Modified Munich Chain-Ladder Method

Michael Merz∗ and Mario V. Wüthrich†‡

August 29, 2014

Abstract

The Munich chain-ladder method was introduced on an axiomatic basis. We analyze these axioms and we define a modified Munich chain-ladder method which is based on an explicit stochastic model. This stochastic model then allows to consider claims prediction and prediction uncertainty for the Munich chain-ladder method in a consistent way.

Keywords. Munich chain-ladder method, claims reserving, prediction uncertainty, mean-square error of prediction, multivariate Gaussian model, claims paid and claims incurred.

1 Introduction

The Munich chain-ladder method was introduced by Quarg and Mack [6] on a pure axiomatic basis, and in 2003 it was awarded the Gauss prize by DAV and DGVFM, see [6]. But still today it is not known whether there is a non-trivial interesting stochastic model that fulfills these axioms, nor is anything known about the prediction uncertainty in the Munich chain-ladder method. The aim of this paper is to study the axioms of the Munich chain-ladder method and to define a modified Munich chain-ladder method which is based on an explicit stochastic model. For this modified version we analyze claims prediction and its uncertainty.

There are two different ways to view the Munich chain-ladder method. The first way is to define a stochastic model which has the required structure of the Munich chain-ladder factors; this is the approach taken in [6]. The second way is to define a general chain-ladder model and derive estimators that have the Munich chain-ladder factor structure; this is the approach taken in [4]. Here we analyze both of these views and we show how the second way leads to a modified Munich chain-ladder method. The first main result is that within the family of multivariate normal models, there is, in general, no interesting model which fulfills the Munich chain-ladder model assumptions, see Theorem 4.3 below. Therefore, the Munich chain-ladder predictor always has an approximation error which is quantified in Theorem 4.4 below. Based on these findings, we define a modified Munich chain-ladder model for which we can derive optimal predictors and the corresponding prediction uncertainty.

∗Faculty of Business Administration, University of Hamburg, 20146 Hamburg, Germany
†ETH Zurich, RiskLab, Department of Mathematics, 8092 Zurich, Switzerland
‡Swiss Finance Institute SFI Professor
Organization of the paper. In the next section we consider stochastic models which simultaneously fulfill the chain-ladder assumptions for cumulative payments and claims incurred. In Theorem 2.2 we will see that such models only allow for rather restricted correlation structures. For these restricted chain-ladder models we then study the optimal one-step ahead prediction in Section 3. This optimal one-step ahead prediction can then directly be compared to the Munich chain-ladder axioms which are introduced in Section 4. In Theorem 4.3 we find that, in general, the Munich chain-ladder axioms are not fulfilled. This leads to a modified Munich chain-ladder method which is presented in Section 5. For this modified version we derive optimal predictors and study prediction uncertainty in Section 6. These results are then compared numerically to other methods in Section 7. This numerical study is based on the original data set of Quarg and Mack [6].

2 Chain-ladder models

We denote cumulative payments of accident year \( i \) and development year \( j \) by \( P_{i,j} \) and the corresponding claims incurred are denoted by \( I_{i,j} \) for \( i = 0, \ldots, J \) and \( j = 0, \ldots, J \). We define the following sets of information

\[
B^P_j = \{ P_{i,k}; k \leq j, 0 \leq i \leq J \}, \quad B^I_j = \{ I_{i,k}; k \leq j, 0 \leq i \leq J \} \quad \text{and} \quad B_j = B^P_j \cup B^I_j.
\]

Assumption 1 (distribution-free chain-ladder model).

(A1) We assume that the random vectors \((P_{i,0}, \ldots, P_{i,J}, I_{i,0}, \ldots, I_{i,J})\) are independent for different accident years \( i = 0, \ldots, J \).

(A2) There exist parameters \( f^P_j, f^I_j, (\sigma^P_j)^2 > 0 \) and \((\sigma^I_j)^2 > 0\) such that for \( 0 \leq j \leq J - 1 \) and \( 0 \leq i \leq J \) we have

\[
\mathbb{E} \left[ P_{i,j+1} \mid B^P_j \right] = f^P_j P_{i,j} \quad \text{and} \quad \text{Var} \left( P_{i,j+1} \mid B^P_j \right) = (\sigma^P_j)^2 P_{i,j}^2, \\
\mathbb{E} \left[ I_{i,j+1} \mid B^I_j \right] = f^I_j I_{i,j} \quad \text{and} \quad \text{Var} \left( I_{i,j+1} \mid B^I_j \right) = (\sigma^I_j)^2 I_{i,j}^2.
\]

These assumptions correspond to \( \text{PE, PV, IE, IV and PIU} \) in [6], except that we make a modification in the variance assumptions \( \text{PV and IV} \). We make this change because it substantially simplifies our considerations (but it does not harm our argumentation). Assumption 1 states that cumulative payments \((P_{i,j})_{i,j}\) and claims incurred \((I_{i,j})_{i,j}\) fulfill distribution-free chain-ladder model assumptions simultaneously. Our first aim is to show that there is a non-trivial stochastic model that fulfills the chain-ladder model assumptions simultaneously for cumulative payments and claims incurred. To this end we define an explicit distributional model. The distributions are chosen such that the analysis becomes as simple as possible. We will see that assumption (A2) requires a sophisticated consideration.

Choose a continuous and strictly increasing link function \( g \) with range \( \mathbb{R} \). The standard example that we use in the sequel is the log-link given by

\[
g(x) = \log x \quad \text{for} \ x > 0. \quad (2.1)
\]
For given link function \( g \) we define the transformed age-to-age ratios for \( 0 \leq j \leq J \) and \( 0 \leq i \leq J \) by
\[
\xi_{i,j}^P = g\left( \frac{P_{i,j}}{P_{i,j-1}} \right) \quad \text{and} \quad \xi_{i,j}^I = g\left( \frac{I_{i,j}}{I_{i,j-1}} \right),
\]
where we set fixed initial values \( P_{i,-1} = I_{i,-1} = \nu_i \) according to given volume measures \( \nu_i > 0 \).

To simplify the outline we introduce vector notation, for \( 0 \leq i \leq J \) we set
\[
\Xi_i = (\xi_{i,0}^P, \ldots, \xi_{i,J}^P, \xi_{i,0}^I, \ldots, \xi_{i,J}^I)',
\]

**Assumption 2** (multivariate (log-)normal chain-ladder model I).

**(B1)** We assume that the random vectors \( \Xi_i \) are independent for different accident years \( i = 0, \ldots, J \).

**(B2)** There exists a parameter vector \( \theta = (\theta_0^P, \ldots, \theta_J^P, \theta_0^I, \ldots, \theta_J^I)' \in \mathbb{R}^{(J+1)} \) and a positive definite covariance matrix \( \Sigma \in \mathbb{R}^{2(J+1) \times 2(J+1)} \) such that we have for \( 0 \leq i \leq J \)
\[
\Xi_i \sim N(\theta, \Sigma).
\]

Note that we have
\[
\sigma \{ P_{i,k}; \; k \leq j, \; 0 \leq i \leq J \} = \sigma \{ \xi_{i,k}^P; \; k \leq j, \; 0 \leq i \leq J \}.
\]

Therefore, by an abuse of notation, we use \( B_0^P \) for both sets of information, and analogously for \( B_j^I \) and \( B_j \). From this we immediately see that assumptions (A1) and (B1) coincide. Due to this independence of different accident years \( i \) we have for \( * \in \{ P, I \} \)
\[
\xi_{i,j+1}^*|B_j^* \overset{(d)}{=} \xi_{i,j+1}^*|\{\xi_{i,0}^*\ldots\xi_{i,j}^*\},
\]

\[
(\xi_{i,j+1}^P, \xi_{i,j+1}^I)|B_j \overset{(d)}{=} (\xi_{i,j+1}^P, \xi_{i,j+1}^I)|\{\xi_{i,0}^P\ldots\xi_{i,j}^P, \xi_{i,0}^I\ldots\xi_{i,j}^I\}.
\]

For \( * \in \{ P, I \} \) we denote by \( \theta_{[j]}^* = (\theta_0^*, \ldots, \theta_j^*)' \in \mathbb{R}^{j+1} \) and let \( \Sigma_{[j]}^* \in \mathbb{R}^{(j+1) \times (j+1)} \) be the (positive definite) covariance matrix of the random vector \( \xi_{i,[j]}^* = (\xi_{i,0}^*, \ldots, \xi_{i,j}^*)' \). Moreover, let \( \Sigma_{j,j+1}^* \in \mathbb{R}^{j+1} \) denote the covariance vector between \( \xi_{i,[j]}^* \) and \( \xi_{i,j+1}^* \), and let \( (s_{j+1}^*)^2 \in \mathbb{R}_+ \) be the variance of component \( \xi_{i,j+1}^* \).

**Lemma 2.1.** Under Assumption 2 we have for \( * \in \{ P, I \}, \; 0 \leq j \leq J - 1 \) and \( 0 \leq i \leq J \)
\[
\xi_{i,j+1}^*|B_j^* \sim N\left( \theta_{[j+1]}^* + (\Sigma_{j,j+1}^*)' \left( \Sigma_{[j]}^* \right)^{-1} \left( \xi_{i,[j]}^* - \theta_{[j]}^* \right), (s_{j+1}^*)^2 \right),
\]

with \( (s_{j+1}^*)^2 = (s_{j+1}^*)^2 - (\Sigma_{j,j+1}^*)' \left( \Sigma_{[j]}^* \right)^{-1} \Sigma_{j,j+1}^* \).

**Proof.** This is a standard result for multivariate Gaussian distributions, see Result 4.6 in [3]. \( \square \)
or, equivalently, if and only if the matrix

\[ A \]

for an appropriate matrix \( A \) symmetric matrix

**Lemma 2.3.**

\[
\begin{align*}
\mathbb{E} [ P_{i,j+1} | B_j^P ] &= P_{i,j} \mathbb{E} \left[ g^{-1} \left( g \left( \frac{P_{i,j+1}}{P_{i,j}} \right) \right) \right] B_j^P = P_{i,j} \mathbb{E} \left[ g^{-1} \left( \xi_{i,j+1}^P \right) \right] B_j^P, \\
\mathbb{E} [ I_{i,j+1} | B_j^I ] &= I_{i,j} \mathbb{E} \left[ g^{-1} \left( g \left( \frac{I_{i,j+1}}{I_{i,j}} \right) \right) \right] B_j^I = I_{i,j} \mathbb{E} \left[ g^{-1} \left( \xi_{i,j+1}^I \right) \right] B_j^I.
\end{align*}
\]

We have assumed that \( \Sigma \) is positive definite. This implies that also \((\Sigma^*_j)^{-1}\) is positive definite for \(* \in \{P, I\}\). We then see from Lemma 2.1 that, in general, the last terms in (2.2) and (2.3) depend on \( \xi_{i,j}^P \) and \( \xi_{i,j}^I \), respectively. Therefore, these last terms are not constant, and Assumption 1 (A2) is not fulfilled unless both \( \Sigma_{j,j+1}^P \) and \( \Sigma_{j,j+1}^I \) are equal to the zero vector. This immediately gives the next theorem.

**Theorem 2.2.** Assume that Assumption 2 is fulfilled. The model fulfills Assumption 1 if and only if

\[
\Sigma_{[j]}^P = \text{diag} \left( (s_0^P)^2, \ldots, (s_J^P)^2 \right) \quad \text{and} \quad \Sigma_{[j]}^I = \text{diag} \left( (s_0^I)^2, \ldots, (s_J^I)^2 \right).
\]

In that case we have for the log-link \( g(x) = \log x \) and for \( 0 \leq j \leq J - 1 \) and \( 0 \leq i \leq J \)

\[
\begin{align*}
\mathbb{E} [ P_{i,j+1} | B_j^P ] &= P_{i,j} \exp \left\{ \theta_{j+1}^P + (s_{j+1}^P)^2/2 \right\}, \\
\text{Var} \left( P_{i,j+1} | B_j^P \right) &= P_{i,j}^2 \exp \left\{ 2\theta_{j+1}^P + (s_{j+1}^P)^2 \right\} \left( \exp \left\{ (s_{j+1}^P)^2 \right\} - 1 \right).
\end{align*}
\]

Analogous statements hold true for claims incurred \( I_{i,j+1}, \) conditioned on \( B_j^I \).

We see that under the assumptions of Theorem 2.2 the process \( (P_{i,j})_{0 \leq j \leq J} \) has the Markov property, and we obtain chain-ladder parameters for the log-link \( g(x) = \log x \)

\[
f_j^* = \exp \left\{ \theta_{j+1}^* + (s_{j+1}^*)^2/2 \right\} \quad \text{and} \quad (\sigma_j^*)^2 = (f_j^*)^2 \left( \exp \left\{ (s_{j+1}^*)^2 \right\} - 1 \right),
\]

with \(* \in \{P, I\}\). Moreover, the covariance matrix \( \Sigma \) under Theorem 2.2 is given by

\[
\Sigma = \begin{pmatrix}
\Sigma_{[j]}^P &=& \text{diag} \left( (s_0^P)^2, \ldots, (s_J^P)^2 \right) \\
A' & \Sigma_{[j]}^I &=& \text{diag} \left( (s_0^I)^2, \ldots, (s_J^I)^2 \right)
\end{pmatrix},
\]

for an appropriate matrix \( A \in \mathbb{R}^{(J+1) \times (J+1)} \) such that \( \Sigma \) is positive definite.

**Lemma 2.3.** A symmetric matrix \( \Sigma \) of the form (2.5) is positive definite if and only if the matrix

\[
S_{[j]}^P = \Sigma_{[j]}^P - A' \left( \Sigma_{[j]}^P \right)^{-1} A \quad \text{is positive definite},
\]

or, equivalently, if and only if the matrix

\[
S_{[j]}^I = \Sigma_{[j]}^I - A' \left( \Sigma_{[j]}^I \right)^{-1} A \quad \text{is positive definite}.
\]
The matrices $S_{[j]}^*$ are called Schur complements of $\Sigma_{[j]}^*$ in $\Sigma$, for $* \in \{P, I\}$. One may still choose more structure in matrix $A = (a_{k,l})_{0 \leq k, l \leq J}$, for instance, a lower-left-triangular matrix is often a reasonable choice, i.e. $a_{k,l} = 0$ for all $k < l$. For the time-being we allow for any matrix $A$ such that $\Sigma$ is positive definite. This leads to the following model assumptions.

**Assumption 3** (multivariate (log-)normal chain-ladder model II).

(C1) We assume that the random vectors $\Xi_i$ are independent for different accident years $i = 0, \ldots, J$.

(C2) There exists a parameter vector $\theta = (\theta_0^P, \ldots, \theta_j^P, \theta_0^I, \ldots, \theta_j^I)' \in \mathbb{R}^{2(J+1)}$ and a matrix $\Sigma$ of the form (2.5) with positive definite Schur complements $S_{P,j}^*$ and $S_{I,j}^*$ such that we have for $0 \leq i \leq J$

$$\Xi_i \sim \mathcal{N}(\theta, \Sigma).$$

**Corollary 2.4.** The model of Assumption 3 fulfills the distribution-free chain-ladder model of Assumption 1. The chain-ladder parameters are given by (2.4) for the log-link function (2.1).

The previous corollary states that we have found a class of non-trivial stochastic models that fulfill the distribution-free chain-ladder assumptions simultaneously for cumulative payments and claims incurred. Note that a reasonable choice of matrix $A$ in (2.5) allows for dependence between cumulative payments and claims incurred, this will be crucial in the sequel.

## 3 One-step ahead prediction

Theorem 2.2 provides the best prediction of $P_{i,j+1}$ based on $B_{j+1}^P$ and the best prediction of $I_{i,j+1}$ based on $B_{j}^I$, respectively, under Assumption 3. The idea in the Munich chain-ladder method is to consider best predictions based on both sets of information $B_j = B_{j}^P \cup B_{j}^I$. This is similar to the considerations in [4]. In this section we start with the special case of “one-step ahead prediction”, the general case is presented in Section 6, below. We denote by $\theta_{[j]} = (\theta_0^P, \ldots, \theta_j^P, \theta_0^I, \ldots, \theta_j^I)' \in \mathbb{R}^{2(j+1)}$ and let $\Sigma_{[j]} \in \mathbb{R}^{2(j+1) \times 2(j+1)}$ be the (positive definite) covariance matrix of the random vector $\xi_{i,[j]} = (\xi_{i,0}^P, \ldots, \xi_{i,j}^P, \xi_{i,0}^I, \ldots, \xi_{i,j}^I)'$. Moreover, let $\Sigma_{j,j+1}^{(s)} \in \mathbb{R}^{2(j+1)}$ denote the covariance vector between $\xi_{i,[j]}$ and $\xi_{i,j+1}^s$ for $* \in \{P, I\}$. Note that in contrast to Lemma 2.1 we replace $\Sigma_{j,j+1}^s$ by $\Sigma_{j,j+1}^{(s)}$, i.e. we set the upper index in brackets.

**Lemma 3.1.** Under Assumption 3 we have for $* \in \{P, I\}$, $0 \leq j \leq J - 1$ and $0 \leq i \leq J$

$$
\xi_{i,j+1}^s | B_i \sim \mathcal{N} \left( \theta_{j+1}^s + (\Sigma_{j,j+1}^{(s)\prime})^{-1} \Sigma_{[j]}^{-1} (\xi_{i,[j]} - \theta_{[j]}), (\Sigma_{j,j+1}^{(s)\prime})^{-1} \right),
$$

with $(\Sigma_{j,j+1}^{(s)\prime})^{-1} = (\Sigma_{j,j+1}^s)^2 - (\Sigma_{j,j+1}^{(s)\prime}) \Sigma_{[j]}^{-1} \Sigma_{j,j+1}^{(s)}$. 

Proof. This lemma is a standard result in linear algebra about Schur complements, see Section C.4.1 in [1]. □
The previous lemma shows that the conditional expectation of $\xi_{i,j+1}$, given $\mathcal{B}_j$, is linear in the observations $\xi_{i,[j]}$. This will be crucial. An easy consequence of the previous lemma is the following corollary for the the log-link.

**Corollary 3.2** (one-step ahead prediction for log-link). Under Assumption 3 we have prediction for log-link $g(x) = \log x$ and for $0 \leq j \leq J - 1$ and $0 \leq i \leq J$

$$
E[\{P_{i,j+1}|\mathcal{B}_j\}] = \begin{bmatrix} P_{i,j} \exp\left\{ \theta_{j+1}^P + (\Sigma_{j+1}^{(P)})' \Sigma_{[j]}^{-1} (\xi_{i,[j]} - \theta_{[j]}) + \left( s_{j+1}^{(P),\text{post}} \right)^2 / 2 \right\} \\
\gamma_j^P(\xi_{i,[j]}) = E\left[ P_{i,j+1}|\mathcal{B}_j^P \right] \gamma_j^P(\xi_{i,[j]}) \end{bmatrix}
$$

with for $* \in \{P, I\}$

$$
\gamma_j^*(\xi_{i,[j]}) = \begin{cases} \beta_j^*(\xi_{i,[j]}) - (\Sigma_{j+1}^{(P)})' \Sigma_{[j]}^{-1} \Sigma_{j+1}^{(P)} / 2 & \text{for log-link} \\
(\Sigma_{j+1}^{(P)})' \Sigma_{[j]}^{-1} (\xi_{i,[j]} - \theta_{[j]}) & \text{for } g(x) \end{cases}
$$

Analogous statements hold true for claims incurred $I_{i,j+1}$.

Note that $\gamma_j^P(\xi_{i,[j]})$ gives the correction if we experience not only $\mathcal{B}_j^P$ but also $\mathcal{B}_j^I$. This increased information leads also to a reduction of prediction uncertainty of size

$$(s_{j+1}^P)^2 \implies (s_{j+1}^{(P),\text{post}})^2 = (s_{j+1}^P)^2 - (\Sigma_{j+1}^{(P)})' \Sigma_{[j]}^{-1} \Sigma_{j+1}^{(P)} \leq (s_{j+1})^2.$$  

**Example 3.3** (log-link). The analysis of the correction term $\gamma_j^P(\xi_{i,[j]})$ is not straightforward. Therefore, we consider an explicit example for the case $J = 2$ and $j = 0, 1$. In this case the covariance matrix $\Sigma$ under Assumption 3 is given by

$$\Sigma = \Sigma_{[2]} = \begin{pmatrix}
(s_0^P)^2 & 0 & 0 & a_{0,0} & a_{0,1} & a_{0,2} \\
0 & (s_1^P)^2 & 0 & a_{1,0} & a_{1,1} & a_{1,2} \\
0 & 0 & (s_2^P)^2 & a_{2,0} & a_{2,1} & a_{2,2} \\
a_{0,0} & a_{1,0} & a_{2,0} & (s_0^I)^2 & 0 & 0 \\
a_{0,1} & a_{1,1} & a_{2,1} & 0 & (s_1^I)^2 & 0 \\
a_{0,2} & a_{1,2} & a_{2,2} & 0 & 0 & (s_2^I)^2
\end{pmatrix}.$$

- **Case $j = 0$.** We start the analysis for $j = 0$, i.e. given information $\mathcal{B}_0$.

$$\Sigma_{[1]} = \begin{pmatrix}
(s_0^P)^2 & 0 & a_{0,0} & a_{0,1} \\
0 & (s_1^P)^2 & a_{1,0} & a_{1,1} \\
a_{0,0} & a_{1,0} & (s_0^I)^2 & 0 \\
a_{0,1} & a_{1,1} & 0 & (s_1^I)^2
\end{pmatrix} \quad \text{and} \quad \Sigma_{[0]}^{-1} = \frac{1}{(s_0^P s_0^I)^2 - a_{0,0}^2} \begin{pmatrix}
(s_0^P)^2 & -a_{0,0} \\
-a_{0,0} & (s_0^P)^2
\end{pmatrix}.$$

Moreover, $\Sigma_{0,1}^{(P)} = (0, a_{1,0})'$. This provides credibility weight $\alpha_{0,0}^{(P)} \in \mathbb{R}_2$ given by

$$\alpha_{0,0}^{(P)} = \Sigma_{0,1}^{(P)} \Sigma_{[0]}^{-1} = \frac{1}{(s_0^P s_0^I)^2 - a_{0,0}^2} \left( -a_{0,0} a_{1,0}, a_{1,0} (s_0^P)^2 \right).$$
and posterior variance

\[ (s_1^{(P,\text{post})})^2 = (s_1^P)^2 - (\Sigma_{0,1}^{(P)})' \Sigma_{0,1}^{-1} \Sigma_{0,1}^{(P)} = (s_1^P)^2 - \frac{(a_{1,0}s_0^{(P)})^2}{(s_0^P s_0^T)^2 - a_{0,0}^2}. \]

Observe that \( a_{1,0} = \text{Cov}(\xi_{i,1}^P, \xi_{i,0}^I) \) is the crucial term in the credibility weight \( \alpha_{0,0}^P \). If these two random variables \( \xi_{i,1}^P \) and \( \xi_{i,0}^I \) are uncorrelated, then \( a_{1,0} = 0 \) and we cannot learn from observation \( \xi_{i,0}^I \) to improve prediction \( \xi_{i,1}^P \). The predictor for log-link \( g(x) = \log x \) is given by

\[ \mathbb{E}[P_{i,1} | \mathcal{B}_0] = P_{i,0} \exp \left\{ \theta_1^P + \beta_0^P (\xi_{i,0}^P) + (s_1^{(P,\text{post})})^2 / 2 \right\} = f_0^P P_{i,0} \gamma_0^P (\xi_{i,0}^P). \]

with

\[ \beta_0^P (\xi_{i,0}^P) = \alpha_{0,0}^P (\xi_{i,0}^P - \theta_0) = -\frac{a_{1,0} a_{0,0}}{(s_0^P s_0^T)^2 - a_{0,0}^2} (\xi_{i,0}^P - \theta_0) + \frac{a_{1,0} (s_0^{(P)})^2}{(s_0^P s_0^T)^2 - a_{0,0}^2} (\xi_{i,0}^I - \theta_0). \]

Remarkable is that also observation \( \xi_{i,0}^I \) is used to improve prediction of \( \xi_{i,1}^P \), though these two random variables are uncorrelated under Assumption 3. This comes from the fact that if \( a_{0,0} \neq 0 \) then \( \xi_{i,0}^I \) is used to adjust \( \xi_{i,0}^I \).

- **Case** \( j = 1 \). This case is more involved. Set

\[
\begin{align*}
  b_{0,0} &= (s_0^I)^2 - a_{0,0}^2/(s_0^P)^2 - a_{1,0}^2/(s_1^P)^2, \\
  b_{0,1} &= -a_{0,0}^2a_{0,0}/(s_0^P)^2 - a_{1,1}a_{1,0}/(s_1^P)^2, \\
  b_{1,1} &= (s_1^I)^2 - a_{0,1}^2/(s_0^P)^2 - a_{1,1}^2/(s_1^P)^2, \\
  c_{0,0} &= \frac{b_{0,0}b_{1,1} - b_{0,1}^2}{b_{0,0}b_{1,1} - b_{0,1}^2}(s_0^P)^2, \quad c_{0,1} = \frac{-b_{0,0}a_{0,0} + b_{0,1}a_{0,0}}{b_{0,0}b_{1,1} - b_{0,1}^2}(s_0^P)^2, \\
  c_{1,0} &= \frac{b_{0,0}b_{1,1} - b_{0,1}^2}{b_{0,0}b_{1,1} - b_{0,1}^2}(s_1^P)^2, \quad c_{1,1} = \frac{-b_{0,0}a_{1,1} + b_{0,1}a_{1,1}}{b_{0,0}b_{1,1} - b_{0,1}^2}(s_1^P)^2.
\end{align*}
\]

We have the following inverse matrix for \( \Sigma_{[1]} \), see Appendix B for the full inverse matrix,

\[
\Sigma_{[1]}^{-1} = \begin{pmatrix}
  * & * & c_{0,0} & c_{0,1} \\
  * & * & c_{1,0} & c_{1,1} \\
  c_{0,0} & c_{1,0} & \frac{b_{1,1}}{b_{0,0}b_{1,1} - b_{0,1}^2} & \frac{-b_{0,1}}{b_{0,0}b_{1,1} - b_{0,1}^2} \\
  c_{0,1} & c_{1,1} & \frac{-b_{0,1}}{b_{0,0}b_{1,1} - b_{0,1}^2} & \frac{b_{1,1}}{b_{0,0}b_{1,1} - b_{0,1}^2}
\end{pmatrix}.
\]

Moreover, \( \Sigma_{[1,2]} = (0, 0, a_{2,0}, a_{2,1})' \) is the covariance vector between \( \xi_{i,[1]} \) and \( \xi_{i,[2]}^P \). This provides credibility weight \( (a_{[1,j]})' = (\Sigma_{[1,2]} \Sigma_{[1]}^{-1})' = \Sigma_{[1]}^{-1} \Sigma_{[1,2]} \in \mathbb{R}^4 \) given by

\[
a_{[1,1]} = \begin{pmatrix}
  a_{2,0}c_{0,0} + a_{2,1}c_{0,1}, \\
  a_{2,0}c_{1,0} + a_{2,1}c_{1,1}, \\
  b_{1,1}a_{2,0} - b_{0,1}a_{2,1}, \\
  -b_{0,0}a_{2,0} + b_{0,0}a_{2,1}
\end{pmatrix},
\]

and posterior variance

\[
(s_2^{(P,\text{post})})^2 = (s_2^P)^2 - (\Sigma_{1,2}^P)' \Sigma_{[1]}^{-1} \Sigma_{1,2}^P = (s_2^P)^2 - \frac{b_{1,1}a_{2,0}^2 - 2b_{0,1}a_{2,1}a_{2,0} + b_{0,0}a_{2,1}^2}{b_{0,0}b_{1,1} - b_{0,1}^2}.
\]

We again see that the crucial terms are \( a_{2,0} = \text{Cov}(\xi_{i,2}^P, \xi_{i,0}^I) \) and \( a_{2,1} = \text{Cov}(\xi_{i,2}^P, \xi_{i,1}^I) \). If these two covariances are zero then claims incurred observation is not helpful to improve prediction.
of \( \xi_{i,2} \). Therefore, we assume that at least one of these two covariances is different from zero. The predictor for the log-link \( g(x) = \log x \) is given by

\[
E[P_{i,2}|B_1] = P_{i,1} \exp \left\{ \beta_2^P + \beta_1^P (\xi_{i,[1]}^P) + (\xi_{2,\text{post}}^P)^2/2 \right\} = f_1^P P_{i,1} \gamma_1^P (\xi_{i,[1]}^P).
\]

with

\[
\beta_1^P (\xi_{i,[1]}^P) = (a_{2,0}c_{0,0} + a_{2,1}c_{0,1}) (\xi_{i,0}^P - \theta_0^P) + (a_{2,0}c_{1,0} + a_{2,1}c_{1,1}) (\xi_{i,1}^P - \theta_1^P) \tag{3.1}
\]

\[+ \frac{b_{1,1}a_{2,0} - b_{0,1}a_{2,1}}{b_{0,0}b_{1,1} - b_{0,1}^2} (\xi_{i,0}^P - \theta_0^P) + \frac{b_{0,1}a_{2,0} + b_{0,0}a_{2,1}}{b_{0,0}b_{1,1} - b_{0,1}^2} (\xi_{i,1}^P - \theta_1^P).\]

Again \( \xi_{i,0}^P \) and \( \xi_{i,1}^P \) are used to adjust \( \xi_{i,0}^I \) and \( \xi_{i,1}^I \) through \( a_{0,0}, a_{0,1} \) and \( a_{1,0}, a_{1,1} \), respectively, which are integrated into \( c_{0,0}, c_{0,1} \) and \( c_{1,0}, c_{1,1} \), respectively.

4 Munich chain-ladder model

In Corollary 3.2 we have derived the best prediction under Assumption 3 for the log-link. Note that this best prediction is understood relative to the mean-square error of prediction and it crucially depends on the choice of the link function \( g \). Since this model also fulfills the chain-ladder model Assumption 1 for any link function \( g \), see Corollary 2.4, it can also be considered as the best prediction for given information \( B_j \) in the distribution-free chain-ladder model (for the chosen link function \( g \)). The Munich chain-ladder method takes a different viewpoint in that it extends the distribution-free chain-ladder model Assumption 1 and then derives prediction under these additional assumptions. We will define this extended model in Assumption 4, below, and then study under which circumstances our distributional model from Assumption 3 fulfills these Munich chain-ladder model assumptions. Define the residuals

\[
\varepsilon_{i,j}^P = I_{i,j} - E[I_{i,j}|B_j^P] \quad \text{and} \quad \varepsilon_{i,j}^P = P_{i,j} - E[P_{i,j}|B_j^I].
\]

The Munich chain-ladder assumptions of Quarg and Mack [6] are given by:

**Assumption 4** (Munich chain-ladder model). Assume in addition to Assumption 1 that there exist constants \( \lambda^P, \lambda^I \in (-1,1) \) such that for \( 0 \leq j \leq J - 1 \) and \( 0 \leq i \leq J \)

\[
E[P_{i,j+1}|B_j] = f_j^P P_{i,j} + \lambda^P \text{ Var } (P_{i,j+1}|B_j^P)^{1/2} \varepsilon_{i,j}^P,
\]

and

\[
E[I_{i,j+1}|B_j] = f_j^I I_{i,j} + \lambda^I \text{ Var } (I_{i,j+1}|B_j^I)^{1/2} \varepsilon_{i,j}^P.
\]

The tower property for conditional expectations \( E[|P_{i,j+1}|B_j^P] = E[E[P_{i,j+1}|B_j]|B_j^P] \) implies

\[
E[P_{i,j+1}|B_j^P] = f_j^P P_{i,j} + \lambda^P \text{ Var } (P_{i,j+1}|B_j^P)^{1/2} E[\xi_{i,j}^P|B_j^P] = f_j^P P_{i,j}.
\]

8
Therefore, Assumption 4 does not contradict Assumption 1. We now analyze Assumption 4 from the viewpoint of the multivariate (log-)normal chain-ladder model of Assumption 3. We therefore need to analyze the correction term defined in the Munich chain-ladder model

\[ \lambda^P \text{Var} \left( P_{i,j+1} | B_j^P \right)^{1/2} \varepsilon_{i,j}^P = \lambda^P \sigma_j^P \varepsilon_{i,j}^P, \]

and compare it to the optimal correction term obtained from Lemma 3.1. We start with log-link \( g(x) = \log x \) and then provide the general result in Theorem 4.3, below. For the log-link we have

\[ I_{i,j} = \nu_i \exp \left\{ \sum_{l=0}^j \xi_{i,l}^l \right\} . \]

Therefore, for \( \varepsilon_{i,j}^P \) we need to determine the conditional distribution of \( \sum_{l=0}^j \xi_{i,l}^l \), given \( \xi_{i,[j]}^P \).

**Lemma 4.1.** Under Assumption 3 we have

\[ \sum_{l=0}^j \xi_{i,l}^l | B_j^P \sim \mathcal{N} \left( \sum_{l=0}^j \theta_l^l + (a_{0,j})' \left( \Sigma_{[j]}^P \right)^{-1} \left( \xi_{i,[j]}^P - \theta_{[j]}^P \right), (s_{0,j}^{\text{post}})^2 \right), \]

with covariance vector \( a_{0,j}^l = (a_{0,l}, \ldots, a_{0,J})^l \in \mathbb{R}^{j+1} \) for \( A = (a_{k,l})_{0 \leq k, l \leq J} \), and posterior variance \( (s_{0,j}^{\text{post}})^2 = \sum_{l=0}^j (s_l^j)^2 - (a_{0,j}^l)' \left( \Sigma_{[j]}^P \right)^{-1} a_{0,j}^l \).

**Proof.** This is a standard result for multivariate Gaussian distributions, see Result 4.6 in [3].

**Example 4.2 (log-link).** We consider log-link \( g(x) = \log x \). In this case we have from (4.2) and using Lemma 4.1 for the residual of the correction term

\[ \varepsilon_{i,j}^P = \frac{\exp \left\{ \sum_{l=0}^j \xi_{i,l}^l \right\} - \mathbb{E} \left[ \exp \left\{ \sum_{l=0}^j \xi_{i,l}^l \right\} | B_j^P \right]}{\text{Var} \left( \exp \left\{ \sum_{l=0}^j \xi_{i,l}^l \right\} | B_j^P \right)^{1/2}} = \frac{\exp \left\{ \sum_{l=0}^j (\xi_{i,l}^l - \theta_l^l) - (a_{0,j}^l)' \left( \Sigma_{[j]}^P \right)^{-1} \left( \xi_{i,[j]}^P - \theta_{[j]}^P \right) - (s_{0,j}^{\text{post}})^2/2 \right\} - 1}{\left( \exp \left\{ (s_{0,j}^{\text{post}})^2 \right\} - 1 \right)^{1/2}}. \]

This implies for the Munich chain-ladder model Assumption 4, we also use (2.4),

\[ f_j^P P_{i,j} + \lambda^P \text{Var} \left( P_{i,j+1} | B_j^P \right)^{1/2} \varepsilon_{i,j}^P = f_j^P P_{i,j} + \lambda^P \sigma_j^P P_{i,j} \varepsilon_{i,j}^P = f_j^P P_{i,j} \gamma_j^{P,\text{MCL}}(\xi_{i,[j]}^P), \]

with Munich chain-ladder correction factor defined by

\[ \gamma_j^{P,\text{MCL}}(\xi_{i,[j]}^P) = 1 + \lambda^P \sqrt{\frac{e^{(s_{0,j}^{\text{post}})^2}/2 - 1}{e^{(s_{0,j}^{\text{post}})^2} - 1}} \left( \sum_{l=0}^j (\xi_{i,l}^l - \theta_l^l) - (a_{0,j}^l)' \left( \Sigma_{[j]}^P \right)^{-1} \left( \xi_{i,[j]}^P - \theta_{[j]}^P \right) - (s_{0,j}^{\text{post}})^2/2 \right). \]

(4.3)
We analyze this Munich chain-ladder correction factor for \( j = 1 \). It is given by

\[
\gamma_1^{P,MCL}(\xi_{i,1}) = 1 + \lambda^P \sqrt{\frac{e(s_i^P)^2 - 1}{e(\bar{s}_{0,1}^{\text{post}})^2 - 1}} \times \left( e(\xi_{i,0}^P - \theta_0^P) + (\xi_{i,1}^P - \theta_1^P) - \frac{a_2,0 + a_2,1}{2(b_{0,0}b_{1,1} - b_{0,1}^2)} \left[ -b_{0,1} (\xi_{i,0}^P - \theta_0^P) + b_{0,0} (\xi_{i,1}^P - \theta_1^P) \right] + c_{0,1} (\xi_{i,0}^P - \theta_0^P) + c_{1,1} (\xi_{i,1}^P - \theta_1^P) \right) - \frac{b_{0,0}a_{2,1}^2}{2(b_{0,0}b_{1,1} - b_{0,1}^2)}. \tag{4.4}
\]

We compare this to the best prediction in the case \( j = 1 \) characterized by (3.1) and under the additional assumptions that \( a_{2,0} = 0 \) and \( a_{2,1} \neq 0 \). In this case we obtain from (3.1) and (2.4) correction term

\[
\gamma_1^P(\xi_{i,1}) = \exp \left( \beta_1^P(\xi_{i,1}) - \frac{b_{0,0}a_{2,1}^2}{2(b_{0,0}b_{1,1} - b_{0,1}^2)} \right)
\]

\[
= \exp \left\{ \frac{a_{2,1}}{b_{0,0}b_{1,1} - b_{0,1}^2} \left[ -b_{0,1} (\xi_{i,0}^P - \theta_0^P) + b_{0,0} (\xi_{i,1}^P - \theta_1^P) \right] + a_{2,1} [c_{0,1} (\xi_{i,0}^P - \theta_0^P) + c_{1,1} (\xi_{i,1}^P - \theta_1^P)] - \frac{b_{0,0}a_{2,1}^2}{2(b_{0,0}b_{1,1} - b_{0,1}^2)} \right\}.
\]

Note that this differs from (4.4). This can, for instance, be seen because all terms in the sum \( \sum_{l=0}^j (\xi_{i,l}^P - \theta_1^P) \) in \( \gamma_1^{P,MCL}(\xi_{i,1}) \) are equally weighted, whereas for the best predictor we consider a weighted sum \(-b_{0,1} (\xi_{i,0}^P - \theta_0^P) + b_{0,0} (\xi_{i,1}^P - \theta_1^P)\) in \( \gamma_1^P(\xi_{i,1}) \). We conclude that the Munich chain-ladder estimator in general is non-optimal under Assumption 3.

The (disappointing) conclusion from Example 4.2 is that within the family of models fulfilling Assumption 3 with log-link \( g(x) = \log x \) there is no interesting example satisfying the Munich chain-ladder model Assumption 4. Exceptions can only be found for rather artificial covariance matrices \( \Sigma \), for instance, a choice with \( A = 0 \) would fulfill the Munich chain-ladder Assumption 4. But this latter choice is not of interest because it requires \( \lambda^P = \lambda^l = 0 \). This result can be generalized to any link function as the next theorem shows.

**Theorem 4.3.** Assume that cumulative payments \( P_{i,j} \) and claims incurred \( I_{i,j} \) fulfill Assumption 3 for a given link function \( g \). In general, this model does not fulfill the Munich chain-ladder model Assumption 4, except for special choices of \( \Sigma \).

**Proof.** The optimal one-step ahead prediction for given link function \( g \) is given by, see also Lemma 3.1,

\[
E[P_{i,j+1}|B_j] = P_{i,j} E \left[ g^{-1} \left( \xi_{i,j+1}^P \right) \bigg| B_j \right],
\]

with

\[
\xi_{i,j+1}^P | B_j \sim N \left( \theta^P_{i,j+1} + \left( \Sigma^P_{j+1} \right)^{-1} \left( \sum_{[j]} \left( \xi_{i,[j]} - \theta_{[j]} \right) \right), \left( \Sigma^P_{j+1} \right)^{-1} \right).
\]

From the latter we observe that observation \( \xi_{i,[j]} \) is considered in a linear fashion \( c^T \xi_{i,[j]} \) for an appropriate vector \( c \in \mathbb{R}^{j+1} \), which typically is different from zero (for \( A \neq 0 \)) and which does not point into the direction of \((1, \ldots, 1)^T \in \mathbb{R}^{j+1}\), i.e. we consider a weighted sum of the components of \( \xi_{i,[j]} \) (with non-identical weights).
On the other hand, the correction terms from the Munich chain-ladder assumption for a given link function \( g \) are given by, see also (4.1),

\[
\lambda^P \sigma_j^P P_{i,j} \varepsilon_{i,j}^{lP} = \lambda^P \sigma_j^P P_{i,j} \frac{g^{-1}(\nu_i) \prod_{l=0}^i g^{-1}(\xi_{i,l}) - \mathbb{E} \left[ I_{i,j} \mid B_j^P \right]}{\text{Var} \left( I_{i,j} \mid B_j^P \right)^{1/2}} = \lambda^P \sigma_j^P P_{i,j} \frac{g^{-1}(\nu_i) \exp \left\{ \sum_{l=0}^i \log \left( g^{-1}(\xi_{i,l}) \right) \right\} - \mathbb{E} \left[ I_{i,j} \mid B_j^P \right]}{\text{Var} \left( I_{i,j} \mid B_j^P \right)^{1/2}}.
\]

Thus, the only link function \( g \) which considers the components of \( \xi_{i,j}^l \) in a linear fashion is the log-link \( g(x) = \log x \). For the log-link we get

\[
\lambda^P \sigma_j^P P_{i,j} \varepsilon_{i,j}^{lP} = \lambda^P \sigma_j^P P_{i,j} \frac{\exp \left\{ \nu_i + \sum_{l=0}^i \xi_{i,l} \right\} - \mathbb{E} \left[ I_{i,j} \mid B_j^P \right]}{\text{Var} \left( I_{i,j} \mid B_j^P \right)^{1/2}}.
\]

From this we see that all components of \( \xi_{i,j}^l \) are considered with identical weights, and therefore it differs from the optimal one-step ahead prediction (if the latter uses non-identical weights). This is exactly what we have seen in Example 4.2.

In Theorem 4.1 of [4] the Munich chain-ladder structure has been found as a best linear approximation to \( \mathbb{E} \left[ P_{i,j+1} \mid B_j \right] \) in the following way

\[
\mathbb{E}_{\text{linear}} \left[ P_{i,j+1} \mid B_j \right] = \arg\min_{X = c_1 P_{i,j+1} + c_2 I_{i,j}; \ c_1, c_2 \in \mathcal{L}(B_j^P)} \mathbb{E} \left[ (X - P_{i,j+1})^2 \mid B_j^P \right]
\]

\[
= P_{i,j}^P \lambda_j^P \text{Corr} \left( P_{i,j+1}, I_{i,j} \mid B_j^P \right) \text{Var} \left( P_{i,j+1} \mid B_j^P \right)^{1/2} \varepsilon_{i,j}^{lP},
\]

where \( \mathcal{L}(B_j^P) \) is the space of \( B_j^P \)-measurable random variables. Note that this approximates the exact conditional expectation \( \mathbb{E} \left[ P_{i,j+1} \mid B_j \right] \) and it gives an explicit meaning to parameter \( \lambda^P \in (-1, 1) \) (which typically is non-constant in \( j \)), see also Section 2.2.2 in [6].

**Theorem 4.4** (approximation error of MCL predictor). Under Assumption 3 and the log-link choice \( g(x) = \log x \) we have approximation error for the Munich chain-ladder predictor \( \mathbb{E}_{\text{linear}} \left[ P_{i,j+1} \mid B_j \right] \) given by

\[
\mathbb{E}_{\text{linear}} \left[ P_{i,j+1} \mid B_j \right] - \mathbb{E} \left[ P_{i,j+1} \mid B_j \right] = f_j^P P_{i,j} \left( \gamma_j^P \text{MCL}(\xi_{i,j}) - \gamma_j^P(\xi_{i,j}) \right),
\]

where \( \gamma_j^P \text{MCL}(\xi_{i,j}) \) is given in (4.3) with \( \lambda^P \) replaced by \( \text{Corr}(P_{i,j+1}, I_{i,j} \mid B_j^P) \) and \( \gamma_j^P(\xi_{i,j}) \) is given in Corollary 3.2.

**Proof.** This proof follows from Example 4.2.

\]

5 The modified Munich chain-ladder method

In the sequel we concentrate on the model of Assumption 3 with log-link function \( g(x) = \log x \). This provides the chain-ladder model specified in Theorem 2.2 and the one-step ahead prediction given in Corollary 3.2. The issues that we still need to consider are the following: (i) Would
like to extend the one-step ahead prediction to get the ultimate claim prediction, i.e. prediction of all future periods. (ii) Typically, model parameters are not known and need to be estimated. (iii) We should specify the prediction uncertainty. In order to achieve these goals we choose a Bayesian modeling framework.

**Assumption 5** (modified Munich chain-ladder model). Choose the log-link \( g(x) = \log x \) and assume the following: There is a fixed covariance matrix \( \Sigma \) of the form (2.5) given having positive definite Schur complements \( S^P_{[J]} \) and \( S^I_{[J]} \).

- Conditionally, given parameter vector \( \Theta = (\Theta^P_0, \ldots, \Theta^P_J, \Theta^I_0, \ldots, \Theta^I_J)' \), the random vectors \( \Xi_i \) are independent for different accident years \( i = 0, \ldots, J \) with
  \[
  \Xi_i|\Theta \sim N(\Theta, \Sigma).
  \]

- The parameter vector \( \Theta \) has prior distribution
  \[
  \Theta \sim N(\theta, T),
  \]
  with prior mean \( \theta \in \mathbb{R}^{2(J+1)} \) and symmetric positive definite prior covariance matrix \( T \in \mathbb{R}^{2(J+1) \times 2(J+1)} \).

We first merge all accident years \( i = 0, \ldots, J \) to one random vector

\[
\Xi = (\Xi'_0, \ldots, \Xi'_J)',
\]

which has conditional distribution

\[
\Xi|\Theta \sim N(A\Theta, \Sigma^+),
\]

for an appropriate matrix \( A \in \mathbb{R}^{2(J+1)^2 \times 2(J+1)} \) and covariance matrix \( \Sigma^+ = \text{diag}(\Sigma_0, \ldots, \Sigma) \). The following lemma is crucial, for the proof we refer to Theorem 10.17 in [7].

**Lemma 5.1.** Under Assumption 5 the random vector \( \zeta = (\Xi', \Theta)\)' has a multivariate Gaussian distribution given by

\[
\zeta = \begin{pmatrix} \Xi \\ \Theta \end{pmatrix} \sim N\left(\begin{pmatrix} A\theta \\ \theta \end{pmatrix}, S = \begin{pmatrix} \Sigma^+ + AA' \\ TA' \\ TA \\ T \end{pmatrix}\right).
\]

An easy consequence of Lemma 5.1 is the following marginal distribution

\[
\Xi \sim N(A\theta, \Sigma^+ + AA').
\]

This shows that in the Bayesian multivariate normal model with Gaussian priors we can completely "integrate out" the hierarchy of parameters \( \Theta \). However, we will keep the hierarchy of parameters in order to obtain Bayesian parameter estimates for \( \Theta \).

Denote the dimension of \( \zeta \) by \( n = 2(J+1)^2 + 2(J+1) \). Choose \( t, v \in \mathbb{N} \) with \( t + v = n \). Denote by \( P_t \in \mathbb{R}^{t \times n} \) and \( P_v \in \mathbb{R}^{v \times n} \) the projections such that we obtain a disjoint decomposition of the components of \( \zeta \)

\[
\zeta \mapsto (\zeta_t, \zeta_v) = (P_t\zeta, P_v\zeta).
\]
The random vector \((\zeta_t', \zeta_v')\) has a multivariate Gaussian distribution with expected values
\[
\mu_t = \mathbb{E}[\zeta_t] = P_t \mu \quad \text{and} \quad \mu_v = \mathbb{E}[\zeta_v] = P_v \mu,
\]
and with covariance matrices
\[
S_t = \text{Cov}(\zeta_t) = P_t S'P_t', \quad S_v = \text{Cov}(\zeta_v) = P_v S'P_v', \quad S_{v,t} = \text{Cov}(\zeta_t, \zeta_v) = P_t S'P_v'.
\]
The projections in (5.1) only describe a permutation of the components of \(\zeta\). In complete analogy to Lemma 2.1 we have the following lemma.

**Lemma 5.2.** Under Assumption 5 the random vector \(\zeta_v|\{\zeta_t\}\) has a multivariate Gaussian distribution with the first two conditional moments given by
\[
\mu_{\text{post}}^v = \mathbb{E}[\zeta_v|\zeta_t] = \mu_v + S_{v,t} (S_t)^{-1} (\zeta_t - \mu_t),
\]
\[
S_{\text{post}}^v = \text{Cov}(\zeta_v|\zeta_t) = S_v - S_{v,t} (S_t)^{-1} S_{t,v}.
\]
This lemma now allows for parameter estimation and prediction at time \(J\), conditionally given observations
\[
\mathcal{D}_J^P = \{P_{i,j}; 0 \leq i \leq J, 0 \leq j \leq J; i + j \leq J\},
\]
\[
\mathcal{D}_J^I = \{I_{i,j}; 0 \leq i \leq J, 0 \leq j \leq J; i + j \leq J\},
\]
\[
\mathcal{D}_J = \mathcal{D}_J^P \cup \mathcal{D}_J^I.
\]
Choose \(t = |\mathcal{D}_J|\) and \(v = n - t\) and denote by \(P_t\) the projection of \(\zeta\) onto the components \(\xi_{i,j}^P\) and \(\xi_{i,j}^I\) with \(i + j \leq J\). These are exactly the components that generate information \(\mathcal{D}_J\).

Lemma 5.2 allows to calculate the posterior distribution of \(\zeta_v\), conditionally given \(\mathcal{D}_J\). We split this calculation into two parts, one for parameter estimation and one for claims prediction. We consider therefore the following projection
\[
Q_\Theta \in \mathbb{R}^{2(J+1) \times v} \text{ with } Q_\Theta \zeta_v = \Theta.
\]
This projection extracts the parameter vector \(\Theta\) from the unobserved components \(\zeta_v\).

**Corollary 5.3 (parameter estimation).** Under Assumption 5 the Bayesian estimator for the parameter vector \(\Theta\) is at time \(J\) given by
\[
\theta^{\text{post}} = \mathbb{E}[\Theta|\mathcal{D}_J] = Q_\Theta \mu_{\text{post}}^v.
\]
This can now be compared to the individual estimates
\[
\theta^{(*),\text{post}} = \mathbb{E}[\Theta|\mathcal{D}_J^*], \quad (5.2)
\]
where for \(* \in \{P, I\}\) we either condition on \(\mathcal{D}_J^P\) or on \(\mathcal{D}_J^I\).
6 Claims prediction and prediction uncertainty

For the prediction of the ultimate claim we have two different possibilities, either we predict the ultimate claim of cumulative payments \( P_{i,J} \) or the ultimate claim of claims incurred \( I_{i,J} \). Naturally, these two predictors will differ, unless we make a similar (additional) assumption as in [5]. We refrain from making an additional assumption in order to keep the predictor comparable to the Munich chain-ladder method of [6]. Assume we have chosen the log-link \( g(x) = \log x \), then we need to calculate for \( i = 1, \ldots, J \)

\[
\mathbb{E} [P_{i,J} | D_J] = P_{i,J-i} \mathbb{E} \left[ \exp \left\{ \sum_{l = J-i+1}^{J} \xi^P_{i,l} \right\} | D_J \right],
\]

and

\[
\mathbb{E} [I_{i,J} | D_J] = I_{i,J-i} \mathbb{E} \left[ \exp \left\{ \sum_{l = J-i+1}^{J} \xi^I_{i,l} \right\} | D_J \right].
\]

Assume again that \( \zeta \) exactly corresponds to the observations in \( D_J \). Then we define for \( i = 1, \ldots, J \) and \( * \in \{ P, I \} \) the linear maps

\[ Q^*_i \in \mathbb{R}^{1 \times v} \text{ with } Q^*_i \zeta_v = \sum_{l = J-i+1}^{J} \xi^*_i l. \]

This is the sum of the unobserved components of accident year \( i \) at time \( J \) for cumulative payments and claims incurred, respectively.

**Theorem 6.1** (modified Munich chain-ladder (mMCL) predictors). Under Assumption 5 the Bayesian predictors for the ultimate claims of accident years \( i = 1, \ldots, J \) at time \( J \) are given by

\[
\hat{P}^{\text{mMCL}}_{i,J} = \mathbb{E} [P_{i,J} | D_J] = P_{i,J-i} \exp \left\{ Q^P_i \mu^\text{post}_v + Q^P_i S^\text{post}_v (Q^P_i)' / 2 \right\},
\]

and

\[
\hat{I}^{\text{mMCL}}_{i,J} = \mathbb{E} [I_{i,J} | D_J] = I_{i,J-i} \exp \left\{ Q^I_i \mu^\text{post}_v + Q^I_i S^\text{post}_v (Q^I_i)' / 2 \right\}.
\]

The conditional mean-square error of predictions are given by

\[
\text{msep}_{\sum_{i=1}^{J} P_{i,J} | D_J} \left( \sum_{i=1}^{J} \hat{P}^{\text{mMCL}}_{i,J} \right) = \text{Var} \left( \sum_{i=1}^{J} \hat{P}^{\text{mMCL}}_{i,J} | D_J \right)
\]

\[
= \sum_{i,k=1}^{J} \hat{P}^{\text{mMCL}}_{i,J} \hat{P}^{\text{mMCL}}_{k,J} \left( \exp \{ Q^P_i S^\text{post}_v (Q^P_k)' \} - 1 \right)
\]

and analogously for claims incurred \( \text{msep}_{\sum_{i=1}^{J} I_{i,J} | D_J} \left( \sum_{i=1}^{J} \hat{I}^{\text{mMCL}}_{i,J} \right) \).

This can now again be compared to the individual predictors

\[
\hat{P}^{\text{CL}}_{i,J} = \mathbb{E} [P_{i,J} | D^P_J] \quad \text{and} \quad \hat{I}^{\text{CL}}_{i,J} = \mathbb{E} [I_{i,J} | D^I_J],
\]

and the corresponding conditional mean-square errors of prediction. Note that these individual predictors correspond to the predictors in the model of [2] under Gaussian prior assumption for the (unknown) mean parameters. Predictors and prediction uncertainty of (6.1) can (easily) be obtained from Theorem 6.1 using the particular choice \( A = 0 \) in \( \Sigma \).
7 Example

We provide an explicit example for which we calculate the chain-ladder (CL) reserves according to (6.1), the reserves in the modified Munich chain-ladder (mMCL) method of Theorem 6.1, the (non-optimal) Munich chain-ladder (MCL) reserves (according to Assumption 4), as well as the paid-incurred chain (PIC) reserves derived in [5]. In order to have comparability between these different methods we choose for each method a Bayesian framework with non-informative priors for the mean parameters.

We choose the original data of Quarg and Mack [6], they are provided in the appendix. We also provide the choices of \( s_j^p \) and \( s_j^I \) for the log-link \( g(x) = \log x \) in the appendix (for these parameters we simply choose the sample standard deviations with the usual exponential extrapolation for the last period \( j = 6 \)). We then calculate the chain-ladder parameters \( \hat{f}_j^P \) and \( \hat{f}_j^I \) according to Corollary 2.4 and (5.2), i.e. in (2.4) we replace \( \theta_j^{\ast} + 1 \) by \( \theta_j^{\ast\text{post}} \) and \( (s_j^{\ast} + 1)^2 \) by \( (s_j^{\ast} + 1)^2(1 + 1/(J - j)) \), the latter being the posterior variance parameters in the non-informative prior case for \( A = 0 \) in \( \Sigma \). We also provide these numerical values in the appendix. From these parameters we can then calculate the chain-ladder reserves from the chain-ladder predictors (6.1), which are defined by

\[
\hat{R}_{CL,\text{paid}}^i = E\left[P_{i,J} | D_j^P\right] - P_{i,J} - i \quad \text{and} \quad \hat{R}_{CL,\text{inc}}^i = E\left[I_{i,J} | D_j^I\right] - P_{i,J} - i.
\]

The results are provided in Table 1. The main observation is that there are quite substantial differences between the chain-ladder reserves from cumulative payments \( \hat{R}^i_{CL,\text{paid}} \) and the ones from claims incurred \( \hat{R}^i_{CL,\text{inc}} \), see Table 1. To bridge this gap we study the other reserving methods.

| a.y. | chain-ladder | MCL | mMCL |
|------|--------------|-----|------|
|      |  | \( \hat{R}_{CL,\text{paid}}^i \) | \( \hat{R}_{MCL,\text{paid}}^i \) | \( \hat{R}_{mMCL,\text{paid}}^i \) |
| 1    | 32  | 35  | 8   |
| 2    | 157 | 92  | 96  | 124 |
| 3    | 337 | 262 | 384 | 286 | 300 |
| 4    | 416 | 289 | 365 | 201 | 345 |
| 5    | 925 | 656 | 858 | 459 | 643 |
| 6    | 4'339 | 5'395 | 4'817 | 5'381 |
| total| 6'205 | 7'730 | 6'528 | 7'731 | 6'839 |

Table 1: Resulting reserves from the chain-ladder method \( \hat{R}_{CL,\text{paid}}^i \) and \( \hat{R}_{CL,\text{inc}}^i \), from the Munich chain-ladder method \( \hat{R}_{MCL,\text{paid}}^i \) and \( \hat{R}_{MCL,\text{inc}}^i \), from the modified Munich chain-ladder method \( \hat{R}_{mMCL,\text{paid}}^i \) and \( \hat{R}_{mMCL,\text{inc}}^i \) and from the paid-incurred chain method \( \hat{R}_{PIC}^i \).

We start with the Munich chain-ladder method of [6] with changed variance functions according to our Assumption 1 (A2). These changes of the variances also lead to slightly different parameter estimates compared to [6]. For the correlation parameters defined in Assumption 4 we obtain
estimates $\hat{\lambda}^P = 49\%$ and $\hat{\lambda}^I = 45\%$ (if we use the estimators of Section 3.1.2 in [6] with changed variance functions). Using these estimates we can then calculate the reserves in the Munich chain-ladder method. As in [6] we obtain the two values $\tilde{R}_i^{MCL,\text{paid}}$ and $\tilde{R}_i^{MCL,\text{inc}}$ for the reserves based on cumulative payments with claims incurred corrections $\varepsilon_{i,j}^P$ and for the ones based on claims incurred with cumulative payments corrections $\varepsilon_{i,j}^I$, respectively (see also Assumption 4). The results are provided in Table 1. We observe that the gap between cumulative payment reserves and claims incurred reserves becomes more narrow due to the correction factors. Both reserves were derived under model Assumption 5 and, henceforth, are non-optimal within this model (as has been seen in Theorems 4.3 and 4.4). Moreover, there is no sensible estimate for the prediction uncertainty. Therefore, we study the optimal estimators within Assumption 4 next.

We calculate the reserves in the modified Munich chain-ladder method of Assumption 5, see Theorem 6.1. We therefore need to specify the off-diagonal matrix $A = (a_{k,l})_{0 \leq k,l \leq J}$, see (2.5).

A first idea to calibrate this matrix $A$ is to use correlation estimates $\hat{\lambda}^P = 49\%$ and $\hat{\lambda}^I = 45\%$ from the Munich chain-ladder method. A crude approximation using Theorem 4.4 provides

$$
49\% = \hat{\lambda}^P \approx \text{Corr}(P_{i,j+1}, I_{i,j} | B_{i,j}) \approx \frac{\sum_{k=0}^{j} a_{j+1,k} \sigma_{j+1}^P}{\left(\sum_{k=0}^{j} \sigma_k^I\right)^{1/2}} = \frac{\sum_{k=0}^{j} \text{Corr}(\xi_{i,j+1}, \xi_{i,k}) \sigma_k^I}{\left(\sum_{k=0}^{j} (\sigma_k^I)^2\right)^{1/2}}.
$$

From this we see that in our numerical example we need comparatively high correlations, for instance, $\text{Corr}(\xi_{i,j+1}, \xi_{i,k}) \geq 40\%$ would be in line with $\hat{\lambda}^P = 49\%$. The difficulty with this choice is that the resulting matrix $\Sigma$ of type (2.5) is not positive definite! Therefore, we need to choose smaller correlations. We do the following choice for all $i, j \geq 0$

$$
\text{Corr}(\xi_{i,j+m}, \xi_{i,j}) = \begin{cases} 
40\% & \text{for } m = 0, \\
30\% & \text{for } m = 1, \\
20\% & \text{for } m = 2, \\
10\% & \text{for } m = 3,
\end{cases}
$$

and 0\% otherwise. This provides a positive definite choice for $\Sigma$ of type (2.5) in our example. This choice means that we can learn from claims incurred observations $\xi_{i,j}^I$ for cumulative payments observations $\xi_{i,j+m}^P$ with development lags $m = 0, 1, 2, 3$, but no other conclusions can be drawn from observations. The resulting modified Munich chain-ladder reserves

$$
\tilde{R}_i^{mMCL,\text{paid}} = \tilde{P}_i^{mMCL} - P_{i,J-i} \quad \text{and} \quad \tilde{R}_i^{mMCL,\text{inc}} = \tilde{P}_i^{mMCL} - P_{i,J-i},
$$

according to Theorem 6.1, are then provided in Table 1. We observe that the reserves $\tilde{R}_i^{mMCL,\text{paid}}$ based on cumulative payments are closer to the claims incurred reserves $\tilde{R}_i^{mMCL,\text{inc}}$. This is due to correlation choices (7.1). On the other hand, claims incurred reserves remain almost constant, i.e. $\tilde{R}_i^{mMCL,\text{inc}} \approx \tilde{R}_i^{mMCL,\text{inc}}$. This is due to the fact that $a_{k,l} = 0$ for $k < l$ and therefore cumulative payments observations have only a minor influence (via parameter estimation) on claims incurred reserves.

If we want to completely close the gap between cumulative payment reserves and claims incurred reserves, we need to make an additional assumption in Assumption 5. This additional
The assumption can be of similar nature as the one in the paid-incurred chain reserving method presented in [5], namely $P_{i,j} = I_{i,j}$, $\mathbb{P}$-a.s. If this assumption is made, then ultimate claims are identical, $\mathbb{P}$-a.s., and there is only one reserve $\hat{R}_{i}^{\text{PIC}}$ which is based on the entire information $D_J = D_J^P \cup D_J^L$, see [5]. The numerical result of paid-incurred chain model of [5] is provided in the last column of Table 1.

Finally, we analyze the prediction uncertainty measured by the square-rooted conditional mean square error of prediction. The results are provided in Table 2. The prediction uncertainties

| Method                                      | Reserves | msep$^{1/2}$ |
|---------------------------------------------|----------|--------------|
| chain-ladder paid $\hat{R}_{i}^{\text{CL, paid}}$ | 6'205    | 1'249        |
| chain-ladder incurred $\hat{R}_{i}^{\text{CL, inc}}$ | 7'730    | 1'565        |
| Munich chain-ladder paid $\hat{R}_{i}^{\text{MCL, paid}}$ | 6'729    | 1'239*       |
| Munich chain-ladder incurred $\hat{R}_{i}^{\text{MCL, inc}}$ | 7'140    | 1'673*       |
| modified Munich chain-ladder paid $\hat{R}_{i}^{\text{mMCL, paid}}$ | 6'528    | 1'223        |
| modified Munich chain-ladder incurred $\hat{R}_{i}^{\text{mMCL, inc}}$ | 7'731    | 1'565        |
| paid-incurred chain method $\hat{R}_{i}^{\text{PIC}}$ | 6'839    | 976          |

Table 2: Resulting reserves and square-rooted conditional mean square error of prediction from the chain-ladder method $\hat{R}_{i}^{\text{CL, paid}}$ and $\hat{R}_{i}^{\text{CL, inc}}$, from the Munich chain-ladder method $\hat{R}_{i}^{\text{MCL, paid}}$ and $\hat{R}_{i}^{\text{MCL, inc}}$, from the modified Munich chain-ladder method $\hat{R}_{i}^{\text{mMCL, paid}}$ and $\hat{R}_{i}^{\text{mMCL, inc}}$ and from the paid-incurred chain method $\hat{R}_{i}^{\text{PIC}}$. The values highlighted with * are calculated from formula (7.2).

of the chain-ladder reserves and of the modified Munich chain-ladder reserves were calculated according to Theorem 6.1. For the former (chain-ladder reserves) we simply need to set $A = 0$. We see that the uncertainties in the modified version for cumulative payments are reduced because correlations (7.1) imply that we can learn from incurred claims for cumulative payments. For claims incurred they remain invariant because of choices $a_{k,l} = 0$ for $k < l$.

We can now also calculate the prediction uncertainty for the Munich chain-ladder method (which is still an open problem). Within Assumption 5 we know that the modified Munich chain-ladder predictor is optimal, therefore, we obtain prediction uncertainty for the Munich chain-ladder method

$$
msep_{\sum_i P_{i,j} | D_J} \left( \sum_i \hat{P}_{i,j}^{\text{MCL}} \right) = msep_{\sum_i P_{i,j} | D_J} \left( \sum_i \hat{P}_{i,j}^{\text{mMCL}} \right) + \left( \sum_i \hat{P}_{i,j}^{\text{MCL}} - \sum_i \hat{P}_{i,j}^{\text{mMCL}} \right)^2,
$$

and similarly for claims incurred. The second term in (7.2) is the approximation error because the Munich chain-ladder predictor is non-optimal within Assumption 5.

Finally, we provide the prediction uncertainty in the paid-incurred chain method. In this case we fully benefit from cumulative payment data and claims incurred data, therefore the square-rooted conditional mean square error of prediction is roughly $976 \approx \frac{1}{2}(1'249 + 1'565)/\sqrt{2} = 995$.

We have performed these 4 methods on various different data sets. It has turned out that the
level of reserves depends in a rather sensitive way on the quality of claims incurred data. For instance, changes in the estimation philosophy of claims incurred data lead to diagonal effects in claims development triangles. These diagonal effects in claims incurred data often lead to unreasonable reserves in the (modified) Munich chain-ladder method and, therefore, in such cases one should either rely on cumulative payment data only or one should use other reserving methods such as the paid-incurred chain method of [5].

8 Conclusions

We have studied the Munich chain-ladder axioms of Quarg and Mack [6]. In Theorem 4.3 we conclude that, in general, chain-ladder models do not fulfill these axioms. Therefore, we introduce a modified Munich chain-ladder method which is fully consistent with our stochastic model assumptions. Within this new framework we derive best-estimate reserves and the corresponding prediction uncertainties. These results also allow to analyze the prediction uncertainty in the classical Munich chain-ladder method (which was still an open problem). Our concluding example proposes that the paid-incurred chain method of [5] provides more stable results compared to the (modified) Munich chain-ladder method.

A Data of Quarg and Mack [6]

Observed cumulative payments $P_{i,j}$, $i + j \leq 6$, and parameter choices.

| a.y. $i$/d.y. $j$ | 0    | 1    | 2    | 3    | 4    | 5    | 6    |
|-------------------|------|------|------|------|------|------|------|
| 0                 | 576  | 1'804| 1'970| 2'024| 2'074| 2'102| 2'131|
| 1                 | 866  | 1'948| 2'162| 2'232| 2'284| 2'348|
| 2                 | 1'412| 3'758| 4'252| 4'416| 4'494|
| 3                 | 2'286| 5'292| 5'724| 5'850|
| 4                 | 1'868| 3'778| 4'648|
| 5                 | 1'442| 4'010|
| 6                 | 2'044|

| $\theta_j^{(P),post}$ | 7.2195 | 0.9163 | 0.1203 | 0.0296 | 0.0216 | 0.0205 | 0.0137 |
|------------------------|--------|--------|--------|--------|--------|--------|--------|
| $s_j^P$                | 0.4972 | 0.1600 | 0.0515 | 0.0069 | 0.0036 | 0.0101 | 0.0036 |
| $f_j^P$                | 1'573  | 2.5376 | 1.1296 | 1.0301 | 1.0219 | 1.0208 | 1.0138 |

Observed claims incurred $I_{i,j}$, $i + j \leq 6$, and parameter choices.

| a.y. $i$/d.y. $j$ | 0    | 1    | 2    | 3    | 4    | 5    | 6    |
|-------------------|------|------|------|------|------|------|------|
| 0                 | 978  | 2'104| 2'134| 2'144| 2'174| 2'182| 2'174|
| 1                 | 1'844| 2'552| 2'466| 2'480| 2'508| 2'454|
| 2                 | 2'904| 4'354| 4'698| 4'600| 4'644|
| 3                 | 3'502| 5'958| 6'070| 6'142|
| 4                 | 2'812| 4'882| 4'852|
| 5                 | 2'642| 4'406|
| 6                 | 5'022|

| $\theta_j^{(I),post}$ | 7.8404 | 0.5151 | 0.0137 | 0.0003 | 0.0115 | -0.0090 | -0.0037 |
|------------------------|--------|--------|--------|--------|--------|---------|---------|
| $s_j^I$                | 0.5182 | 0.1503 | 0.0406 | 0.0146 | 0.0022 | 0.0180  | 0.0022  |
| $f_j^I$                | 2'963  | 1.6959 | 1.0148 | 1.0004 | 1.0116 | 0.9912  | 0.9963  |
B Inverse matrix $\Sigma_{[1]}$

Consider the matrix

$$
\Sigma_{[1]} = \begin{pmatrix}
(s_0^p)^2 & 0 & a_{0,0} & a_{0,1} \\
0 & (s_1^p)^2 & a_{1,0} & a_{1,1} \\
a_{0,0} & a_{1,0} & (s_0^d)^2 & 0 \\
a_{0,1} & a_{1,1} & 0 & (s_1^d)^2
\end{pmatrix}.
$$

Set

$$
b_{0,0} = (s_0^d)^2 - a_{0,0}/(s_0^p)^2 - a_{1,0}/(s_1^p)^2, $$

$$b_{1,1} = (s_1^d)^2 - a_{0,1}/(s_0^p)^2 - a_{1,1}/(s_1^p)^2, $$

$$b_{0,1} = -a_{0,0}a_{0,1}/(s_0^p)^2 - a_{1,1}a_{1,0}/(s_1^p)^2. $$

The inverse matrix of $\Sigma_{[1]}$ is given by

$$
(\Sigma_{[1]})^{-1} = \begin{pmatrix}
\frac{1}{(s_0^p)^2} + \frac{b_{0,0}a_{0,0}a_{0,1}^2 - 2b_{0,0}a_{0,0}a_{0,1} + b_{0,1}a_{0,0}^2}{(b_{0,0}b_{1,1} + b_{0,1}^2)(s_0^d)^4} & \frac{b_{0,0}a_{0,1}a_{1,1} - b_{1,1}(a_{0,0}a_{0,1} + a_{1,0}a_{0,1} - b_{1,1}1a_{0,0} + b_{1,1}a_{0,0}a_{0,0})}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_0^d)^2(s_1^d)^2} & \frac{b_{0,0}a_{0,1}a_{1,1} - b_{1,1}(a_{0,0}a_{0,1} + a_{1,0}a_{0,1} - b_{1,1}1a_{0,0} + b_{1,1}a_{0,0}a_{0,0})}{(b_{0,0}b_{1,1} + b_{0,1}^2)(s_0^d)^2(s_1^d)^2} & -\frac{b_{0,0}a_{0,1} + b_{0,1}a_{0,0}}{b_{0,0}a_{0,1} - b_{0,1}a_{0,0}} \\
\frac{1}{(s_1^p)^2} + \frac{b_{0,0}a_{0,0}a_{0,1}^2 - 2b_{0,0}a_{0,0}a_{0,1} + b_{0,1}a_{0,0}^2}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_0^d)^4} & \frac{b_{0,0}a_{0,1}a_{1,1} - b_{1,1}(a_{0,0}a_{0,1} + a_{1,0}a_{0,1} - b_{1,1}1a_{0,0} + b_{1,1}a_{0,0}a_{0,0})}{(b_{0,0}b_{1,1} + b_{0,1}^2)(s_0^d)^2(s_1^d)^2} & \frac{b_{0,0}a_{0,1}a_{1,1} - b_{1,1}(a_{0,0}a_{0,1} + a_{1,0}a_{0,1} - b_{1,1}1a_{0,0} + b_{1,1}a_{0,0}a_{0,0})}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_0^d)^2(s_1^d)^2} & \frac{b_{0,0}a_{0,1} + b_{0,1}a_{0,0}}{b_{0,0}a_{0,1} - b_{0,1}a_{0,0}} \\
\frac{1}{(s_0^d)^2} & \frac{b_{0,0}a_{0,0}a_{0,1}^2 - 2b_{0,0}a_{0,0}a_{0,1} + b_{0,1}a_{0,0}^2}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_0^d)^4} & \frac{b_{0,0}a_{0,1}a_{1,1} - b_{1,1}(a_{0,0}a_{0,1} + a_{1,0}a_{0,1} - b_{1,1}1a_{0,0} + b_{1,1}a_{0,0}a_{0,0})}{(b_{0,0}b_{1,1} + b_{0,1}^2)(s_0^d)^2(s_1^d)^2} & \frac{b_{0,0}a_{0,1} + b_{0,1}a_{0,0}}{b_{0,0}a_{0,1} - b_{0,1}a_{0,0}} \\
\frac{1}{(s_1^d)^2} & \frac{b_{0,0}a_{0,0}a_{0,1}^2 - 2b_{0,0}a_{0,0}a_{0,1} + b_{0,1}a_{0,0}^2}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_0^d)^4} & \frac{b_{0,0}a_{0,1}a_{1,1} - b_{1,1}(a_{0,0}a_{0,1} + a_{1,0}a_{0,1} - b_{1,1}1a_{0,0} + b_{1,1}a_{0,0}a_{0,0})}{(b_{0,0}b_{1,1} + b_{0,1}^2)(s_0^d)^2(s_1^d)^2} & \frac{b_{0,0}a_{0,1} + b_{0,1}a_{0,0}}{b_{0,0}a_{0,1} - b_{0,1}a_{0,0}}
\end{pmatrix}.
$$
References

[1] Boyd, S., Vandenberghe, L. (2004). Convex Optimization. Cambridge University Press.

[2] Hertig, J. (1985). A statistical approach to the IBNR-reserves in marine insurance. ASTIN Bulletin 15/2, 171-183.

[3] Johnson, R.A., Wichern, D.W. (1988). Applied Multivariate Statistical Analysis. 2nd edition. Prentice-Hall.

[4] Merz, M., Wüthrich, M.V. (2006). A credibility approach to the Munich chain-ladder method. Blätter DGVFM XXVII, 619-628.

[5] Merz, M., Wüthrich, M.V. (2010). Paid-incurred chain claims reserving method. Insurance: Mathematics Economics 46/3, 568-579.

[6] Quarg, G., Mack, T. (2004). Munich chain ladder. Blätter DGVFM XXVI, 597-630.

[7] Wüthrich, M.V., Merz, M. (2013). Financial Modeling, Actuarial Valuation and Solvency in Insurance. Springer.