

Multivariate Bühlmann-Straub Credibility Model Applied to Claims Reserving for Correlated Run-off Triangles

by Sebastian Happ, Ramona Maier, and Michael Merz

ABSTRACT

In the present paper we consider the claims reserving problem in a multivariate context. More precisely, we apply the multivariate generalization of the well-known credibility model proposed by Bühlmann and Straub (1970) to claims reserving. This multivariate model allows for a simultaneous study of N correlated run-off portfolios and enables the derivation of an estimator for the conditional mean square error of prediction (MSEP) for the credibility predictor of the ultimate claim of the total portfolio. Thereby, we apply multivariate credibility predictors which reflect the correlation structure between the N portfolios and which are optimal in terms of a classical optimality criterion. We illustrate the results by means of an example and compare it to the results derived by the multivariate chain-ladder method and the multivariate additive loss reserving method proposed by Merz and Wüthrich (2008a, b).

KEYWORDS

Multivariate Bühlmann-Straub model, credibility, claims reserving, ultimate claim prediction, mean square error of prediction

1. Introduction and motivation

A non-life insurance company needs to hold sufficient reserves on its balance sheet in order to meet the future claims payment cashflow. Therefore, given the available information about the past claims payment cash flow and the claims settlement process as well as external knowledge from experts and prior information (e.g., premium, number of contracts, data from similar run-off portfolios, market statistics), the prediction of the outstanding loss liabilities and the quantification of the uncertainties in these predictions is a major task in actuarial practice and science. It is the basis for proving solvency on the one hand and it allows for reliable premium calculations on the other hand (see, e.g., CAS 2001 and Teugels and Sundt 2004).

In this paper we consider the claims reserving problem in a multivariate context. That is, we consider a portfolio consisting of several correlated run-off portfolios (e.g., subportfolios of certain lines of business) and we apply the multivariate generalization of the well-known credibility model proposed by Bühlmann and Straub (1970) for predicting outstanding loss liabilities. This simultaneous study of several individual run-off triangles is motivated by several important facts (cf. Merz and Wüthrich 2008b).

Since in actuarial practice the conditional mean square error of prediction (MSEP) is the most popular risk measure to quantify the uncertainties in claims reserves, we provide an estimator of the conditional MSEP. Such studies of uncertainties for correlated run-off portfolios are especially crucial in the development of new solvency guidelines for the quantification of risk profiles for different insurance companies. However, they do not provide a complete picture of the uncertainty associated with the predictor of the claims reserves for the total portfolio. This can only be provided by the whole predictive distribution of the claims reserves calculated under very restrictive model assumptions or by applying numerical algorithms such as bootstrapping methods and Markov chain Monte Carlo (MCMC) methods (cf. England and Verrall 2007 and Wüthrich and Merz 2008). However, in practical applications and solvency considerations, estimates

for second moments such as the conditional MSEP and its components conditional process variance/estimation error are often sufficient, since in most cases one fits an analytic overall distribution using these first two moments by the method of moments. Moreover, analytic solutions are important because they allow for explicit interpretations in terms of the parameters involved and enable sensitivity analysis with respect to parameter changes.

1.1. Claims reserving methods and credibility theory in a multivariate context

The calculation of the conditional MSEP for the predictor of the ultimate claim for a whole portfolio of several correlated run-off portfolios is more sophisticated than for only one run-off portfolio. Holmberg (1994) was probably the first one to investigate the problem of dependence between run-off portfolios of different lines of business. Later Halliwell (1999) and Quarg and Mack (2004) proposed the first bivariate models which express the dependence between the paid and incurred losses of a single run-off portfolio. Braun (2004) and Merz and Wüthrich (2008a, b) generalized the well-known univariate chain-ladder model of Mack (1993) to the bivariate and the general multivariate case, respectively, by incorporating correlations between different run-off portfolios. They derived an estimate of the conditional MSEP for the predictor of the ultimate claim of the total portfolio. Merz and Wüthrich (2009a) study a special case of the multivariate additive loss reserving model proposed by Hess, Schmidt, and Zocher (2006) and derive an estimate for the conditional MSEP. Moreover, Dahms (2012) presented a general class of models that contains most models mentioned above.

In the present paper we give a credibility approach to the claims reserving problem in a multivariate context. Univariate and multivariate credibility methods are widely used in insurance pricing, but for claims reserving they are less used (though they are also useful in this context). In the claims reserving context univariate credibility methods can, e.g., be found in Benktander (1976), De Vylder (1982), Neuhaus (1992), Mack (2000), Witting (1987), and Gisler and

Wüthrich (2008). More recently, Dahms and Happ (forthcoming) presented in a credibility framework a very general class of (multivariate) claims reserving methods, which contains most of the methods mentioned above as special cases. In the present paper we choose a multivariate claims reserving method different from Dahms and Happ (forthcoming) and apply the multivariate generalization of the credibility model of Bühlmann and Straub (1970) to the (multivariate) claims reserving problem.

1.2. Claims development triangle and notation

Throughout this paper all random variables are square integrable random variables defined on a common probability space (Ω, \mathcal{F}, P) . We consider the situation where we have $N \geq 1$ portfolios. The associated losses of each portfolio are represented by run-off triangles (claim development triangles). We assume for the reason of simplicity that all run-off triangles are of the same size (the whole theory presented in this paper can be easily generalized to different sizes but the notation then becomes more complicated) and we denote by I (J) the last the accident (development) year. The claims development data have the structure shown in Figure 1. Thereby we denote by $X_{ij}^{(n)}$ the incremental claim payments for accident year $i \in \{0, \dots, I\}$ and development year $j \in \{0, \dots, J\}$ of run-off portfolio $n \in \{1, \dots, N\}$.

The cumulative claims payments of triangle n for accident year i up to development year j are denoted by

$$C_{i,j}^{(n)} = \sum_{k=0}^j X_{i,k}^{(n)}.$$

Figure 1. Claims development triangle number n .

accident year i	development year j				
	0	...	j	...	J
0	realizations of r.v. $C_{i,j}^{(n)}$ i.e. observations $\mathcal{D}_I^{(n)}$				
\vdots					
$I-j$					
\vdots					
I					

For simplicity, we always assume that $I = J$, i.e., we deal with development triangles (for $I \geq J$ we have development trapezoids), but all results also hold true under slight modifications for the case $I \geq J$.

Usually, at time I we have the sets of observations (σ -algebras)

$$\mathcal{D}_I^{(n)} = \sigma(C_{i,j}^{(n)}; i+j \leq I) \subseteq \mathcal{F}$$

for all run-off portfolios $n \in \{1, \dots, N\}$. The total of observation over all run-off portfolios is then given by

$$\mathcal{D}_I^N = \sigma\left(\bigcup_{n=1}^N \mathcal{D}_I^{(n)}\right).$$

For the following derivations it is convenient to write the data of the N run-off portfolios in vector form. Thus we define the N -dimensional random vectors

$$\mathbf{X}_{i,j} = (X_{i,j}^{(1)}, \dots, X_{i,j}^{(N)})'$$

and

$$\mathbf{C}_{i,j} = (C_{i,j}^{(1)}, \dots, C_{i,j}^{(N)})'$$

of incremental and cumulative claims payments, respectively. The vector of outstanding claims payments for accident year $i \in \{1, \dots, I\}$ is defined by

$$\mathbf{R}_i = (R_i^{(1)}, \dots, R_i^{(N)})' = \mathbf{C}_{i,I} - \mathbf{C}_{i,I-i} = \sum_{j=I-i+1}^I \mathbf{X}_{i,j}.$$

Furthermore, we define the N -dimensional column vector consisting of one's by $\mathbf{1} = (1, \dots, 1)' \in \mathbb{R}^N$ and the $N \times N$ -dimensional identity matrix by \mathbf{I} .

2. Multivariate Bühlmann-Straub credibility model

Bayesian methods are often an appropriate tool to combine data with expert opinion, or, in other words, to combine internal data (observations) with external given prior information. However, in most Bayesian models the derivation of the posterior distribution is infeasible and numerical methods such as

MCMC or numerical integration have to be applied. Analytical posterior distributions can only be calculated under very restrictive (distributional) model assumptions, for example, if one restricts to distributions from the exponential dispersion family with associated conjugated prior (cf. Bühlmann and Gisler 2005). For an example of this strategy applied in claims reserving we refer to Hashorva, Merz, and Wüthrich (2013). However, in many models it is impossible to express the Bayesian predictor in an analytical closed form. In this case the best we can do is to restrict the class of possible predictors to the class of so-called credibility predictors, which are affine-linear functions of the observations with minimum MSEP; see (2.7). These predictors have the big advantage that prior knowledge and data can be combined for the prediction and that they can be calculated under less restrictive model assumptions than Bayesian predictors. For a detailed introduction and more details on credibility predictors, see Bühlmann and Gisler (2005).

Applied to our multivariate claims reserving problem, this means that we are interested in the predictor of the ultimate claim $C_{i,J}$ for accident year $i \in \{1, \dots, I\}$, which is the best affine-linear function of the components of the observations $\mathbf{X}_{i,0}, \dots, \mathbf{X}_{i,I-i}$ at time I with respect to the MSEP. We study this problem in the framework of the multivariate Bühlmann-Straub model (cf. Bühlmann and Gisler 2005). In order to formulate the model assumptions of the multivariate Bühlmann-Straub model, we introduce latent random variables Θ_i , which describe the risk characteristics of the different accident years $i = 0, \dots, I$. Moreover, we assume that there are (known) volume measures $\mu_i^{(n)}$ and (unknown) incremental loss development patterns (cash flow patterns) $(\gamma_j^{(n)})_{j=0, \dots, J} \subset \mathbb{R}_+$ for the N development triangles $n \in \{1, \dots, N\}$, such that

$$E[X_{i,j}^{(n)}] = \gamma_j^{(n)} \mu_i^{(n)} \quad (2.1)$$

for all $i = 0, \dots, I$. This leads to the normalized incremental claims payments given by

$$Y_{i,j}^{(n)} = \frac{X_{i,j}^{(n)}}{\gamma_j^{(n)} \mu_i^{(n)}}$$

for all $i = 0, \dots, I$ and $j = 0, \dots, J$. To shorten notation we define the cumulative loss development patterns $(\beta_j^{(n)})_{j=0, \dots, J} \subset \mathbb{R}_+$ by

$$\beta_0^{(n)} = \gamma_0^{(n)} > 0 \text{ and } \beta_j^{(n)} - \beta_{j-1}^{(n)} = \gamma_j^{(n)} > 0 \\ \text{for } j = 1, \dots, J. \quad (2.2)$$

In vector form we have for $i = 0, \dots, I$ and $j = 0, \dots, J$:

$$\boldsymbol{\gamma}_j = (\gamma_j^{(1)}, \dots, \gamma_j^{(N)})' \\ \boldsymbol{\beta}_j = (\beta_j^{(1)}, \dots, \beta_j^{(N)})' \\ \boldsymbol{\mu}_i = (\mu_i^{(1)}, \dots, \mu_i^{(N)})' \\ \mathbf{Y}_{i,j} = (Y_{i,j}^{(1)}, \dots, Y_{i,j}^{(N)})'$$

In the following we denote by

$$\mathbf{D}(\mathbf{a}) = \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_N \end{pmatrix} \text{ and } \mathbf{D}(\mathbf{a})^b = \begin{pmatrix} a_1^b & & 0 \\ & \ddots & \\ 0 & & a_N^b \end{pmatrix}$$

the $N \times N$ -diagonal matrices of the N -dimensional vectors $\mathbf{a} = (a_1, \dots, a_N)' \in \mathbb{R}^N$ and $(a_1^b, \dots, a_N^b)' \in \mathbb{R}^N$ for an admissible exponent $b \in \mathbb{R}$, respectively. Then we have for the normalized incremental claims

$$\mathbf{Y}_{i,j} = \mathbf{D}(\mathbf{w}_{i,j})^{-1} \mathbf{X}_{i,j}, \quad (2.3)$$

where $\mathbf{w}_{i,j} = (w_{i,j}^{(1)}, \dots, w_{i,j}^{(N)})'$ with $w_{i,j}^{(n)} = \gamma_j^{(n)} \mu_i^{(n)}$ for all $i = 0, \dots, I$, $j = 0, \dots, J$ and $n = 1, \dots, N$.

Having this notation the multivariate Bühlmann-Straub model is then given by:

Model Assumptions 2.1 (Multivariate Bühlmann-Straub model)

- Conditionally, given Θ_i , the normalized incremental claims $\mathbf{Y}_{i,0}, \dots, \mathbf{Y}_{i,J}$ are independent with

$$E[\mathbf{Y}_{i,j} | \Theta_i] = \boldsymbol{\mu}(\Theta_i) \quad (2.4)$$

$$\begin{aligned} \text{Var}(\mathbf{Y}_{i,j}|\Theta_i) &= \mathbf{D}(\mathbf{w}_{i,j,\xi,\delta})^{-1/2} \cdot \Sigma(\Theta_i) \cdot \mathbf{D}(\mathbf{w}_{i,j,\xi,\delta})^{-1/2} \\ &= \begin{pmatrix} \frac{\sigma_1^2(\Theta_i)}{w_{i,j,\xi,\delta}^{(1)}} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\sigma_N^2(\Theta_i)}{w_{i,j,\xi,\delta}^{(N)}} \end{pmatrix} \end{aligned} \quad (2.5)$$

where $w_{i,j,\xi,\delta}^{(n)} = \gamma_j^{(n)\xi} \mu_i^{(n)\delta}$ with $\xi \in [0, 2]$ and $\delta \geq 0$. The matrix $\Sigma(\Theta_i) = \mathbf{D}(\sigma_1^2(\Theta_i), \dots, \sigma_N^2(\Theta_i))$ is the $N \times N$ -diagonal matrix of $(\sigma_1^2(\Theta_i), \dots, \sigma_N^2(\Theta_i))'$.

- The pairs $(\Theta_i, \mathbf{Y}_i' = (\mathbf{Y}_{i,0}', \dots, \mathbf{Y}_{i,I}'))$ for $i = 0, \dots, I$ are independent and the latent variables $\Theta_0, \dots, \Theta_I$ are identically distributed.

Remarks

- From (2.4) it follows that the normalized incremental claim payments $\mathbf{Y}_{i,j}$ of accident year i are higher or lower than the normalized incremental claim payments $\mathbf{Y}_{k,j}$ of another accident year k . This means there are accident years which are systematically better or worse than other ones.
- We assume that the prior volumes $\boldsymbol{\mu}_i = (\mu_i^{(1)}, \dots, \mu_i^{(N)})'$ are known and that the incremental loss development pattern $\boldsymbol{\gamma}_j = (\gamma_j^{(1)}, \dots, \gamma_j^{(N)})'$ is unknown.
- The weights $\xi \in [0, 2]$ and $\delta > 0$ reflect the relation of the (conditional) expected value Equation 2.4 and its variance Equation 2.5. For a discussion and a motivation of choices $\xi = 0$ and $\xi = 1$ we refer to Mack (2002). Although the cases $\xi = 0$ and $\xi = 1$ can be clearly interpreted, we allow for $\xi \in [0, 2]$. In Section 4 we show how appropriate choices for ξ and δ can be derived.
- The parameter Θ_i tells us whether we have a good or a bad accident year i . For a more detailed explanation in the framework of tariffication and pricing we refer to Bühlmann and Gisler (2005).
- It is straightforward to show that in the case of one-dimensional observations (i.e., $N = 1$), the assumptions of the (classical) one-dimensional Bühlmann-Straub model are satisfied.

- For the normalized incremental claims and the cumulative claims we obtain

$$E[\mathbf{Y}_{i,j}] = E[\boldsymbol{\mu}(\Theta_i)] = \mathbf{1}$$

and, respectively,

$$E[\mathbf{C}_{i,j}|\Theta_i] = \mathbf{D}(\boldsymbol{\beta}_j) \mathbf{D}(\boldsymbol{\mu}_i) \boldsymbol{\mu}(\Theta_i)$$

$$E[\mathbf{C}_{i,j}] = \mathbf{D}(\boldsymbol{\beta}_j) \boldsymbol{\mu}_i$$

for all $i = 0, \dots, I$ and $j = 0, \dots, J$. Moreover, we obtain for the claims reserves

$$E[\mathbf{R}_i|\Theta_i] = (\mathbf{D}(\boldsymbol{\beta}_J) - \mathbf{D}(\boldsymbol{\beta}_{I-I})) \mathbf{D}(\boldsymbol{\mu}_i) \boldsymbol{\mu}(\Theta_i)$$

$$E[\mathbf{R}_i] = (\mathbf{D}(\boldsymbol{\beta}_J) - \mathbf{D}(\boldsymbol{\beta}_{I-I})) \boldsymbol{\mu}_i \quad (2.6)$$

for all $i = 1, \dots, I$.

In the following we define the mean square error of prediction (MSEP) of a N -dimensional predictor $\widehat{\mathbf{X}} = (\widehat{X}_1, \dots, \widehat{X}_N)'$ for a N -dimensional random variable $\mathbf{X} = (X_1, \dots, X_N)'$ by

$$\text{mse}_x(\widehat{\mathbf{X}}) = \sum_{n=1}^N E[(\widehat{X}_n - X_n)^2]. \quad (2.7)$$

In the multidimensional credibility theory one looks now for a predictor $\widehat{\boldsymbol{\mu}}(\Theta_i)$ of $\boldsymbol{\mu}(\Theta_i)$ that minimizes the MSEP (2.7) among all N -dimensional predictors $\widehat{\mathbf{X}} = (\widehat{X}_1, \dots, \widehat{X}_N)'$ whose components \widehat{X}_k are affine-linear in the components of the N -dimensional observations $\mathbf{X}_{i,0}, \dots, \mathbf{X}_{i,I-i}$ with $i = 0, \dots, I$. That is, one has to solve the optimization problem

$$\widehat{\boldsymbol{\mu}}(\Theta_i)^{\text{cred}} = \underset{\widehat{\mathbf{X}} \in L(D_i^N, 1)}{\text{argmin}} \text{mse}_{\boldsymbol{\mu}(\Theta_i)}(\widehat{\mathbf{X}}),$$

where

$$\begin{aligned} L(\mathcal{D}_I^N, 1) \\ = \left\{ \widehat{\mathbf{X}}; \widehat{X}_k = a + \sum_{i=0}^I \sum_{j=0}^{I-i} \sum_{n=1}^N a_{i,j}^{(n)} X_{i,j}^{(n)} \text{ with } a, a_{i,j}^{(n)} \in \mathbb{R} \right\}. \end{aligned} \quad (2.8)$$

We define the structural parameter matrices

$$S = E[\Sigma(\Theta_i)]$$

$$T = \text{Var}(\boldsymbol{\mu}(\Theta_i))$$

and obtain:

Theorem 2.2 (Bühlmann-Straub predictor)

Under Model Assumptions 2.1 the optimal affine-linear predictor of $\mu(\Theta_i)$ is given by

$$\widehat{\mu(\Theta_i)}^{cred} = A_i \mathbf{K}_i + (I - A_i) \mathbf{1} \quad (2.9)$$

for $1 \leq i \leq I$, where

$$\mathbf{K}_i = \left(\sum_{j=0}^{I-i} \frac{\gamma_j^{(1)\xi-1}}{\beta_{I-i,\xi}^{(1)} \mu_i^{(1)}} X_{i,j}^{(1)}, \dots, \sum_{j=0}^{I-i} \frac{\gamma_j^{(N)\xi-1}}{\beta_{I-i,\xi}^{(N)} \mu_i^{(N)}} X_{i,j}^{(N)} \right)$$

$$A_i = T \left(T + \mathbf{D}(\beta_{I-i,\xi})^{-1/2} \mathbf{D}(\mu_i)^{-\delta/2} S \mathbf{D}(\mu_i)^{-\delta/2} \mathbf{D}(\beta_{I-i,\xi})^{-1/2} \right)^{-1}$$

$$\beta_{I-i,\xi}^{(n)} = \sum_{j=0}^{I-i} \gamma_j^{(n)\xi}$$

$$\beta_{I-i,\xi} = (\beta_{I-i,\xi}^{(1)}, \dots, \beta_{I-i,\xi}^{(N)})'$$

Proof: The normalized incremental claim payments $\mathbf{Y}_{i,j}$ fulfill the model assumptions of the multidimensional Bühlmann-Straub model. Hence, Theorem 2.2 is a direct consequence of Theorem 7.8 in Bühlmann and Gisler (2005). \square

Remarks

- It is usual to compress the data $\mathbf{X}_{i,0}, \dots, \mathbf{X}_{i,I-i}$ in an appropriate manner so that we have a single observation vector \mathbf{K}_i , which has the same dimension as $\mu(\Theta_i)$:

$$\mathbf{K}_i = \left(\sum_{j=0}^{I-i} \frac{w_{i,j,\xi,\delta}^{(1)}}{\sum_{j=0}^{I-i} w_{i,j,\xi,\delta}^{(1)}} Y_{i,j}^{(1)}, \dots, \sum_{j=0}^{I-i} \frac{w_{i,j,\xi,\delta}^{(N)}}{\sum_{j=0}^{I-i} w_{i,j,\xi,\delta}^{(N)}} Y_{i,j}^{(N)} \right)$$

$$= \left(\sum_{j=0}^{I-i} \frac{\gamma_j^{(1)\xi-1}}{\beta_{I-i,\xi}^{(1)} \mu_i^{(1)}} X_{i,j}^{(1)}, \dots, \sum_{j=0}^{I-i} \frac{\gamma_j^{(N)\xi-1}}{\beta_{I-i,\xi}^{(N)} \mu_i^{(N)}} X_{i,j}^{(N)} \right).$$

Observe that the compressed vector \mathbf{K}_i only depends on the observations of accident year i . This is a consequence of the independence assumption between different accident years. Vector \mathbf{K}_i contains all information which is relevant for accident year i and its n -th component is defined as the weighted average of the normalized incremental claims $Y_{i,j}^{(n)}$ over all observed development years $j = 0, \dots, I - i$.

- Note that

$$\mathbf{D}(\beta_{I-i,\xi})^{-1/2} \mathbf{D}(\mu_i)^{-\delta/2} S \mathbf{D}(\mu_i)^{-\delta/2} \mathbf{D}(\beta_{I-i,\xi})^{-1/2}$$

$$= \begin{pmatrix} \frac{\sigma_1^2}{\beta_{I-i,\xi}^{(1)} \mu_i^{(1)\delta}} & & 0 \\ & \ddots & \\ 0 & & \frac{\sigma_N^2}{\beta_{I-i,\xi}^{(N)} \mu_i^{(N)\delta}} \end{pmatrix},$$

where $\sigma_n^2 = E[\sigma_n^2(\Theta_i)]$ for $n = 1, \dots, N$.

- Credibility predictor (2.9) is unbiased for the prior mean $E[\mu(\Theta_i)] = 1$.
- In the case of one-dimensional observations and $\xi = 1$ predictor (2.9) reduces to the one-dimensional credibility predictor applied to the claims reserving problem (cf. Wüthrich and Merz 2008).

To obtain a predictor for the outstanding claim payments we have to determine an estimator for the parameter γ_j and β_j , respectively. An unbiased estimator for γ_j is given by

$$\hat{\gamma}_j = \mathbf{D} \left(\sum_{i=0}^{I-j} \mu_i \right)^{-1} \sum_{i=0}^{I-j} \mathbf{X}_{i,j} \quad (2.10)$$

for all $j = 0, \dots, J$. Thus the estimator

$$\hat{\beta}_j = \sum_{k=0}^j \hat{\gamma}_k \quad (2.11)$$

is an unbiased estimator for β_j for all $j = 0, \dots, J$. This leads to the following predictor:

Predictor 2.3 Under Model Assumptions 2.1 we have the following predictors for the ultimate claims

$$\widehat{\mathbf{C}}_{i,J}^{cred} = \mathbf{C}_{i,I-i} + \mathbf{D}(\hat{\beta}_J - \hat{\beta}_{I-i}) \mathbf{D}(\mu_i) \widehat{\mu(\Theta_i)}^{cred} \quad (2.12)$$

for $1 \leq i \leq I$.

For the numerical calculation of the predictor (2.12) we have to estimate the structure parameter matrices S and T . This will be done in Section 4.

Under Model Assumptions 2.1 the predicted outstanding claim payments are given by

$$\begin{aligned}\widehat{\mathbf{R}}_i^{cred} &= \left(\widehat{R}_i^{(1)cred}, \dots, \widehat{R}_i^{(N)cred} \right)' \\ &= \mathbf{D}(\widehat{\boldsymbol{\beta}}_j - \widehat{\boldsymbol{\beta}}_{I-i}) \mathbf{D}(\boldsymbol{\mu}_i) \widehat{\boldsymbol{\mu}}(\Theta_i)^{cred}\end{aligned}\quad (2.13)$$

for $1 \leq i \leq I$.

For the derivation of the (conditional) MSEP in the next section the following lemma on the quadratic loss matrices of the multidimensional credibility predictors will be used:

Lemma 2.4 *In the multidimensional Bühlmann-Straub Model (2.1), the quadratic loss matrices for the credibility predictors are given by*

$$E\left[\left(\widehat{\boldsymbol{\mu}}(\Theta_i)^{cred} - \boldsymbol{\mu}(\Theta_i)\right) \cdot \left(\widehat{\boldsymbol{\mu}}(\Theta_i)^{cred} - \boldsymbol{\mu}(\Theta_i)\right)'\right] = (\mathbf{I} - \mathbf{A}_i) \mathbf{T} \quad (2.14)$$

for $1 \leq i \leq I$.

Proof: The stated quadratic loss of the credibility predictor $\widehat{\boldsymbol{\mu}}(\Theta_i)^{cred}$ in Lemma 2.4 is a direct consequence of Theorem 7.5 in Bühlmann and Gisler (2005). \square

3. Conditional mean square error of prediction

In the last section we have provided predictors for the ultimate claims and the outstanding claim payments. In this section we quantify the prediction uncertainty of these predictions for single and aggregated accident years in terms of second moments. More precisely, our goal is to derive an estimate for the conditional MSEP of the predicted outstanding claim payments for single as well as aggregated accident years

$$\sum_{n=1}^N \widehat{R}_i^{(n)cred} = \mathbf{1}' \widehat{\mathbf{R}}_i^{cred} \quad \sum_{i=1}^I \sum_{n=1}^N \widehat{R}_i^{(n)cred} = \sum_{i=1}^I \mathbf{1}' \widehat{\mathbf{R}}_i^{cred}.$$

We derive these estimators under the assumption that all parameters in the Bühlmann-Straub credibility predictor $\widehat{\boldsymbol{\mu}}(\Theta_i)^{cred}$ in $\widehat{R}_i^{(n)cred}$ are known. Afterwards, all parameters are replaced by their estimates. This approach is generally applied for such kind of questions and is known as “empirical credibility approach.”

3.1. Single accident years

The conditional MSEP for a single year $i \in \{1, \dots, I\}$, given \mathcal{D}_I^N , is defined by

$$\begin{aligned}\text{mseP}_{\sum_n R_i^{(n)} | \mathcal{D}_I^N} \left(\sum_{n=1}^N \widehat{R}_i^{(n)cred} \right) &= E \left[\left(\sum_{n=1}^N \widehat{R}_i^{(n)cred} - \sum_{n=1}^N R_i^{(n)} \right)^2 \middle| \mathcal{D}_I^N \right] \\ &= \mathbf{1}' E \left[\left(\widehat{\mathbf{R}}_i^{cred} - \mathbf{R}_i \right) \left(\widehat{\mathbf{R}}_i^{cred} - \mathbf{R}_i \right)' \middle| \mathcal{D}_I^N \right] \mathbf{1}.\end{aligned}\quad (3.1)$$

It can be decomposed into conditional process variance and conditional estimation error:

$$\begin{aligned}\text{mseP}_{\sum_n R_i^{(n)} | \mathcal{D}_I^N} \left(\sum_{n=1}^N \widehat{R}_i^{(n)cred} \right) &= \underbrace{\mathbf{1}' \text{Var}(\mathbf{R}_i | \mathcal{D}_I^N) \mathbf{1}}_{\text{conditional process variance}} \\ &+ \underbrace{\mathbf{1}' \left(\widehat{\mathbf{R}}_i^{cred} - E[\mathbf{R}_i | \mathcal{D}_I^N] \right) \left(\widehat{\mathbf{R}}_i^{cred} - E[\mathbf{R}_i | \mathcal{D}_I^N] \right)' \mathbf{1}}_{\text{conditional estimation error}}.\end{aligned}\quad (3.2)$$

Note that for this decoupling we use the fact that $\widehat{\mathbf{R}}_i^{cred}$ is known/observable at time I (i.e., $\widehat{\mathbf{R}}_i^{cred}$ is \mathcal{D}_I^N -measurable). For the conditional process variance we obtain the following result:

Lemma 3.1 *Under Model Assumptions 2.1 the conditional process variance for a single accident year $i \in \{1, \dots, I\}$ is given by*

$$\begin{aligned}\mathbf{1}' \text{Var}(\mathbf{R}_i | \mathcal{D}_I^N) \mathbf{1} &= \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i)^{2-\delta} E[\Sigma(\Theta_i) | \mathcal{D}_I^N] \sum_{j=I-i+1}^J \mathbf{D}(\boldsymbol{\gamma}_j)^{2-\xi} \mathbf{1} \\ &+ \mathbf{1}' \mathbf{D}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{I-i}) \mathbf{D}(\boldsymbol{\mu}_i) \text{Var}(\boldsymbol{\mu}(\Theta_i) | \mathcal{D}_I^N) \\ &\quad \mathbf{D}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{I-i}) \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{1}.\end{aligned}\quad (3.3)$$

Proof. See Appendix. \square

The conditional process variance $\mathbf{1}' \text{Var}(\mathbf{R}_i | \mathcal{D}_I^N)$ originates from stochastic movements of \mathbf{R}_i . If we approximate $E[\Sigma(\Theta_i) | \mathcal{D}_I^N]$ and $\text{Var}(\boldsymbol{\mu}(\Theta_i) | \mathcal{D}_I^N)$ by S and T , respectively, and if we replace the development pattern $\boldsymbol{\gamma}_j$ as well as the structure parameters S and T by their corresponding estimators (see Section 4) we

obtain the following estimator of the conditional process variance for a single accident year:

Estimator 3.2 (Conditional process variance for single accident years)

Under Model Assumptions 2.1 we have the following estimator for the conditional process variance of a single accident year $i \in \{1, \dots, I\}$:

$$\begin{aligned} \mathbf{1}' \widehat{\text{Var}}(\mathbf{R}_i | \mathcal{D}_i^N) \mathbf{1} &= \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i)^{2-\delta} \hat{S} \sum_{j=i+1}^I \mathbf{D}(\hat{\gamma}_j)^{2-\xi} \mathbf{1} \\ &\quad + \mathbf{1}' \mathbf{D}(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}_{i-1}) \mathbf{D}(\boldsymbol{\mu}_i) \times \\ &\quad \hat{T} \mathbf{D}(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}_{i-1}) \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{1} \end{aligned} \quad (3.4)$$

(see Section 4 for estimates of the structure parameter matrices S and T .)

The conditional estimation error in (3.2) reflects the uncertainty in the prediction of the conditional expectation (mean value) $E[\mathbf{R}_i | \mathcal{D}_i^N]$ by $\widehat{\mathbf{R}}_i^{\text{cred}}$. After some calculations (see Appendix) and replacing the unknown parameters by their estimates (see Section 4) we obtain the following result:

Estimator 3.3 (Conditional estimation error for single accident years)

Under Model Assumptions 2.1 we have the following estimator for the conditional estimation error of a single accident year $i \in \{1, \dots, I\}$:

$$\begin{aligned} \mathbf{1}' \widehat{\text{Var}}(\widehat{\mathbf{R}}_i^{\text{cred}} | \mathcal{D}_i^N) \mathbf{1} &= -\mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D}(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}_{i-1}) \hat{A}_i \hat{T} \mathbf{D}(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}_{i-1}) \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{1} \\ &\quad + \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D}(\widehat{\boldsymbol{\mu}}(\widehat{\Theta}_i)^{\text{cred}}) \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}_{i-1}) \times \\ &\quad \mathbf{D}(\widehat{\boldsymbol{\mu}}(\widehat{\Theta}_i)^{\text{cred}}) \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{1}, \end{aligned} \quad (3.5)$$

where

$$\begin{aligned} \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}_{i-1}) &= \sum_{j,l>i} \mathbf{D}\left(\sum_{m=0}^{l-j} \boldsymbol{\mu}_m\right)^{-1} \left[\delta_{jl} \mathbf{D}(\hat{\gamma}_j)^{2-\xi} \cdot \sum_{k=0}^{l-j} \mathbf{D}(\boldsymbol{\mu}_k)^{2-\delta} \hat{S} \right. \\ &\quad \left. + \mathbf{D}(\hat{\gamma}_j) \sum_{k=0}^{l-\max\{j,l\}} \mathbf{D}(\boldsymbol{\mu}_k) \hat{T} \mathbf{D}(\boldsymbol{\mu}_k) \mathbf{D}(\hat{\gamma}_l) \right] \mathbf{D}\left(\sum_{m=0}^{l-j} \boldsymbol{\mu}_m\right)^{-1} \end{aligned} \quad (3.6)$$

$$\widehat{\boldsymbol{\mu}}(\widehat{\Theta}_i)^{\text{cred}} = \hat{A}_i \hat{\mathbf{K}}_i + (I - \hat{A}_i) \mathbf{1} \quad (3.7)$$

$$\hat{\mathbf{K}}_i = \left(\sum_{j=0}^{l-i} \frac{(\hat{\gamma}_j^{(1)})^{\xi-1}}{\hat{\beta}_{l-i,\xi}^{(1)} \boldsymbol{\mu}_i^{(1)}} X_{i,j}^{(1)}, \dots, \sum_{j=0}^{l-i} \frac{(\hat{\gamma}_j^{(N)})^{\xi-1}}{\hat{\beta}_{l-i,\xi}^{(N)} \boldsymbol{\mu}_i^{(N)}} X_{i,j}^{(N)} \right)' \quad (3.8)$$

$$\hat{A}_i = \hat{T} \left(\hat{T} + \mathbf{D}(\hat{\boldsymbol{\beta}}_{l-i,\xi})^{-1/2} \mathbf{D}(\boldsymbol{\mu}_i)^{-\delta/2} \hat{S} \mathbf{D}(\boldsymbol{\mu}_i)^{-\delta/2} \mathbf{D}(\hat{\boldsymbol{\beta}}_{l-i,\xi})^{-1/2} \right) \quad (3.9)$$

and $\delta_{jl} = 1$ if $j = l$ and $\delta_{jl} = 0$ else (see Section 4 for estimates of the structure parameter matrices S and T .)

Combining Estimators 3.2 and 3.3 leads to the following estimator for the conditional MSEP of the predicted outstanding claim payments for a single accident year:

Estimator 3.4 (Conditional MSEP for single accident years)

Under Model Assumptions 2.1 we have the following estimator for the conditional MSEP of a single accident year $i \in \{1, \dots, I\}$:

$$\begin{aligned} \widehat{\text{msep}}_{\sum_n \mathbf{R}_i^{(n)} | \mathcal{D}_i^N} \left(\sum_{n=1}^N \widehat{\mathbf{R}}_i^{(n)} \right)^{\text{cred}} &= \mathbf{1}' \widehat{\text{Var}}(\mathbf{R}_i | \mathcal{D}_i^N) \mathbf{1} + \mathbf{1}' \widehat{\text{Var}}(\widehat{\mathbf{R}}_i^{\text{cred}} | \mathcal{D}_i^N) \mathbf{1}, \end{aligned} \quad (3.10)$$

where the two terms on the right-hand side of (3.10) are given by (3.4) and (3.5), respectively.

Remarks

- Predictor (3.7) is the so-called empirical credibility predictor, which results from the credibility predictor (2.9) by replacing the structural parameters by their estimates.
- For $N = 1$ (i.e., only one run-off portfolio) estimator (3.10) leads to the following estimator

$$\begin{aligned} \widehat{\text{msep}}_{\mathbf{R}_i | \mathcal{D}_i} (\widehat{\mathbf{R}}_i^{\text{cred}}) &= \underbrace{\mu_i^{2-\delta} \hat{S} \sum_{j=i+1}^I \hat{\gamma}_j^{2-\xi} + \mu_i^2 \left(\sum_{j=i+1}^I \hat{\gamma}_j \right)^2 \hat{T}}_{\text{conditional process variance}} \\ &\quad - \underbrace{\mu_i^2 \left(\sum_{j=i+1}^I \hat{\gamma}_j \right)^2 \hat{A}_i \hat{T} + \mu_i^2 \left(\widehat{\boldsymbol{\mu}}(\widehat{\Theta}_i)^{\text{cred}} \right)^2 \widehat{\text{Var}}\left(\sum_{j=i+1}^I \hat{\gamma}_j \right)}_{\text{conditional estimation error}}, \end{aligned} \quad (3.11)$$

where

$$\widehat{\text{Var}}\left(\sum_{j=l-i+1}^l \hat{\gamma}_j\right) = \sum_{j,l>l-i} \left(\sum_{m=0}^{l-j} \mu_m\right)^{-1} \left(\sum_{m=0}^{l-j} \mu_m\right)^{-1} \left[\delta_{jl} \hat{\gamma}_j^{2-\xi} \sum_{k=0}^{l-j} \mu_k^{2-\xi} \hat{S} + \hat{\gamma}_j \hat{\gamma}_l \hat{T} \sum_{k=0}^{l-\max\{j,l\}} \mu_k^2\right] \quad (3.12)$$

and

$$\widehat{\mu}(\Theta_i)^{\text{cred}} = \hat{A}_i \hat{K}_i + (1 - \hat{A}_i) 1$$

with

$$\hat{K}_i = \frac{\sum_{j=0}^{l-i} \hat{\gamma}_j^{\xi-1}}{\hat{\mu}_{l-i,\xi} \mu_i} X_{i,j} \quad \text{and} \quad \hat{A}_i = \frac{\hat{\beta}_{l-i,\xi}}{\hat{\beta}_{l-i,\xi} + \frac{\hat{S}}{\mu_i^\delta \hat{T}}}.$$

3.2. Aggregated accident years

At first we consider the case of two different accident years $1 \leq i < k \leq l$. We have to be careful if we aggregate the estimators $\widehat{\mathbf{R}}_i^{\text{cred}}$ and $\widehat{\mathbf{R}}_k^{\text{cred}}$ because they use the same observations for estimating the parameters γ_j and β_j , respectively. Therefore, they are not independent. We define the conditional MSEP of two aggregated accident years i and k by

$$\begin{aligned} \text{mse}_{\sum_n \mathbf{R}_i^{(n)} + \sum_n \mathbf{R}_k^{(n)} | \mathcal{D}_l^N} \left(\sum_{n=1}^N \widehat{\mathbf{R}}_i^{(n)\text{cred}} + \sum_{n=1}^N \widehat{\mathbf{R}}_k^{(n)\text{cred}} \right) \\ = E \left[\left(\sum_{n=1}^N \left(\widehat{\mathbf{R}}_i^{(n)\text{cred}} + \widehat{\mathbf{R}}_k^{(n)\text{cred}} \right) - \sum_{n=1}^N (\mathbf{R}_i^{(n)} + \mathbf{R}_k^{(n)}) \right)^2 \middle| \mathcal{D}_l^N \right]. \end{aligned}$$

As for a single accident year we have the decomposition

$$\begin{aligned} \text{mse}_{\sum_n \mathbf{R}_i^{(n)} + \sum_n \mathbf{R}_k^{(n)} | \mathcal{D}_l^N} \left(\sum_{n=1}^N \widehat{\mathbf{R}}_i^{(n)\text{cred}} + \sum_{n=1}^N \widehat{\mathbf{R}}_k^{(n)\text{cred}} \right) \\ = \mathbf{1}' \text{Var}(\mathbf{R}_i + \mathbf{R}_k | \mathcal{D}_l^N) \mathbf{1} \\ + \mathbf{1}' \left(\widehat{\mathbf{R}}_i^{\text{cred}} + \widehat{\mathbf{R}}_k^{\text{cred}} - E[\mathbf{R}_i + \mathbf{R}_k | \mathcal{D}_l^N] \right) \times \\ \left(\widehat{\mathbf{R}}_i^{\text{cred}} + \widehat{\mathbf{R}}_k^{\text{cred}} - E[\mathbf{R}_i + \mathbf{R}_k | \mathcal{D}_l^N] \right)' \mathbf{1}. \end{aligned}$$

Using the independence of different accident years, the conditional MSEP of two aggregated accident years can be represented as follows:

$$\begin{aligned} \text{mse}_{\sum_n \mathbf{R}_i^{(n)} + \sum_n \mathbf{R}_k^{(n)} | \mathcal{D}_l^N} \left(\sum_{n=1}^N \widehat{\mathbf{R}}_i^{(n)\text{cred}} + \sum_{n=1}^N \widehat{\mathbf{R}}_k^{(n)\text{cred}} \right) \\ = \text{mse}_{\sum_n \mathbf{R}_i^{(n)} | \mathcal{D}_l^N} \left(\sum_{n=1}^N \widehat{\mathbf{R}}_i^{(n)\text{cred}} \right) + \text{mse}_{\sum_n \mathbf{R}_k^{(n)} | \mathcal{D}_l^N} \left(\sum_{n=1}^N \widehat{\mathbf{R}}_k^{(n)\text{cred}} \right) \\ + 2 \cdot \mathbf{1}' \left(\widehat{\mathbf{R}}_i^{\text{cred}} - E[\mathbf{R}_i | \mathcal{D}_l^N] \right) \left(\widehat{\mathbf{R}}_k^{\text{cred}} - E[\mathbf{R}_k | \mathcal{D}_l^N] \right)' \mathbf{1}. \end{aligned} \quad (3.13)$$

Since we have already derived an estimator for the first and second term on the right-hand side of (3.13) (cf. Estimator 3.4) we only have to derive an estimator for the third term to obtain an estimator for the MSEP (3.13). After some calculations (see Appendix) and replacing the parameters by their estimates (see Section 4) we obtain for the third term the following estimator:

$$2 \cdot \mathbf{1}' \mathbf{D}(\mu_i) \mathbf{D}(\widehat{\mu}(\Theta_i)^{\text{cred}}) \widehat{\text{Cov}}(\hat{\beta}_j - \hat{\beta}_{l-i}, \hat{\beta}_j - \hat{\beta}_{l-k}) \times \mathbf{D}(\widehat{\mu}(\Theta_k)^{\text{cred}}) \mathbf{D}(\mu_k) \mathbf{1}, \quad (3.14)$$

where

$$\begin{aligned} \widehat{\text{Cov}}(\hat{\beta}_j - \hat{\beta}_{l-i}, \hat{\beta}_j - \hat{\beta}_{l-k}) \\ = \widehat{\text{Var}}(\hat{\beta}_j - \hat{\beta}_{l-i}) + \sum_{\substack{l-i < j \\ l-k < l \leq l-i}} \mathbf{D} \left(\sum_{m=0}^{l-j} \mu_m \right)^{-1} \mathbf{D}(\hat{\gamma}_j) \times \\ \sum_{s=0}^{l-\max\{j,l\}} \mathbf{D}(\mu_s) \hat{T} \mathbf{D}(\mu_s) \mathbf{D}(\hat{\gamma}_l) \mathbf{D} \left(\sum_{m=0}^{l-l} \mu_m \right)^{-1} \end{aligned} \quad (3.15)$$

and $\widehat{\text{Var}}(\hat{\beta}_j - \hat{\beta}_{l-i})$ is given by (3.6). For more details to the derivation see the Appendix.

For the generalization on more than two accident years we use the decomposition

$$\begin{aligned} \text{mse}_{\sum_{i,n} \mathbf{R}_i^{(n)} | \mathcal{D}_l^N} \left(\sum_{i=1}^l \sum_{n=1}^N \widehat{\mathbf{R}}_i^{(n)\text{cred}} \right) = \sum_{i=1}^l \text{mse}_{\sum_n \mathbf{R}_i^{(n)} | \mathcal{D}_l^N} \left(\sum_{n=1}^N \widehat{\mathbf{R}}_i^{(n)\text{cred}} \right) \\ + 2 \sum_{1 \leq i < k \leq l} \mathbf{1}' \left(\widehat{\mathbf{R}}_i^{\text{cred}} - E[\mathbf{R}_i | \mathcal{D}_l^N] \right) \left(\widehat{\mathbf{R}}_k^{\text{cred}} - E[\mathbf{R}_k | \mathcal{D}_l^N] \right)' \mathbf{1}. \end{aligned}$$

This leads to the following estimator for the conditional MSEP of aggregated years:

Estimator 3.5 (Conditional MSEP for aggregated accident years)

Under Model Assumptions 2.1 we have the following estimator for the conditional MSEP for aggregated accident years:

$$\begin{aligned} \widehat{\text{mse}}_{\sum_i \sum_n R_i^{(n)} | \mathcal{D}_I^N} \left(\sum_{i=1}^I \sum_{n=1}^N \widehat{R}_i^{(n) \text{ cred}} \right) \\ = \sum_{i=1}^I \widehat{\text{mse}}_{\sum_n R_i^{(n)} | \mathcal{D}_I^N} \left(\sum_{n=1}^N \widehat{R}_i^{(n) \text{ cred}} \right) \\ + 2 \sum_{1 \leq i < k \leq I} \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D} \left(\widehat{\boldsymbol{\mu}}(\Theta_i) \right) \\ \widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}_J - \hat{\boldsymbol{\beta}}_{I-i}, \hat{\boldsymbol{\beta}}_J - \hat{\boldsymbol{\beta}}_{I-k}) \mathbf{D} \left(\widehat{\boldsymbol{\mu}}(\Theta_k) \right) \mathbf{D}(\boldsymbol{\mu}_k) \mathbf{1}, \end{aligned} \quad (3.16)$$

with $\widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}_J - \hat{\boldsymbol{\beta}}_{I-i}, \hat{\boldsymbol{\beta}}_J - \hat{\boldsymbol{\beta}}_{I-k})$ given by (3.15).

Remark

For $N = 1$ (i.e., only one run-off portfolio) Estimator 3.5 leads to the estimator

$$\begin{aligned} \widehat{\text{mse}}_{\sum_i R_i | \mathcal{D}_I^N} \left(\sum_{i=1}^I \widehat{R}_i^{\text{ cred}} \right) = \sum_{i=1}^I \widehat{\text{mse}}_{R_i | \mathcal{D}_I^N} \left(\widehat{R}_i^{\text{ cred}} \right) \\ + 2 \sum_{1 \leq j < k \leq I} \boldsymbol{\mu}_i \boldsymbol{\mu}_k \widehat{\boldsymbol{\mu}}(\Theta_i) \widehat{\boldsymbol{\mu}}(\Theta_k) \widehat{\text{Cov}} \left(\sum_{j=I-i+1}^J \hat{\gamma}_j, \sum_{j=I-k+1}^J \hat{\gamma}_j \right), \end{aligned}$$

where $\widehat{\text{mse}}_{R_i | \mathcal{D}_I^N} \left(\widehat{R}_i^{\text{ cred}} \right)$ is given in (3.11) and

$$\begin{aligned} \widehat{\text{Cov}} \left(\sum_{j=I-i+1}^J \hat{\gamma}_j, \sum_{j=I-k+1}^J \hat{\gamma}_j \right) = \widehat{\text{Var}} \left(\sum_{j=I-i+1}^J \hat{\gamma}_j \right) \\ + \sum_{\substack{I-i < j \\ I-k < j < I-i}} \left(\sum_{m=0}^{I-j} \boldsymbol{\mu}_m \right)^{-1} \left(\sum_{m=0}^{I-l} \boldsymbol{\mu}_m \right)^{-1} \hat{\gamma}_j \hat{\gamma}_l \hat{T} \sum_{s=0}^{I-\max\{j,l\}} \boldsymbol{\mu}_s^2, \end{aligned}$$

with $\widehat{\text{Var}} \left(\sum_{j=I-i+1}^J \hat{\gamma}_j \right)$ given in (26).

4. Parameter estimation

The estimators of the development pattern $\boldsymbol{\gamma}_j$ and $\boldsymbol{\beta}_j$ are given in (2.10) and (2.11). In this section we first give estimators for the structure parameter matrices S and T for known $\boldsymbol{\gamma}_j$ and $\boldsymbol{\beta}_j$. Then, we replace $\boldsymbol{\gamma}_i$ and

$\boldsymbol{\beta}_j$ by their estimators (2.10) and (2.11) in order to get the final estimators. Under Model Assumptions 2.1 S is a diagonal matrix, where the diagonal elements $\sigma_n^2 = E[\sigma_n^2(\Theta_i)]$, $n = 1, \dots, N$ can be estimated by (see Bühlmann and Gisler 2005):

$$\hat{\sigma}_n^2 = \frac{1}{I} \sum_{i=0}^{I-1} \frac{1}{I-i} \sum_{j=0}^{I-1} \gamma_j^{\xi(n)} \mu_i^{\delta(n)} \left(\frac{X_{i,j}^{(n)}}{\gamma_j^{(n)} \mu_i^{(n)}} - K_i^{(n)} \right)^2. \quad (4.1)$$

We replace $\gamma_j^{(n)}$ by $\hat{\gamma}_j^{(n)}$ in (4.1) and thus the estimator of the structure parameter matrix S is given by

$$\hat{S} = \begin{pmatrix} \hat{\sigma}_n^2 & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \hat{\sigma}_n^2 \end{pmatrix}.$$

For the estimation of the diagonal elements of T we define

$$\hat{T}_{n,n}^{\circ} = c \cdot \left(\sum_{i=0}^I \frac{\beta_{\xi, I-i}^{(n)} \mu_i^{\delta(n)}}{\sum_k \beta_{\xi, I-k}^{(n)} \mu_k^{\delta(n)}} \cdot \left(K_i^{(n)} - \bar{K}^{(n)} \right)^2 - \frac{I \cdot \hat{\sigma}_n^2}{\sum_k \beta_{\xi, I-k}^{(n)} \mu_k^{\delta(n)}} \right)$$

with

$$c = \left(\sum_{i=0}^I \frac{\beta_{\xi, I-i}^{(n)} \mu_i^{\delta(n)}}{\sum_k \beta_{\xi, I-k}^{(n)} \mu_k^{\delta(n)}} \cdot \left(1 - \frac{\beta_{\xi, I-i}^{(n)} \mu_i^{\delta(n)}}{\sum_k \beta_{\xi, I-k}^{(n)} \mu_k^{\delta(n)}} \right) \right)^{-1}$$

and

$$\bar{K}^{(n)} = \sum_{i=0}^I \frac{\beta_{\xi, I-i}^{(n)} \mu_i^{\delta(n)}}{\sum_k \beta_{\xi, I-k}^{(n)} \mu_k^{\delta(n)}} \cdot K_i^{(n)} = \sum_{i=0}^I \frac{\mu_i^{\delta-1(n)} \sum_{j \leq I-1} \gamma_j^{\xi-1(n)} X_{i, I-i}^{(n)}}{\sum_k \beta_{\xi, I-k}^{(n)} \mu_k^{\delta(n)}}.$$

Observe, that for $\delta = 1$ we have $\bar{K}^{(n)} = 1$ for all $n = 1, \dots, N$. Since $\hat{T}_{n,n}^{\circ}$ could be negative, we take

$$\hat{T}_{n,n} = \max(\hat{T}_{n,n}^{\circ}, 0) \quad (4.2)$$

as estimator for the diagonal elements of T . An estimator for the non-diagonal elements of T (i.e., $T_{n,m}$ with $n \neq m$) is given by (see also Bühlmann and Gisler 2005)

$$\hat{T}_{n,m} = \text{sgn}\left(\frac{\hat{T}_{n,m}^a + \hat{T}_{n,m}^b}{2}\right) \cdot \min\left(\frac{|\hat{T}_{n,m}^a + \hat{T}_{n,m}^b|}{2}, \sqrt{\hat{T}_{n,n} \cdot \hat{T}_{m,m}}\right), \quad (4.3)$$

where

$$\hat{T}_{n,m}^a = c_a \cdot \left(\sum_{i=0}^I \frac{\beta_{\xi, I-i}^{(n)} \mu_i^{\delta^{(n)}}}{\sum_k \beta_{\xi, I-k}^{(n)} \mu_k^{\delta^{(n)}}} \cdot (K_i^{(n)} - \bar{K}^{(n)}) \cdot (K_i^{(m)} - \bar{K}^{(m)}) \right)$$

with

$$c_a = \left(\sum_{i=0}^I \frac{\beta_{\xi, I-i}^{(n)} \mu_i^{\delta^{(n)}}}{\sum_k \beta_{\xi, I-k}^{(n)} \mu_k^{\delta^{(n)}}} \cdot \left(1 - \frac{\beta_{\xi, I-i}^{(n)} \mu_i^{\delta^{(n)}}}{\sum_k \beta_{\xi, I-k}^{(n)} \mu_k^{\delta^{(n)}}} \right) \right)^{-1}$$

and

$$\hat{T}_{n,m}^b = c_b \cdot \left(\sum_{i=0}^I \frac{\beta_{\xi, I-i}^{(m)} \mu_i^{\delta^{(m)}}}{\sum_k \beta_{\xi, I-k}^{(m)} \mu_k^{\delta^{(m)}}} \cdot (K_i^{(n)} - \bar{K}^{(n)}) \cdot (K_i^{(m)} - \bar{K}^{(m)}) \right)$$

with

$$c_b = \left(\sum_{i=0}^I \frac{\beta_{\xi, I-i}^{(m)} \mu_i^{\delta^{(m)}}}{\sum_k \beta_{\xi, I-k}^{(m)} \mu_k^{\delta^{(m)}}} \cdot \left(1 - \frac{\beta_{\xi, I-i}^{(m)} \mu_i^{\delta^{(m)}}}{\sum_k \beta_{\xi, I-k}^{(m)} \mu_k^{\delta^{(m)}}} \right) \right)^{-1}.$$

The estimator $\hat{T}_{n,m}$ in (4.3) takes the value zero if $\hat{T}_{n,n}$ or $\hat{T}_{m,m}$ is zero. In this case, we have an estimator of T which is not invertible, which leads to an estimator of A_i that is also not invertible. Alternatively to (4.3) we can take as estimator for the non-diagonal elements of T

$$\hat{T}_{n,m} = \frac{\hat{T}_{n,m}^a + \hat{T}_{n,m}^b}{2}. \quad (4.4)$$

By replacing $\beta_{\xi, I-i}^{(n)}$ with their estimators in (4.2) and (4.3) we get an estimator of T , which we denote by $\hat{T} = (\hat{T}_{n,m})_{n,m=1,\dots,N}$.

The estimators $\hat{\sigma}_n^2$, $n = 1, \dots, N$ are unbiased estimators for the components of S and the estimators $\hat{T}_{n,n}^o$, $\hat{T}_{n,n}^a$ and $\hat{T}_{n,n}^b$ are unbiased estimators for the components of T . However, the estimators $\hat{T}_{n,n}$ and $\hat{T}_{n,m}$ are no longer unbiased. Apart from that, we cannot state anything about the unbiasedness of \hat{S} and \hat{T} .

Finally, we get an estimator of A_i by replacing all structure parameters by their estimators, that is

$$\hat{A}_i = \hat{T} \left(\hat{T} + \mathbf{D}(\hat{\beta}_{\xi, I-i})^{-1/2} \mathbf{D}(\mu_i)^{-\delta/2} \cdot \hat{S} \cdot \mathbf{D}(\hat{\beta}_{\xi, I-i})^{-1/2} \mathbf{D}(\mu_i)^{-\delta/2} \right)^{-1}.$$

For the specific choice of the weights $\xi \in [0, 2]$ and $\delta > 0$ we propose the method of minimum sum of squared residuals. That means we estimate ξ and δ by

$$(\hat{\xi}, \hat{\delta}) = \underset{(\xi, \delta)}{\text{argmin}} \{SSE(\xi, \delta)\},$$

with

$$SSE(\xi, \delta) = \sum_{0 \leq i+j \leq I} \sum_{n=1}^N (X_{i,j}^{(n)} - \hat{X}_{i,j}^{(n)})^2. \quad (4.5)$$

Another possible way to determine ξ by means of exploratory data analysis is given in Mack (2002).

5. Example

We consider two portfolios A and B (i.e., $N = 2$) from General Liability Reinsurance and Auto Liability Reinsurance containing incremental claim payments with $I = 16$ accident and $J = 10$ development years. The corresponding data sets are provided in the end of this section. In this case the last accident year is greater than the last development year, i.e., $I > J$. However, all results we presented except the parameter estimate (4.1) also hold for this case. The estimator (4.1) has to be adapted as follows:

$$\hat{\sigma}_n^2 = \frac{1}{I} \sum_{i=I-J}^I \frac{1}{J} \sum_{j=0}^J \gamma_j^{(n)} \mu_i^{\delta^{(n)}} \left(\frac{X_{i,j}^{(n)}}{\gamma_j^{(n)} \mu_i^{(n)}} - K_i^{(n)} \right)^2 + \frac{1}{I} \sum_{i=I-J}^{I-1} \frac{1}{I-i} \sum_{j=0}^{I-i} \gamma_j^{(n)} \mu_i^{\delta^{(n)}} \left(\frac{X_{i,j}^{(n)}}{\gamma_j^{(n)} \mu_i^{(n)}} - K_i^{(n)} \right)^2. \quad (5.1)$$

Table 1. Prior means $\mu_i^{(n)}$

i	7	8	9	10	11	12	13	14	15	16
$\mu_i^{(1)}$	36,824	34,498	42,154	41,681	36,807	36,708	53,947	34,469	34,721	32,377
$\mu_i^{(2)}$	29,864	31,711	39,496	32,810	32,365	39,905	32,526	30,360	35,155	31,751

Table 2. Estimated loss development pattern $\widehat{\gamma}_j^{(n)}$ in Portfolio A and B

$\xi = 0$	0	1	2	3	4	5	6	7	8	9	10
$\widehat{\gamma}_j^{(1)}$	0.54435	0.30141	0.07132	0.03068	0.01890	0.01379	0.01234	0.00359	0.00100	0.00251	0.00010
$\widehat{\gamma}_j^{(2)}$	0.56160	0.30554	0.04912	0.03138	0.01579	0.01456	0.00262	0.00219	0.01283	0.00363	0.00073

We assume different prior means for the different accident years and use the prior means given in Table 1 which result as the ultimate claim predictions of the classical CL method.

Next we calculate the estimator (2.10) for the incremental development pattern $(\gamma_j^{(n)})_{j=0,\dots,J} \subset \mathbb{R}_+$ (note that this estimator is independent from the specific values of ξ and δ) which are given in Table 2. We see that about 60% to 80% of the expected claim payments are due in the first two development years. Moreover, $\widehat{\gamma}_j^{(n)} > 0$ holds for all development years $j = 0, \dots, 10$ fulfilling Model Assumptions 2.1. However, by briefly studying the incremental claims payment pattern of the two portfolios we obtain some “untypical” accident years. Unlike the other accident years, the accident years 2 and 13 show an extremely slow decline, whereas accident year 3 shows a fast decline of the incremental claim payments.

We calculate the credibility predictors (2.9) in order to determine the claims reserves (2.13) for every accident year $i \in \{0, \dots, I\}$. For illustrative purposes we restrict the analysis to four explicit parameter choices $\xi \in \{0, 2\}$ and $\delta \in \{0, 2\}$.

The credibility predictors (2.13) for these four parameter choices are given in Table 3. In a second step we choose the parameters $\xi \in \{0, 2\}$ and $\delta \in \{0, 2\}$, which provide the best model fit to the data.

Goodness-of-Fit

In order to choose specific ξ and δ , we use as a goodness-of-fit criterion the $SSE(\xi, \delta)$, see (4.5), and take the parameter combination $\xi \in \{0, 2\}$, $\delta \in \{0, 2\}$ with minimum $SSE(\xi, \delta)$. This method is also proposed in Chapter 11 in Wüthrich and Merz (2008) for comparing the fit to the data of different claims reserving methods. Table 4 shows that $\xi = 2$ provides a much

Table 3. Credibility predictor $\widehat{\mu}(\Theta_i)^{cred}$ for $\xi \in \{0, 2\}$ and $\delta \in \{0, 2\}$

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$*\widehat{\mu}(\Theta_i)^{cred}$	1.03	0.96	2.33	0.73	0.83	0.64	0.88	0.44	1.22	0.85	0.72	0.80	1.13	1.35	0.97	1.00	1.00
$\xi = 0 \delta = 0$	0.73	0.87	1.74	1.16	0.84	0.75	0.91	0.64	1.31	0.94	0.78	0.84	1.19	1.15	0.97	1.00	1.00
$*\widehat{\mu}(\Theta_i)^{cred}$	0.97	0.94	2.39	0.84	0.81	0.61	0.87	0.44	1.25	0.85	0.70	0.80	1.20	1.50	0.97	1.00	1.00
$\xi = 0 \delta = 2$	0.91	0.93	1.72	1.10	0.87	0.77	0.93	0.69	1.21	0.94	0.82	0.87	1.17	1.23	0.98	1.00	1.00
$*\widehat{\mu}(\Theta_i)^{cred}$	1.01	1.02	0.96	1.04	1.04	1.01	0.99	1.03	1.01	1.02	1.03	1.01	0.96	0.89	1.00	1.00	1.00
$\xi = 2 \delta = 0$	1.12	1.06	0.96	0.96	1.01	1.02	1.01	1.01	0.98	0.98	1.05	1.03	0.91	0.98	1.02	0.99	1.00
$*\widehat{\mu}(\Theta_i)^{cred}$	1.02	1.02	0.96	1.02	1.03	1.01	0.99	1.03	1.00	1.02	1.03	1.01	0.97	0.87	1.00	1.00	1.00
$\xi = 2 \delta = 2$	1.07	1.03	0.97	0.96	1.01	1.01	1.01	1.01	0.98	0.98	1.04	1.03	0.91	0.98	1.02	0.99	1.00

Table 4. Estimated loss development pattern $\widehat{\gamma}_j^{(n)}$ in Portfolio A and B

	$\xi = 0, \delta = 0$	$\xi = 0, \delta = 2$	$\xi = 2, \delta = 0$	$\xi = 2, \delta = 2$
$\sqrt{SSE(\xi, \delta)}$	57955	60104	19996	19862

better fit than $\xi = 0$ what can be explained as follows: In Model Assumptions 2.1 the incremental claim payments are of the type $\mathbf{Y}_{ij}|\Theta_i \sim \boldsymbol{\mu}(\Theta_i) + \boldsymbol{\varepsilon}_{ij}$ with $E[\boldsymbol{\varepsilon}_{ij}|\Theta_i] = 0$ and $\text{Var}(\boldsymbol{\varepsilon}_{ij}|\Theta_i) = \mathbf{D}(\mathbf{w}_{ij, \xi, \delta})^{-1/2} \cdot \boldsymbol{\Sigma}(\Theta_i) \cdot \mathbf{D}(\mathbf{w}_{ij, \xi, \delta})^{-1/2}$. In the case of $\xi = 0$ the conditional variance $\text{Var}(\boldsymbol{\varepsilon}_{ij}|\Theta_i)$ does not depend on development year j and we assume homogeneous variances of $(\mathbf{Y}_{ij}|\Theta_i)_{0 \leq j \leq J}$, although this assumption seems quite unrealistic. Then the weights of the standardized incremental claim payments \mathbf{Y}_{ij} in the compressed data vector \mathbf{K}_i in Theorem 2.2 also do not depend on the development year j . This leads to high credibility predictors for the untypical accident years (2 and 13) and small predictors for accident year 7

for $\xi = 0$ (see Table 3) and consequently to a high $SSE(0, \delta)$. The $\sqrt{SSE(0, \delta)}$ decreases by about 40% if these “untypical” accident years (2, 7, and 13) are left out in the calculation of $\sqrt{SSE(0, \delta)}$. Heterogeneous variances are respected in the case $\xi = 2$ leading to a much better model fit (see Table 4). Figure 2 shows the predicted claim payments versus the corresponding residuals. There is no clear trend in the plot, and the data are (approximately) centered. However, there remains the problem of untypical accident years which cannot be fitted appropriately. This problem is visualized in Figure 2, where in Portfolio A and B the highest and lowest values (outliers) result from accident years 7 and 13. The $\sqrt{SSE(2, 0)}$ decreases by about 57% (from 19,996 to 8,512) if the untypical accident years (2, 7 and 13) are ignored.

For $\xi = 2$ and $\delta = 0$ we analyse the standardized residuals given by

$$r_{i,j} = \mathbf{D}(\boldsymbol{\mu}_i)^{-2} \boldsymbol{\Sigma}(\Theta_i)^{-1} (\mathbf{X}_{i,j} - \mathbf{D}(\boldsymbol{\gamma}_j) \mathbf{D}(\boldsymbol{\mu}_i) \boldsymbol{\mu}(\Theta_i)). \quad (5.2)$$

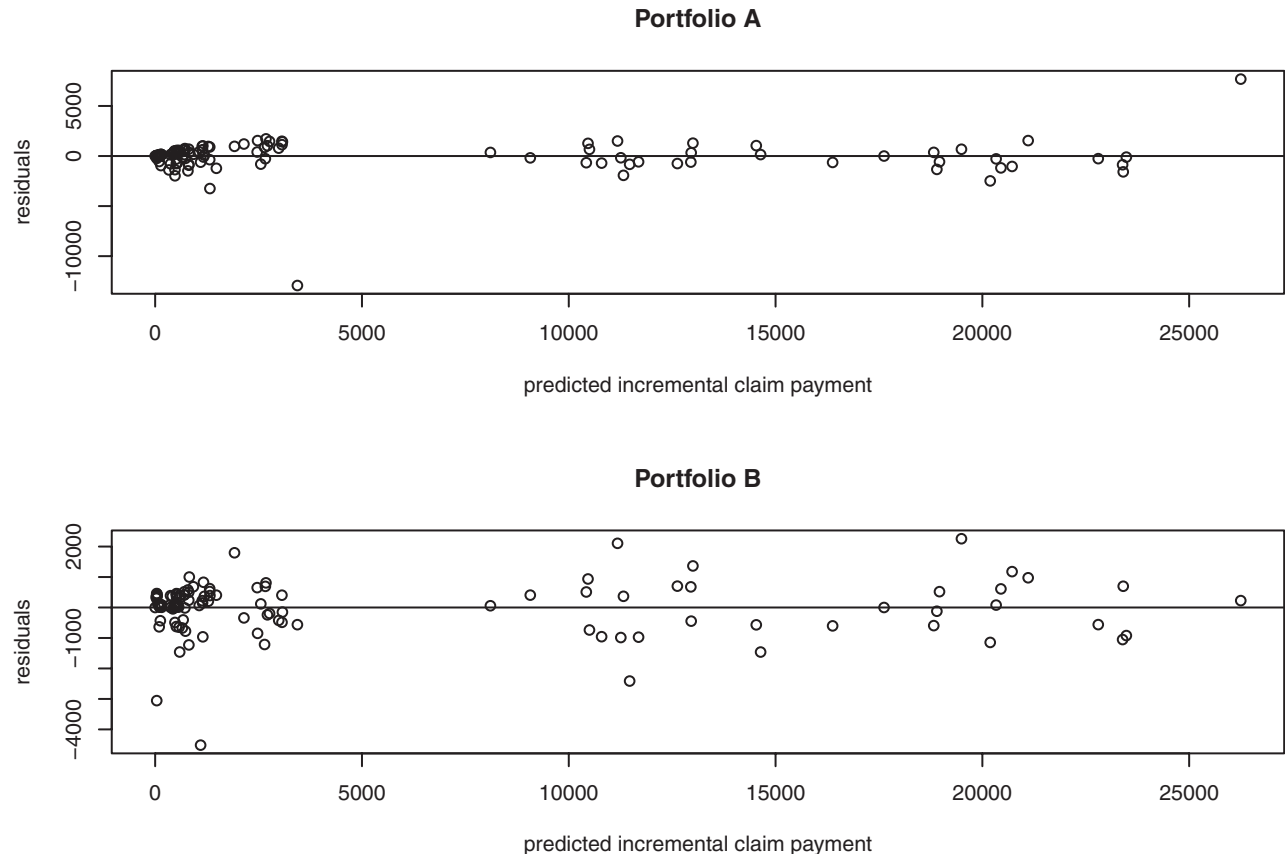
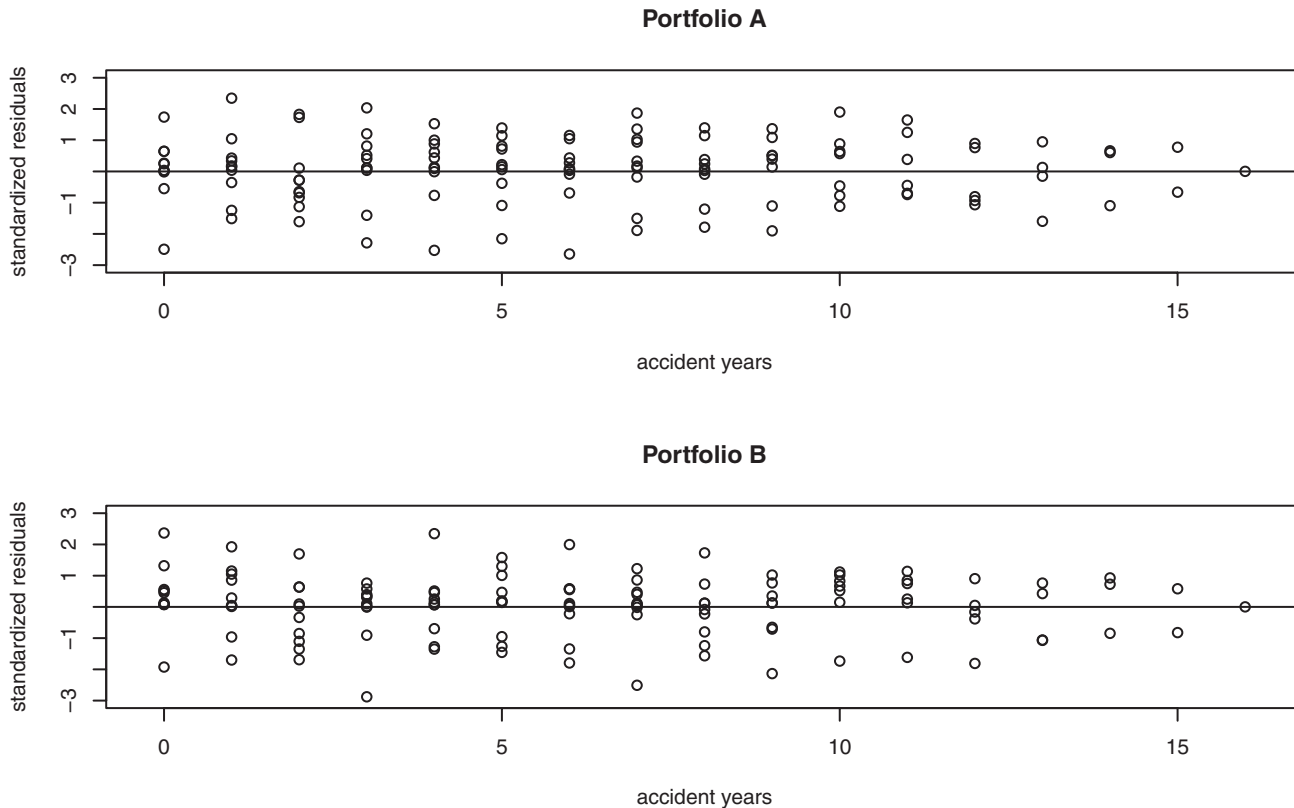
Figure 2. Empirical residuals vs. predicted incremental payments ($\xi = 2$ and $\delta = 0$).

Figure 3. Empirical standardized residuals of Portfolio A and B by accident year

These standardized residuals have, conditionally given Θ_i , zero mean and diagonal identity matrix as covariance matrix. Replacing all unknown parameters in (5.2) by their estimates leads to the empirical standardized residuals $\hat{r}_{i,j}$. For Portfolio A and B, Figure 3 shows the empirical standardized residuals which are (approximately) centered and do not seem to have a trend.

We consider for the case $\xi = 2$ and $\delta = 0$ the corresponding reserves 2.13 as well as the associated prediction uncertainty. Table 5 shows the estimates for the aggregated claims reserves, conditional process standard deviation, squared conditional estimation error and conditional standard error of prediction for the aggregated reserves over all accident years resulting by the multivariate Bühlmann-Straub ($\xi = 2$ and $\delta = 0$), chain-ladder and additive loss reserving model. The last two columns contain the estimates of the fourth iteration for the multivariate chain-ladder and additive loss reserving model, respec-

tively (see Merz and Wüthrich (2008), Chapter 8, for more details to the models). We obtain in the Bühlmann-Straub model higher reserves as well as a higher prediction uncertainty compared to the other two models. The difference can partly be explained by the fact that allowing for different “accident years qualities” (through $\widehat{\mu}(\Theta_i)^{cred}$) in the multivariate Bühlmann-Straub model increases the parameter uncertainty and the process variance of the model. Tables 6 and 7 show the observed incremental claim payments in Portfolios A and B.

Table 5. Results for the whole portfolio for aggregated accident years

	Bühlmann-Straub model	CL model	ALR model
Estimated reserves	54350	52,734	54,042
Process std. deviation	14063	9,697	7,539
$\sqrt{\text{Estimation error}}$	12454	5,549	4,749
Prediction std. error	18.785	11,172	8,910

Table 6. Observed incremental claim payments $C_{ij}^{(1)}$ in Portfolio A

	0	1	2	3	4	5	6	7	8	9	10
0	14.492	7.746	949	467	814	234	1.718	104	15	49	4
1	17.017	9.251	945	750	33	196	1.144	18	10	1	11
2	19.563	12.265	1.302	1.099	967	1.237	1.838	1.093	276	643	0
3	21.632	13.249	963	136	25	172	27	39	7	0	0
4	22.672	9.677	1.779	153	95	1.265	31	1	45	0	8
5	23.062	13.375	2.200	327	2.273	20	23	0	0	2	0
6	23.588	11.713	1.660	4.569	1.662	256	0	39	0	32	2
7	21.758	12.300	1.685	1.266	-28	-41	-49	-45	-26	0	
8	20.233	9.197	932	644	754	2.452	161	34	0		
9	24.984	12.632	1.931	415	1.730	163	113	33			
10	24.260	13.555	1.585	1.679	123	147	32				
11	20.616	11.430	2.932	516	558	37					
12	18.814	11.499	3.363	1.711	97						
13	18.563	13.492	16.370	2.704							
14	18.457	11.089	2.064								
15	19.533	9.833									
16	17.620										

Table 7. Observed incremental claim payments $C_{ij}^{(2)}$ in Portfolio B

	0	1	2	3	4	5	6	7	8	9	10
0	16.651	8.206	-468	-152	5	4	6	0	1	0	0
1	16.292	8.129	1.713	195	426	35	68	50	55	0	11
2	16.658	10.566	1.736	618	1.272	442	572	15	38	749	2
3	19.715	10.690	968	2.098	154	197	100	63	3.519	24	149
4	21.220	8.815	2.969	896	151	1.146	26	20	0	0	2
5	21.302	10.582	2.237	952	7	1.183	14	0	26	70	-9
6	17.201	7.493	1.574	518	1.685	25	2	56	-8	71	7
7	15.835	11.668	1.728	122	493	-10	4	1	1	0	
8	17.560	8.550	2.373	908	1.163	389	127	501	2		
9	21.051	12.279	2.387	690	372	2.024	14	0			
10	20.368	9.832	1.285	460	43	186	0				
11	18.623	11.160	942	892	8	28					
12	18.112	12.040	1.662	5.654	979						
13	17.744	10.346	2.134	599							
14	17.993	8.956	869								
15	19.082	11.403									
16	17.809										

Appendix: Proofs and derivations

In order to simplify the notation, we omit the superscript “cred” in the following derivations.

A.1. Proof of Lemma 3.1

We have, using the conditional independence of the incremental claims, given Θ_i ,

$$\begin{aligned} \mathbf{1}' \text{Var}(\mathbf{R}_i | \mathcal{D}_i^N) \mathbf{1} &= \mathbf{1}' E[\text{Var}(\mathbf{R}_i | \mathcal{D}_i^N, \Theta_i) | \mathcal{D}_i^N] \mathbf{1} \\ &\quad + \mathbf{1}' \text{Var}(E[\mathbf{R}_i | \mathcal{D}_i^N, \Theta_i] | \mathcal{D}_i^N) \mathbf{1} \\ &= \mathbf{1}' E[\text{Var}(\mathbf{R}_i | \Theta_i) | \mathcal{D}_i^N] \mathbf{1} \\ &\quad + \mathbf{1}' \text{Var}(E[\mathbf{R}_i | \Theta_i] | \mathcal{D}_i^N) \mathbf{1} \quad (\text{A.1}) \end{aligned}$$

for $i \in \{1, \dots, I\}$. Using (2.6) we obtain for the second term on the right-hand side

$$\begin{aligned} \mathbf{1}' \text{Var}(E[\mathbf{R}_i | \Theta_i] | \mathcal{D}_i^N) \mathbf{1} &= \mathbf{1}' \text{Var}(\mathbf{D}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{l-i}) \mathbf{D}(\boldsymbol{\mu}_i) \boldsymbol{\mu}(\Theta_i) | \mathcal{D}_i^N) \mathbf{1} \\ &= \mathbf{1}' \mathbf{D}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{l-i}) \mathbf{D}(\boldsymbol{\mu}_i) \text{Var}(\boldsymbol{\mu}(\Theta_i) | \mathcal{D}_i^N) \mathbf{D}(\boldsymbol{\mu}_i) \times \\ &\quad \mathbf{D}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{l-i}) \mathbf{1}. \end{aligned}$$

Using the conditional independence of the normalized incremental claims $\mathbf{Y}_{i,l-i+1}, \dots, \mathbf{Y}_{i,J}$ and (2.5) we obtain for the first term on the right-hand side of (A.1)

$$\begin{aligned} \mathbf{1}' E[\text{Var}(\mathbf{R}_i | \Theta_i) | \mathcal{D}_i^N] \mathbf{1} &= \mathbf{1}' E \left[\sum_{j=l-i+1}^J \mathbf{D}(\mathbf{w}_{i,j}) \text{Var}(\mathbf{Y}_{i,j} | \Theta_i) \mathbf{D}(\mathbf{w}_{i,j}) | \mathcal{D}_i^N \right] \mathbf{1} \\ &= \mathbf{1}' E \left[\sum_{j=l-i+1}^J \mathbf{D}(\mathbf{w}_{i,j}) \mathbf{D}(\mathbf{w}_{i,j,\xi,\delta})^{-1/2} \boldsymbol{\Sigma}(\Theta_i) \times \right. \\ &\quad \left. \mathbf{D}(\mathbf{w}_{i,j,\xi,\delta})^{-1/2} \mathbf{D}(\mathbf{w}_{i,j}) | \mathcal{D}_i^N \right] \mathbf{1} \\ &= \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i)^{2-\delta} E \left[\sum_{j=l-i+1}^J \mathbf{D}(\gamma_j)^{2-\xi} \boldsymbol{\Sigma}(\Theta_i) | \mathcal{D}_i^N \right] \mathbf{1}. \end{aligned}$$

This finishes the proof of Lemma 3.1.

A.2. Derivation of conditional estimation error for single accident years

Using the conditional independence of the normalized incremental claims, given Θ_i , and (2.13) we obtain

$$\begin{aligned} E[\mathbf{R}_i | \mathcal{D}_i^N] &= \sum_{j=l-i+1}^J E[E[\mathbf{X}_{i,j} | \Theta_i, \mathcal{D}_i^N] | \mathcal{D}_i^N] \\ &= \sum_{j=l-i+1}^J E[\mathbf{D}(\mathbf{w}_{i,j}) E[\mathbf{Y}_{i,j} | \Theta_i] | \mathcal{D}_i^N] \\ &= \mathbf{D} \left(\sum_{j>l-i} \gamma_j \right) \mathbf{D}(\boldsymbol{\mu}_i) E[\boldsymbol{\mu}(\Theta_i) | \mathcal{D}_i^N] \end{aligned}$$

and

$$\widehat{\mathbf{R}}_i = \mathbf{D} \left(\sum_{j>l-i} \hat{\gamma}_j \right) \mathbf{D}(\boldsymbol{\mu}_i) \widehat{\boldsymbol{\mu}(\Theta_i)}$$

respectively. This leads to

$$\begin{aligned} \widehat{\mathbf{R}}_i - E[\mathbf{R}_i | \mathcal{D}_i^N] &= \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D} \left(\sum_{j>l-i} \gamma_j \right) (\widehat{\boldsymbol{\mu}(\Theta_i)} - E[\boldsymbol{\mu}(\Theta_i) | \mathcal{D}_i^N]) \\ &\quad - \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D} \left(\sum_{j>l-i} \gamma_j (\gamma_j - \hat{\gamma}_j) \right) \widehat{\boldsymbol{\mu}(\Theta_i)}. \end{aligned}$$

Hence, we obtain the following formula for the conditional estimation error

$$\begin{aligned} \mathbf{1}' (\widehat{\mathbf{R}}_i - E[\mathbf{R}_i | \mathcal{D}_i^N]) (\widehat{\mathbf{R}}_i - E[\mathbf{R}_i | \mathcal{D}_i^N])' \mathbf{1} &= \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D} \left(\sum_{j>l-i} \gamma_j \right) (\widehat{\boldsymbol{\mu}(\Theta_i)} - E[\boldsymbol{\mu}(\Theta_i) | \mathcal{D}_i^N]) \times \\ &\quad (\widehat{\boldsymbol{\mu}(\Theta_i)} - E[\boldsymbol{\mu}(\Theta_i) | \mathcal{D}_i^N])' \mathbf{D} \left(\sum_{j>l-i} \gamma_j \right) \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{1} \\ &\quad - \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D} \left(\sum_{j>l-i} (\gamma_j - \hat{\gamma}_j) \right) \widehat{\boldsymbol{\mu}(\Theta_i)} \times \\ &\quad (\widehat{\boldsymbol{\mu}(\Theta_i)} - E[\boldsymbol{\mu}(\Theta_i) | \mathcal{D}_i^N])' \mathbf{D} \left(\sum_{j>l-i} \gamma_j \right) \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{1} \\ &\quad - \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D} \left(\sum_{j>l-i} \gamma_j \right) (\widehat{\boldsymbol{\mu}(\Theta_i)} - E[\boldsymbol{\mu}(\Theta_i) | \mathcal{D}_i^N]) \end{aligned}$$

$$\begin{aligned}
 & \widehat{\boldsymbol{\mu}}(\Theta_i)' \mathbf{D} \left(\sum_{j>I-i} (\gamma_j - \hat{\gamma}_j) \right) \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{1} \\
 & - \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D} \left(\sum_{j>I-i} (\gamma_j - \hat{\gamma}_j) \right) \widehat{\boldsymbol{\mu}}(\Theta_i) \widehat{\boldsymbol{\mu}}(\Theta_i)' \times \\
 & \mathbf{D} \left(\sum_{j>I-i} (\gamma_j - \hat{\gamma}_j) \right) \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{1}. \quad (\text{A.2})
 \end{aligned}$$

To get an estimator for the conditional estimation error we have to determine an estimator of each term on the right-hand side of (A.2). Using (2.14) and the approximation $E[\text{Var}(\boldsymbol{\mu}(\Theta_i) | \mathcal{D}_I^N)] \approx T$ we obtain the approximation

$$\begin{aligned}
 & E \left[\left(\widehat{\boldsymbol{\mu}}(\Theta_i) - E[\boldsymbol{\mu}(\Theta_i) | \mathcal{D}_I^N] \right) \left(\widehat{\boldsymbol{\mu}}(\Theta_i) - E[\boldsymbol{\mu}(\Theta_i) | \mathcal{D}_I^N] \right)' \right] \\
 & = E \left[E \left[\left(\boldsymbol{\mu}(\Theta_i) - \widehat{\boldsymbol{\mu}}(\Theta_i) \right) \left(\boldsymbol{\mu}(\Theta_i) - \widehat{\boldsymbol{\mu}}(\Theta_i) \right)' \middle| \mathcal{D}_I^N \right] \right. \\
 & \quad \left. - \text{Var}(\boldsymbol{\mu}(\Theta_i) | \mathcal{D}_I^N) \right] \\
 & \approx (I - A_i) T - T \\
 & = -A_i T. \quad (\text{A.3})
 \end{aligned}$$

This leads to the following approximation for the first term:

$$-\mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D} \left(\sum_{j>I-i} \gamma_j \right) A_i T \mathbf{D} \left(\sum_{j>I-i} \gamma_j \right) \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{1}. \quad (\text{A.4})$$

Since different accident years are independent, the expectation of the second and third term disappears. Using the fact that it holds

$$\mathbf{D}(\mathbf{b}) \mathbf{a} \mathbf{c}' \mathbf{D}(\mathbf{d}) = \mathbf{D}(\mathbf{a}) \mathbf{b} \mathbf{d}' \mathbf{D}(\mathbf{c})$$

for all N -dimensional vectors \mathbf{a} , \mathbf{b} , \mathbf{c} and \mathbf{d} , we obtain for the fourth term on the right-hand side of (A.2)

$$\begin{aligned}
 & \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D}(\widehat{\boldsymbol{\mu}}(\Theta_i)) \sum_{j>I-i} (\gamma_j - \hat{\gamma}_j) \sum_{j>I-i} (\gamma_j - \hat{\gamma}_j)' \times \\
 & \mathbf{D}(\widehat{\boldsymbol{\mu}}(\Theta_i)) \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{1}. \quad (\text{A.5})
 \end{aligned}$$

In the following we approximate (A.5) by

$$\begin{aligned}
 & \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D}(\widehat{\boldsymbol{\mu}}(\Theta_i)) E \left[\sum_{j>I-i} (\gamma_j - \hat{\gamma}_j) \sum_{j>I-i} (\gamma_j - \hat{\gamma}_j)' \right] \times \\
 & \mathbf{D}(\widehat{\boldsymbol{\mu}}(\Theta_i)) \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{1} \\
 & = \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D}(\widehat{\boldsymbol{\mu}}(\Theta_i)) \text{Var} \left(\sum_{j>I-i} \hat{\gamma}_j \right) \times \\
 & \mathbf{D}(\widehat{\boldsymbol{\mu}}(\Theta_i)) \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{1}. \quad (\text{A.6})
 \end{aligned}$$

For this end we have to calculate

$$\begin{aligned}
 \text{Var} \left(\sum_{j>I-i} \hat{\gamma}_j \right) & = E \left[\text{Var} \left(\sum_{j>I-i} \hat{\gamma}_j | \Theta \right) \right] \\
 & + \text{Var} \left(E \left[\sum_{j>I-i} \hat{\gamma}_j | \Theta \right] \right), \quad (\text{A.7})
 \end{aligned}$$

where $\Theta = (\Theta_0, \dots, \Theta_I)$. Using the independence of the accident years and the conditional independence of the normalized incremental claims $\mathbf{Y}_{k,0}, \dots, \mathbf{Y}_{k,J}$, given Θ_k , we obtain for the first term on the right-hand side of (A.7)

$$\begin{aligned}
 & E \left[\text{Var} \left(\sum_{j>I-i} \hat{\gamma}_j | \Theta \right) \right] = \sum_{j>I-i} E \left[\text{Var}(\hat{\gamma}_j | \Theta) \right] \\
 & = \sum_{j>I-i} E \left[\mathbf{D} \left(\sum_{m=0}^{I-j} \boldsymbol{\mu}_m \right)^{-1} \sum_{k=0}^{I-j} \text{Var}(\mathbf{D}(\mathbf{w}_{i,j}) \mathbf{Y}_{k,j} | \Theta_k) \times \right. \\
 & \quad \left. \mathbf{D} \left(\sum_{m=0}^{I-j} \boldsymbol{\mu}_m \right)^{-1} \right] \\
 & = \sum_{j>I-i} E \left[\mathbf{D} \left(\sum_{m=0}^{I-j} \boldsymbol{\mu}_m \right)^{-1} \sum_{k=0}^{I-j} \mathbf{D}(\boldsymbol{\mu}_k)^{2-\delta} \times \right. \\
 & \quad \left. \mathbf{D}(\gamma_j)^{2-\xi} \sum (\Theta_k) \mathbf{D} \left(\sum_{m=0}^{I-j} \boldsymbol{\mu}_m \right)^{-1} \right] \\
 & = \sum_{j>I-i} \mathbf{D} \left(\sum_{m=0}^{I-j} \boldsymbol{\mu}_m \right)^{-1} \mathbf{D}(\gamma_j)^{2-\xi} \sum_{k=0}^{I-j} \mathbf{D}(\boldsymbol{\mu}_k)^{2-\delta} S \mathbf{D} \left(\sum_{m=0}^{I-j} \boldsymbol{\mu}_m \right)^{-1} \quad (\text{A.8})
 \end{aligned}$$

For the second term on the right-hand side of (A.7) we obtain

$$\begin{aligned} \text{Var}\left(E\left[\sum_{j>l-i}\hat{\gamma}_j|\Theta\right]\right) &= \text{Var}\left(\sum_{j>l-i}E\left[\mathbf{D}\left(\sum_{m=0}^{l-j}\boldsymbol{\mu}_m\right)^{-1}\sum_{k=0}^{l-j}\mathbf{X}_{k,j}|\Theta\right]\right) \\ &= \text{Var}\left(\sum_{j>l-i}\mathbf{D}\left(\sum_{m=0}^{l-j}\boldsymbol{\mu}_m\right)^{-1}\mathbf{D}(\gamma_j)\sum_{k=0}^{l-j}\mathbf{D}(\boldsymbol{\mu}_k)\boldsymbol{\mu}(\Theta_k)\right). \end{aligned}$$

We define

$$\mathbf{Z}_j = \mathbf{D}\left(\sum_{m=0}^{l-j}\boldsymbol{\mu}_m\right)^{-1}\mathbf{D}(\gamma_j)\sum_{k=0}^{l-j}\mathbf{D}(\boldsymbol{\mu}_k)\boldsymbol{\mu}(\Theta_k). \quad (\text{A.9})$$

Using the independence of $\Theta_0, \dots, \Theta_l$ we obtain

$$\begin{aligned} &= \sum_{j>l-i}\text{Var}(\mathbf{Z}_j) + 2 \sum_{l-i<j<l}\text{Cov}(\mathbf{Z}_j, \mathbf{Z}_l) \\ &= \sum_{j>l-i}\mathbf{D}\left(\sum_{m=0}^{l-j}\boldsymbol{\mu}_m\right)^{-1}\mathbf{D}(\gamma_j)\sum_{k=0}^{l-j}\mathbf{D}(\boldsymbol{\mu}_k)T\mathbf{D}(\boldsymbol{\mu}_k)\mathbf{D}(\gamma_j) \\ &\quad \mathbf{D}\left(\sum_{m=0}^{l-j}\boldsymbol{\mu}_m\right)^{-1} + 2 \sum_{l-i<j<l}\mathbf{D}\left(\sum_{m=0}^{l-j}\boldsymbol{\mu}_m\right)^{-1}\mathbf{D}(\gamma_j) \\ &\quad \sum_{k=0}^{l-\max\{j,l\}}\mathbf{D}(\boldsymbol{\mu}_k)T\mathbf{D}(\boldsymbol{\mu}_k)\mathbf{D}(\gamma_l)\mathbf{D}\left(\sum_{m=0}^{l-l}\boldsymbol{\mu}_m\right)^{-1}. \quad (\text{A.10}) \end{aligned}$$

This leads to the term

$$\begin{aligned} \text{Var}\left(\sum_{j>l-i}\hat{\gamma}_j\right) &= \sum_{j>l-i}\mathbf{D}\left(\sum_{m=0}^{l-j}\boldsymbol{\mu}_m\right)^{-1} \\ &\quad \left(\delta_{jl}\mathbf{D}(\gamma_j)^{2-\xi} \cdot \sum_{k=0}^{l-j}\mathbf{D}(\boldsymbol{\mu}_k)^{2-\delta}S + \mathbf{D}(\gamma_j) \times \right. \\ &\quad \left. \sum_{k=0}^{l-\max\{j,l\}}\mathbf{D}(\boldsymbol{\mu}_k)T\mathbf{D}(\boldsymbol{\mu}_k)\mathbf{D}(\gamma_l)\mathbf{D}\left(\sum_{m=0}^{l-l}\boldsymbol{\mu}_m\right)^{-1}\right), \quad (\text{A.11}) \end{aligned}$$

where δ_{jl} is the Kronecker delta, which is 1, when $j = l$ and 0 else.

Putting (A.4), (A.6), and (A.11) together and replacing all parameters by their estimates leads to Estimator 3.3.

A.3. Derivation of conditional MSEF for aggregated accident years

To obtain an estimator of the conditional MSEF for two aggregated accident years i and k we have to derive an estimator for the third term on the right-hand side of equation (3.13). Analogously to (A.2) we have for $1 \leq i < k \leq I$

$$\begin{aligned} &1'(\widehat{\mathbf{R}}_i - E[\mathbf{R}_i|\mathcal{D}_I^N])(\widehat{\mathbf{R}}_k - E[\mathbf{R}_k|\mathcal{D}_I^N])' \mathbf{1} \\ &= 1' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D}\left(\sum_{j>l-i}\gamma_j\right)(\widehat{\boldsymbol{\mu}(\Theta_i)} - E[\boldsymbol{\mu}(\Theta_i)|\mathcal{D}_I^N]) \times \\ &\quad (\widehat{\boldsymbol{\mu}(\Theta_k)} - E[\boldsymbol{\mu}(\Theta_k)|\mathcal{D}_I^N])' \mathbf{D}\left(\sum_{j>l-k}\gamma_j\right) \mathbf{D}(\boldsymbol{\mu}_k) \mathbf{1} \\ &\quad - 1' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D}\left(\sum_{j>l-i}(\gamma_j - \hat{\gamma}_j)\right) \widehat{\boldsymbol{\mu}(\Theta_i)} \times \\ &\quad (\widehat{\boldsymbol{\mu}(\Theta_k)} - E[\boldsymbol{\mu}(\Theta_k)|\mathcal{D}_I^N])' \mathbf{D}\left(\sum_{j>l-k}\gamma_j\right) \mathbf{D}(\boldsymbol{\mu}_k) \mathbf{1} \\ &\quad - 1' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D}\left(\sum_{j>l-i}\gamma_j\right) (\widehat{\boldsymbol{\mu}(\Theta_i)} - E[\boldsymbol{\mu}(\Theta_i)|\mathcal{D}_I^N]) \times \\ &\quad \widehat{\boldsymbol{\mu}(\Theta_k)}' \mathbf{D}\left(\sum_{j>l-k}(\gamma_j - \hat{\gamma}_j)\right) \mathbf{D}(\boldsymbol{\mu}_k) \mathbf{1} + 1' \mathbf{D}(\boldsymbol{\mu}_i) \\ &\quad \mathbf{D}(\widehat{\boldsymbol{\mu}(\Theta_i)}) \left(\sum_{j>l-k}(\gamma_j - \hat{\gamma}_j)\right) \sum_{j>l-k}(\gamma_j - \hat{\gamma}_j)' \times \\ &\quad \mathbf{D}(\widehat{\boldsymbol{\mu}(\Theta_k)}) \mathbf{D}(\boldsymbol{\mu}_k) \mathbf{1}. \quad (\text{A.12}) \end{aligned}$$

As for single accident years, we determine for each term on the right-hand side of (A.12) an estimator. Due to the fact, that different accident years are independent, the expectation of the first and second term disappears. Using again the independence of different accident years we obtain

$$\begin{aligned} &E\left[(\widehat{\boldsymbol{\mu}(\Theta_i)} - E[\boldsymbol{\mu}(\Theta_i)|\mathcal{D}_I^N]) \widehat{\boldsymbol{\mu}(\Theta_i)}' \mathbf{D}\left(\sum_{j>l-k}(\gamma_j - \hat{\gamma}_j)\right)\right] \\ &= E\left[(\widehat{\boldsymbol{\mu}(\Theta_i)} - E[\boldsymbol{\mu}(\Theta_i)|\mathcal{D}_I^N]) \times \right. \\ &\quad \left. \sum_{l-k<j\leq l-i}(\gamma_j - \hat{\gamma}_j)' \mathbf{D}(\widehat{\boldsymbol{\mu}(\Theta_k)})\right] \\ &= -E\left[(\widehat{\boldsymbol{\mu}(\Theta_i)} - E[\boldsymbol{\mu}(\Theta_i)|\mathcal{D}_I^N]) \sum_{l-k<j\leq l-i}\hat{\gamma}_j'\right] \\ &\quad \cdot E[\mathbf{D}(\widehat{\boldsymbol{\mu}(\Theta_k)})] \quad (\text{A.13}) \end{aligned}$$

As mentioned in the remarks to Theorem 2.2, the credibility predictor $\widehat{\boldsymbol{\mu}}(\Theta_i)$ can be constructed using orthogonal projections in Hilbert spaces. Since $\widehat{\boldsymbol{\mu}}(\Theta_i)$ is the orthogonal projection of the so-called Bayes-Predictor $E[\boldsymbol{\mu}(\Theta_i)|\mathcal{D}_l^N]$ on the Hilbert space $L(\mathcal{D}_l^N, 1)$, it fulfills the orthogonality condition

$$E\left[\left(\widehat{\boldsymbol{\mu}}(\Theta_i) - E[\boldsymbol{\mu}(\Theta_i)|\mathcal{D}_l^N]\right) \cdot X'_{l,m}\right]$$

for all l, m (see Bühlmann and Gisler (2005), p. 182f.). Thus, the expectation from equation (A.13) disappears, so that we have zero as estimator of the third term. We approximate the fourth term by

$$\begin{aligned} & \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D}(\widehat{\boldsymbol{\mu}}(\Theta_i)) E\left[\sum_{j>l-i} (\gamma_j - \hat{\gamma}_j) \sum_{j>l-k} (\gamma_j - \hat{\gamma}_j)'\right] \times \\ & \quad \mathbf{D}(\widehat{\boldsymbol{\mu}}(\Theta_k)) \mathbf{D}(\boldsymbol{\mu}_k) \mathbf{1} \\ &= \mathbf{1}' \mathbf{D}(\boldsymbol{\mu}_i) \mathbf{D}(\widehat{\boldsymbol{\mu}}(\Theta_i)) \text{Cov}\left[\sum_{j>l-i} \hat{\gamma}_j, \sum_{j>l-k} \hat{\gamma}_j\right] \times \\ & \quad \mathbf{D}(\widehat{\boldsymbol{\mu}}(\Theta_k)) \mathbf{D}(\boldsymbol{\mu}_k) \mathbf{1}. \end{aligned} \quad (\text{A.14})$$

So, we have to calculate

$$\begin{aligned} & \text{Cov}\left(\sum_{j>l-i} \hat{\gamma}_j, \sum_{j>l-k} \hat{\gamma}_j\right) \\ &= E\left[\text{Cov}\left(\sum_{j>l-i} \hat{\gamma}_j, \sum_{j>l-k} \hat{\gamma}_j \middle| \Theta\right)\right] + \text{Cov}\left(E\left[\sum_{j>l-i} \hat{\gamma}_j \middle| \Theta\right], \right. \\ & \quad \left. E\left[\sum_{j>l-k} \hat{\gamma}_j \middle| \Theta\right]\right) \end{aligned} \quad (\text{A.15})$$

where $\Theta = (\Theta_0, \dots, \Theta_l)$. Using the independence of different accident years and the conditional independence of the normalized incremental claims $\mathbf{Y}_{k,0}, \dots, \mathbf{Y}_{k,l}$, given Θ_k , we obtain for the first term on the right-hand side of (A.15)

$$\begin{aligned} E\left[\text{Cov}\left(\sum_{j>l-i} \hat{\gamma}_j, \sum_{j>l-k} \hat{\gamma}_j \middle| \Theta\right)\right] &= E\left[\sum_{j>l-i} \text{Var}(\hat{\gamma}_j | \Theta)\right] \\ &= E\left[\text{Var}\left(\sum_{j>l-i} \hat{\gamma}_j \middle| \Theta\right)\right], \end{aligned}$$

which is given in (A.8). For the calculation of the second term on the right-hand side of (A.15) we use the variables \mathbf{Z}_j defined in (A.9) and obtain

$$\begin{aligned} & \text{Cov}\left(E\left[\sum_{j>l-i} \hat{\gamma}_j \middle| \Theta\right], E\left[\sum_{j>l-k} \hat{\gamma}_j \middle| \Theta\right]\right) = \text{Cov}\left(\sum_{j>l-i} \mathbf{Z}_j, \sum_{j>l-k} \mathbf{Z}_j\right) \\ &= \sum_{j>l-i} \text{Var}(\mathbf{Z}_j) + 2 \sum_{\substack{l-i < j \\ l-k < l \leq l-i}} \text{Cov}(\mathbf{Z}_j, \mathbf{Z}_l) \\ &+ \sum_{\substack{l-i < j \\ l-k < l \leq l-i}} \text{Cov}(\mathbf{Z}_j, \mathbf{Z}_l) = \text{Var}\left(E\left[\sum_{j>l-i} \hat{\gamma}_j \middle| \Theta\right]\right) \\ &+ \sum_{\substack{l-i < j \\ l-k < l \leq l-i}} \mathbf{D}\left(\sum_{m=0}^{l-j} \boldsymbol{\mu}_m\right)^{-1} \mathbf{D}(\gamma_j) \sum_{s=0}^{l-\max\{j,l\}} \mathbf{D}(\boldsymbol{\mu}_s) T \mathbf{D}(\boldsymbol{\mu}_s) \times \\ & \quad \mathbf{D}(\gamma_l) \mathbf{D}\left(\sum_{m=0}^{l-l} \boldsymbol{\mu}_m\right)^{-1}, \end{aligned}$$

where $\text{Var}(E[\sum_{j>l-i} \hat{\gamma}_j | \Theta])$ is given in (A.10). Hence, we obtain

$$\begin{aligned} & \text{Cov}\left(\sum_{j>l-i} \hat{\gamma}_j, \sum_{j>l-k} \hat{\gamma}_j\right) = \text{Var}\left(\sum_{j>l-i} \hat{\gamma}_j\right) \\ &+ \sum_{\substack{l-i < j \\ l-k < l \leq l-i}} \mathbf{D}\left(\sum_{m=0}^{l-j} \boldsymbol{\mu}_m\right)^{-1} \mathbf{D}(\gamma_j) \times \\ & \quad \sum_{s=0}^{l-\max\{j,l\}} \mathbf{D}(\boldsymbol{\mu}_s) T \mathbf{D}(\boldsymbol{\mu}_s) \mathbf{D}(\gamma_l) \mathbf{D}\left(\sum_{m=0}^{l-l} \boldsymbol{\mu}_m\right)^{-1}, \end{aligned} \quad (\text{A.16})$$

where $\text{Var}(\sum_{j>l-i} \hat{\gamma}_j)$ is given in (A.11).

Putting (A.14) and (A.16) together and replacing all parameters by their estimates leads to Estimator 3.14.

References

- Benktander, G., "An Approach to Credibility in Calculating IBNR for Casualty Excess Reinsurance," *Actuarial Review*, April 1976, p. 7.
- Braun, C. "The Prediction Error of the Chain Ladder Method Applied to Correlated Run Off Triangles," *ASTIN Bulletin* 34/2, 2004, pp. 399–423.
- Bühlmann, H., and A. Gisler, *A Course in Credibility Theory and its Applications*, Berlin: Springer-Verlag, 2005.

- Bühlmann, H., and E. Straub, "Glaubwürdigkeit für Schadensätze," *Bulletin of Swiss Association of Actuaries*, 1970, pp. 111–133.
- CAS (Casualty Actuarial Society), *Foundations of Casualty Actuarial Science*, fourth edition, Arlington, VA: CAS, 2001.
- Dahms, R., "Linear Stochastic Reserving Methods," *ASTIN Bulletin* 42:1, 2012, pp. 1–34.
- Dahms, R., and S. Happ, "Credibility for the Linear Stochastic Reserving Methods," forthcoming.
- De Vylder, F., "Estimation of IBNR Claims by Credibility Theory," *Insurance: Mathematics and Economics* 1, 1982, pp. 35–40.
- England, P.D., and R.J. Verrall, "Predictive Distributions of Outstanding Liabilities in General Insurance," *Annals of Actuarial Science* 1:2, 2007, pp. 221–270.
- Gisler, A., and M. V. Wüthrich, "Credibility for the Chain Ladder Reserving Method," *ASTIN Bulletin* 38:2, 2008, pp. 565–600.
- Halliwell, L., "Conjoint Prediction of Paid and Incurred Losses," *CAS Forum* (Summer) 1, 1999, pp. 241–379.
- Hashorva, E., M. Merz, and M. V. Wüthrich, "Dependence Modeling in Multivariate Claims Run-off Triangles," *Annals of Actuarial Science* 7: 1, 2013, pp. 3–25.
- Hess, K. T., K. D. Schmidt, and M. Zocher, "Multivariate Loss Prediction in the Multivariate Additive Model," *Insurance, Mathematics and Economics* 39, 2006, pp. 185–191.
- Holmberg, R. D., "Correlation and the Measurement of Loss Reserve Variability," *CAS Forum* (Spring) 1994, pp. 247–278.
- Mack, T., *Schadenversicherungsmathematik*. Karlsruhe: Verlag Versicherungswirtschaft, 2002.
- Mack, T., "Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates," *ASTIN Bulletin* 23, 1993, pp. 213–225.
- Mack, T., "Credible Claims Reserves: The Benktander Method," *ASTIN Bulletin* 30:2, 2000, pp. 333–347.
- Merz, M., and M. V. Wüthrich, "Prediction Error of the Chain Ladder Reserving Method Applied to Correlated Run off Triangles," *Annals of Actuarial Science* 2:1, 2008a, pp. 25–50.
- Merz, M., and M. V. Wüthrich, "Prediction Error of the Multivariate Chain Ladder Reserving Method," *North American Actuarial Journal* 12:2, 2008b, pp. 175–197.
- Merz, M., and Wüthrich, "Prediction Error of the Multivariate Additive Loss Reserving Method for Dependent Lines of Business," *Variance* 3, 2009, pp. 131–151.
- Neuhaus, W., "Another Pragmatic Loss Reserving Method or Bornhuetter/Ferguson Revisited," *Scandinavian Actuarial Journal* 2, 1992, pp. 151–162.
- Quarg, G., and T. Mack, "Munich Chain Ladder," *Blätter DGVFM*, Volume 26, 2004, pp. 597–630.
- Schmidt, K. D., "Convergence of Bayes and Credibility Premiums," *ASTIN Bulletin* 20, 1990, pp. 167–172.
- Teugels, J. L., and B. Sundt, *Encyclopedia of Actuarial Science*, Vol. 1, Chichester: Wiley, 2004.
- Witting, T., "Kreditabilitätsschätzungen für die Anzahl IBNR-Schäden," *Blätter der DGVFM* Band 18, 1987, pp. 45–58.
- Wüthrich, M. V., and M. Merz, *Stochastic Claims Reserving Methods in Insurance*, Hoboken, NJ: Wiley, 2008.