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 cashflow 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
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 (\( \sigma \)-algebras)

\[
\mathcal{D}_I^n = \sigma (C^{(n)}_{ij}; i + j \leq I) \subseteq \mathcal{F}
\]

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

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

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

\[
X = (X^{(n)}_{ij}; n \in \{1, \ldots, N\}, i \in \{0, \ldots, I\}, j \in \{0, \ldots, J\})',
\]

and

\[
C = (C^{(n)}_{ij}; n \in \{1, \ldots, N\}, i \in \{0, \ldots, I\}, j \in \{0, \ldots, J\})'.
\]

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

\[
R_i = (R^{(1)}_i, \ldots, R^{(N)}_i)' = C_{ij} - C_{i,j-1} = \sum_{j=1}^{J} X_{ij}.
\]

Furthermore, we define the \( N \)-dimensional column vector consisting of one’s by

\[
1' = (1, \ldots, 1)' \in \mathbb{R}^N \quad \text{and} \quad \mathbb{I} = \mathbb{I}_N \times \mathbb{I}_N.
\]

### 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 $i \in \{1, \ldots, I\}$, which is the best affine-linear function of the components of the observations $X_{i,0}, \ldots, X_{i,J}$ 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 $\Theta_i$, which describe the risk characteristics of the different accident years $i = 0, \ldots, I$. Moreover, we assume that there are (known) volume measures $\mu^{(n)}_i$ and (unknown) incremental loss development patterns (cash flow patterns) $(\gamma^{(n)}_{i,j})_{j=0,\ldots,J} \subset \mathbb{R}$ for the $N$ development triangles $n \in \{1, \ldots, N\}$, such that

$$E[X^{(n)}_{i,j}] = \gamma^{(n)}_{i,j} \mu^{(n)}_i \quad (2.1)$$

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

$$Y^{(n)}_{i,j} = \frac{X^{(n)}_{i,j}}{\gamma^{(n)}_{i,j} \mu^{(n)}_i}$$

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

$$\beta^{(n)}_0 = \gamma^{(n)}_{0} > 0 \text{ and } \beta^{(n)}_j - \beta^{(n)}_j = \gamma^{(n)}_{j} > 0 \quad \forall j = 1, \ldots, J. \quad (2.2)$$

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

$$\gamma_i = (\gamma^{(n)}_{i,0}, \ldots, \gamma^{(n)}_{i,J})'$$
$$\beta_i = (\beta^{(n)}_{0}, \ldots, \beta^{(n)}_{J})'$$
$$\mu_i = (\mu^{(n)}_{i,0}, \ldots, \mu^{(n)}_{i,J})'$$
$$Y_{i,j} = (Y^{(n)}_{i,0}, \ldots, Y^{(n)}_{i,J})'.$$

In the following we denote by

$$D(a) = \begin{pmatrix} a_i & 0 & \cdots & 0 \\ 0 & a_j & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & a_k \end{pmatrix}$$
$$D(a) = \begin{pmatrix} a_i^b & 0 & \cdots & 0 \\ 0 & a_j^b & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & a_k^b \end{pmatrix}$$
the $N \times N$-diagonal matrices of the $N$-dimensional vectors $a = (a_i, \ldots, a_k)' \in \mathbb{R}^N$ and $(a_i^b, \ldots, a_k^b)' \in \mathbb{R}^N$ for an admissible exponent $b \in \mathbb{R}$, respectively. Then we have for the normalized incremental claims

$$Y_{i,j} = D(w_{i,j})^{-1} X_{i,j}, \quad (2.3)$$

where $w_{i,j} = (w^{(1)}_{i,j}, \ldots, w^{(N)}_{i,j})'$ with $w^{(n)}_{i,j} = \gamma^{(n)}_{i,j} \mu^{(n)}_i$ for all $i = 0, \ldots, I$, $j = 0, \ldots, J$ and $n = 1, \ldots, N$.

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

**Model Assumptions 2.1**

(Multivariate Bühlmann-Straub model)

- Conditionally, given $\Theta_i$, the normalized incremental claims $Y_{i,0}, \ldots, Y_{i,J}$ are independent with

$$E[Y_{i,J} | \Theta_i] = \mu(\Theta_i) \quad (2.4)$$
Var \( (Y_{i,j} | \Theta) = D(\mathbf{w}_{i,j,k})^{-1/2} \cdot \Sigma(\Theta) \cdot D(\mathbf{w}_{i,j,k})^{-1/2} \)

\[
\begin{pmatrix}
\sigma^2_i(\Theta) & 0 & \ldots & 0 \\
0 & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \ldots & 0 & \sigma^2_i(\Theta) \\
\end{pmatrix}
\]

(2.5)

where \( w_{i,j,k}^{(n)} = \gamma_{ij}^{(n)} \eta_{ki}^{(n)} \) with \( \xi \in [0, 2] \) and \( \delta \geq 0 \). The matrix \( \Sigma(\Theta) = D(\sigma^2_i(\Theta), \ldots, \sigma^2_i(\Theta)) \) is the \( N \times N \)-diagonal matrix of \( (\sigma^2_i(\Theta), \ldots, \sigma^2_i(\Theta))' \).

- The pairs \( (\Theta_i, Y_i') = (Y_i', \ldots, Y_i') \) for \( i = 0, \ldots, I \) are independent and the latent variables \( \Theta_o, \ldots, \Theta_i \) are identically distributed.

**Remarks**

- From (2.4) it follows that the normalized incremental claim payments \( Y_{i,j} \) of accident year \( i \) are higher or lower than the normalized incremental claim payments \( Y_{i,k} \) 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 \( \mu_i = (\mu_1^{(i)}, \ldots, \mu_N^{(i)})' \) are known and that the incremental loss development pattern \( \gamma_i = (\gamma_1^{(i)}, \ldots, \gamma_N^{(i)})' \) 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 \( \xi \) and \( \delta \) can be derived.

- The parameter \( \Theta_i \) 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

\[
E[Y_{i,j}] = E[\mu(\Theta)] = 1
\]

and, respectively,

\[
E[C_{i,j}] = D(\beta_n) D(\mu_n) \mu(\Theta)
\]

for all \( i = 0, \ldots, I \) and \( j = 0, \ldots, J \). Moreover, we obtain for the claims reserves

\[
E[R_{i,j}] = (D(\beta_n) - D(\beta_n)) D(\mu_n) \mu(\Theta)
\]

for all \( i = 1, \ldots, I \).

In the following we define the mean square error of prediction (MSEP) of a \( N \)-dimensional predictor \( \hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_N)' \) for a \( N \)-dimensional random variable \( \mathbf{x} = (x_1, \ldots, x_N)' \) by

\[
\text{msep}_\mathbf{x}(\hat{\mathbf{x}}) = \sum_{i=1}^N E[(\hat{x}_i - x_i)^2]
\]

(2.7)

In the multidimensional credibility theory one looks now for a predictor \( \hat{\mu}(\Theta) \) or \( \mu(\Theta) \) that minimizes the MSEP (2.7) among all \( N \)-dimensional predictors \( \hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_N)' \) whose components \( \hat{x}_i \) are affine-linear in the components of the \( N \)-dimensional observations \( x_{i,0}, \ldots, x_{i,j-1} \) with \( i = 0, \ldots, I \). That is, one has to solve the optimization problem

\[
\hat{\mu}(\Theta) = \arg\min_{\Theta \in (D^+_n \cup D^-_n)} \text{msep}_{\mathbf{mu}(\Theta)}(\hat{\mathbf{x}}),
\]

where

\[
L(D^+_n, 1) = \left\{ \hat{\mathbf{x}}; \hat{x}_i = a + \sum_{i=1}^N \sum_{j=1}^N a_{ij}^{(n)} X_{ij}^{(n)} \text{ with } a, a_{ij}^{(n)} \in \mathbb{R} \right\}
\]

(2.8)

We define the structural parameter matrices

\[
S = E[\mu(\Theta)]
\]

\[
T = \text{Var}(\mu(\Theta))
\]

and obtain:
Theorem 2.2 (Bühlmann-Straub predictor)

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

\[
\hat{\mu}(\Theta) = A \mathbf{K}_i + (I - A_i) \mathbf{1}
\]

for \( 1 \leq i \leq I \), where

\[
\mathbf{K}_i = \left( \sum_{j=0}^{I-1} \gamma_{ij} X_{ij}^{(1)}, \ldots, \sum_{j=0}^{I-1} \gamma_{ij} X_{ij}^{(N)} \right)^{\prime}
\]

\[
A_i = T(T + D(\mathbf{B}_{t+i})^{-1} D(\mu_\ast) - D(\mu_\ast) D(\mathbf{B}_{t+i})^{-1})
\]

\[
\beta_{t+i} = (\beta_{t+i}^{(1)}, \ldots, \beta_{t+i}^{(N)})^{\prime}
\]

**Proof:** The normalized incremental claim payments \( Y_{ij} \) 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). □

**Remarks**

- It is usual to compress the data \( X_{ij}, \ldots, X_{j,k} \) in an appropriate manner so that we have a single observation vector \( \mathbf{K}_i \), which has the same dimension as \( \mu(\Theta) \):

\[
\mathbf{K}_i = \left( \sum_{j=0}^{I-1} \frac{W_{ij}^{(1)}}{\sum_{j=0}^{I-1} W_{ij}^{(1)}} Y_{ij}^{(1)}, \ldots, \sum_{j=0}^{I-1} \frac{W_{ij}^{(N)}}{\sum_{j=0}^{I-1} W_{ij}^{(N)}} Y_{ij}^{(N)} \right)^{\prime}
\]

\[
= \left( \sum_{j=0}^{I-1} \frac{Y_{ij}^{(1)} \mu_\ast^{(1)}}{\beta_{t+i}^{(1)} \mu_\ast^{(1)}} X_{ij}^{(1)}, \ldots, \sum_{j=0}^{I-1} \frac{Y_{ij}^{(N)} \mu_\ast^{(N)}}{\beta_{t+i}^{(N)} \mu_\ast^{(N)}} X_{ij}^{(N)} \right)^{\prime}
\]

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 \( Y_{ij}^{(n)} \) over all observed development years \( j = 0, \ldots, I - i \).

- Note that

\[
D(\beta_{t+i})^{-1} D(\mu_\ast) S D(\mu_\ast) D(\beta_{t+i})^{-1} = \begin{pmatrix}
\sigma_0^2 & 0 \\
0 & \frac{\sigma_n^2}{\beta_{t+i}^{(n)} \mu_\ast^{(n)}}
\end{pmatrix}
\]

where \( \sigma_n^2 = E[\sigma_n^2(\Theta)] \) for \( n = 1, \ldots, N \).

- Credibility predictor (2.9) is unbiased for the prior mean \( E[\mu(\Theta)] = \mathbf{1} \).

- 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 \( \gamma_i \) and \( \beta_i \), respectively. An unbiased estimator for \( \gamma_i \) is given by

\[
\gamma_i = D \left( \sum_{j=0}^{I-1} \mu_\ast^{(1)} \right)^{\prime} \sum_{j=0}^{I-1} X_{ij}
\]

for all \( j = 0, \ldots, J \). Thus the estimator

\[
\hat{\beta}_j = \sum_{i=0}^{J-1} \gamma_i
\]

is an unbiased estimator for \( \beta_j \) for all \( j = 0, \ldots, J \). This leads to the following predictor:

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

\[
\hat{C}_{i,j}^{\text{cred}} = C_{i,j} + D(\hat{\beta}_j - \hat{\beta}_i) D(\mu_\ast) \mu(\Theta)^{\text{cred}}
\]

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
\[ \tilde{R}_i^{\text{cred}} = \left( \tilde{R}_1^{(1)}, \ldots, \tilde{R}_n^{(n)} \right) \]

\[ = D(\hat{\beta}_i - \bar{\beta}_n) D(\mu_0) \mu(\Theta) \tag{2.13} \]

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

\[ E[\left( \mu(\Theta) - \mu(\hat{\Theta}) \right) \cdot \left( \mu(\Theta) - \mu(\hat{\Theta}) \right)^T] = (I - A)T \tag{2.14} \]

for \( 1 \leq i \leq I. \)

**Proof:** The stated quadratic loss of the credibility predictor \( \mu(\hat{\Theta}) \) in Lemma 2.4 is a direct consequence of Theorem 7.5 in Bühlmann and Gisler (2005). □

### 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

\[ \sum_{i=1}^{n} \tilde{R}_i^{(i) \text{cred}} = 1' \tilde{R}_i^{\text{cred}} \quad \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{R}_i^{(i) \text{cred}} = \sum_{i=1}^{n} 1' \tilde{R}_i^{(i) \text{cred}}. \]

We derive these estimators under the assumption that all parameters in the Bühlmann-Straub credibility predictor \( \mu(\hat{\Theta}) \) in \( \tilde{R}_i^{\text{cred}} \) 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 \( i \in \{1, \ldots, I\} \), given \( D_i \), is defined by

\[ \text{mse} \sum_{i=1}^{n} \epsilon_i^2 \left( \tilde{R}_i^{(i) \text{cred}} \right) = E \left[ \left( \sum_{i=1}^{n} \tilde{R}_i^{(i) \text{cred}} - \sum_{i=1}^{n} R_i^{(i)} \right)^2 \right] D_i \]

\[ = 1' E [\tilde{R}_i^{(i) \text{cred}} - R_i^{(i)}] [\tilde{R}_i^{(i) \text{cred}} - R_i^{(i)}] D_i \tag{3.1} \]

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

\[ \text{mse} \sum_{i=1}^{n} \epsilon_i^2 \left( \tilde{R}_i^{(i) \text{cred}} \right) = 1' \text{Var}(R_i | D_i) 1 + 1' \left( \tilde{R}_i^{(i) \text{cred}} - E[R_i | D_i] \right) \left( \tilde{R}_i^{(i) \text{cred}} - E[R_i | D_i] \right)' 1. \tag{3.2} \]

Note that for this decoupling we use the fact that \( \tilde{R}_i^{(i) \text{cred}} \) is known/observable at time \( i \) (i.e., \( \tilde{R}_i^{(i) \text{cred}} \) is \( D_i \)-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 \( i \in \{1, \ldots, I\} \) is given by

\[ 1' \text{Var}(R_i | D_i) 1 \]

\[ = 1' D(\mu_0)^{-2} E[\Sigma(\Theta) | D_i] \sum_{j=1}^{I} D(\gamma_i)^{2-j} 1 \]

\[ + 1' D(\bar{\beta}_n - \bar{\beta}_0) D(\mu_0) \text{Var}(\mu(\Theta) | D_i) \]

\[ + D(\beta_i - \beta_0) D(\mu_0) 1. \tag{3.3} \]

**Proof:** See Appendix. □

The conditional process variance \( 1' \text{Var}(R_i | D_i) \) originates from stochastic movements of \( R_i \). If we approximate \( E[\Sigma(\Theta) | D_i] \) and \( \text{Var}(\mu(\Theta) | D_i) \) by \( S \) and \( T \), respectively, and if we replace the development pattern \( \gamma_i \) 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 $i \in \{1, \ldots, I\}$:

$$1' \tilde{\text{Var}}(R_i | D^R_i) 1 = 1' D(\mu) \sum_{j,l \in I} D(\hat{\gamma}_j) 1 + 1' D(\hat{\beta}_j - \hat{\beta}_{r.s.}) D(\mu) \times \hat{T} D(\hat{\beta}_j - \hat{\beta}_{r.s.}) D(\mu) 1$$

(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) $E[R_i | D^R_i]$ by $\tilde{R}_i^{\text{cred}}$. 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 $i \in \{1, \ldots, I\}$:

$$1' \tilde{\text{Var}}(R_i, cred | D^R_i) 1 = -1' D(\mu) D(\hat{\beta}_j - \hat{\beta}_{r.s.}) \hat{A} \hat{T} D(\hat{\beta}_j - \hat{\beta}_{r.s.}) D(\mu) 1$$

$$+ 1' D(\mu) D(\hat{\mu}(\Theta)) \tilde{\text{Var}}(\hat{\beta}_j - \hat{\beta}_{r.s.}) \times D(\hat{\mu}(\Theta)) D(\mu) 1,$$

(3.5)

where

$$\tilde{\text{Var}}(\hat{\beta}_j - \hat{\beta}_{r.s.})$$

$$= \sum_{j,l \in I} D(\sum_{a=0}^{l} \mu_a) \left[ \sum_{j,l \in I} D(\hat{\gamma}_j) 1 + D(\hat{\gamma}_j) \sum_{k=0}^{l} D(\mu_k) \times \hat{T} D(\hat{\gamma}_j) 1 \right] D(\sum_{a=0}^{l} \mu_a)$$

(3.6)
where

\[
\hat{\text{Var}} \left( \sum_{i=1}^{l} \hat{y}_i \right) = \sum_{i=1}^{l} \left( \sum_{k=0}^{l-i} \hat{\mu}_k \right)^{-1} \left( \sum_{k=0}^{l-i} \hat{\mu}_k \right)^{-1}
\]

\[
\left[ \delta \sum_{k=0}^{l-i} \hat{\mu}_k - \delta \sum_{k=0}^{l-i} \hat{\mu}_k \right] \left[ \sum_{k=0}^{l-i} \hat{\mu}_k S^T \sum_{k=0}^{l-i} \hat{\mu}_k \right] \]  

(3.12)

and

\[
\hat{\mu}(\Theta) = \hat{A} \hat{K} + (1 - \hat{A}) \hat{1}
\]

with

\[
\hat{K} = \frac{\sum_{i=0}^{l} \hat{\gamma}_k^i X_{i}}{\hat{b}} \quad \text{and} \quad \hat{A} = \frac{b}{\hat{b} + \hat{S}^T}.
\]

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 \( \hat{R}_i \) and \( \hat{R}_k \) because they use the same observations for estimating the parameters \( \gamma \) and \( \beta \), respectively. Therefore, they are not independent. We define the conditional MSEP of two aggregated accident years \( i \) and \( k \) by

\[
\text{mse} \sum_{s_i} \gamma \sum_{s_k} \mu = \sum_{s_i} \left( \sum_{s_k} \mu + \sum_{s_k} \mu \right) \left( \sum_{s_i} \mu + \sum_{s_k} \mu \right)
\]

\[
= \text{E} \left[ \left( \sum_{s_i} \left( \hat{R}_i^{\text{cond}} + \hat{R}_k^{\text{cond}} \right) - \sum_{s_i} \left( \hat{R}_i^{\text{cond}} + \hat{R}_k^{\text{cond}} \right) \right)^2 \right]  
\]

As for a single accident year we have the decomposition

\[
\text{mse} \sum_{s_i} \gamma \sum_{s_k} \mu = \text{I} \text{Var} (\hat{R} + \hat{R} | \mathcal{D}^\mu) \text{I} + \text{I} \text{Cov} (\hat{R} + \hat{R} | \mathcal{D}^\mu) \text{I}
\]

\[
= \left( \hat{R}_i^{\text{cond}} + \hat{R}_k^{\text{cond}} - \text{E} [\hat{R} + \hat{R} | \mathcal{D}^\mu] \right) \times
\]

\[
\left( \hat{R}_i^{\text{cond}} + \hat{R}_k^{\text{cond}} - \text{E} [\hat{R} + \hat{R} | \mathcal{D}^\mu] \right) \text{I}
\]

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

\[
\text{mse} \sum_{s_i} \gamma \sum_{s_k} \mu = \sum_{s_i} \left( \sum_{s_k} \mu + \sum_{s_k} \mu \right) \left( \sum_{s_i} \mu + \sum_{s_k} \mu \right)
\]

\[
= \text{mse} \sum_{s_i} \gamma \sum_{s_k} \mu + \text{mse} \sum_{s_i} \gamma \sum_{s_k} \mu \left( \sum_{s_i} \mu + \sum_{s_k} \mu \right)
\]

\[
= \text{mse} \sum_{s_i} \gamma \sum_{s_k} \mu \left( \sum_{s_i} \mu + \sum_{s_k} \mu \right)
\]

\[
+ 2 \sum_{s_i} \gamma \sum_{s_k} \mu \left( \sum_{s_i} \mu + \sum_{s_k} \mu \right) \text{I}
\]

(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:

\[
2 \cdot \text{I} \text{D} \left( \hat{\mathbf{\mu}}_1 \right) \text{D} \left( \hat{\mathbf{\Theta}}_1 \right) \text{D} \left( \hat{\mathbf{\mu}}_1 \right) \text{I}
\]

\[
\text{D} \left( \hat{\mathbf{\Theta}}_1 \right) \text{D} \left( \hat{\mathbf{\mu}}_1 \right) \text{I},
\]

(3.14)

where

\[
\text{Cov} (\hat{\beta}_i - \hat{\beta}_j, \hat{\beta}_i - \hat{\beta}_j)
\]

\[
= \text{Var} (\hat{\beta}_i - \hat{\beta}_j) + \sum_{l \neq \mu \in \mathcal{D}^\mu} \text{D} \left( \hat{\mathbf{\mu}}_1 \right) \text{D} \left( \hat{\mathbf{\mu}}_1 \right) \text{I}
\]

\[
\times \sum_{l \neq \mu \in \mathcal{D}^\mu} \text{D} \left( \hat{\mathbf{\mu}}_1 \right) \text{D} \left( \hat{\mathbf{\mu}}_1 \right) \text{I}
\]

\[
(3.15)
\]

and \( \hat{\text{Var}} (\hat{\beta}_i - \hat{\beta}_j) \) 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

\[
\text{mse} \sum_{s_i} \gamma \sum_{s_k} \mu = \text{mse} \sum_{s_i} \gamma \sum_{s_k} \mu \left( \sum_{s_i} \mu + \sum_{s_k} \mu \right)
\]

\[
+ 2 \sum_{s_i \neq s_k} \text{I} \left( \hat{R}_i^{\text{cond}} - \text{E} [\hat{R} | \mathcal{D}^\mu] \right) \left( \hat{R}_k^{\text{cond}} - \text{E} [\hat{R} | \mathcal{D}^\mu] \right) \text{I}
\]

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:

$$\hat{\text{msep}} \sum_{i=1}^{I} \hat{\gamma}_{i} = \sum_{i=1}^{I} \text{msep} \sum_{i=1}^{I} \hat{R}_{i}^{(\text{cond})}$$

$$= \sum_{i=1}^{I} \text{msep} \sum_{i=1}^{I} \hat{R}_{i}^{(\text{cond})}$$

$$+ 2 \sum_{i=1}^{I} \sum_{j=1}^{I} \hat{\gamma}_{i} \hat{\gamma}_{j} \hat{\mu}_{i} \hat{\mu}_{j} \hat{\gamma}_{i} \hat{\gamma}_{j}$$

$$= \hat{\text{cov}}(\hat{\beta}_{1}, \ldots, \hat{\beta}_{I}, -\hat{\beta}_{I}, -\hat{\beta}_{I}) \hat{\text{cov}}(\hat{\mu}_{1}, \ldots, \hat{\mu}_{I}) \hat{\text{cov}}(\hat{\gamma}_{1}, \ldots, \hat{\gamma}_{I}),$$

with \( \hat{\text{cov}}(\hat{\beta}_{1}, \ldots, \hat{\beta}_{I}, -\hat{\beta}_{I}, -\hat{\beta}_{I}) \) given by (3.15).

Remark

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

$$\hat{\text{msep}} \sum_{i=1}^{I} \hat{\gamma}_{i} = \sum_{i=1}^{I} \text{msep} \sum_{i=1}^{I} \hat{R}_{i}^{(\text{cond})}$$

where \( \hat{\text{msep}} \sum_{i=1}^{I} \hat{R}_{i}^{(\text{cond})} \) is given in (3.11) and

$$\hat{\text{cov}}(\hat{\gamma}_{1}, \ldots, \hat{\gamma}_{I}) = \hat{\text{var}}(\hat{\gamma}_{i}) + \sum_{i=1}^{I} \sum_{j=1}^{I} (\hat{\gamma}_{i}) (\hat{\gamma}_{j}) (\hat{\mu}_{i}) (\hat{\mu}_{j}) (\hat{\gamma}_{i}) (\hat{\gamma}_{j})$$

with \( \hat{\text{var}}(\hat{\gamma}_{i}) \) given in (26).

4. Parameter estimation

The estimators of the development pattern \( \gamma \) and \( \hat{\beta} \) 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 \( \gamma \) and \( \hat{\beta} \). Then, we replace \( \gamma \) and \( \hat{\beta} \) 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 \( \sigma_{n}^{2} = E_{n}^{2}(\Theta_{n}) \), \( n = 1, \ldots, N \) can be estimated by (see Bühlmann and Gisler 2005):

$$\hat{\sigma}_{n}^{2} = \frac{1}{I} \sum_{i=1}^{I} \frac{1}{I-1} \sum_{j=1}^{I} \hat{\gamma}_{i}^{(n)} \hat{\gamma}_{j}^{(n)} \left( \frac{X_{i}^{(n)}}{\hat{\gamma}_{j}^{(n)} \hat{\gamma}_{i}^{(n)}} - K_{j}^{(n)} \right)^{2},$$

(4.1)

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

$$\hat{S} = \begin{pmatrix} \hat{\sigma}_{1}^{2} & 0 & \cdots & 0 \\ 0 & \ddots & \vdots \\ \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \hat{\sigma}_{n}^{2} \end{pmatrix}.$$

For the estimation of the diagonal elements of \( T \) we define

$$\hat{T}_{i}^{(n)} = c \left( \sum_{i=1}^{I} \frac{\hat{\beta}_{i}^{(n)} \hat{\mu}_{i}^{(n)}}{\sum_{j=1}^{I} \hat{\beta}_{j}^{(n)} \hat{\mu}_{j}^{(n)}} \right) \left( \hat{K}_{i}^{(n)} - \hat{\bar{K}}_{i}^{(n)} \right) \left( \sum_{i=1}^{I} \frac{\hat{\beta}_{i}^{(n)} \hat{\mu}_{i}^{(n)}}{\sum_{j=1}^{I} \hat{\beta}_{j}^{(n)} \hat{\mu}_{j}^{(n)}} \right)^{-1},$$

with

$$c = \left( \sum_{i=1}^{I} \frac{\hat{\beta}_{i}^{(n)} \hat{\mu}_{i}^{(n)}}{\sum_{j=1}^{I} \hat{\beta}_{j}^{(n)} \hat{\mu}_{j}^{(n)}} \right)^{-1}$$

and

$$\hat{\bar{K}}_{i}^{(n)} = \sum_{i=1}^{I} \frac{\hat{\beta}_{i}^{(n)} \hat{\mu}_{i}^{(n)}}{\sum_{j=1}^{I} \hat{\beta}_{j}^{(n)} \hat{\mu}_{j}^{(n)}} \sum_{j=1}^{I} \hat{\gamma}_{j}^{(n)} X_{i-j}^{(n)}.$$
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)

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

(4.3)

where

$$
\hat{T}_{n,m} = c_n \cdot \left( \sum_{i=0}^N \sum_{j=0}^N (\beta_n^{(i)} \mu_j^{(i)})^2 \cdot (K_i^{(i)} - \bar{K}_n^{(i)}) \cdot (K_j^{(i)} - \bar{K}_m^{(i)}) \right)
$$

with

$$
c_n = \left( \sum_{i=0}^N \sum_{j=0}^N (\beta_n^{(i)} \mu_j^{(i)})^2 \cdot 1 \cdot \sum_{i=0}^N \sum_{j=0}^N (\beta_n^{(i)} \mu_j^{(i)})^2 \right)^{-\frac{1}{2}}.
$$

and

$$
\hat{T}_{m,n} = c_m \cdot \left( \sum_{i=0}^N \sum_{j=0}^N (\beta_m^{(i)} \mu_j^{(i)})^2 \cdot (K_i^{(i)} - \bar{K}_m^{(i)}) \cdot (K_j^{(i)} - \bar{K}_n^{(i)}) \right)
$$

with

$$
c_m = \left( \sum_{i=0}^N \sum_{j=0}^N (\beta_m^{(i)} \mu_j^{(i)})^2 \cdot 1 \cdot \sum_{i=0}^N \sum_{j=0}^N (\beta_m^{(i)} \mu_j^{(i)})^2 \right)^{-\frac{1}{2}}.
$$

The estimator $\hat{T}_{n,m}$ in (4.3) takes the value zero if $\hat{T}_{n,m}$ or $\hat{T}_{m,n}$ 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 $T$

$$
\hat{T}_{n,m} = \frac{\hat{T}_{n,m}^2 + \hat{T}_{m,n}^2}{2},
$$

(4.4)

By replacing $\beta_n^{(i)}$ with their estimators in (4.2) and (4.3) we get an estimator of $T$, which we denote by

$$
\hat{T} = (\hat{T}_{n,m})_{n,m=1}^N.
$$

The estimators $\hat{S}_n, n = 1, \ldots, N$ are unbiased estimators for the components of $S$ and the estimators $\hat{T}_{n,s}, \hat{T}_{n,t}$ are unbiased estimators for the components of $T$. However, the estimators $\hat{T}_{n,s}$ and $\hat{T}_{n,t}$ 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

$$
\hat{A}_i = \hat{T} \cdot D(\hat{\beta}_{i,j})^{-\frac{1}{2}} D(\mu_i)^{-\frac{1}{2}} \cdot \hat{S} \cdot D(\hat{\beta}_{i,j})^{-\frac{1}{2}} D(\mu_i)^{-\frac{1}{2}}.
$$

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 $\xi$ and $\delta$ by

$$
\hat{\xi}, \hat{\delta} = \arg\min_{\{\xi, \delta\}} \{ \text{SSE}(\hat{\xi}, \hat{\delta}) \},
$$

with

$$
\text{SSE}(\xi, \delta) = \sum_{i=0}^N \sum_{j=0}^N (X_{i,j} - \hat{X}_{i,j})^2.
$$

(4.5)

Another possible way to determine $\xi$ 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:

$$
\hat{\sigma}_n = \frac{1}{I - i} \sum_{i=J+1} \sum_{j=0}^J \gamma_i^{(n)} \mu_j^{(n)} \left( X_{i,j}^{(n)} - \bar{K}_i^{(n)} \right) + \frac{1}{I - i} \sum_{i=J+1} \sum_{j=0}^J \gamma_i^{(n)} \mu_j^{(n)} \left( X_{i,j}^{(n)} - \bar{K}_i^{(n)} \right).
$$

(5.1)
We calculate the credibility predictors (2.9) in order to determine the claims reserves (2.13) for every accident year \( i \in \{0, \ldots, I\} \). For illustrative purposes we restrict the analysis to four explicit parameter choices \( x \in \{0, 2\} \) and \( d \in \{0, 2\} \).

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

Goodness-of-Fit

In order to choose specific \( x \) and \( d \), 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 better fit.

<table>
<thead>
<tr>
<th>Table 1. Prior means ( \mu^{(i)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( i )</td>
</tr>
<tr>
<td>( \mu^{(1)} )</td>
</tr>
<tr>
<td>( \mu^{(2)} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Estimated loss development pattern ( \hat{\gamma}^{(i)} ) in Portfolio A and B</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \xi = 0 )</td>
</tr>
<tr>
<td>( \hat{\gamma}^{(1)} )</td>
</tr>
<tr>
<td>( \hat{\gamma}^{(2)} )</td>
</tr>
</tbody>
</table>

Table 3. Credibility predictor \( \mu(\xi) \) \( \xi \in \{0, 2\} \) and \( \delta \in \{0, 2\} \) for \( \xi \in \{0, 2\} \) and \( \delta \in \{0, 2\} \)

| \( \xi = 0, \delta = 0 \) | \( 0.73 \) | \( 0.87 \) | \( 1.74 \) | \( 1.16 \) | \( 0.84 \) | \( 0.75 \) | \( 0.91 \) | \( 0.64 \) | \( 1.31 \) | \( 0.94 \) | \( 0.78 \) | \( 0.84 \) | \( 1.19 \) | \( 1.15 \) | \( 0.97 \) | \( 1.00 \) | \( 1.00 \) |
| \( \xi = 0, \delta = 2 \) | \( 0.91 \) | \( 0.93 \) | \( 1.72 \) | \( 1.10 \) | \( 0.87 \) | \( 0.77 \) | \( 0.93 \) | \( 0.69 \) | \( 1.21 \) | \( 0.94 \) | \( 0.82 \) | \( 0.87 \) | \( 1.17 \) | \( 1.23 \) | \( 0.98 \) | \( 1.00 \) | \( 1.00 \) |
| \( \xi = 2, \delta = 0 \) | \( 1.01 \) | \( 1.02 \) | \( 0.96 \) | \( 1.04 \) | \( 1.04 \) | \( 1.01 \) | \( 0.99 \) | \( 1.03 \) | \( 1.02 \) | \( 1.03 \) | \( 1.01 \) | \( 0.96 \) | \( 0.89 \) | \( 1.00 \) | \( 1.00 \) | \( 1.00 \) |
| \( \xi = 2, \delta = 2 \) | \( 1.02 \) | \( 1.02 \) | \( 0.96 \) | \( 1.02 \) | \( 1.03 \) | \( 1.01 \) | \( 0.99 \) | \( 1.03 \) | \( 1.02 \) | \( 1.03 \) | \( 1.01 \) | \( 0.97 \) | \( 0.87 \) | \( 1.00 \) | \( 1.00 \) | \( 1.00 \) |

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 \( (\gamma^{(i)}_{\mu_{\xi}})_{\mu_{\xi} \in \mathbb{R}} \) (note that this estimator is independent from the specific values of \( \xi \) and \( \delta \) 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, \( \hat{\gamma}^{(i)} > 0 \) holds for all development years \( j = 0, \ldots, 10 \) 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.
Table 4. Estimated loss development pattern $\gamma^{\hat{\mu}}_j$ in Portfolio A and B

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>$\xi = 0$, $\delta = 0$</th>
<th>$\xi = 0$, $\delta = 2$</th>
<th>$\xi = 2$, $\delta = 0$</th>
<th>$\xi = 2$, $\delta = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sqrt{\text{SSE} (\xi, \delta)}$</td>
<td>57955</td>
<td>60104</td>
<td>19996</td>
<td>19862</td>
</tr>
</tbody>
</table>

For $\xi = 0$ what can be explained as follows: In Model Assumptions 2.1 the incremental claim payments are of the type $Y_{ij} | Q_i = \mu(Q_i) + \epsilon_{ij}$ with $E[\epsilon_{ij} | \Theta] = 0$ and $\text{Var}(\epsilon_{ij} | \Theta) = D(w_{ij,x,d})^{-1/2} \cdot \Sigma(\Theta) \cdot D(w_{ij,x,d})^{-1/2}$. In the case of $\xi = 0$ the conditional variance $\text{Var}(\epsilon_{ij} | \Theta)$ does not depend on development year $j$ and we assume homogeneous variances of $(Y_{ij} | Q_i)_{0 \leq j \leq J}$, although this assumption seems quite unrealistic. Then the weights of the standardized incremental claim payments $Y_{ij}$ in the compressed data vector $K_i$ 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 $\sqrt{\text{SSE}(0, \delta)}$. The $\sqrt{\text{SSE}(0, \delta)}$ decreases by about 40% if these “untypical” accident years (2, 7, and 13) are left out in the calculation of $\sqrt{\text{SSE}(0, \delta)}$. 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 $\sqrt{\text{SSE}(2, 0)}$ 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

$$r_{ij} = D(\mu_i) \cdot (\Sigma(\Theta_i)^{1/2} \cdot (X_{ij} - D(\gamma_{ij}) \cdot D(\mu_i) \cdot (\Theta_i))).$$

Figure 2. Empirical residuals vs. predicted incremental payments ($\xi = 2$ and $\delta = 0$).
Variance Advancing the Science of Risk

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

These standardized residuals have, conditionally given $\Theta_n$, 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 $\hat{r}_{ij}$. 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, respectively (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 $\mu(\Theta)$) 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

<table>
<thead>
<tr>
<th></th>
<th>Bühlmann-Straub model</th>
<th>CL model</th>
<th>ALR model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated reserves</td>
<td>54350</td>
<td>52,734</td>
<td>54,042</td>
</tr>
<tr>
<td>Process std. deviation</td>
<td>14063</td>
<td>9,697</td>
<td>7,539</td>
</tr>
<tr>
<td>$\sqrt{\text{Estimation error}}$</td>
<td>12454</td>
<td>5,549</td>
<td>4,749</td>
</tr>
<tr>
<td>Prediction std. error</td>
<td>18.785</td>
<td>11,172</td>
<td>8,910</td>
</tr>
</tbody>
</table>
### Table 6. Observed incremental claim payments $C_{i,j}^{(1)}$ in Portfolio A

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

### Table 7. Observed incremental claim payments $C_{i,j}^{(2)}$ in Portfolio B

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>16.651</td>
<td>8.206</td>
<td>−468</td>
<td>−152</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>16.292</td>
<td>8.129</td>
<td>1.713</td>
<td>195</td>
<td>426</td>
<td>35</td>
<td>68</td>
<td>50</td>
<td>55</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>16.658</td>
<td>10.566</td>
<td>1.736</td>
<td>618</td>
<td>1.272</td>
<td>442</td>
<td>572</td>
<td>15</td>
<td>38</td>
<td>749</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>19.715</td>
<td>10.690</td>
<td>968</td>
<td>2.098</td>
<td>154</td>
<td>197</td>
<td>100</td>
<td>63</td>
<td>3.519</td>
<td>24</td>
<td>149</td>
</tr>
<tr>
<td>4</td>
<td>21.220</td>
<td>8.815</td>
<td>2.969</td>
<td>896</td>
<td>151</td>
<td>1.146</td>
<td>26</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>21.302</td>
<td>10.882</td>
<td>2.237</td>
<td>952</td>
<td>7</td>
<td>1.183</td>
<td>14</td>
<td>0</td>
<td>26</td>
<td>70</td>
<td>−9</td>
</tr>
<tr>
<td>6</td>
<td>17.201</td>
<td>7.493</td>
<td>1.574</td>
<td>518</td>
<td>1.685</td>
<td>25</td>
<td>2</td>
<td>56</td>
<td>−8</td>
<td>71</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>15.835</td>
<td>11.668</td>
<td>1.728</td>
<td>122</td>
<td>493</td>
<td>−10</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>17.560</td>
<td>8.550</td>
<td>2.373</td>
<td>908</td>
<td>1.163</td>
<td>389</td>
<td>127</td>
<td>501</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>21.051</td>
<td>12.279</td>
<td>2.387</td>
<td>690</td>
<td>372</td>
<td>2.024</td>
<td>14</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>20.368</td>
<td>9.832</td>
<td>1.285</td>
<td>460</td>
<td>43</td>
<td>186</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>18.623</td>
<td>11.160</td>
<td>942</td>
<td>892</td>
<td>8</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>18.112</td>
<td>12.040</td>
<td>1.662</td>
<td>5.654</td>
<td>979</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>17.744</td>
<td>10.346</td>
<td>2.134</td>
<td>599</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>17.993</td>
<td>8.956</td>
<td>869</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>19.082</td>
<td>11.403</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>17.809</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 \( \Theta \),

\[
1' \text{Var}(R | D^y) 1 = 1' E[\text{Var}(R | D^y, \Theta) | D^y] 1 \\
+ 1' \text{Var}(E[R | D^y, \Theta] | D^y) 1 \\
= 1' E[\text{Var}(R, \Theta) | D^y] 1 \\
+ 1' \text{Var}(E[R, \Theta] | D^y) 1 \\
\]

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

\[
1' \text{Var}(E[R, \Theta] | D^y) 1 \\
= 1' \text{Var}(D(\beta, \beta_{\star}) D(\mu, \mu(\Theta)) | D^y) 1 \\
= 1' D(\beta, \beta_{\star}) D(\mu, \mu(\Theta)) D(\mu) D(\beta, \beta_{\star}) 1. \\
\]

Using the conditional independence of the normalized incremental claims \( Y_{i, \ell+1}, \ldots, Y_{\ell, j} \) and (2.5) we obtain for the first term on the right-hand side of (A.1)

\[
1' E[\text{Var}(R, \Theta) | D^y] 1 \\
= 1' E \left[ \sum_{j=1}^{\ell} D(w_{i,j}, \Theta) \text{Var}(Y_{i,j} | \Theta) D(w_{i,j}) | D^y \right] 1 \\
= 1' E \left[ \sum_{j=1}^{\ell} D(w_{i,j}) D(w_{i,j}, \Theta) \Sigma(\Theta) \times \right. \\
\left. D(w_{i,j}) | D^y \right] 1 \\
= 1' D(\mu) \sum_{j=1}^{\ell} E \left[ \sum_{j=1}^{\ell} D(\gamma_{\ell,j}) \Sigma(\Theta) | D^y \right] 1. \\
\]

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 \( \Theta \), and (2.13) we obtain

\[
E[R, \Theta] | D^y = \sum_{j=1}^{\ell} E[\theta_{i,j} | \Theta, D^y] \\
= \sum_{j=1}^{\ell} E[D(w_{i,j}) E[Y_{i,j} | \Theta, D^y] \\
= D(\sum_{j=1}^{\ell} \gamma_{j}) D(\mu) E[\mu(\Theta) | D^y] \\
\]

and

\[
\hat{R} = D(\sum_{j=1}^{\ell} \gamma_{j}) D(\mu) \mu(\Theta). \\
\]

This leads to

\[
\hat{R} - E[R, \Theta] | D^y = D(\mu) \mu(\Theta) - E[\mu(\Theta) | D^y] \\
- D(\mu) \mu(\Theta) - E[\mu(\Theta) | D^y]. \\
\]

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

\[
1' (\hat{R} - E[R, \Theta] | D^y) (\hat{R} - E[R, \Theta] | D^y)^\prime 1 \\
= 1' D(\mu) \left[ \sum_{j=1}^{\ell} \gamma_{j} \right] \left[ \mu(\Theta) - E[\mu(\Theta) | D^y] \right] \times \left[ \mu(\Theta) - E[\mu(\Theta) | D^y] \right]^\prime D(\mu) 1 \\
- 1' D(\mu) \left[ \sum_{j=1}^{\ell} \gamma_{j} \right] \mu(\Theta) \times \left[ \mu(\Theta) - E[\mu(\Theta) | D^y] \right]^\prime D(\mu) 1 \\
- 1' D(\mu) \left[ \sum_{j=1}^{\ell} \gamma_{j} \right] \left[ \mu(\Theta) - E[\mu(\Theta) | D^y] \right] \times \left[ \mu(\Theta) - E[\mu(\Theta) | D^y] \right]^\prime D(\mu) 1 \\
- 1' D(\mu) \left[ \sum_{j=1}^{\ell} \gamma_{j} \right] \left[ \mu(\Theta) - E[\mu(\Theta) | D^y] \right]. \\
\]
\[
\hat{\mu}(\Theta) \cdot D \left( \sum_{j=1}^{\nu} (\gamma_j - \hat{\gamma}_j) \right) D(\mu) \cdot 1 \\
- 1 \cdot D(\mu) \cdot D \left( \sum_{j=1}^{\nu} (\gamma_j - \hat{\gamma}_j) \right) \mu(\Theta) \cdot \mu(\Theta)^{\prime} \\
D \left( \sum_{j=1}^{\nu} (\gamma_j - \hat{\gamma}_j) \right) D(\mu) \cdot 1. \tag{A.2}
\]

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 \( E[\text{Var}(\mu(\Theta) | D_\nu)] = T \) we obtain the approximation

\[
E \left[ \left( \mu(\Theta) - E[\mu(\Theta) | D_\nu] \right) \left( \mu(\Theta) - E[\mu(\Theta) | D_\nu] \right)^{\prime} \right] \\
= E \left[ \left( \mu(\Theta) - \mu(\Theta) \right) \left( \mu(\Theta) - \mu(\Theta) \right)^{\prime} \right] | D_\nu \\
- \text{Var}(\mu(\Theta) | D_\nu) \\
= (I - A) \cdot T - T \\
= -A \cdot T. \tag{A.3}
\]

This leads to the following approximation for the first term:

\[
-1 \cdot D(\mu) \cdot D \left( \sum_{j=1}^{\nu} \gamma_j \right) A \cdot T \cdot D \left( \sum_{j=1}^{\nu} \gamma_j \right) D(\mu) \cdot 1. \tag{A.4}
\]

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

\[
D(b) \cdot a \cdot c^{\prime} \cdot D(d) = D(a) \cdot b \cdot d^{\prime} \cdot D(c)
\]

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

\[
1' \cdot D(\mu) \cdot D(\mu(\Theta)) \left( \sum_{j=1}^{\nu} (\gamma_j - \hat{\gamma}_j) \sum_{j=1}^{\nu} (\gamma_j - \hat{\gamma}_j)^{\prime} \right) \\
D(\mu(\Theta)) \cdot D(\mu) \cdot 1. \tag{A.5}
\]

In the following we approximate (A.5) by

\[
1' \cdot D(\mu) \cdot D(\mu(\Theta)) \left( \sum_{j=1}^{\nu} (\gamma_j - \hat{\gamma}_j) \sum_{j=1}^{\nu} (\gamma_j - \hat{\gamma}_j)^{\prime} \right) \\
D(\mu(\Theta)) \cdot D(\mu) \cdot 1 \\
= 1' \cdot D(\mu) \cdot D(\mu(\Theta)) \text{Var} \left( \sum_{j=1}^{\nu} \gamma_j \right) \times \\
D(\mu(\Theta)) \cdot D(\mu) \cdot 1. \tag{A.6}
\]

For this end we have to calculate

\[
\text{Var} \left( \sum_{j=1}^{\nu} \gamma_j \right) = E \left[ \text{Var} \left( \sum_{j=1}^{\nu} \gamma_j | \Theta \right) \right] \\
+ \text{Var} \left( E \left[ \sum_{j=1}^{\nu} \gamma_j | \Theta \right] \right), \tag{A.7}
\]

where \( \Theta = (\Theta_0, \ldots, \Theta_\nu) \). Using the independence of the accident years and the conditional independence of the normalized incremental claims \( Y_{\nu,0}, \ldots, Y_{\nu,\nu} \), given \( \Theta_\nu \), we obtain for the first term on the right-hand side of (A.7)

\[
E \left[ \text{Var} \left( \sum_{j=1}^{\nu} \gamma_j | \Theta \right) \right] = \sum_{j=1}^{\nu} E \left[ \text{Var}(\gamma_j | \Theta) \right] \\
= \sum_{j=1}^{\nu} E \left[ D \left( \sum_{k=0}^{j} \mu_k \right)^{\prime} \sum_{k=0}^{j} \text{Var}(D(\mu_k) Y_{\nu,k} | \Theta) \times \\
D \left( \sum_{k=0}^{j} \mu_k \right) \right] \\
= \sum_{j=1}^{\nu} E \left[ D \left( \sum_{k=0}^{j} \mu_k \right)^{\prime} D(\gamma_j)^{\prime} \sum_{k=0}^{j} D(\mu_k) \times \\
D(\gamma_j) \cdot D \left( \sum_{k=0}^{j} \mu_k \right) \right] \\
= \sum_{j=1}^{\nu} D \left( \sum_{k=0}^{j} \mu_k \right) \times \\
D(\gamma_j)^{\prime} \sum_{k=0}^{j} D(\mu_k) \times \\
D \left( \sum_{k=0}^{j} \mu_k \right)^{\prime} \times \\
D \left( \sum_{k=0}^{j} \mu_k \right)^{\prime} \tag{A.8}
\]
For the second term on the right-hand side of (A.7) we obtain

\[ \text{Var} \left( E \left[ \sum_{j=1}^{t} \gamma_j \mid \Theta \right] \right) = \text{Var} \left( \sum_{j=1}^{t} E \left[ D \left( \sum_{k=0}^{l} \mu_k \right) \gamma_j \right] \mid \Theta \right) \]

\[ = \text{Var} \left( \sum_{j=1}^{t} D \left( \sum_{k=0}^{l} \mu_k \right) \gamma_j \sum_{j=1}^{t} D \left( \mu_k \right) \mu(\Theta) \right) \]

We define

\[ Z_i = D \left( \sum_{k=0}^{l} \mu_k \right) \gamma_j \sum_{j=1}^{t} D \left( \mu_k \right) \mu(\Theta) \]

(A.9)

Using the independence of \( \Theta_1, \ldots, \Theta_t \) we obtain

\[ = \sum_{j=1}^{t} \text{Var} (Z_i) + 2 \sum_{j=1}^{t} \text{Cov} (Z_i, Z_i) \]

\[ = \sum_{j=1}^{t} D \left( \sum_{k=0}^{l} \mu_k \right)^2 \gamma_j \sum_{j=1}^{t} D \left( \mu_k \right) \gamma_j \sum_{j=1}^{t} D \left( \mu_k \right) \mu(\Theta) \]

\[ + 2 \sum_{j=1}^{t} D \left( \sum_{k=0}^{l} \mu_k \right)^2 \gamma_j \sum_{j=1}^{t} D \left( \mu_k \right) \gamma_j \sum_{j=1}^{t} D \left( \mu_k \right) \mu(\Theta) \]

\[ \sum_{j=1}^{t} D \left( \sum_{k=0}^{l} \mu_k \right) \gamma_j \sum_{j=1}^{t} D \left( \mu_k \right) \mu(\Theta) \]

(A.10)

This leads to the term

\[ \text{Var} \left( \sum_{j=1}^{t} \gamma_j \right) = \sum_{j=1}^{t} D \left( \sum_{k=0}^{l} \mu_k \right)^2 \gamma_j \sum_{j=1}^{t} D \left( \mu_k \right) \gamma_j \sum_{j=1}^{t} D \left( \mu_k \right) \mu(\Theta) \]

\[ \gamma_j \sum_{j=1}^{t} D \left( \mu_k \right) \mu(\Theta) \]

where \( \delta_{ij} \) 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 MSEP for aggregated accident years

To obtain an estimator of the conditional MSEP 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 \( 1 \leq i < k \leq I \)

\[ \left( \hat{\mathbf{R}}_i - \mathbf{R}_i \right) \left( \hat{\mathbf{R}}_k - \mathbf{R}_k \right) \]

\[ = \left( \mathbf{D}(\mu_i) \right) \left( \mathbf{D}(\mu_k) \right) \sum_{j=1}^{t} D(\gamma_j) \sum_{j=1}^{t} \gamma_j \sum_{j=1}^{t} D(\mu_k) \mu(\Theta) \]

\[ - \left( \mathbf{D}(\mu_i) \right) \left( \mathbf{D}(\mu_k) \right) \sum_{j=1}^{t} D(\gamma_j) \sum_{j=1}^{t} \gamma_j \sum_{j=1}^{t} D(\mu_k) \mu(\Theta) \]

\[ - \left( \mathbf{D}(\mu_i) \right) \left( \mathbf{D}(\mu_k) \right) \sum_{j=1}^{t} D(\gamma_j) \sum_{j=1}^{t} \gamma_j \sum_{j=1}^{t} D(\mu_k) \mu(\Theta) \]

\[ \sum_{j=1}^{t} D(\gamma_j) \sum_{j=1}^{t} \gamma_j \sum_{j=1}^{t} D(\mu_k) \mu(\Theta) \]

(A.12)

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

\[ E \left[ \left( \mathbf{D}(\mu_i) \right) \left( \mathbf{D}(\mu_k) \right) \sum_{j=1}^{t} D(\gamma_j) \sum_{j=1}^{t} \gamma_j \sum_{j=1}^{t} D(\mu_k) \mu(\Theta) \right] \]

\[ = E \left[ \left( \mathbf{D}(\mu_i) \right) \left( \mathbf{D}(\mu_k) \right) \sum_{j=1}^{t} D(\gamma_j) \sum_{j=1}^{t} \gamma_j \sum_{j=1}^{t} D(\mu_k) \mu(\Theta) \right] \]

\[ \sum_{j=1}^{t} D(\gamma_j) \sum_{j=1}^{t} \gamma_j \sum_{j=1}^{t} D(\mu_k) \mu(\Theta) \]

\[ \sum_{j=1}^{t} D(\gamma_j) \sum_{j=1}^{t} \gamma_j \sum_{j=1}^{t} D(\mu_k) \mu(\Theta) \]

\[ \sum_{j=1}^{t} D(\gamma_j) \sum_{j=1}^{t} \gamma_j \sum_{j=1}^{t} D(\mu_k) \mu(\Theta) \]

\[ \sum_{j=1}^{t} D(\gamma_j) \sum_{j=1}^{t} \gamma_j \sum_{j=1}^{t} D(\mu_k) \mu(\Theta) \]

(A.13)
As mentioned in the remarks to Theorem 2.2, the credibility predictor $\hat{\mu}(\hat{\Theta})$ can be constructed using orthogonal projections in Hilbert spaces. Since $\hat{\mu}(\hat{\Theta})$ is the orthogonal projection of the so-called Bayes-Predictor $E[\hat{\mu}(\Theta)|D^n]$ on the Hilbert space $L(D^n, 1)$, it fulfills the orthogonality condition

$$E[(\hat{\mu}(\Theta) - E[\hat{\mu}(\Theta)|D^n])\cdot X_{\gamma}]$$

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

$$1' D(\mu_{ij})D(\hat{\mu}(\Theta))E\left[\sum_{j=l}^{m} (\gamma_{j} - \hat{\gamma}_{j}) \sum_{j=1}^{m} (\gamma_{j} - \hat{\gamma}_{j})'\right] \times$$

$$D(\hat{\mu}(\Theta))D(\mu_{ij})1$$

$$= 1' D(\mu_{ij})D(\hat{\mu}(\Theta))Cov\left[\sum_{j=l}^{m} \hat{\gamma}_{j}, \sum_{j=1}^{m} \hat{\gamma}_{j}\right] \times$$

$$D(\hat{\mu}(\Theta))D(\mu_{ij})1.$$  \hspace{1cm} (A.14)

So, we have to calculate

$$\text{Cov}\left[\sum_{j=l}^{m} \hat{\gamma}_{j}, \sum_{j=1}^{m} \hat{\gamma}_{j}\right]$$

$$= E\left[\text{Cov}\left(\sum_{j=l}^{m} \hat{\gamma}_{j}, \sum_{j=1}^{m} \hat{\gamma}_{j}|\Theta\right)\right] + \text{Cov}\left(E\left[\sum_{j=l}^{m} \hat{\gamma}_{j}|\Theta\right]\right),$$

$$E\left[\sum_{j=l}^{m} \hat{\gamma}_{j}|\Theta\right]$$ \hspace{1cm} (A.15)

where $\Theta = (\Theta_0, \ldots, \Theta_t)$. Using the independence of different accident years and the conditional independence of the normalized incremental claims $Y_{k,0}, \ldots, Y_{k,t}$ given $\Theta_{t}$, we obtain for the first term on the right-hand side of (A.15)

$$E\left[\text{Cov}\left(\sum_{j=l}^{m} \hat{\gamma}_{j}, \sum_{j=1}^{m} \hat{\gamma}_{j}|\Theta\right)\right] = E\left[\text{Var}(\hat{\gamma}_{j}|\Theta)\right] = E\left[\text{Var}(\sum_{j=l}^{m} \hat{\gamma}_{j}|\Theta)\right]$$

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 $Z_{j}$ defined in (A.9) and obtain

$$\text{Cov}\left(E\left[\sum_{j=l}^{m} \hat{\gamma}_{j}|\Theta\right], E\left[\sum_{j=1}^{m} \hat{\gamma}_{j}|\Theta\right]\right) = \text{Cov}\left(\sum_{j=l}^{m} Z_{j}, \sum_{j=1}^{m} Z_{j}\right)$$

$$= \sum_{j=l}^{m} \text{Var}(Z_{j}) + 2 \sum_{l<j} \text{Cov}(Z_{l}, Z_{j})$$

$$+ \sum_{l<j} \text{Var}(Z_{l}, Z_{j}) = \text{Var}\left(E\left[\sum_{j=l}^{m} \hat{\gamma}_{j}|\Theta\right]\right)$$

$$+ \sum_{l<j} \left(D(\sum_{n=0}^{l-1} \mu_{n})^{-1} D(\gamma_{l}) D(\sum_{n=0}^{l-1} \mu_{n})\right) T T D(\mu_{l}) \times$$

$$D(\gamma_{l}) D(\sum_{n=0}^{l-1} \mu_{n})^{-1},$$

where $\text{Var}(E[\sum_{j=l}^{m} \hat{\gamma}_{j}|\Theta])$ is given in (A.10). Hence, we obtain

$$\text{Cov}\left(\sum_{j=l}^{m} \hat{\gamma}_{j}, \sum_{j=1}^{m} \hat{\gamma}_{j}\right) = \text{Var}\left(\sum_{j=l}^{m} \hat{\gamma}_{j}\right)$$

$$+ \sum_{l<j} \left(D(\sum_{n=0}^{l-1} \mu_{n})^{-1} D(\gamma_{l}) \times$$

$$\sum_{n=0}^{l-1} D(\mu_{n}) T D(\mu_{l}) D(\gamma_{l}) D(\sum_{n=0}^{l-1} \mu_{n})^{-1}\right),$$ \hspace{1cm} (A.16)

where $\text{Var}(\sum_{j=l}^{m} \hat{\gamma}_{j})$ 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**


