i_0}^{J-j} C_{i,j}\left(\frac{C_{i,j+1}}{C_{i,j}} - \widehat{f}_j\right)^2, \quad j = 1, \ldots, J-2,
$$

and similarly for $\widehat{\sigma}_{J-1}^2$ if $i_0 \leq 0$. Notice also that

$$
\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_j \mid B_j) = \sigma_j^2\left(\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{A}}_j\right)^{-1} = \frac{\sigma_j^2}{\sum_{i=i_0}^{J-j} C_{i,j}} = \mathrm{Var}(\widehat{f}_j \mid B_j).
$$

Using the tower property of conditional expectations together with (18) and (20) it is straightforward to verify that

$$h_i(\boldsymbol{f}; \mathcal{F}_0) := \mathbb{E}[U_i \mid \mathcal{F}_0] = C_{i,J-i+1} \prod_{j=J-i+1}^{J-1} f_j,$$

$$h(\boldsymbol{f}; \mathcal{F}_0) := \mathbb{E}[U \mid \mathcal{F}_0] = \sum_{i=2}^{J} h_i(\boldsymbol{f}; \mathcal{F}_0) = \sum_{i=2}^{J} C_{i,J-i+1} \prod_{j=J-i+1}^{J-1} f_j. \quad (21)$$

In order to calculate MSEP for the ultimate claim amount following Lemma 2.1, we need to obtain expressions for process (co)variances and the $Q_{i,j}$s given by

$$Q_{i,j}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) = \nabla h_i(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) \nabla h_j(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0).$$

The process variances are given in (Mack, 1993, Thm. 3, Cor.) and follows by using variance decomposition, the tower property of conditional expectations, (18), (19) and (20), and may, after simplifications, be expressed as

$$\mathrm{Var}(U \mid \mathcal{F}_0) = \sum_{i=2}^{J} \mathrm{Var}(U_i \mid \mathcal{F}_0)$$

$$= \sum_{i=2}^{J} C_{i,J+1-i} \sum_{k=J+1-i}^{J-1} f_{J+1-i} \cdots f_{k-1} \sigma_k^2 f_{k+1}^2 \cdots f_{J-1}^2. \quad (22)$$

For detailed calculations, see (Mack, 1993, Thm. 3, Cor.). Further, letting

$$\widehat{U}_i := h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0) = \widehat{C}_{i,J}, \quad \widehat{C}_{i,j} := C_{i,J-j+1} \prod_{k=J-i+1}^{j-1} \widehat{f}_k,$$

it follows that

$$\mathrm{Var}(U \mid \mathcal{F}_0)(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\sigma}}^2) = \sum_{i=2}^{J} \widehat{U}_i^2 \sum_{k=J-i+1}^{J-1} \frac{\widehat{\sigma}_k^2}{\widehat{f}_k^2 \widehat{C}_{i,k}}. \quad (23)$$

Thus, if we set

$$\widehat{\Gamma}_{i,J}^{U} := \sum_{k=J-i+1}^{J-1} \frac{\widehat{\sigma}_k^2}{\widehat{f}_k^2 \widehat{C}_{i,k}}, \quad (24)$$

we see that

$$\mathrm{Var}(U \mid \mathcal{F}_0)(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\sigma}}^2) = \sum_{i=2}^{J} \widehat{U}_i^2 \widehat{\Gamma}_{i,J}^U.$$

If we turn to the calculation of $Q_{i,j}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)$ we see that

$$\{\nabla_{\boldsymbol{f}} h_i(\boldsymbol{f}; \mathcal{F}_0)\}_j = \frac{\partial}{\partial f_j} h_i(\boldsymbol{f}; \mathcal{F}_0) = 1_{\{J-i+1 \leq j\}} C_{i,J-i+1} \frac{1}{f_j} \prod_{l=J-i+1}^{J-1} f_l,$$

for $i = 2, \ldots, J$ and $j = 1, \ldots, J-1$ and that

$$\{\boldsymbol{\Lambda}(\boldsymbol{\sigma}; \mathcal{F}_0)\}_{j,j} = \mathrm{Var}(\widehat{f}_j \mid B_j) = \frac{\sigma_j^2}{\sum_{i=i_0}^{J-j} C_{i,j}}, \tag{25}$$

where $\{\boldsymbol{\Lambda}(\boldsymbol{\sigma}; \mathcal{F}_0)\}_{i,j} = 0$ for all $i \neq j$. Hence,

$$\{\nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)\}_j = 1_{\{J-i+1 \leq j\}} \frac{\widehat{U}_i}{\widehat{f}_j}, \tag{26}$$

and it follows by direct calculations that

$$\begin{aligned}
Q_{i,i}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) &= \nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0) \nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0) \\
&= \sum_{k=J-i+1}^{J-1} \frac{\widehat{U}_i^2 \widehat{\sigma}_k^2}{\widehat{f}_k^2 \sum_{l=i_0}^{J-k} C_{l,k}} \\
&= \widehat{U}_i^2 \widehat{\Delta}_{i,J}^U,
\end{aligned}$$

where

$$\widehat{\Delta}_{i,J}^U := \sum_{k=J-i+1}^{J-1} \frac{\widehat{\sigma}_k^2}{\widehat{f}_k^2 \sum_{l=i_0}^{J-k} C_{l,k}}. \tag{27}$$

Thus, from Lemma 2.1 it follows that for a single accident year i,

$$\begin{aligned}
\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(U_i, \widehat{U}_i) &= \mathrm{Var}(U_i \mid \mathcal{F}_0)(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\sigma}}^2) + \nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0) \nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0) \\
&= \widehat{U}_i^2 (\widehat{\Gamma}_{i,J}^U + \widehat{\Delta}_{i,J}^U) \\
&= \widehat{U}_i^2 \sum_{k=J-i+1}^{J-1} \frac{\widehat{\sigma}_k^2}{\widehat{f}_k^2} \Big(\frac{1}{\widehat{C}_{i,k}} + \frac{1}{\sum_{l=i_0}^{J-k} C_{l,k}} \Big),
\end{aligned}$$

which is equivalent to (Mack, 1993, Thm. 3). We state this result together with the corresponding result for the total ultimate claim amount in the following proposition:

Proposition 5.1. *In the setting of Mack's distribution-free chain ladder,*

$$\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(U_i, \widehat{U}_i) = \widehat{U}_i^2(\widehat{\Gamma}_{i,J}^U + \widehat{\Delta}_{i,J}^U)$$

$$= \widehat{U}_i^2 \sum_{k=J-i+1}^{J-1} \frac{\widehat{\sigma}_k^2}{\widehat{f}_k^2}\Big(\frac{1}{\widehat{C}_{i,k}} + \frac{1}{\sum_{l=i_0}^{J-k} C_{l,k}}\Big),$$

$$\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(U, \widehat{U}) = \sum_{i=2}^{J} \widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(U_i, \widehat{U}_i) + 2 \sum_{2 \le i < k \le J} \widehat{U}_i \widehat{U}_k \widehat{\Delta}_{i,J}^U,$$

where $\widehat{\Gamma}_{i,J}^U$ is given by (24) and $\widehat{\Delta}_{i,J}^U$ is given by (27).

The remaining part of the proof is given in Appendix C and amounts, due to Lemma 2.1, to identifying $Q_{i,k}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) = \widehat{U}_i \widehat{U}_k \widehat{\Delta}_{i,J}^U$ and noting that all covariances are 0.

By comparing Proposition 5.1 with Mack's estimator in (Mack, 1993, Thm. 3, Cor.) for the chain ladder model it is clear that the formulas coincide. Moreover, following the discussion in Section 4.1 it is clear from Propositions 4.2 and B.1 that

(i) the conditional specification of $\widehat{\boldsymbol{f}}^*$ provides an unbiased estimator of the computationally intractable unconditional (co)variances of the parameter estimators,

(ii) the two covariance specifications are asymptotically equal.

The figures in Appendix D illustrate the differences between conditional MSEP estimates when using the conditional specification and when using the unconditional specification of $\widehat{\boldsymbol{f}}^*$. The illustrations are based on simulations and data from Mack (1993). It is seen that the disfferences are essentially indistinguishable.

Before ending the discussion of conditional MSEP estimation for the ultimate claim amount, recall that the conditional MSEP can be split into one process variance part and one estimation error part. In Mack (1993) all process variances are calculated without using any approximations, and the estimation error is calculated exactly up until a final step where, p. 219 in Mack (1993), "*...we replace S_k^2 with $E(S_k^2 \mid B_k)$ and $S_j S_k, j < k$, with $E(S_j S_k \mid B_k)$*". This last step may, as noted already in Buchwalder et al. (2006), be seen as a specific choice of $\widehat{\boldsymbol{f}}^*$, following the general approach in the present paper. Given this specific choice of $\widehat{\boldsymbol{f}}^*$, the calculations carried out in Mack (1993) are exact. However, the implicit choice of $\widehat{\boldsymbol{f}}^*$ used in Mack (1993) is different from the one used in the present paper, since

Proposition 5.1 relies on a certain Taylor approximation. In Buchwalder et al. (2006) an exact MSEP calculation for the ultimate claim amount is carried out using a choice of $\widehat{\boldsymbol{f}}^*$ which is identical with that used in the present paper. Moreover, from the calculations in Buchwalder et al. (2006) it is clear that the Taylor approximation used in Proposition 5.1 will result in under estimation, w.r.t. the specific choice of $\widehat{\boldsymbol{f}}^*$ used in the current paper. For further details, see Buchwalder et al. (2006) as well as the discussion in Mack et al. (2006).

We will now provide the necessary building blocks needed in order to be able to arrive at the estimator of conditional MSEP for the CDR following Section 3.2 using Definition 2.3. This will be done using the same notion of conditional MSEP for both the ultimate claim amount and for CDR which, as introduced in Section 2, is the $\mathcal{F}_0$-conditional expectation of the squared distance between a random variable and its $\mathcal{F}_0$-measurable predictor, as well as the same estimation procedures.

We now proceed with the derivation of the estimator of conditional MSEP for the CDR in the chain ladder setting, in complete analogy with the corresponding derivation of the estimator of conditional MSEP for the ultimate claim amount. Note that many of the partial results needed for the computation of our suggested estimator of conditional MSEP for the CDR can be found in Merz and Wüthrich (2007); Wüthrich and Merz (2008a); Wüthrich et al. (2009). The results in the mentioned papers do, however, use a different indexation than that used in Mack (1993), which is the indexation used in the present paper. Due to this, we have rephrased all results for the CDR-calculations in terms of the indexation used in Mack (1993).

As before, let $h(\boldsymbol{\theta}; \mathcal{F}_0)$ denote the theoretical predictor, but now w.r.t. CDR:

$$h_i(\boldsymbol{f}; \mathcal{F}_0) := \mathbb{E}[\text{CDR}_i \mid \mathcal{F}_0], \quad h(\boldsymbol{f}; \mathcal{F}_0) := \mathbb{E}[\text{CDR} \mid \mathcal{F}_0] = \sum_{i=2}^{J} h_i(\boldsymbol{f}; \mathcal{F}_0).$$

It follows from Lemma 3.3 in Wüthrich and Merz (2008a) that

$$h_i(\boldsymbol{f}; \mathcal{F}_0) = C_{i,J-i+1} \Big(\prod_{j=J-i+1}^{J-1} \widehat{f}_j - f_{J-i+1} \prod_{j=J-i+2}^{J-1} \Big(\frac{S_j^0}{S_j^1} \widehat{f}_j + f_j \frac{C_{J-j+1,j}}{S_j^1} \Big) \Big),$$

(28)

where $S_j^k = \sum_{i=i_0}^{J-j+k} C_{i,j}$ for $k = 0, 1$. Notice that, $h(\widehat{\boldsymbol{f}}; \mathcal{F}_0) = 0$ and consequently it follows that

$$\text{MSEP}_{\mathcal{F}_0}(\text{CDR}, \widehat{\text{CDR}}) = \text{MSEP}_{\mathcal{F}_0}(\text{CDR}, 0),$$

which is referred to as the "observable" CDR in e.g. Wüthrich et al. (2009).

In order to calculate conditional MSEP for the CDR, we again make use of Lemma 2.1. The plug-in estimator of the process variance for a single accident year, one of the two terms of the estimator of conditional MSEP, is derived in Wüthrich et al. (2009), see (Wüthrich and Merz, 2008b, Result 3.3, Eq. (3.17)):

$$\mathrm{Var}(\mathrm{CDR}_i \mid \mathcal{F}_0)(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\sigma}}^2) := \widehat{U}_i^2 \widehat{\Gamma}_{i,J}^{\mathrm{CDR}}, \tag{29}$$

where

$$\widehat{\Gamma}_{i,J}^{\mathrm{CDR}} := \left(\left(1 + \frac{\widehat{\sigma}_{J-i+1}^2}{\widehat{f}_{J-i+1}^2 C_{i,J-i+1}}\right) \prod_{j=J-i+2}^{J-1} \left(1 + \frac{\widehat{\sigma}_j^2}{\widehat{f}_j^2 C_{J-j+1,j}} \left(\frac{C_{J-j+1,j}}{S_j^1}\right)^2\right)\right) - 1. \tag{30}$$

The process variance for all accident years is given by

$$\mathrm{Var}(\mathrm{CDR} \mid \mathcal{F}_0)(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\sigma}}^2) := \sum_{i=2}^J \mathrm{Var}(\mathrm{CDR}_i \mid \mathcal{F}_0)(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\sigma}}^2)$$
$$+ 2 \sum_{2 \le i < k \le J} \widehat{U}_i \widehat{U}_k \widehat{\Xi}_{i,J}^{\mathrm{CDR}}, \tag{31}$$

where

$$\widehat{\Xi}_{i,J}^{\mathrm{CDR}} := \left(\left(1 + \frac{\widehat{\sigma}_{J-i+1}^2}{\widehat{f}_{J-i+1}^2 S_{J-i+1}^1}\right) \prod_{j=J-i+2}^{J-1} \left(1 + \frac{\widehat{\sigma}_j^2}{\widehat{f}_j^2 C_{J-j+1,j}} \left(\frac{C_{J-j+1,j}}{S_j^1}\right)^2\right)\right) - 1 \tag{32}$$

which follows from (Wüthrich et al., 2009, Result 3.3, Eq. (3.18)). Notice that $\widehat{U}_i \widehat{U}_k \widehat{\Xi}_{i,J}^{\mathrm{CDR}}$ corresponds to covariance terms, which did not appear in the calculation of the process variance for the ultimate claim amount due to independence between accident years.

Further, based on Lemma 2.1, what remains to be determined are the $Q_{i,j}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)$s. From the definition of $h(\boldsymbol{f}; \mathcal{F}_0)$ above, it immediately follows that

$$\{\nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)\}_j = \begin{cases} -C_{i,J-i+1} \frac{C_{J-j+1,j}}{\widehat{f}_j S_j^1} \prod_{l=J-i+2}^{J-1} \widehat{f}_l, & j > J-i+1, \\ -C_{i,J-i+1} \prod_{l=J-i+2}^{J-1} \widehat{f}_l, & j = J-i+1, \end{cases}$$

which may be written as

$$\{\nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)\}_j = \begin{cases} -\widehat{U}_i \frac{C_{J-j+1,j}}{\widehat{f}_j S_j^1}, & j > J-i+1, \\ -\widehat{U}_i \frac{1}{\widehat{f}_{J-i+1}}, & j = J-i+1. \end{cases} \tag{33}$$

28

Hence, it follows that

$$Q_{i,i}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) = \nabla h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0) \nabla h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)$$

$$= \widehat{U}_i^2 \Big(\frac{\widehat{\sigma}_{J-i+1}^2}{\widehat{f}_{J-i+1}^2 S_{J-i+1}^0} + \sum_{j=J-i+2}^{J-1} \frac{\widehat{\sigma}_j^2}{\widehat{f}_j^2 S_j^0} \Big(\frac{C_{J-j+1,j}}{S_j^1} \Big)^2 \Big),$$

where

$$\mathrm{Var}(\widehat{f}_j \mid B_j) = \frac{\sigma_j^2}{S_j^0}.$$

and $\boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0)$ is diagonal with jth diagonal element $\widehat{\sigma}_j^2 / S_j^0$. If we set

$$\widehat{\Delta}_{i,J}^{\mathrm{CDR}} := \frac{\widehat{\sigma}_{J-i+1}^2}{\widehat{f}_{J-i+1}^2 S_{J-i+1}^0} + \sum_{j=J-i+2}^{J-1} \frac{\widehat{\sigma}_j^2}{\widehat{f}_j^2 S_j^0} \Big(\frac{C_{J-j+1,j}}{S_j^1} \Big)^2, \qquad (34)$$

which corresponds to (Wüthrich and Merz, 2008a, Eq. (3.4)), then

$$\nabla h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0) \nabla h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0) = \widehat{U}_i^2 \widehat{\Delta}_{i,J}^{\mathrm{CDR}}.$$

Combining the above, using Lemma 2.1, gives that $\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(\mathrm{CDR}_i, \widehat{\mathrm{CDR}}_i)$, given by Definition 2.3, simplifies to

$$\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(\mathrm{CDR}_i, \widehat{\mathrm{CDR}}_i) := \mathrm{Var}(\mathrm{CDR}_i \mid \mathcal{F}_0)(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\sigma}}^2) + Q_{i,i}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)$$

$$= \widehat{U}_i^2 \Big(\widehat{\Gamma}_{i,J}^{\mathrm{CDR}} + \widehat{\Delta}_{i,J}^{\mathrm{CDR}} \Big). \qquad (35)$$

Note that by using the linearisation of the process variance used in (Wüthrich and Merz, 2008a, Eq. (A.1)) it follows that

$$\widehat{\Gamma}_{i,J}^{\mathrm{CDR}} \approx \frac{\widehat{\sigma}_{J-i+1}^2}{\widehat{f}_{J-i+1}^2 C_{i,J-i+1}} + \sum_{j=J-i+2}^{J-1} \frac{\widehat{\sigma}_j^2}{\widehat{f}_j^2 C_{J-j+1,j}} \Big(\frac{C_{J-j+1,j}}{S_j^1} \Big)^2,$$

it in turn follows that (35) reduces to Result 3.1, Eq. (3.9), in Wüthrich and Merz (2008a). Notice that our estimator of conditional MSEP coincides with that in Wüthrich and Merz (2008a) despite the quite different logics of the two approaches for deriving the estimator. The derivation of Result 3.1 in Wüthrich and Merz (2008a) is based on perturbing the **initial** $\widehat{f}_j$s, i.e. the $\widehat{f}_j^{(0)}$, that in our setting are a part of the basis of prediction and therefore may not be perturbed. That the two approaches give estimators that coincide is due to the underlying symmetry $\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(\mathrm{CDR}_i, \widehat{\mathrm{CDR}}_i) =$

$\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(\widehat{\mathrm{CDR}}_i, \mathrm{CDR}_i)$ and the fact that the CDR-quantities are multilinear in the model parameters.

Furthermore, the MSEP calculations for the CDR aggregated over all accident years follow the same steps as those used for the derivation of the corresponding MSEP calculations for the ultimate claim amount verbatim. The only resulting difference is the necessity to keep track of covariance terms across accident years. That is, we will get contributions of the form

$$
\begin{aligned}
Q_{i,k}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) &= \nabla h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0) \nabla h_k(\widehat{\boldsymbol{f}}; \mathcal{F}_0) \\
&= \widehat{U}_i \widehat{U}_k \Big(\frac{\widehat{\sigma}_{J-i+1}^2 C_{i,J-i+1}}{\widehat{f}_{J-i+1}^2 S_{J-i+1}^0 S_{J-i+1}^1} + \sum_{j=J-i+2}^{J-1} \frac{\widehat{\sigma}_j^2}{\widehat{f}_j^2 S_j^0} \Big(\frac{C_{J-j+1,j}}{S_j^1} \Big)^2 \Big)
\end{aligned}
$$

when $i < k$, which by introducing

$$
\widehat{\chi}_{i,J}^{\mathrm{CDR}} := \frac{\widehat{\sigma}_{J-i+1}^2 C_{i,J-i+1}}{\widehat{f}_{J-i+1}^2 S_{J-i+1}^0 S_{J-i+1}^1} + \sum_{j=J-i+2}^{J-1} \frac{\widehat{\sigma}_j^2}{\widehat{f}_j^2 S_j^0} \Big(\frac{C_{J-j+1,j}}{S_j^1} \Big)^2, \qquad (36)
$$

allows us to summarize the results obtained in the following proposition:

Proposition 5.2. *In the setting of Mack's distribution-free chain ladder,*

$$
\begin{aligned}
\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(\mathrm{CDR}_i, \widehat{\mathrm{CDR}}_i) &:= \widehat{U}_i^2 \big(\widehat{\Gamma}_{i,J}^{\mathrm{CDR}} + \widehat{\Delta}_{i,J}^{\mathrm{CDR}} \big), \\
\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(\mathrm{CDR}, \widehat{\mathrm{CDR}}) &:= \sum_{i=2}^{J} \widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(\mathrm{CDR}_i, \widehat{\mathrm{CDR}}_i) \\
&\quad + 2 \sum_{2 \le i < k \le J} \widehat{U}_i \widehat{U}_k \big(\widehat{\Xi}_{i,J}^{\mathrm{CDR}} + \widehat{\chi}_{i,J}^{\mathrm{CDR}} \big),
\end{aligned}
$$

where $\widehat{\Gamma}_{i,J}^{\mathrm{CDR}}$, $\widehat{\Delta}_{i,J}^{\mathrm{CDR}}$, $\widehat{\Xi}_{i,J}^{\mathrm{CDR}}$ and $\widehat{\chi}_{i,J}^{\mathrm{CDR}}$ are given by (30), (34), (32) and (36), respectively.

As noted in the discussion leading up to Proposition 5.2, the proof is identical to that of Proposition 5.1 in all aspects, except for the covariance terms, see Appendix C for details. Again, in analogy with the situation for a single accident year, using the process (co)variance approximation following (Wüthrich and Merz, 2008a, Eq. (A.1)), it is seen that Proposition 5.2 will coincide with Result 3.3 in Wüthrich and Merz (2008a). Even though the results from Proposition 5.2, given the mentioned approximation, will coincide with those obtained in (Wüthrich and Merz, 2008a, Result 3.3), the underlying estimation procedures differ. The procedure advocated here

for the CDR is consistent with that for the ultimate claim amount and is straightforward to apply.

As mentioned in Section 3.2, the primary purpose with the current section was to illustrate how the introduced methods can be applied to different functions of the future development of the underlying stochastic process — here the ultimate claim amount and the CDR. In the next, and final, section, we illustrate how the general approach to calculate conditional MSEP introduced in the present paper applies to other reserving methods.

6   Applications to non-sequential reserving models

We will now demonstrate that the general approach to estimation of conditional MSEP presented in Section 2 also applies when the model is quite different from the sequential conditional linear models considered in Section 4. We will show how to compute conditional MSEP estimates for the ultimate claim amount for the over-dispersed Poisson chain ladder model, see e.g. Mack (1991); England and Verrall (1999). The overdispersed Poisson chain ladder model is based on the following assumptions:

$$\mathbb{E}[I_{i,j}] = \mu_{i,j}, \quad \mathrm{Var}(I_{i,j}) = \phi\mu_{i,j}, \quad \log(\mu_{i,j}) = \eta + \alpha_i + \beta_j,$$

where $i, j = 1, \ldots, J$ and $\alpha_1 = \beta_1 = 0$. The model parameters may be estimated using standard quasi-likelihood theory and the natural predictor for the ultimate claim amount for accident year i is given by

$$h_i(\boldsymbol{\theta}; \mathcal{F}_0) = \mathbb{E}[U_i \mid \mathcal{F}_0] = C_{i,J-i} + \sum_{j=J-i+1}^{J} \mu_{i,j} = C_{i,J-i} + g(\boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = (\eta, \{\alpha_i\}, \{\beta_k\})$. We may use Lemma 2.1 to calculate conditional MSEP for the ultimate claim amount. Firstly, due to independence across all indices,

$$\mathrm{Var}(U_i \mid \mathcal{F}_0) = \mathrm{Var}(R_i) = \phi \sum_{j=J-i+1}^{J} \mu_{i,j}, \quad \mathrm{Var}(U \mid \mathcal{F}_0) = \sum_{i=2}^{J} \mathrm{Var}(R_i).$$

31

Secondly, in order to determine the $Q_{i,j}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)$s we need the partial derivatives of $h_i(\boldsymbol{\theta}; \mathcal{F}_0)$ which are given by

$$\frac{\partial}{\partial \eta} h_i(\boldsymbol{\theta}; \mathcal{F}_0) = \frac{\partial}{\partial \eta} g_i(\boldsymbol{\theta}) = \sum_{j=I-i+1}^{J} \mu_{i,j},$$

$$\frac{\partial}{\partial \alpha_k} h_i(\boldsymbol{\theta}; \mathcal{F}_0) = \frac{\partial}{\partial \alpha_k} g_i(\boldsymbol{\theta}) = \sum_{j=I-k+1}^{J} \mu_{k,j},$$

$$\frac{\partial}{\partial \beta_k} h_i(\boldsymbol{\theta}; \mathcal{F}_0) = \frac{\partial}{\partial \beta_k} g_i(\boldsymbol{\theta}) = 1_{\{I-i+1 \leq k\}} \mu_{i,k}.$$

Hence,

$$\nabla h(\boldsymbol{\theta}; \mathcal{F}_0) = \nabla g(\boldsymbol{\theta}), \quad \nabla h_i(\boldsymbol{\theta}; \mathcal{F}_0) = \nabla g_i(\boldsymbol{\theta}), \quad i = 1, \ldots, J$$

are independent of $\mathcal{F}_0$, and in particular

$$Q_{i,k}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) = Q_{i,k}(\widehat{\boldsymbol{\theta}}) = \nabla g_i(\widehat{\boldsymbol{\theta}})' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}) \nabla g_k(\widehat{\boldsymbol{\theta}}).$$

By combining the above relations together with Lemma 2.1 it follows that the estimator of conditional MSEP in Definition 2.3, applied to the ultimate claim amount, is given by

$$\widehat{\text{MSEP}}_{\mathcal{F}_0}(U, \widehat{U}) = \widehat{\text{MSEP}}(R, \widehat{R})$$

and takes the form

$$\sum_{i=2}^{J} \text{Var}(R_i)(\widehat{\boldsymbol{\theta}}) + \sum_{i=2}^{J} \nabla g_i(\widehat{\boldsymbol{\theta}})' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}) \nabla g_i(\widehat{\boldsymbol{\theta}}) + 2 \sum_{2 \leq i < k \leq J} \nabla g_i(\widehat{\boldsymbol{\theta}})' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}) \nabla g_k(\widehat{\boldsymbol{\theta}}).$$
$$(37)$$

What remains for having a computable estimator of conditional MSEP for the ultimate claim amount is to compute the covariance matrix $\boldsymbol{\Lambda}(\boldsymbol{\theta}) = \text{Cov}(\widehat{\boldsymbol{\theta}})$. Notice that the estimator (37) corresponds to the general conditional MSEP estimator upon choosing $\widehat{\boldsymbol{\theta}}^*$ as an independent copy $\widehat{\boldsymbol{\theta}}^{\perp}$ of $\widehat{\boldsymbol{\theta}}$, which gives

$$\boldsymbol{\Lambda}(\boldsymbol{\theta}; \mathcal{F}_0) := \text{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0) = \text{Cov}(\widehat{\boldsymbol{\theta}}).$$

Notice also since the overdispersed Poisson chain ladder model relies on quasi-likelihood theory we do not have access to an explicit expression for

the covariance of the parameter estimators. However, no such explicit expression is needed since a numerical approximation is easily obtained as output of a standard quasi-Poisson GLM-fit. That is, using standard numerical procedures for approximating the covariance matrix, e.g. GLM-fitting procedures, one obtains a non-simulation based procedure for estimation of the conditional MSEP for the ultimate claim amount. Further, since quasi-likelihood estimators are M-estimators, see e.g. (Van der Vaart, 2000, Ch. 5), these can be shown to be consistent given certain regularity conditions. This motivates neglecting possible bias when using Definition 2.3. Another alternative is, of course, to introduce a bias correction, see e.g. Lindholm et al. (2017). Another observation concerning the conditional MSEP estimator (37) for the overdispersed Poisson chain ladder model is the following:

Proposition 6.1. *The estimator* (37) *of conditional* MSEP *for the ultimate claim amount for the overdispersed Poisson chain ladder model coincides with the one derived in (Renshaw, 1994, Sec. 4.4).*

The proof follows by noting that all $\nabla g_i(\widehat{\boldsymbol{\theta}})$ are functions of $\nabla \mu_{i,j}(\widehat{\boldsymbol{\theta}})$s and

$$\nabla \mu_{i,j}(\widehat{\boldsymbol{\theta}})' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}) \nabla \mu_{k,l}(\widehat{\boldsymbol{\theta}}) = \mu_{i,j}(\widehat{\boldsymbol{\theta}}) \operatorname{Cov}(\widehat{\eta}_{i,j}, \widehat{\eta}_{k,l}) \mu_{k,l}(\widehat{\boldsymbol{\theta}}),$$

where $\eta_{i,j} := \log(\mu_{i,j})$. See also (England and Verrall, 1999, Eq. (3.4) & (3.5)).

Notice that due to Lemma 2.1 the semi-analytical estimator (37) is valid for any non-sequential GLM-based reserving model.

The above example of calculating a semi-analytical expression for the estimator of conditional MSEP for the ultimate claim amount according to Definition 2.3 for the overdispersed Poisson chain ladder model can of course be extended to more complex models as long as it is possible to compute

(i) $h(\boldsymbol{\theta}; \mathcal{F}_0)$ together with its partial derivatives,

(ii) (an approximation of) a suitable, conditional or unconditional, covariance matrix of $\widehat{\boldsymbol{\theta}}$.

One example of a more complex GLM-based reserving model is the one introduced in Verrall et al. (2010), which is based on one triangle for observed counts and one triangle for incremental payments. In this model the counts are modelled as an overdispersed Poisson chain ladder model, and the incremental payments are modelled as a quasi-Poisson GLM model conditional on counts. Due to the overall quasi-Poisson structure of the model it is possible to obtain explicit expressions for the predictor of the ultimate claim

amount, together with the corresponding process variance, but where $\mathcal{F}_0$ now also contains information concerning observed counts. The conditional MSEP for the ultimate claim amount can again be calculated using Lemma 2.1.

Furthermore, the general exposition of the methods introduced in the present paper do not rely on that the data generating process is defined in terms of runoff triangles. Examples of another type of models are the continuous time point process models treated in e.g. Norberg (1993); Antonio and Plat (2014). These models rely on extensive stochastic simulations in order to be used in practice. One simple example of a special case of a point process model for which the quantities needed for the calculation of a semi-analytical MSEP estimator for the ultimate claim amount according to Definition 2.3 is possible is the model described in Section 8.A in Norberg (1993). Hence, it is again possible to use Lemma 2.1 to calculate the conditional MSEP of the ultimate claim amount.

The above examples provide semi-analytical MSEP estimators which only rely on that we are able to calculate certain expected values and (co)variances. One advantage of this approach is that there is no need for simulation based techniques in order to carry out MSEP calculations.

7 Acknowledgment

References

Akaike, H. (1969). Fitting autoregressive models for prediction. *Ann. Inst. Statist. Math.*, 21:243–247.

Akaike, H. (1970). Statistical predictor identification. *Ann. Inst. Statist. Math.*, 22:203–217.

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

Bhansali, R. J. and Downham, D. Y. (1977). Some properties of the order of an autoregressive model selected by a generalization of Akaike's FPE criterion. *Biometrika*, 64(3):547–551.

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

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

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap.* CRC press.

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

Gisler, A. (2006). The estimation error in the chain-ladder reserving method: a Bayesian approach. *ASTIN Bulletin*, 36(2):554–565.

Kallenberg, O. (2002). *Foundations of Modern Probability.* Springer.

Kremer, E. (1984). A class of autoregressive models for predicting the final claims amount. *Insurance: Mathematics and Economics*, 3(2):111–119.

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

Mack, T. (1991). A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bulletin*, 21(1):93–109.

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

Mack, T., Quarg, G., and Braun, C. (2006). The mean square error of prediction in the chain ladder reserving method–a comment. *ASTIN Bulletin*, 36(2):543–552.

Merz, M. and Wüthrich, M. V. (2007). Prediction error of the expected claims development result in the chain ladder method. *Bulletin of Swiss Association of Actuaries*, 1(2007):117–137.

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

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

Renshaw, A. E. (1994). On the second moment properties and the implementation of certain GLIM based stochastic claims reserving models. City University Actuarial Research Paper No. 65.

Röhr, A. (2016). Chain ladder and error propagation. *ASTIN Bulletin*, 46(2):293–330.

Seber, G. A. F. and Lee, A. J. (2003). *Linear regression analysis, Second Ed.* John Wiley & Sons.

Speed, T. P. and Yu, B. (1993). Model selection and prediction: normal regression. *Ann. Inst. Statist. Math.*, 45(1):35–54.

Taylor, G. C. and Ashe, F. (1983). Second moments of estimates of outstanding claims. *Journal of Econometrics*, 23(1):37–61.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

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

Wüthrich, M. V. and Merz, M. (2008a). Modelling the claims development result for solvency purposes. In *Casualty Actuarial Society E-Forum*, pages 542–568.

Wüthrich, M. V. and Merz, M. (2008b). *Stochastic claims reserving methods in insurance*, volume 435. John Wiley & Sons.

Wüthrich, M. V., Merz, M., and Lysenko, N. (2009). Uncertainty of the claims development result in the chain ladder method. *Scandinavian Actuarial Journal*, 2009(1):63–84.

A   Special cases of the model class from Section 4

Here we present assumptions that may be imposed on the structure of the conditional mean values in the general development-year dynamics (11). Model assumptions prescribing autoregressive structures for the conditional means are commonly encountered and enable explicit calculations.

Assumption A.1 (Cumulative model). *For $(i,j) \in \mathcal{I} \times \mathcal{J}$, $Y_{i,j} = C_{i,j}$ and*

$$C_{i,j+1} = \beta_{j,1} C_{i,j} + \cdots + \beta_{j,p_j} C_{i,j-p_j+1} + \sigma_j \{\boldsymbol{\Sigma}_j\}_{i,i}^{1/2} \{\boldsymbol{e}_j\}_i, \quad p_j \leq j, \quad (38)$$

or

$$C_{i,j+1} = \beta_{j,1} + \beta_{j,2} C_{i,j} + \cdots + \beta_{j,p_j} C_{i,j-p_j+2} + \sigma_j \{\boldsymbol{\Sigma}_j\}_{i,i}^{1/2} \{\boldsymbol{e}_j\}_i, \quad p_j \leq j+1. \tag{39}$$

Assumption A.2 (Incremental model). *For $(i,j) \in \mathcal{I} \times \mathcal{J}$, $Y_{i,j} = I_{i,j}$ and*

$$I_{i,j+1} = \beta_{j,1} I_{i,j} + \cdots + \beta_{j,p_j} I_{i,j-p_j+1} + \sigma_j \{\boldsymbol{\Sigma}_j\}_{i,i}^{1/2} \{\boldsymbol{e}_j\}_i, \quad p_j \leq j, \quad (40)$$

or

$$I_{i,j+1} = \beta_{j,1} + \beta_{j,2} I_{i,j} + \cdots + \beta_{j,p_j} I_{i,j-p_j+2} + \sigma_j \{\boldsymbol{\Sigma}_j\}_{i,i}^{1/2} \{\boldsymbol{e}_j\}_i, \quad p_j \leq j+1. \tag{41}$$

Remark 3. *The models with intercepts defined by (39) and (41) require that the payment data is normalized by an exposure measure before any statistical analysis. The normalization may correspond to dividing all payments stemming from a given accident year by the number of written insurance contracts that accident year.*

Remark 4. *Under Assumption A.1, using the tower property of conditional expectations,*

$$\mathbb{E}[U \mid \mathcal{F}_0] = \sum_{i=2}^{J} \mathbb{E}[C_{i,J} \mid \mathcal{F}_0] = \sum_{i=2}^{J} \left(a_{i,0} + \sum_{j=1}^{J-i+1} a_{i,j} C_{i,j} \right),$$

where each coefficient $a_{i,j}$ is either 0 or a finite product of distinct β-parameters β_{jk} for $j \in \{1, \ldots, J-1\}$ and $k \in \{1, \ldots, p_j\}$. In particular, $\mathbb{E}[U \mid \mathcal{F}_0]$ is an $\mathcal{F}_0$-measurable multi-affine function in the parameters β_{jk}, an expression of the form $c + d\beta_{jk}$. Under Assumption A.2, using the tower property of conditional expectations,

$$\mathbb{E}[U \mid \mathcal{F}_0] = \sum_{i=2}^{J} \left(\sum_{j=1}^{J-i+1} I_{i,j} + \sum_{j=J-i+2}^{J} \mathbb{E}[I_{i,j} \mid \mathcal{F}_0] \right)$$

$$= \sum_{i=2}^{J} \left(\sum_{j=1}^{J-i+1} I_{i,j} + b_{i,0} + \sum_{j=1}^{J-i+1} b_{i,j} I_{i,j} \right),$$

where each coefficient $b_{i,j}$ is either 0 or a finite product of distinct β-parameters β_{jk} for $j \in \{1, \ldots, J-1\}$ and $k \in \{1, \ldots, p_j\}$. In particular, $\mathbb{E}[U \mid \mathcal{F}_0]$ is again an $\mathcal{F}_0$-measurable multi-affine function in the parameters β_{jk}, an expression of the form $c + d\beta_{jk}$.

It is clear that each of Assumption A.1 and A.2 implies that the statement in Assumption 4.1 holds.

B Asymptotic properties of conditional weighted least squares estimators

The following result motivates the approximation of $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_j)$ by $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_j \mid B_j)$, and hence also the approximation of $\mathrm{Cov}(\widehat{\boldsymbol{\beta}})$ by $\boldsymbol{\Lambda}(\boldsymbol{\sigma}; \mathcal{F}_0)$, by asymptotic arguments, corresponding to letting the number of accident years in the available data set tend to infinity.

Proposition B.1. *Let $|\mathcal{I}| := J - i_0$ be the number of fully or partially developed accident years. For $j \in \{1, \ldots, J-1\}$, suppose the following statements hold:*

(i) For $i, k \in \{1, \ldots, p_j\}$, $\{|\mathcal{I}|\{(\boldsymbol{A}_j' \boldsymbol{\Sigma}_j^{-1} \boldsymbol{A}_j)^{-1}\}_{i,k}\}_{|\mathcal{I}|}$ is uniformly integrable.

(ii) For $i, k \in \{1, \ldots, p_j\}$,

$$\lim_{|\mathcal{I}| \to \infty} \sup_{l \leq |\mathcal{I}|} \mathrm{Var}\left(\{\boldsymbol{\Sigma}_j^{-1}\}_{l,l} \{\boldsymbol{A}_j\}_{l,i} \{\boldsymbol{A}_j\}_{l,k}\right) < \infty.$$

(iii) There exists an invertible $p_j \times p_j$ matrix $\boldsymbol{\nu}_j$ such that

$$\lim_{|\mathcal{I}| \to \infty} \frac{1}{|\mathcal{I}|} \mathbb{E}[\boldsymbol{A}_j' \boldsymbol{\Sigma}_j^{-1} \boldsymbol{A}_j] = \boldsymbol{\nu}_j.$$

Then $\lim_{|\mathcal{I}| \to \infty} |\mathcal{I}| \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_j) = \boldsymbol{\nu}_j^{-1}$ and $|\mathcal{I}| \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_j \mid B_j) \overset{a.s.}{\to} \boldsymbol{\nu}_j^{-1}$ as $|\mathcal{I}| \to \infty$.

The proof of Proposition B.1 is given in the appendix and relies on that the conditional covariance may be written in the form of weighted sums of independent random variables.

Remark 5. *Conditions (i)-(iii) are technical conditions that can be verified given additional mild assumptions, essentially existence of higher order moments, on the development-year dynamics in (11). The conditions can be simplified if it is assumed that the development-year dynamics for different*

accident years are identical, corresponding to identically distributed rows for $\boldsymbol{A}_j$ *and* $\boldsymbol{\Sigma}_j$. *Condition (iii) is equivalent to the existence of an invertible* $p_j \times p_j$ *matrix* $\boldsymbol{\nu}_j$ *such that*

$$\lim_{|\mathcal{I}| \to \infty} \frac{1}{|\mathcal{I}|} \sum_{l=1}^{|\mathcal{I}|} \mathbb{E}[\{\boldsymbol{\Sigma}_j^{-1}\}_{ll}\{\boldsymbol{A}_j\}_{li}\{\boldsymbol{A}_j\}_{lk}] = \{\boldsymbol{\nu}_j\}_{ik}.$$

If the rows of $\boldsymbol{A}_j$ *and* $\boldsymbol{\Sigma}_j$ *are identically distributed, then*

$$\frac{1}{|\mathcal{I}|} \sum_{l=1}^{|\mathcal{I}|} \mathbb{E}[\{\boldsymbol{\Sigma}_j^{-1}\}_{ll}\{\boldsymbol{A}_j\}_{li}\{\boldsymbol{A}_j\}_{lk}] = \mathbb{E}[\{\boldsymbol{\Sigma}_j^{-1}\}_{11}\{\boldsymbol{A}_j\}_{1i}\{\boldsymbol{A}_j\}_{1k}]$$

so (iii) automatically holds if the $p_j \times 1$ *vector* $\{\boldsymbol{\Sigma}_j^{-1/2}\}_{11}\{\boldsymbol{A}_j\}'_{1.}$ *has an invertible covariance matrix.*

Remark 6. *Proposition B.1 provides the asymptotic behavior of* $\mathrm{Cov}(\widehat{\boldsymbol{\beta}})$ *and* $\boldsymbol{\Lambda}(\boldsymbol{\sigma}; \mathcal{F}_0)$ *as the number of accident years in the available data set tends to infinity. Proposition B.1 can be extended to also address the asymptotic behavior of* $\boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0)$ *by considering conditions ensuring consistency and a certain rate of convergence for the estimators* $\widehat{\sigma}_j^2$. *We will not analyze such conditions in this paper.*

Combining Markov's inequality and Propositions 4.1 and B.1 immediately gives consistency of the weighted least-squares estimator $\widehat{\boldsymbol{\beta}}$ as the number of fully or partially developed accident years tends to infinity: Moreover, combining Proposition B.1 with either Assumption A.1 or A.2 allows the asymptotic behavior of the term in Definition 2.3 accounting for estimation error to be analyzed. We state these facts as a corollary to Proposition B.1:

Corollary B.1. *Let* $|\mathcal{I}| := J - i_0$ *denote the number of fully or partially developed accident years. If the conditions of Proposition B.1 hold, then* $\widehat{\boldsymbol{\beta}} \overset{\mathbb{P}}{\to} \boldsymbol{\beta}$ *as* $|\mathcal{I}| \to \infty$. *Moreover, if in addition either Assumption A.1 or A.2 holds, and* $\widehat{\sigma}_j^2 \overset{\mathbb{P}}{\to} \sigma_j^2$ *as* $|\mathcal{I}| \to \infty$ *for* $j = 1, \ldots, J - 1$, *then*

$$|\mathcal{I}| \nabla_{\boldsymbol{\beta}} h(\widehat{\boldsymbol{\beta}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0) \nabla_{\boldsymbol{\beta}} h(\widehat{\boldsymbol{\beta}}; \mathcal{F}_0) \overset{\mathbb{P}}{\to} c \quad as \ |\mathcal{I}| \to \infty$$

for some constant $c < \infty$.

C Proofs

Proof of Lemma 2.1. Recall from Definition 2.3 that it is possible to split the conditional MSEP approximation into a process variance part and an

estimation error part. Thus, given that $X = \sum_{i \in \mathcal{I}} X_i$, it follows that the process variance may be expressed as

$$\mathrm{Var}(X \mid \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}) = \sum_{i \in \mathcal{I}} \mathrm{Var}(X_i \mid \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}) + 2 \sum_{i,j \in \mathcal{I}, i < j} \mathrm{Cov}(X_i, X_j \mid \mathcal{F}_0)(\widehat{\boldsymbol{\theta}}),$$

and, if it in addition holds that $h(\boldsymbol{\theta}; \mathcal{F}_0) = \sum_{i \in \mathcal{I}} h_i(\boldsymbol{\theta}; \mathcal{F}_0)$, the estimation error part of (7) may be re-written according to

$$\nabla h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) \nabla h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) = \Big(\sum_{i \in \mathcal{I}} h_i(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) \Big) \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) \Big(\sum_{i \in \mathcal{I}} h_i(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) \Big)$$

$$= \sum_{i \in \mathcal{I}} h_i(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) h_i(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)$$

$$+ 2 \sum_{i,j \in \mathcal{I}, i < j} h_i(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) h_j(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0).$$

Lemma 2.1 follows by combining the above. $\qquad\square$

Proof of Proposition 4.1. Proof of Statement (i): By construction $\mathbb{E}[\widehat{\boldsymbol{\beta}}_j \mid B_j] = \boldsymbol{\beta}_j$. For $j < k$,

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_j, \widehat{\boldsymbol{\beta}}_k) = \mathbb{E}\Big[\mathrm{Cov}\big(\widehat{\boldsymbol{\beta}}_j, \widehat{\boldsymbol{\beta}}_k \mid B_k\big) \Big] + \mathrm{Cov}\Big(\mathbb{E}[\widehat{\boldsymbol{\beta}}_j \mid B_k], \mathbb{E}[\widehat{\boldsymbol{\beta}}_k \mid B_k] \Big)$$

$$= \mathbb{E}\Big[\mathrm{Cov}\big(\widehat{\boldsymbol{\beta}}_j, \widehat{\boldsymbol{\beta}}_k \mid B_k\big) \Big] + \mathrm{Cov}\Big(\boldsymbol{\beta}_j, \boldsymbol{\beta}_k \Big)$$

$$= \mathbb{E}\Big[\mathrm{Cov}\big(\widehat{\boldsymbol{\beta}}_j, \widehat{\boldsymbol{\beta}}_k \mid B_k\big) \Big]$$

and, since $\widehat{\boldsymbol{\beta}}_j$ is $\sigma(B_k)$-measurable,

$$\mathrm{Cov}\big(\widehat{\boldsymbol{\beta}}_j, \widehat{\boldsymbol{\beta}}_k \mid B_k\big) = \mathbb{E}\Big[\big(\widehat{\boldsymbol{\beta}}_j - \mathbb{E}[\widehat{\boldsymbol{\beta}}_j \mid B_k]\big)\big(\widehat{\boldsymbol{\beta}}_k - [\widehat{\boldsymbol{\beta}}_k \mid B_k]\big)' \mid B_k \Big] = 0.$$

Proof of Statement (ii): Let $\boldsymbol{Z}_{j+1} := \widetilde{\boldsymbol{\Sigma}}_j^{-1/2} \widetilde{\boldsymbol{Y}}_{j+1}$ and $\boldsymbol{C}_j := \widetilde{\boldsymbol{\Sigma}}_j^{-1/2} \widetilde{\boldsymbol{A}}_j$ and rewrite the weighted linear model (12) as $\boldsymbol{Z}_{j+1} = \boldsymbol{C}_j \boldsymbol{\beta}_j + \sigma_j \widetilde{\boldsymbol{e}}_j$. Notice that

$$\widehat{\sigma}_j^2 = \frac{1}{n_j - p_j} (\widetilde{\boldsymbol{Y}}_{j+1} - \widetilde{\boldsymbol{A}}_j \widehat{\boldsymbol{\beta}}_j)' \widetilde{\boldsymbol{\Sigma}}_j^{-1} (\widetilde{\boldsymbol{Y}}_{j+1} - \widetilde{\boldsymbol{A}}_j \widehat{\boldsymbol{\beta}}_j)$$

$$= \frac{1}{n_j - p_j} (\boldsymbol{Z}_{j+1} - \boldsymbol{C}_j \widehat{\boldsymbol{\beta}}_j)' (\boldsymbol{Z}_{j+1} - \boldsymbol{C}_j \widehat{\boldsymbol{\beta}}_j).$$

It now follows from Theorem 3.3 in Seber and Lee (2003) that $\mathbb{E}[\widehat{\sigma}_j^2 \mid B_j] = \sigma_j^2$ holds for $j = 1 \ldots, J - 1$ given that $i_0 \leq J - j - p_j$. $\qquad\square$

Proof of Proposition 4.2. Covariance decomposition together with (14) gives on the one hand

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_j) = \mathbb{E}\big[\,\mathrm{Cov}\,(\widehat{\boldsymbol{\beta}}_j \mid B_j)\big] + \mathrm{Cov}\,\big(\mathbb{E}[\widehat{\boldsymbol{\beta}}_j \mid B_j]\big) = \mathbb{E}\big[\,\mathrm{Cov}\,(\widehat{\boldsymbol{\beta}}_j \mid B_j)\big].$$

On the other hand, using Proposition 4.1 (ii), i.e. that $\mathbb{E}[\widehat{\sigma}_j^2 \mid B_j] = \sigma_j^2$,

$$\mathbb{E}\Big[\widehat{\sigma}_j^2\big(\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{A}}_j\big)^{-1}\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\widehat{\sigma}_j^2\big(\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{A}}_j\big)^{-1} \mid B_j\Big]\Big] = \mathbb{E}\big[\,\mathrm{Cov}\,(\widehat{\boldsymbol{\beta}}_j \mid B_j)\big].$$

Therefore, $\widehat{\sigma}_j^2(\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{A}}_j)^{-1}$ is an unbiased estimator of $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_j)$ and, since $\mathrm{Cov}(\widehat{\boldsymbol{\beta}})$ is block diagonal, $\boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0)$ is an unbiased estimator of $\mathrm{Cov}(\widehat{\boldsymbol{\beta}})$. $\quad\square$

Proof of Proposition B.1. The constant parameter σ_j is irrelevant for the argument of the proof and therefore here set to 1. Notice that, for $i, k \in \{1, \ldots, p_j\}$,

$$\{\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{A}}_j\}_{ik} = \sum_{l=1}^{n_j}\{\widetilde{\boldsymbol{\Sigma}}_j^{-1}\}_{ll}\{\widetilde{\boldsymbol{A}}_j\}_{li}\{\widetilde{\boldsymbol{A}}_j\}_{lk}$$

where the terms are independent since $\boldsymbol{A}_j$ and $\boldsymbol{\Sigma}_j$ have independent rows. Further, by assumption (ii) it follows that, for $i, k \in \{1, \ldots, p_j\}$,

$$\sum_l l^{-2}\,\mathrm{Var}\left(\{\widetilde{\boldsymbol{\Sigma}}_j^{-1}\}_{ll}\{\widetilde{\boldsymbol{A}}_j\}_{li}\{\widetilde{\boldsymbol{A}}_j\}_{lk}\right) < \infty.$$

This allows us to use Corollary 4.22 in Kallenberg (2002), i.e. that, for $i, k \in \{1, \ldots, p_j\}$,

$$\frac{1}{n_j}\sum_{l=1}^{n_j}\{\widetilde{\boldsymbol{\Sigma}}_j^{-1}\}_{ll}\{\widetilde{\boldsymbol{A}}_j\}_{li}\{\widetilde{\boldsymbol{A}}_j\}_{lk} \overset{a.s.}{\to} \{\boldsymbol{\nu}_j\}_{ik} \quad \text{as } n_j \to \infty$$

which is equivalent to

$$\frac{1}{n_j}\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{A}}_j \overset{a.s.}{\to} \boldsymbol{\nu}_j \quad \text{as } n_j \to \infty.$$

Since $\boldsymbol{\nu}_j$ is invertible, the latter convergence implies $n_j(\widetilde{\boldsymbol{A}}_j'\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{A}}_j)^{-1} \overset{a.s.}{\to} \boldsymbol{\nu}_j^{-1}$ as $n_j \to \infty$, i.e.

$$n_j\,\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_j \mid B_j) \overset{a.s.}{\to} \boldsymbol{\nu}_j^{-1} \quad \text{as } n_j \to \infty.$$

From the proof of Proposition 4.2 we know that $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_j) = \mathbb{E}\big[\mathrm{Cov}\big(\widehat{\boldsymbol{\beta}}_j \mid B_j\big)\big]$. The assumed uniform integrability and Proposition 4.12 Kallenberg (2002) give

$$n_j \, \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_j) = n_j \mathbb{E}\big[\,\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_j \mid B_j)\big] \to \boldsymbol{\nu}_j^{-1} \quad \text{as } n_j \to \infty.$$

$\square$

Proof of Corollary B.1. We start by proving that $\widehat{\boldsymbol{\beta}} \overset{\mathbb{P}}{\to} \boldsymbol{\beta}$ as $|\mathcal{I}| \to \infty$. By Proposition 4.1, $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$. Now Markov's inequality combined with Proposition B.1 immediately gives consistency: for $k \in \{1, \ldots, p_j\}$ and any $\varepsilon > 0$,

$$\mathbb{P}\big(|\{\widehat{\boldsymbol{\beta}}_j\}_k - \{\boldsymbol{\beta}_j\}_k| > \varepsilon\big) \leq \frac{\mathrm{Var}(\{\widehat{\boldsymbol{\beta}}_j\}_k)}{\varepsilon^2} \to 0 \quad \text{as } n_j \to \infty$$

since $\lim_{n_j \to \infty} n_j \, \mathrm{Var}(\{\widehat{\boldsymbol{\beta}}_j\}_k) = \{\boldsymbol{\nu}_j\}_{k,k}$. Since $\{\widehat{\boldsymbol{\beta}}_j\}_k \overset{\mathbb{P}}{\to} \{\boldsymbol{\beta}_j\}_k$ as $n_j \to \infty$ for every $j = 1, \ldots, J - 1$ and $k \in \{1, \ldots, p_j\}$ if and only if $\widehat{\boldsymbol{\beta}} \overset{\mathbb{P}}{\to} \boldsymbol{\beta}$ as $|\mathcal{I}| := n_1 \to \infty$ the statement is proved.

We continue by showing that $|\mathcal{I}|\boldsymbol{\Lambda}(\boldsymbol{\sigma}; \mathcal{F}_0)$ converges in probability as $|\mathcal{I}| \to \infty$. First, from Proposition B.1 we know that $|\mathcal{I}|\boldsymbol{\Lambda}(\boldsymbol{\sigma}; \mathcal{F}_0) \overset{\mathbb{P}}{\to} \boldsymbol{C}$ as $|\mathcal{I}| \to \infty$, where $\boldsymbol{C}$ is block diagonal with blocks $\boldsymbol{\nu}_j^{-1}$. From this, (15) and the assumption that $\widehat{\sigma}_j^2 \overset{\mathbb{P}}{\to} \sigma_j^2$ as $|\mathcal{I}| \to \infty$ for all $j = 1, \ldots, J - 1$, an application of Slutsky's theorem yields $|\mathcal{I}|\boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0) \overset{\mathbb{P}}{\to} \boldsymbol{C}$ as $|\mathcal{I}| \to \infty$. Further, h is only a function of elements in either $(I_{ij})_{i \geq 2, j \in \mathcal{J}}$ or $(C_{ij})_{i \geq 2, j \in \mathcal{J}}$ and thus it follows that, for a fixed J, h is independent of $|\mathcal{I}|$. Therefore $\boldsymbol{\beta} \mapsto \nabla_{\boldsymbol{\beta}} h(\boldsymbol{\beta}; \mathcal{F}_0)$ does not depend on $|\mathcal{I}|$. Moreover, from Remark 4, each component of $\nabla_{\boldsymbol{\beta}} h(\widehat{\boldsymbol{\beta}}; \mathcal{F}_0)$ is either constant or a multi-affine function of the components of $\widehat{\boldsymbol{\beta}}$, i.e. a sum of products of the components of $\widehat{\boldsymbol{\beta}}$. Therefore, since $\widehat{\boldsymbol{\beta}} \overset{\mathbb{P}}{\to} \boldsymbol{\beta}$ as $|\mathcal{I}| \to \infty$, we can use the continuous mapping theorem to conclude that

$$\nabla_{\boldsymbol{\beta}} h(\widehat{\boldsymbol{\beta}}; \mathcal{F}_0) \overset{\mathbb{P}}{\to} \nabla_{\boldsymbol{\beta}} h(\boldsymbol{\beta}; \mathcal{F}_0)$$

as $|\mathcal{I}| \to \infty$. Putting it all together we have

$$|\mathcal{I}|\nabla_{\boldsymbol{\beta}} h(\widehat{\boldsymbol{\beta}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0) \nabla_{\boldsymbol{\beta}} h(\widehat{\boldsymbol{\beta}}; \mathcal{F}_0) \overset{\mathbb{P}}{\to} \nabla_{\boldsymbol{\beta}} h(\boldsymbol{\beta}; \mathcal{F}_0)' \boldsymbol{C} \nabla_{\boldsymbol{\beta}} h(\boldsymbol{\beta}; \mathcal{F}_0).$$

$\square$

Proof of Proposition 5.1. The proof of MSEP for the ultimate claim amount for a single accident year is already given in Section 5 in the text leading up to the statement of Proposition 5.1. We will now go through the remaining steps needed in the derivation of MSEP for the ultimate claim amount aggregated over all accident years.

In Section 5 we provided the process variance, see (23), hence, following Lemma 2.1, what remains to determine are the $Q_{i,k}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)$s:

$$Q_{i,k}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) = \nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0) \nabla_{\boldsymbol{f}} h_k(\widehat{\boldsymbol{f}}; \mathcal{F}_0),$$

where $\nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)$ is given by (26), i.e.

$$\{\nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)\}_j = 1_{\{J-i+1 \leq j\}} \frac{\widehat{U}_i}{\widehat{f}_j},$$

and

$$\{\boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0)\}_{j,j} = \frac{\widehat{\sigma}_j^2}{\sum_{i=i_0}^{J-j} C_{i,j}},$$

where $\{\boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0)\}_{i,j} = 0$ for all $i \neq j$. By using the above, for $i \leq k$, it follows that

$$\nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0) \nabla_{\boldsymbol{f}} h_k(\widehat{\boldsymbol{f}}; \mathcal{F}_0) = \sum_{j=1}^{J-1} 1_{\{J-i+1 \leq j\}} 1_{\{J-k+1 \leq j\}} \frac{\widehat{U}_i}{\widehat{f}_j} \frac{\widehat{U}_k}{\widehat{f}_j} \frac{\widehat{\sigma}_j^2}{\sum_{l=i_0}^{J-j} C_{l,j}}$$
$$= \widehat{U}_i \widehat{U}_k \widehat{\Delta}_{i,J}^U,$$

where $\widehat{\Delta}_{i,J}^U$ is given by (27). Given the above, the statement in Proposition 5.1 follows by using Lemma 2.1. $\square$

Proof of Proposition 5.2. As in the proof of Proposition 5.1, the process (co)variances are obtained from the references given in the text leading up to the formulation of Proposition 5.2. Thus, given Lemma 2.1, what remains to determine are the $Q_{i,k}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)$s:

$$Q_{i,k}(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0) = \nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0) \nabla_{\boldsymbol{f}} h_k(\widehat{\boldsymbol{f}}; \mathcal{F}_0),$$

where $\nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)$ is given by (33), which may be expressed as

$$\{\nabla_{\boldsymbol{f}} h_i(\widehat{\boldsymbol{f}}; \mathcal{F}_0)\}_j = \begin{cases} -1_{\{J-i+1<j\}} \widehat{U}_i \frac{C_{J-j+1,j}}{\widehat{f}_j S_j^1}, \\ -1_{\{J-i+1=j\}} \widehat{U}_i \frac{1}{\widehat{f}_{J-i+1}}, \end{cases} \tag{42}$$

43

and

$$\{\mathbf{\Lambda}(\widehat{\boldsymbol{\sigma}};\mathcal{F}_0)\}_{j,j} = \mathrm{Var}(\widehat{f}_j \mid B_j) = \frac{\widehat{\sigma}_j^2}{S_j^0},$$

where $\{\mathbf{\Lambda}(\widehat{\boldsymbol{\sigma}};\mathcal{F}_0)\}_{i,j} = 0$ for all $i \neq j$. Thus, for all $i \leq k$ it holds that

$$\begin{aligned}
\nabla_{\boldsymbol{f}} &h_i(\widehat{\boldsymbol{f}};\mathcal{F}_0)' \mathbf{\Lambda}(\widehat{\boldsymbol{\sigma}};\mathcal{F}_0) \nabla_{\boldsymbol{f}} h_k(\widehat{\boldsymbol{f}};\mathcal{F}_0) \\
&= \sum_{j=1}^{J-1} \widehat{U}_i \Big(1_{\{J-i+1<j\}} \frac{C_{J-j+1,j}}{\widehat{f}_j S_j^1} + 1_{\{J-i+1=j\}} \frac{1}{\widehat{f}_{J-i+1}} \Big) \\
&\quad \cdot \frac{\widehat{\sigma}_j^2}{S_j^0} \widehat{U}_k \Big(1_{\{J-k+1<j\}} \frac{C_{J-j+1,j}}{\widehat{f}_j S_j^1} + 1_{\{J-k+1=j\}} \frac{1}{\widehat{f}_{J-i+1}} \Big) \\
&= \begin{cases} \widehat{U}_i \widehat{U}_k \widehat{\Delta}_{i,J}^{\mathrm{CDR}}, & i = k, \\ \widehat{U}_i \widehat{U}_k \widehat{\chi}_{i,J}^{\mathrm{CDR}}, & i < k, \end{cases}
\end{aligned}$$

where $\widehat{\Delta}_{i,J}^{\mathrm{CDR}}$ is given by (34) and $\widehat{\chi}_{i,J}^{\mathrm{CDR}}$ is given by (36). Finally, Proposition 5.2 follows by combining the above together with the corresponding process (co)variances and Lemma 2.1. □

D Numerical Example

In this section a simulation study is presented whose purpose is to analyze and compare the two estimators of conditional MSEP based on the conditional and unconditional specification of $\widehat{\boldsymbol{\theta}}^*$. The data used is the runoff triangle of Taylor and Ashe (1983), see Table 1, which has been widely used and analyzed, e.g. in Mack (1993).

The performance of the two estimators of conditional MSEP, based on this particular data set, is examined by estimating, through simulations, $\mathbb{E}[\Delta V^2]$ as specified in Section 2.1. The practical relevance of computing these estimators is investigated by comparing the size of the estimation error to the size of the process variance.

The data generating process the simulation study is assumed to be a sequence of general linear models of the form in (11) in Section 4. More specifically, for each $i \in \mathcal{I}$, it is assumed that

$$C_{i,1} = \alpha + \tau e_{i,1}, \quad C_{i,j+1} = f_j C_{ij} + \sigma_j \sqrt{C_{ij}} e_{i,j+1}, \quad j = 1, \dots, J-1.$$

The error terms are given by Remark 2, i.e. by translated log-normal variables, which also holds for the first column by setting $C_{i0} := 1$ for all $i \in \mathcal{I}$.

Table 1: Runoff triangle of aggregated payments of Taylor and Ashe (1983).

i \ j	0	1	2	3	4	5	6	7	8	9
1	451,288	339,519	333,371	144,988	93,243	45,511	25,217	20,406	31,482	1,729
2	448,627	512,882	168,467	130,674	560,44	33,397	56,071	26,522	14,346	
3	693,574	497,737	202,272	120,753	125,046	37,154	27,608	17,864		
4	652,043	546,406	244,474	200,896	106,802	106,753	63,688			
5	566,082	503,970	217,838	145,181	165,519	91,313				
6	606,606	562,543	227,374	153,551	132,743					
7	536,976	472,525	154,205	150,564						
8	554,833	590,880	300,964							
9	537,238	701,111								
10	684,944									

The parameter values used in the simulation study are the ones acquired from fitting this model to the data in Table 1 following the weighted least squares estimation introduced in Section 4, see (13) and (16). As seen in Section 5, this is equivalent to fitting a chain ladder model to this triangle together with estimating an intercept and a variance for the first column (using the sample mean and the unbiased sample variance of the first column). The resulting parameter estimates are taken to be the true parameter values in the simulation study, they are denoted by $\boldsymbol{f}, \boldsymbol{\sigma}^2$, α, and τ^2, and referred to jointly as $\boldsymbol{\theta}$. To be able to use the unbiased estimators of the σ_j^2s, the last column of the triangle is removed. An alternative to this approach could be to use maximum likelihood or some form of extrapolation of the σ_j^2s. Since comparison of methods to estimating tail variances is not the purpose of the simulation study, the former simpler approach is chosen. Based on the above development-year dynamics and $\boldsymbol{\theta}$, $N = 10^6$ new triangles are generated giving rise to $\{\mathcal{F}_0^{(i)}\}_{i=1}^N$. For each such triangle, a chain ladder model is fitted together with an intercept and variance for the first column, as described above, to get the parameter estimator $\widehat{\boldsymbol{\theta}}^{(i)}$. For $i = 1, \dots, N$, the following quantities are computed:

- the (true) process variance $\mathrm{Var}(U^{(i)} \mid \mathcal{F}_0^{(i)})$, given in (22), and the plug-in estimator $\mathrm{Var}(U^{(i)} \mid \mathcal{F}_0^{(i)})(\widehat{\boldsymbol{f}}^{(i)}, (\widehat{\boldsymbol{\sigma}}^{(i)})^2)$ given in (23),

- the (true) conditional expectation of the ultimate claim amount $h(\boldsymbol{f}; \mathcal{F}_0^{(i)})$, given in (21), and the plug-in estimator $h(\widehat{\boldsymbol{f}}^{(i)}; \mathcal{F}_0^{(i)})$,

45

- the plug-in estimator of the gradient $\nabla h(\widehat{\boldsymbol{f}}^{(i)}; \mathcal{F}_0^{(i)})$, given in (26),

- the estimator of the conditional covariance of $\widehat{\boldsymbol{f}}^{*}$ using the conditional specification,
$$\Lambda(\widehat{\boldsymbol{\sigma}}^{(i)}; \mathcal{F}_0^{(i)}) = \widehat{\mathrm{Cov}}(\widehat{\boldsymbol{f}}^{*,c} \mid \mathcal{F}_0^{(i)}),$$
the elements of which are given in (25),

- the estimator of the conditional covariance of $\widehat{\boldsymbol{f}}^{*}$ using the unconditional specification,
$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{f}}) = \widehat{\mathrm{Cov}}(\widehat{\boldsymbol{f}}^{*,u} \mid \mathcal{F}_0^{(i)}),$$

- the two estimators of the estimation error,
$$\nabla h(\widehat{\boldsymbol{f}}^{(i)}; \mathcal{F}_0^{(i)})' \Lambda(\widehat{\boldsymbol{\sigma}}^{(i)}; \mathcal{F}_0^{(i)}) \nabla h(\widehat{\boldsymbol{f}}^{(i)}; \mathcal{F}_0^{(i)}),$$
and
$$\nabla h(\widehat{\boldsymbol{f}}^{(i)}; \mathcal{F}_0^{(i)})' \widehat{\mathrm{Cov}}(\widehat{\boldsymbol{f}}) \nabla h(\widehat{\boldsymbol{f}}^{(i)}; \mathcal{F}_0^{(i)}),$$

- ΔV_i^2 for the two resampling specifications as given in Section 2.1.

As already mentioned, $\mathrm{Cov}(\widehat{\boldsymbol{f}})$ is not analytically tractable and is therefore estimated using simulations. Recall, from Proposition 4.2, that $\mathrm{Cov}(\widehat{\boldsymbol{f}}) = \mathbb{E}[\Lambda(\widehat{\boldsymbol{\sigma}}; \mathcal{F}_0)]$. Therefore, for each $i = 1, \ldots, N$, M_i new triangles are generated based on the parameters $\widehat{\boldsymbol{\theta}}^{(i)}$ yielding $\{\mathcal{F}_0^{(i,j)}\}_{j=1}^{M_i}$. For each i, the unbiased estimator
$$\Sigma_{i,M_i} := \frac{1}{M_i} \sum_{j=1}^{M_i} \Lambda(\widehat{\boldsymbol{\sigma}}^{(i)}; \mathcal{F}_0^{(i,j)}).$$

of $\mathrm{Cov}(\widehat{\boldsymbol{f}})(\widehat{\boldsymbol{\theta}}^{(i)})$ is chosen as an estimator of $\mathrm{Cov}(\widehat{\boldsymbol{f}})(\widehat{\boldsymbol{\theta}}^{(i)})$. The choice of M_i is as follows. For a fixed n, consider the increasing sequence $(2^{k-1}n)_{k \geq 1}$. Conditional on not having stopped for the value k, $2^{k+1}n$ new triangles are generated based on the parameters $\widehat{\boldsymbol{\theta}}^{(i)}$ yielding $\{\mathcal{F}_0^{(i,j)}\}_{j=1}^{2^{k+1}n}$ and the estimators
$$\Sigma_{i,2^k n} := \frac{1}{2^k n} \sum_{j=1}^{2^k n} \Lambda(\widehat{\boldsymbol{\sigma}}^{(i)}; \mathcal{F}_0^{(i,j)}), \quad \widetilde{\Sigma}_{i,2^k n} := \frac{1}{2^k n} \sum_{j=2^k n+1}^{2^{k+1} n} \Lambda(\widehat{\boldsymbol{\sigma}}^{(i)}; \mathcal{F}_0^{(i,j)})$$

are computed as well as

$$x_k := \nabla h(\widehat{\boldsymbol{f}}^{(i)}; \mathcal{F}_0^{(i)})' \Sigma_{i,2^k n} \nabla h(\widehat{\boldsymbol{f}}^{(i)}; \mathcal{F}_0^{(i)})$$
$$\widetilde{x}_k := \nabla h(\widehat{\boldsymbol{f}}^{(i)}; \mathcal{F}_0^{(i)})' \widetilde{\Sigma}_{i,2^k n} \nabla h(\widehat{\boldsymbol{f}}^{(i)}; \mathcal{F}_0^{(i)}).$$

The stopping criterion is

$$\frac{|x_k - \widetilde{x}_k|}{\min(|x_k|, |\widetilde{x}_k|)} < 0.001.$$

Upon stopping the two independent samples of size $2^k n$ are merged. Consequently $M_i = 2^{k+1} n$, where k is the smallest number such that the stopping criterion is satisfied.

The results of the simulation study are the following. In Figure 1 the distribution of the difference between the simulated values of ΔV^2 for the unconditional and the conditional specification of $\widehat{\boldsymbol{\theta}}^*$ is illustrated. The distribution is leptokurtic, has a slight positive skewness and is approximately centered at zero. The mean and the median of this distribution are small relative the scale of the data ($-0.94 \cdot 10^{22}$ and $0.28 \cdot 10^{22}$, respectively). To quantify the uncertainty in these quantities 95% bootstrap confidence intervals are computed based on the percentile method, see Efron and Tibshirani (1994), yielding $[-1.2, -0.7] \cdot 10^{22}$ and $[0.2, 0.3] \cdot 10^{22}$, respectively, using 10^5 bootstrap samples. As a matter of fact, none of the bootstrap samples of the mean are above 0 and none of the samples of the median are below 0. This indicates that the unconditional specification is better on average (the mean is negative), but the conditional specification is better more often (the median is positive). The practical relevance of this is, however, questionable since on the relative scale of the data, the mean and median are both approximately zero, indicating that the difference between the two estimators is negligible and that one should therefore focus on the computability of the estimators. In Figure 2 the ratio between the conditional and the unconditional estimators of the estimation error is shown. From this figure it is clear that the two estimators are comparable and do not deviate from each other by much.

The distribution of the difference between the ΔV^2s is heavy tailed, and one is therefore led to question whether this is due to the log-normally distributed error terms. Therefore, the marginal distributions of the components of $\widehat{\boldsymbol{\theta}}$ are illustrated in Figure 3 (first column parameters), Figure 4 (development factors) and Figure 5 (chain ladder variances). The estimators of the intercept of the first column and the development factors are, for
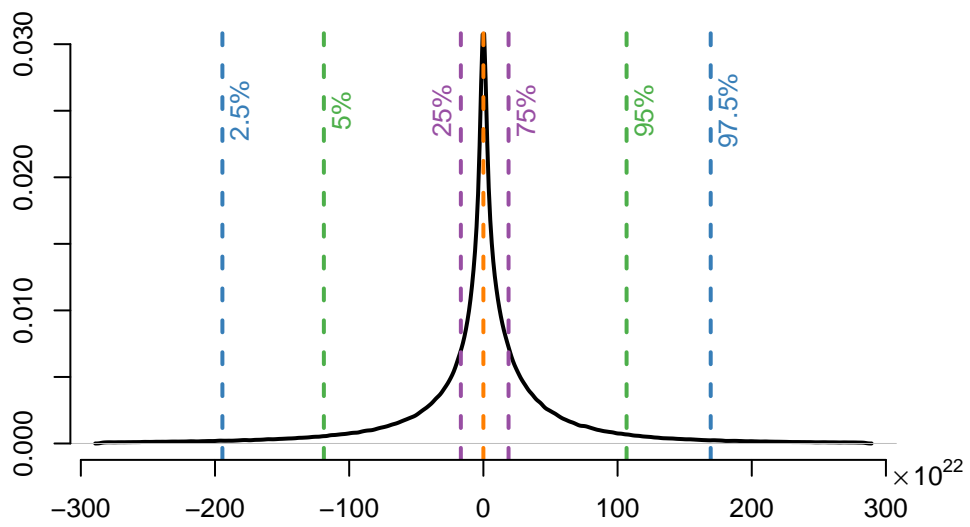
Figure 1: Kernel density estimator of the difference between the simulated values of ΔV^2 for the unconditional and the conditional specification of $\widehat{\boldsymbol{\theta}}^*$ (unconditional minus conditional). Position 0 is market by the orange dashed middle line. The other dashed lines correspond to a chosen set of reference sample quantiles of these differences.
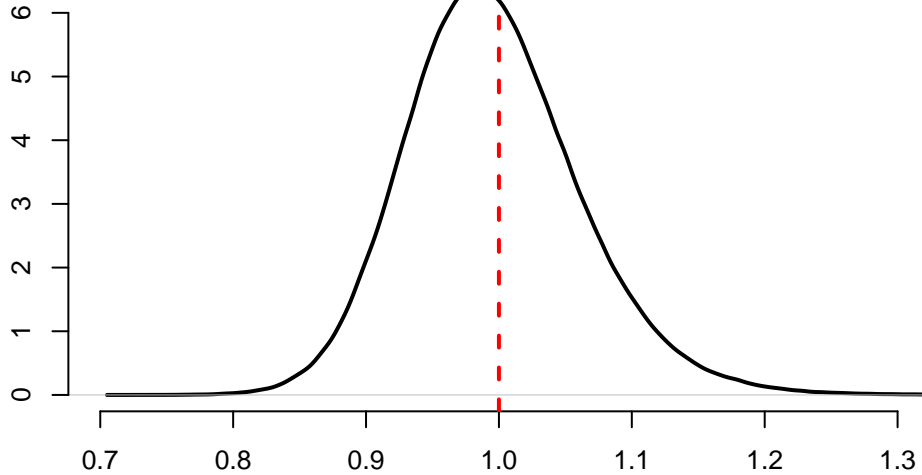
48

Figure 2: Kernel density estimator of the ratio between the conditional and the unconditional estimators of the estimation error. Position 1 is marked by the red dashed line.

all intents and purposes, marginally Gaussian. The variances, however, do have heavier tails (the standard deviations are illustrated in Figure 5). This can have a large effect on the estimated process variance, and thus in turn on the ΔV^2s.

So far the relative performance of the two estimators has been presented. It is of interest to also investigate the absolute performance. Figure 6 shows the distributions of the true estimation error minus the estimated ones based on the conditional and unconditional specification of $\widehat{\boldsymbol{\theta}}^*$. It is seen that there is a tendency to overestimate the true estimation error, although there is a tail to the right indicating that the estimation error will occasionally be greatly underestimated. The mean estimation error in the simulations is $1.9 \cdot 10^{12}$ and the 95% quantiles of the two above distributions are approximately $5 \cdot 10^{12}$. The estimated estimation error will therefore, in the 95% worst case scenario, be underestimated on the scale of, approximately, 2.5 estimation error means.

The practical relevance of estimating the estimation error requires that it is of size comparable to the process variance. Figure 7 shows the distributions of the estimated estimation errors divided by the estimated process variances, together with dashed black vertical lines indicating some of the
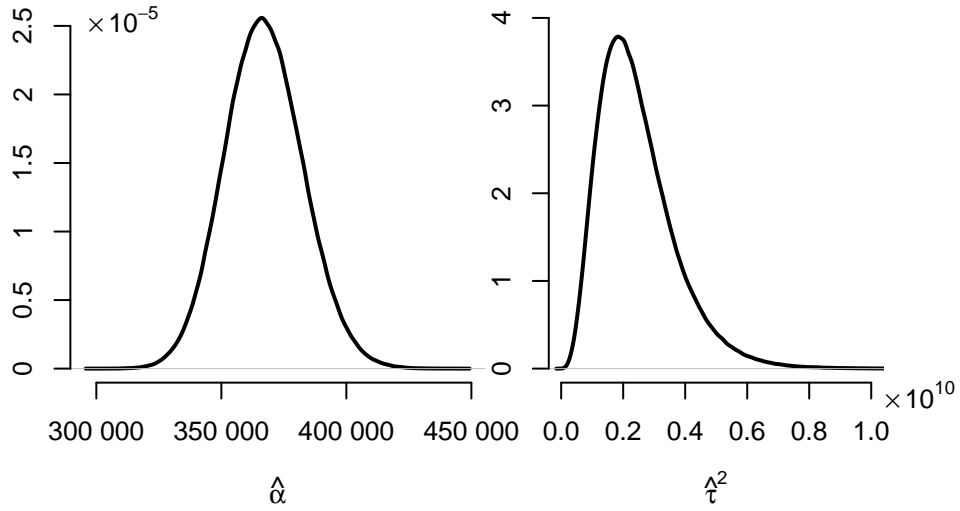
Figure 3: Kernel density estimators of the estimator $\widehat{\alpha}$ of the mean of the first column and the estimator $\widehat{\tau}^2$ of the variance of the first column.
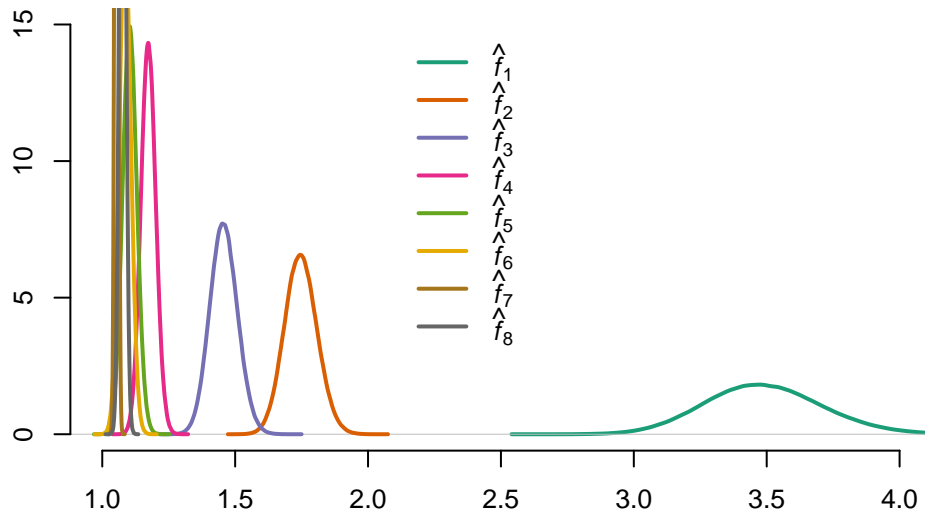

Figure 4: Kernel density estimators of the estimators of the development factors. Some density curves are cut in order to make it easier to visually discriminate between the development factors centered close to 1.
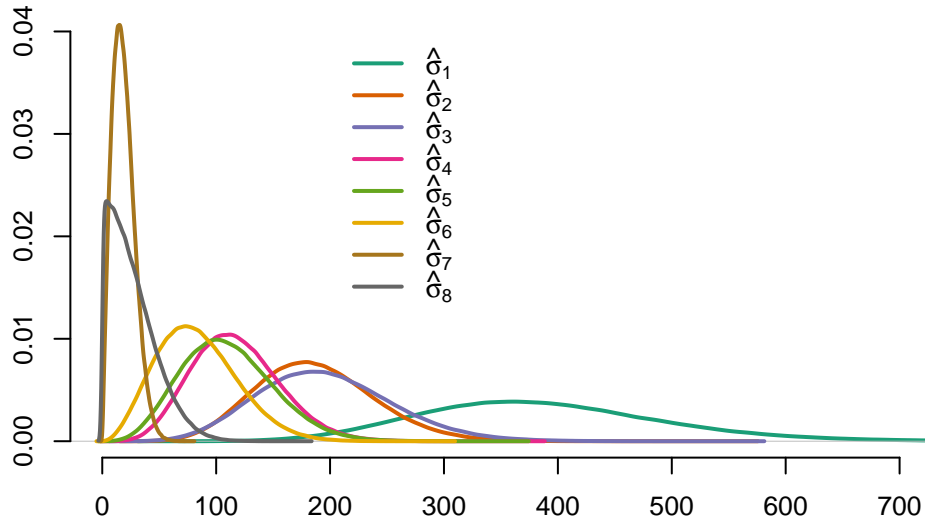
Figure 5: Kernel density estimators of the square roots of the variance estimators $\widehat{\sigma}_j^2$.
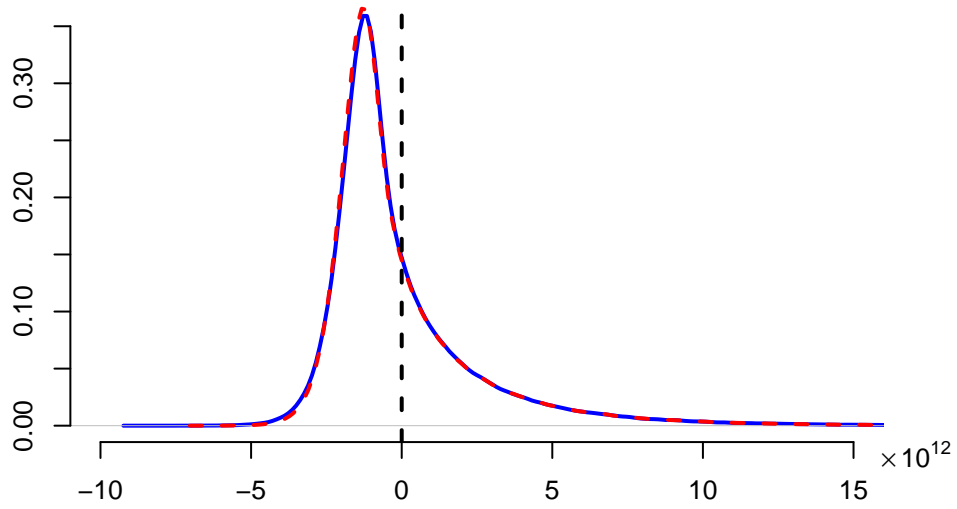

Figure 6: Kernel density estimators of the true estimation error minus the estimated estimation error based on the conditional (blue solid curve) and the unconditional (red dashed curve) specification of $\widehat{\boldsymbol{\theta}}^*$. Position 0 is marked by the black dashed line.
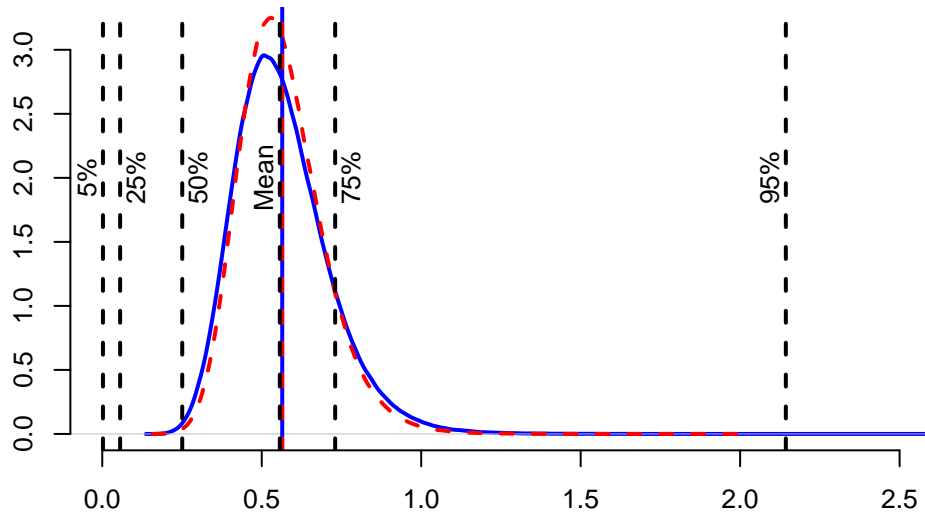
Figure 7: Kernel density estimator of the ratio of the estimation error and the process variance. The blue solid curve is the density for the estimated version of this ratio based on the conditional specification, the red dashed curve is for the unconditional specification. The corresponding vertical lines mark the means of the respective distributions. The black dashed vertical lines mark quantiles and mean of the true distribution of this ratio.

quantiles of the distribution of the true estimation error divided by the true process variance. On average the estimation error is half the size of the process variance, which is also more or less the center of the distributions of the estimated versions. The median, however, of the true distribution lies approximately around 0.25. Therefore, it is as likely that the estimation error is greater than a quarter of the process variance as that it would be less than a quarter of the process variance.

Finally, to illustrate how plug-in estimation of the process variance performs, Figure 8 shows the distribution of the ratio between the estimated process variance (based on plug-in) and the true process variance. Both the mean and the median of this distribution lie close to 1, indicating that on average the estimator yields the correct variance and that we are more or less equally likely to overestimate it as to underestimate it. It is also seen that there are extreme cases where the variance is estimated to be either half or double the true variance.
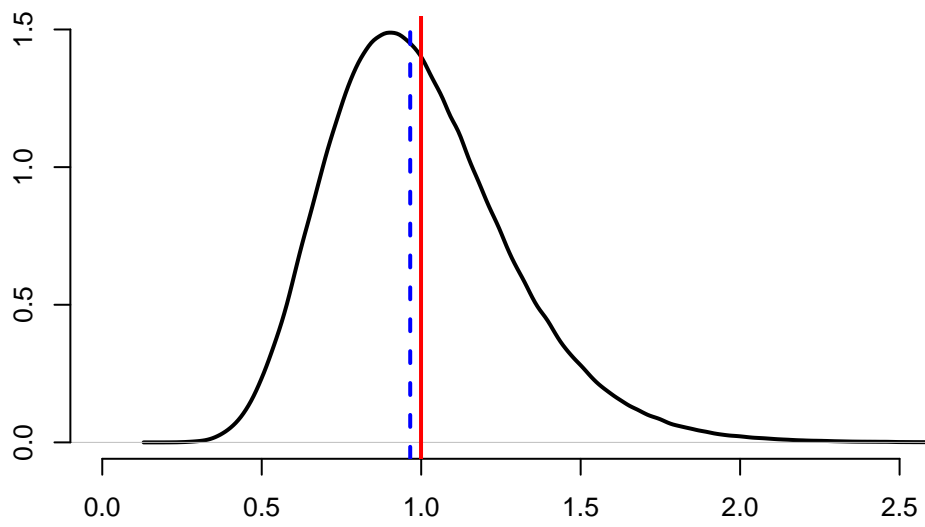
Figure 8: Kernel density estimator of the ratio between the estimated process variance based on plug-in estimation and the true process variance. The red solid line marks the mean of this ratio and the blue dashed line marks the median.