

Article

Modified Munich Chain-Ladder Method

Michael Merz ^{1,*} and Mario V. Wüthrich ^{2,†}

¹ Faculty of Business Administration, University of Hamburg, 20146 Hamburg, Germany

² ETH Zurich, RiskLab, Department of Mathematics, 8092 Zurich, Switzerland;

E-Mail: mario.wuethrich@math.ethz.ch

[†] Swiss Finance Institute SFI Professor.

* Author to whom correspondence should be addressed; E-Mail: michael.merz@uni-hamburg.de;

Tel.: +49-40-42838-5401.

Academic Editor: Alexander Szimayer

Received: 30 September 2015 / Accepted: 1 December 2015 / Published: 21 December 2015

Abstract: The Munich chain-ladder method for claims reserving was introduced by Quarg and Mack 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 us 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; paid and incurred claims

1. Introduction

The Munich chain-ladder method was introduced by Quarg and Mack [1] on a pure axiomatic basis, and in 2003 it was awarded the Gauss prize by Deutsche Aktuarvereinigung (DAV) and Deutsche Gesellschaft für Versicherungs- und Finanzmathematik (DGVMF), see [1]. However, today it is still not known whether there is a non-trivial interesting stochastic model that fulfills these axioms, nor is much known about the prediction uncertainty in the Munich chain-ladder method. Liu and Verrall [2] propose to use bootstrap for the estimation of the prediction uncertainty in the Munich chain-ladder method, however this requires existence of a model that fulfills the Munich chain-ladder axioms. 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. This explicit stochastic model gives a rigorous mathematical foundation for the analysis of 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 [1]. The second way is to define a general chain-ladder model and to derive estimators in this model that have the Munich chain-ladder factor structure; this is the approach taken in [3]. Here, we analyze both of these views and we show how the second way leads to a modified Munich chain-ladder method. This analysis is done within the family of multivariate log-normal models. The first main result is that within this family of models, there is, in general, no interesting Munich chain-ladder model, see Theorem 2 below. The resulting Munich chain-ladder predictor always has an approximation error which is quantified in Theorem 3 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.

Organization of the Paper

In the next section, we consider stochastic models which simultaneously fulfill the chain-ladder assumptions for cumulative payments and incurred claims. In Theorem 1, we see that such models only permit 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 directly be compared to the Munich chain-ladder axioms which are introduced in Section 4. In Theorem 2, we find that, in general, the Munich chain-ladder axioms are not fulfilled in our modeling framework. This motivates 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 compared numerically in Section 7. The numerical study is based on the original data set of Quarg and Mack [1].

2. Chain-Ladder Models

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

$$\mathcal{B}_j^P = \{P_{i,k}; k \leq j, 0 \leq i \leq J\}, \quad \mathcal{B}_j^I = \{I_{i,k}; k \leq j, 0 \leq i \leq J\} \quad \text{and} \quad \mathcal{B}_j = \mathcal{B}_j^P \cup \mathcal{B}_j^I.$$

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

- (A1) We assume that the random vectors $(P_{i,0}, \dots, P_{i,J}, I_{i,0}, \dots, I_{i,J})$ have strictly positive components and are independent for different accident years $i = 0, \dots, J$.
- (A2) There exist parameters $f_j^P, f_j^I, \sigma_j^P, \sigma_j^I > 0$ such that for $0 \leq j \leq J - 1$ and $0 \leq i \leq J$

$$\begin{aligned} \mathbb{E} [P_{i,j+1} | \mathcal{B}_j^P] &= f_j^P P_{i,j} \quad \text{and} \quad \text{Var} (P_{i,j+1} | \mathcal{B}_j^P) = (\sigma_j^P)^2 P_{i,j}^2, \\ \mathbb{E} [I_{i,j+1} | \mathcal{B}_j^I] &= f_j^I I_{i,j} \quad \text{and} \quad \text{Var} (I_{i,j+1} | \mathcal{B}_j^I) = (\sigma_j^I)^2 I_{i,j}^2. \end{aligned}$$

These assumptions correspond to assumptions **PE, PV, IE, IV** and **PIU** in [1], except that we make a modification in the variance assumptions **PV** and **IV**. We make this change because it substantially

simplifies our considerations (we come back to this in Remark 1 below). Assumption 1 states that cumulative payments $(P_{i,j})_{i,j}$ and incurred claims $(I_{i,j})_{i,j}$ fulfill the 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 incurred claims. 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.

We choose a continuous and strictly increasing link function $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ with $\lim_{x \rightarrow 0} g(x) = -\infty$ and $\lim_{x \rightarrow \infty} g(x) = \infty$. The standard example is the log-link function given by

$$g(x) = \log x \quad \text{for } x > 0, \tag{1}$$

but the results derived in this section hold true for general such link functions. The log-link function has the advantage of closed form solutions. For (general) link function g (as introduced above), 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

$$\Xi_i = (\xi_{i,0}^P, \dots, \xi_{i,J}^P, \xi_{i,0}^I, \dots, \xi_{i,J}^I)'$$

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

- (B1) We assume that the random vectors Ξ_i are independent for different accident years $i = 0, \dots, J$.
- (B2) There exists a parameter vector $\theta = (\theta_0^P, \dots, \theta_J^P, \theta_0^I, \dots, \theta_J^I)' \in \mathbb{R}^{2(J+1)}$ and a positive definite covariance matrix $\Sigma \in \mathbb{R}^{2(J+1) \times 2(J+1)}$ such that we have for $0 \leq i \leq J$

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

For log-link Equation (1), we obtain the log-normal chain-ladder model and for a general link g a general link ratio model. We have the following identities for the generated σ -algebras

$$\sigma\{P_{i,k}; k \leq j, 0 \leq i \leq J\} = \sigma\{\xi_{i,k}^P; k \leq j, 0 \leq i \leq J\}.$$

Therefore, by an abuse of notation, we use \mathcal{B}_j^P for both sets of information, and analogously for \mathcal{B}_j^I and \mathcal{B}_j . From this, we immediately see that assumptions (A1) and (B1) agree with each other. Due to the independence of different accident years i we have for $* \in \{P, I\}$

$$\begin{aligned} \xi_{i,j+1}^* | \mathcal{B}_j^* &\stackrel{(d)}{=} \xi_{i,j+1}^* | \{\xi_{i,0}^*, \dots, \xi_{i,j}^*\}, \\ (\xi_{i,j+1}^P, \xi_{i,j+1}^I) | \mathcal{B}_j &\stackrel{(d)}{=} (\xi_{i,j+1}^P, \xi_{i,j+1}^I) | \{\xi_{i,0}^P, \dots, \xi_{i,j}^P, \xi_{i,0}^I, \dots, \xi_{i,j}^I\}. \end{aligned}$$

For $* \in \{P, I\}$ we denote by $\theta_{[j]}^* = (\theta_0^*, \dots, \theta_j^*)' \in \mathbb{R}^{j+1}$ and let $\Sigma_{[j]}^* \in \mathbb{R}^{(j+1) \times (j+1)}$ be the (positive definite) covariance matrix of the random vector $\xi_{i,[j]}^* = (\xi_{i,0}^*, \dots, \xi_{i,j}^*)'$. 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 1. Under Assumption 2 we have for $* \in \{P, I\}$, $0 \leq j \leq J - 1$ and $0 \leq i \leq J$

$$\xi_{i,j+1}^* | \mathcal{B}_j^* \sim \mathcal{N} \left(\theta_{j+1}^* + (\Sigma_{j,j+1}^*)' (\Sigma_{[j]}^*)^{-1} (\xi_{i,[j]}^* - \theta_{[j]}^*), (s_{j+1}^{*,\text{post}})^2 \right),$$

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

Proof. This is a standard result for multivariate Gaussian distributions, see Result 4.6 in [4]. □

Using Lemma 1, we can calculate the conditionally expected claims for given link function g . We have for $0 \leq j \leq J - 1$ and $0 \leq i \leq J$

$$\mathbb{E} [P_{i,j+1} | \mathcal{B}_j^P] = P_{i,j} \mathbb{E} \left[g^{-1} \left(g \left(\frac{P_{i,j+1}}{P_{i,j}} \right) \right) \middle| \mathcal{B}_j^P \right] = P_{i,j} \mathbb{E} [g^{-1} (\xi_{i,j+1}^P) | \mathcal{B}_j^P], \tag{2}$$

$$\mathbb{E} [I_{i,j+1} | \mathcal{B}_j^I] = I_{i,j} \mathbb{E} \left[g^{-1} \left(g \left(\frac{I_{i,j+1}}{I_{i,j}} \right) \right) \middle| \mathcal{B}_j^I \right] = I_{i,j} \mathbb{E} [g^{-1} (\xi_{i,j+1}^I) | \mathcal{B}_j^I]. \tag{3}$$

In a similar way, we obtain for the conditional variances

$$\text{Var} (P_{i,j+1} | \mathcal{B}_j^P) = P_{i,j}^2 \left(\mathbb{E} [g^{-1} ((\xi_{i,j+1}^P)^2) | \mathcal{B}_j^P] - \mathbb{E} [g^{-1} (\xi_{i,j+1}^P) | \mathcal{B}_j^P]^2 \right), \tag{4}$$

$$\text{Var} (I_{i,j+1} | \mathcal{B}_j^I) = I_{i,j}^2 \left(\mathbb{E} [g^{-1} ((\xi_{i,j+1}^I)^2) | \mathcal{B}_j^I] - \mathbb{E} [g^{-1} (\xi_{i,j+1}^I) | \mathcal{B}_j^I]^2 \right). \tag{5}$$

We have assumed that Σ is positive definite. This implies that also $(\Sigma_{[j]}^*)^{-1}$ is positive definite for $* \in \{P, I\}$. We then see from Lemma 1 that, in general, the last terms in Equations (2)–(5) depend on $\xi_{i,[j]}^P$ and $\xi_{i,[j]}^I$, respectively. Therefore, these last terms are not constant w.r.t. information \mathcal{B}_j^P and \mathcal{B}_j^I , respectively, 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 1. Assume that Assumption 2 is fulfilled for general link function g as introduced above. The model fulfills Assumption 1 if and only if

$$\Sigma_{[J]}^P = \text{diag} ((s_0^P)^2, \dots, (s_J^P)^2) \quad \text{and} \quad \Sigma_{[J]}^I = \text{diag} ((s_0^I)^2, \dots, (s_J^I)^2). \tag{6}$$

Under Equation (6), we have in the special case of the log-link $g(x) = \log x$ and for $0 \leq j \leq J - 1$ and $0 \leq i \leq J$

$$\begin{aligned} \mathbb{E} [P_{i,j+1} | \mathcal{B}_j^P] &= P_{i,j} \exp \{ \theta_{j+1}^P + (s_{j+1}^P)^2 / 2 \}, \\ \text{Var} (P_{i,j+1} | \mathcal{B}_j^P) &= P_{i,j}^2 \exp \{ 2\theta_{j+1}^P + (s_{j+1}^P)^2 \} (\exp \{ (s_{j+1}^P)^2 \} - 1). \end{aligned}$$

Analogous statements hold true for incurred claims $I_{i,j+1}$, conditioned on \mathcal{B}_j^I .

Remark 1. The previous theorem says that covariance structure Equation (6) is a necessary condition to obtain the chain-ladder model of Assumption 1. This holds for general link functions g , see Equations (2)–(5), and under Gaussian age-to-age ratios. The resulting variance properties differ from the classical ones of Quarg and Mack [1]. However, our argument does not use the variance assumption in a crucial way (it is already sufficient to consider Equations (2)–(3)), except that under Assumption 2 the analysis receives an analytically tractable closed form solution. Therefore, we expect this result to hold true in broader generality.

Under the assumptions of Theorem 1, the process $(P_{i,j})_{0 \leq j \leq J}$ has the Markov property, and we obtain the following chain-ladder parameters for the log-link $g(x) = \log x$

$$f_j^* = \exp \{ \theta_{j+1}^* + (s_{j+1}^*)^2 / 2 \} \quad \text{and} \quad (\sigma_j^*)^2 = (f_j^*)^2 (\exp \{ (s_{j+1}^*)^2 \} - 1), \tag{7}$$

with $* \in \{P, I\}$. Moreover, the covariance matrix Σ under Theorem 1 is given by

$$\Sigma = \begin{pmatrix} \Sigma_{[J]}^P = \text{diag}((s_0^P)^2, \dots, (s_J^P)^2) & A \\ A' & \Sigma_{[J]}^I = \text{diag}((s_0^I)^2, \dots, (s_J^I)^2) \end{pmatrix}, \tag{8}$$

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

Lemma 2. *A symmetric matrix Σ of the form Equation (8) is positive definite if and only if the matrix*

$$S_{[J]}^P = \Sigma_{[J]}^I - A' (\Sigma_{[J]}^I)^{-1} A \quad \text{is positive definite,}$$

or, equivalently, if and only if the matrix

$$S_{[J]}^I = \Sigma_{[J]}^P - A' (\Sigma_{[J]}^P)^{-1} A \quad \text{is positive definite.}$$

Proof. This lemma is a standard result in linear algebra about Schur complements, see Section C.4.1 in [5]. □

The matrices $S_{[J]}^*$ are called Schur complements of $\Sigma_{[J]}^*$ in Σ , 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 Σ 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 Ξ_i are independent for different accident years $i = 0, \dots, J$.
- (C2) There exists a parameter vector $\theta = (\theta_0^P, \dots, \theta_J^P, \theta_0^I, \dots, \theta_J^I)' \in \mathbb{R}^{2(J+1)}$ and a matrix Σ of the form Equation (8) with positive definite Schur complements $S_{[J]}^P$ and $S_{[J]}^I$ such that we have for $0 \leq i \leq J$

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

Corollary 1. *The model of Assumption 3 fulfills the distribution-free chain-ladder model of Assumption 1 for any link function g (as introduced above). The chain-ladder parameters are given by Equation (7) in the special case of the log-link function Equation (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 incurred claims. Note that an appropriate choice of matrix A in Equation (8) allows us for dependence modeling between cumulative payments and incurred claims, this will be crucial in the sequel.

3. One-Step Ahead Prediction

Formulas (2) and (3) and Theorem 1 provide the best prediction of $P_{i,j+1}$ based on \mathcal{B}_j^P and the best prediction of $I_{i,j+1}$ based on \mathcal{B}_j^I , respectively, under Assumption 3. The basic idea behind the Munich chain-ladder method is to consider best predictions based on both sets of information $\mathcal{B}_j = \mathcal{B}_j^P \cup \mathcal{B}_j^I$, that is, how does prediction of, say, cumulative payments $P_{i,j+1}$ improve by enlarging the information from \mathcal{B}_j^P to \mathcal{B}_j . This is similar to the considerations in [3]. 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 $\boldsymbol{\theta}_{[j]} = (\theta_0^P, \dots, \theta_j^P, \theta_0^I, \dots, \theta_j^I)' \in \mathbb{R}^{2(j+1)}$ and let $\Sigma_{[j]} \in \mathbb{R}^{2(j+1) \times 2(j+1)}$ be the (positive definite) covariance matrix of the random vector $\boldsymbol{\xi}_{i,[j]} = (\xi_{i,0}^P, \dots, \xi_{i,j}^P, \xi_{i,0}^I, \dots, \xi_{i,j}^I)'$. Moreover, let $\Sigma_{j,j+1}^{(*)} \in \mathbb{R}^{2(j+1)}$ denote the covariance vector between $\boldsymbol{\xi}_{i,[j]}$ and $\xi_{i,j+1}^*$ for $* \in \{P, I\}$. Note that in contrast to Lemma 1 we replace $\Sigma_{j,j+1}^*$ by $\Sigma_{j,j+1}^{(*)}$, i.e., we set the upper index in brackets.

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

$$\xi_{i,j+1}^* | \mathcal{B}_j \sim \mathcal{N} \left(\theta_{j+1}^* + (\Sigma_{j,j+1}^{(*)})' \Sigma_{[j]}^{-1} (\boldsymbol{\xi}_{i,[j]} - \boldsymbol{\theta}_{[j]}), (s_{j+1}^{(*),\text{post}})^2 \right),$$

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

Proof. This is a standard result for multivariate Gaussian distributions, see Result 4.6 in [4]. □

The previous lemma shows that the conditional expectation of $\xi_{i,j+1}^P$, given \mathcal{B}_j , is linear in the observations $\boldsymbol{\xi}_{i,[j]}$. This will be crucial. An easy consequence of the previous lemma is the following corollary for the special case of the log-link.

Corollary 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 $0 \leq i \leq J$

$$\begin{aligned} \mathbb{E} [P_{i,j+1} | \mathcal{B}_j] &= P_{i,j} \exp \left\{ \theta_{j+1}^P + (\Sigma_{j,j+1}^{(P)})' \Sigma_{[j]}^{-1} (\boldsymbol{\xi}_{i,[j]} - \boldsymbol{\theta}_{[j]}) + (s_{j+1}^{(P),\text{post}})^2 / 2 \right\} \\ &= f_j^P P_{i,j} \gamma_j^P(\boldsymbol{\xi}_{i,[j]}) = \mathbb{E} [P_{i,j+1} | \mathcal{B}_j^P] \gamma_j^P(\boldsymbol{\xi}_{i,[j]}), \end{aligned}$$

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

$$\begin{aligned} \gamma_j^*(\boldsymbol{\xi}_{i,[j]}) &= \exp \left\{ \beta_j^*(\boldsymbol{\xi}_{i,[j]}) - (\Sigma_{j,j+1}^{(*)})' \Sigma_{[j]}^{-1} \Sigma_{j,j+1}^{(*)} / 2 \right\}, \\ \beta_j^*(\boldsymbol{\xi}_{i,[j]}) &= (\Sigma_{j,j+1}^{(*)})' \Sigma_{[j]}^{-1} (\boldsymbol{\xi}_{i,[j]} - \boldsymbol{\theta}_{[j]}). \end{aligned}$$

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

The term $\gamma_j^P(\boldsymbol{\xi}_{i,[j]})$ 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

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

Example 1 (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 Σ under Assumption 3 is given by*

$$\Sigma = \Sigma_{[2]} = \left(\begin{array}{ccc|ccc} (s_0^P)^2 & 0 & 0 & a_{0,0} & a_{0,1} & a_{0,2} \\ 0 & (s_1^P)^2 & 0 & a_{1,0} & a_{1,1} & a_{1,2} \\ 0 & 0 & (s_2^P)^2 & a_{2,0} & a_{2,1} & a_{2,2} \\ \hline a_{0,0} & a_{1,0} & a_{2,0} & (s_0^I)^2 & 0 & 0 \\ a_{0,1} & a_{1,1} & a_{2,1} & 0 & (s_1^I)^2 & 0 \\ a_{0,2} & a_{1,2} & a_{2,2} & 0 & 0 & (s_2^I)^2 \end{array} \right).$$

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

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

Moreover, $\Sigma_{0,1}^{(P)} = (0, a_{1,0})'$. This provides credibility weight $(\alpha_{[0]}^P)' \in \mathbb{R}^2$ given by

$$\alpha_{[0]}^P = (\Sigma_{0,1}^{(P)})' \Sigma_{[0]}^{-1} = \frac{1}{(s_0^P s_0^I)^2 - a_{0,0}^2} \begin{pmatrix} -a_{1,0} a_{0,0} & a_{1,0} (s_0^P)^2 \end{pmatrix},$$

and posterior variance

$$(s_1^{(P),\text{post}})^2 = (s_1^P)^2 - (\Sigma_{0,1}^{(P)})' \Sigma_{[0]}^{-1} \Sigma_{0,1}^{(P)} = (s_1^P)^2 - \frac{(a_{1,0} s_0^P)^2}{(s_0^P s_0^I)^2 - a_{0,0}^2}.$$

Observe that $a_{1,0} = \text{Cov}(\xi_{i,1}^P, \xi_{i,0}^I)$ is the crucial term in the credibility weight $\alpha_{[0]}^P$. 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

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

with

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

Also remarkable is that observation $\xi_{i,0}^P$ is used to improve the 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 $\xi_{i,0}^P$ is used to adjust $\xi_{i,0}^I$.

- Case $j = 1$. This case is more involved. Set

$$\begin{aligned} b_{0,0} &= (s_0^I)^2 - a_{0,0}^2 / (s_0^P)^2 - a_{1,0}^2 / (s_1^P)^2, & b_{0,1} &= -a_{0,0} a_{0,1} / (s_0^P)^2 - a_{1,1} a_{1,0} / (s_1^P)^2, \\ b_{1,1} &= (s_1^I)^2 - a_{0,1}^2 / (s_0^P)^2 - a_{1,1}^2 / (s_1^P)^2, \\ c_{0,0} &= \frac{b_{0,1} a_{0,1} - b_{1,1} a_{0,0}}{(b_{0,0} b_{1,1} - b_{0,1}^2) (s_0^P)^2}, & c_{0,1} &= \frac{-b_{0,0} a_{0,1} + b_{0,1} a_{0,0}}{(b_{0,0} b_{1,1} - b_{0,1}^2) (s_0^P)^2}, \\ c_{1,0} &= \frac{b_{0,1} a_{1,1} - b_{1,1} a_{1,0}}{(b_{0,0} b_{1,1} - b_{0,1}^2) (s_1^P)^2}, & c_{1,1} &= \frac{-b_{0,0} a_{1,1} + b_{0,1} a_{1,0}}{(b_{0,0} b_{1,1} - b_{0,1}^2) (s_1^P)^2}. \end{aligned}$$

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

$$\Sigma_{[1]}^{-1} = \left(\begin{array}{cc|cc} * & * & c_{0,0} & c_{0,1} \\ * & * & c_{1,0} & c_{1,1} \\ \hline c_{0,0} & c_{1,0} & \frac{b_{1,1}}{b_{0,0}b_{1,1}-b_{0,1}^2} & \frac{-b_{0,1}}{b_{0,0}b_{1,1}-b_{0,1}^2} \\ c_{0,1} & c_{1,1} & \frac{-b_{0,1}}{b_{0,0}b_{1,1}-b_{0,1}^2} & \frac{b_{0,0}}{b_{0,0}b_{1,1}-b_{0,1}^2} \end{array} \right).$$

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

$$\alpha_{[1]}^P = \left(a_{2,0}c_{0,0} + a_{2,1}c_{0,1}, a_{2,0}c_{1,0} + a_{2,1}c_{1,1}, \frac{b_{1,1}a_{2,0} - b_{0,1}a_{2,1}}{b_{0,0}b_{1,1} - b_{0,1}^2}, \frac{-b_{0,1}a_{2,0} + b_{0,0}a_{2,1}}{b_{0,0}b_{1,1} - b_{0,1}^2} \right),$$

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 incurred claims observation is not helpful to improve the prediction of $\xi_{i,2}^P$. 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

$$\mathbb{E}[P_{i,2} | \mathcal{B}_1] = P_{i,1} \exp \left\{ \theta_2^P + \beta_1^P(\xi_{i,[1]}) + (s_2^{(P),\text{post}})^2/2 \right\} = f_1^P P_{i,1} \gamma_1^P(\xi_{i,[1]}).$$

with

$$\begin{aligned} \beta_1^P(\xi_{i,[1]}) &= (a_{2,0}c_{0,0} + a_{2,1}c_{0,1})(\xi_{i,0}^P - \theta_0^P) + (a_{2,0}c_{1,0} + a_{2,1}c_{1,1})(\xi_{i,1}^P - \theta_1^P) \\ &+ \frac{b_{1,1}a_{2,0} - b_{0,1}a_{2,1}}{b_{0,0}b_{1,1} - b_{0,1}^2}(\xi_{i,0}^I - \theta_0^I) + \frac{-b_{0,1}a_{2,0} + b_{0,0}a_{2,1}}{b_{0,0}b_{1,1} - b_{0,1}^2}(\xi_{i,1}^I - \theta_1^I). \end{aligned} \tag{9}$$

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 2, we have derived the best prediction under Assumption 3 for the log-link. 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 our model fulfills the chain-ladder model Assumption 1 for any link function g according to Corollary 1, it can also be considered as the best prediction for given information \mathcal{B}_j in the distribution-free chain-ladder model for other link function choices g . The Munich chain-ladder method tackles the problem from a different viewpoint in that it extends the distribution-free chain-ladder model Assumption 1, so that it enforces the best prediction to have a pre-specified form. We 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

$$\varepsilon_{i,j}^{I|P} = \frac{I_{i,j} - \mathbb{E}[I_{i,j} | \mathcal{B}_j^P]}{\text{Var}(I_{i,j} | \mathcal{B}_j^P)^{1/2}} \quad \text{and} \quad \varepsilon_{i,j}^{P|I} = \frac{P_{i,j} - \mathbb{E}[P_{i,j} | \mathcal{B}_j^I]}{\text{Var}(P_{i,j} | \mathcal{B}_j^I)^{1/2}}.$$

The adapted Munich chain-ladder assumptions of Quarg and Mack [1] 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 $0 \leq i \leq J$

$$\mathbb{E} [P_{i,j+1} | \mathcal{B}_j] = f_j^P P_{i,j} + \lambda^P \text{Var} (P_{i,j+1} | \mathcal{B}_j)^{1/2} \varepsilon_{i,j}^{I|P},$$

and

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

Remark 2. The idea behind these additional assumptions is that one corrects for high and low incurred-paid and paid-incurred ratios via the residuals $\varepsilon_{i,j}^{I|P}$ and $\varepsilon_{i,j}^{P|I}$ because, for instance for cumulative payments, we have

$$\varepsilon_{i,j}^{I|P} = \frac{I_{i,j} - \mathbb{E} [I_{i,j} | \mathcal{B}_j^P]}{\text{Var} (I_{i,j} | \mathcal{B}_j^P)^{1/2}} = \frac{\frac{I_{i,j}}{P_{i,j}} - \mathbb{E} \left[\frac{I_{i,j}}{P_{i,j}} \middle| \mathcal{B}_j^P \right]}{\text{Var} \left(\frac{I_{i,j}}{P_{i,j}} \middle| \mathcal{B}_j^P \right)^{1/2}} = \frac{Q_{i,j}^{-1} - \mathbb{E} [Q_{i,j}^{-1} | \mathcal{B}_j^P]}{\text{Var} (Q_{i,j}^{-1} | \mathcal{B}_j^P)^{1/2}},$$

with incurred-paid ratio $Q_{i,j}^{-1} = I_{i,j}/P_{i,j}$. Therefore, the additional assumptions in Assumption 4 exactly provide **PQ** and **IQ** of Quarg and Mack [1]. If we choose the log-link for Assumption 3 then the incurred-paid ratio $Q_{i,j}^{-1}$ is turned into a difference on the log scale, that is, $\log(Q_{i,j}^{-1}) = \log I_{i,j} - \log P_{i,j} = \sum_{l=0}^j \xi_{i,l}^I - \sum_{l=0}^j \xi_{i,l}^P$. The aim of this section is to analyze under which circumstances these Munich chain-ladder corrections lead to the optimal predictors provided in Corollary 2. Below we will see that the constants λ^P and λ^I are crucial, they measure the (positive) correlation between the cumulative payments and the incurred-paid ratio correction (and similarly for incurred claims), see also Section 2.2.2 in [1]. Moreover, λ^P and λ^I receive an explicit meaning in Theorem 3, below.

The tower property of conditional expectations $\mathbb{E}[P_{i,j+1} | \mathcal{B}_j^P] = \mathbb{E}[\mathbb{E}[P_{i,j+1} | \mathcal{B}_j] | \mathcal{B}_j^P]$ implies under Assumption 4

$$\mathbb{E} [P_{i,j+1} | \mathcal{B}_j^P] = f_j^P P_{i,j} + \lambda^P \text{Var} (P_{i,j+1} | \mathcal{B}_j^P)^{1/2} \mathbb{E} \left[\varepsilon_{i,j}^{I|P} \middle| \mathcal{B}_j^P \right] = f_j^P P_{i,j}.$$

Therefore, Assumption 4 does not contradict Assumption 1. As mentioned in Remark 2, 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

$$\lambda^P \text{Var} (P_{i,j+1} | \mathcal{B}_j^P)^{1/2} \varepsilon_{i,j}^{I|P} = \lambda^P \sigma_j^P P_{i,j} \varepsilon_{i,j}^{I|P}, \tag{10}$$

and compare it to the optimal correction term obtained from Lemma 3 and Corollary 2, respectively. We start with log-link $g(x) = \log x$ and then provide the general result in Theorem 2, below. For the log-link we have representation of incurred claims

$$I_{i,j} = \nu_i \exp \left\{ \sum_{l=0}^j \xi_{i,l}^I \right\}. \tag{11}$$

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

Lemma 4. Under Assumption 3, we have

$$\sum_{l=0}^j \xi_{i,l}^I | \mathcal{B}_j^P \sim \mathcal{N} \left(\sum_{l=0}^j \theta_l^I + (\mathbf{a}_{0:j}^I)' (\Sigma_{[j]}^P)^{-1} (\boldsymbol{\xi}_{i,[j]}^P - \boldsymbol{\theta}_{[j]}^P), (s_{0:j}^{I,\text{post}})^2 \right),$$

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

Proof. This is a standard result for multivariate Gaussian distributions, see Result 4.6 in [4]. □

Example 2 (log-link). We consider log-link $g(x) = \log x$. In this case, we have from Equation (11) and using Lemma 4 for the residual of the correction term

$$\begin{aligned} \varepsilon_{i,j}^{I|P} &= \frac{\exp\{\sum_{l=0}^j \xi_{i,l}^I\} - \mathbb{E} \left[\exp\{\sum_{l=0}^j \xi_{i,l}^I\} \middle| \mathcal{B}_j^P \right]}{\text{Var} \left(\exp\{\sum_{l=0}^j \xi_{i,l}^I\} \middle| \mathcal{B}_j^P \right)^{1/2}} \\ &= \frac{\exp \left\{ \sum_{l=0}^j (\xi_{i,l}^I - \theta_l^I) - (\mathbf{a}_{0:j}^I)' (\Sigma_{[j]}^P)^{-1} (\boldsymbol{\xi}_{i,[j]}^P - \boldsymbol{\theta}_{[j]}^P) - (s_{0:j}^{I,\text{post}})^2 / 2 \right\} - 1}{\left(\exp \left\{ (s_{0:j}^{I,\text{post}})^2 \right\} - 1 \right)^{1/2}}. \end{aligned}$$

This implies for the Munich chain-ladder model Assumption 4, we also use Equation (7),

$$f_j^P P_{i,j} + \lambda^P \text{Var} (P_{i,j+1} | \mathcal{B}_j^P)^{1/2} \varepsilon_{i,j}^{I|P} = f_j^P P_{i,j} + \lambda^P \sigma_j^P P_{i,j} \varepsilon_{i,j}^{I|P} = f_j^P P_{i,j} \gamma_j^{P,\text{MCL}}(\boldsymbol{\xi}_{i,[j]}),$$

with Munich chain-ladder correction factor defined by

$$\gamma_j^{P,\text{MCL}}(\boldsymbol{\xi}_{i,[j]}) = 1 + \lambda^P \sqrt{\frac{e^{(s_j^P)^2} - 1}{e^{(s_{0:j}^{I,\text{post}})^2} - 1}} \left(e^{\sum_{l=0}^j (\xi_{i,l}^I - \theta_l^I) - (\mathbf{a}_{0:j}^I)' (\Sigma_{[j]}^P)^{-1} (\boldsymbol{\xi}_{i,[j]}^P - \boldsymbol{\theta}_{[j]}^P) - \frac{(s_{0:j}^{I,\text{post}})^2}{2}} - 1 \right). \tag{12}$$

We analyze this Munich chain-ladder correction factor for $j = 1$. It is given by

$$\begin{aligned} \gamma_1^{P,\text{MCL}}(\boldsymbol{\xi}_{i,[1]}) &= 1 + \lambda^P \sqrt{\frac{e^{(s_2^P)^2} - 1}{e^{(s_{0:1}^{I,\text{post}})^2} - 1}} \\ &\quad \times \left(e^{(\xi_{i,0}^I - \theta_0^I) + (\xi_{i,1}^I - \theta_1^I) - \frac{a_{0,0} + a_{0,1}}{(s_0^P)^2} (\xi_{i,0}^P - \theta_0^P) - \frac{a_{1,0} + a_{1,1}}{(s_1^P)^2} (\xi_{i,1}^P - \theta_1^P) - \frac{(s_{0:1}^{I,\text{post}})^2}{2}} - 1 \right). \end{aligned} \tag{13}$$

We compare this to the best prediction under Assumption 3 in the case $j = 1$ characterized by Equation (9) and under the additional assumptions that $a_{2,0} = 0$ and $a_{2,1} \neq 0$. In this case we obtain from Equations (7) and (9) correction term

$$\begin{aligned} \gamma_1^P(\boldsymbol{\xi}_{i,[1]}) &= \exp \left(\beta_1^P(\boldsymbol{\xi}_{i,[1]}) - \frac{b_{0,0} a_{2,1}^2}{2(b_{0,0} b_{1,1} - b_{0,1}^2)} \right) \\ &= \exp \left\{ \frac{a_{2,1}}{b_{0,0} b_{1,1} - b_{0,1}^2} [-b_{0,1} (\xi_{i,0}^I - \theta_0^I) + b_{0,0} (\xi_{i,1}^I - \theta_1^I)] \right. \\ &\quad \left. + a_{2,1} [c_{0,1} (\xi_{i,0}^P - \theta_0^P) + c_{1,1} (\xi_{i,1}^P - \theta_1^P)] - \frac{b_{0,0} a_{2,1}^2}{2(b_{0,0} b_{1,1} - b_{0,1}^2)} \right\}. \end{aligned} \tag{14}$$

Note that Equations (13) and (14) differ. This can, for instance, be seen because all terms in the sum $\sum_{l=0}^j (\xi_{i,l}^I - \theta_l^I)$ in $\gamma_1^{P, \text{MCL}}(\xi_{i,[1]})$ are equally weighted, whereas for the best predictor we consider a weighted sum $-b_{0,1}(\xi_{i,0}^I - \theta_0^I) + b_{0,0}(\xi_{i,1}^I - \theta_1^I)$ in $\gamma_1^P(\xi_{i,[1]})$. We conclude that, in general, Assumption 3 does not imply that the Munich chain-ladder model Assumption 4 is fulfilled.

The (disappointing) conclusion from Example 2 is that within the family of models fulfilling Assumption 3 with log-link $g(x) = \log x$ there does not exist (a general) interesting example satisfying the Munich chain-ladder model Assumption 4. Exceptions can only be found for rather artificial covariance matrices Σ , for instance, a choice with $A = 0$ would fulfill the Munich chain-ladder model Assumption 4. But this latter choice is not of interest because it requires $\lambda^P = \lambda^I = 0$ (which does not support the empirical findings of [1] that these correlation parameters should be positive). The result of Example 2 can be generalized to any link function as the next theorem shows.

Theorem 2. Assume that cumulative payments $P_{i,j}$ and incurred claims $I_{i,j}$ fulfill Assumption 3 for a given continuous and strictly increasing link function $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ with $\lim_{x \rightarrow 0} g(x) = -\infty$ and $\lim_{x \rightarrow \infty} g(x) = \infty$. In general, this model does not fulfill the Munich chain-ladder model Assumption 4, except for special choices of Σ .

Proof. The optimal one-step ahead prediction for given link function g is given by, see also Lemma 3,

$$\mathbb{E}[P_{i,j+1} | \mathcal{B}_j] = P_{i,j} \mathbb{E}[g^{-1}(\xi_{i,j+1}^P) | \mathcal{B}_j],$$

with

$$\xi_{i,j+1}^P | \mathcal{B}_j \sim \mathcal{N}\left(\theta_{j+1}^P + (\Sigma_{j,j+1}^{(P)})' \Sigma_{[j]}^{-1} (\xi_{i,[j]} - \theta_{[j]}), (s_{j+1}^{(P), \text{post}})^2\right).$$

From the latter, we observe that observation $\xi_{i,[j]}^I$ is considered in a linear fashion $c' \xi_{i,[j]}^I$ 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 $(1, \dots, 1)' \in \mathbb{R}^{j+1}$, i.e., we consider a weighted sum of the components of $\xi_{i,[j]}^I$ (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 Equation (10),

$$\begin{aligned} \lambda^P \sigma_j^P P_{i,j} \varepsilon_{i,j}^{I|P} &= \lambda^P \sigma_j^P P_{i,j} \frac{g^{-1}(\nu_i) \prod_{l=0}^j g^{-1}(\xi_{i,l}^I) - \mathbb{E}[I_{i,j} | \mathcal{B}_j^P]}{\text{Var}(I_{i,j} | \mathcal{B}_j^P)^{1/2}} \\ &= \lambda^P \sigma_j^P P_{i,j} \frac{g^{-1}(\nu_i) \exp\left\{\sum_{l=0}^j \log(g^{-1}(\xi_{i,l}^I))\right\} - \mathbb{E}[I_{i,j} | \mathcal{B}_j^P]}{\text{Var}(I_{i,j} | \mathcal{B}_j^P)^{1/2}}. \end{aligned}$$

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

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

From this we see that all components of $\xi_{i,[j]}^I$ 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 2 and proves the theorem. \square

In Theorem 4.1 of [3], the Munich chain-ladder structure has been found as a best linear approximation to $\mathbb{E} [P_{i,j+1} | \mathcal{B}_j]$ in the following way

$$\begin{aligned} \widehat{\mathbb{E}}^{\text{linear}} [P_{i,j+1} | \mathcal{B}_j] &= \underset{X=c_1 P_{i,j} + c_2 I_{i,j}; c_1, c_2 \in \mathcal{L}(\mathcal{B}_j^P)}{\text{argmin}} \mathbb{E} [(X - P_{i,j+1})^2 | \mathcal{B}_j^P] \\ &= f_j^P P_{i,j} + \text{Corr} (P_{i,j+1}, I_{i,j} | \mathcal{B}_j^P) \text{Var} (P_{i,j+1} | \mathcal{B}_j^P)^{1/2} \varepsilon_{i,j}^{I|P}, \end{aligned} \tag{15}$$

where $\mathcal{L}(\mathcal{B}_j^P)$ is the space of \mathcal{B}_j^P -measurable random variables. Note that this approximates the exact conditional expectation $\mathbb{E} [P_{i,j+1} | \mathcal{B}_j]$ 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 [1].

Theorem 3 (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 $\widehat{\mathbb{E}}^{\text{linear}} [P_{i,j+1} | \mathcal{B}_j]$ given by the difference*

$$\widehat{\mathbb{E}}^{\text{linear}} [P_{i,j+1} | \mathcal{B}_j] - \mathbb{E} [P_{i,j+1} | \mathcal{B}_j] = f_j^P P_{i,j} \left(\gamma_j^{P, \text{MCL}}(\boldsymbol{\xi}_{i,[j]}) - \gamma_j^P(\boldsymbol{\xi}_{i,[j]}) \right),$$

where $\gamma_j^{P, \text{MCL}}(\boldsymbol{\xi}_{i,[j]})$ is given in Equation (12) with λ^P replaced by $\text{Corr}(P_{i,j+1}, I_{i,j} | \mathcal{B}_j^P)$ and $\gamma_j^P(\boldsymbol{\xi}_{i,[j]})$ is given in Corollary 2.

Proof. This proof follows from Example 2. □

Remark 3. *In Theorem 2, we have seen that, in general, the Munich chain-ladder model Assumption 4 is not fulfilled for chain-ladder models satisfying Assumption 3. If, nevertheless, we would like to use an estimator that has Munich chain-ladder structure, we should use it in the sense of best-linear approximation Equation (15) to the best prediction $\mathbb{E} [P_{i,j+1} | \mathcal{B}_j]$. Theorem 3 gives the approximation error of this approach for the log-link choice.*

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 the second part of Theorem 1 and the one-step ahead prediction given in Corollary 2. The issues that we still need to consider are the following: (i) We would like to extend the one-step ahead predictions to get the predictions of $P_{i,J}$ and $I_{i,J}$, i.e., the final values of each accident year $i = 1, \dots, J$; (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.

We remark that we consider tasks (ii) and (iii) in a Bayesian framework which turns out to be rather straightforward. Alternatively, one could also consider these questions from a frequentist’s viewpoint. In this case, (ii) is solved by maximum likelihood estimation and (iii) can be assessed either with bootstrap methods or by (asymptotic) results for maximum likelihood estimates. Our experience is that in many cases these different assessments lead to rather similar values if one uses non-informative priors in the Bayesian approach.

Assumption 5 ((Bayesian) modified Munich chain-ladder model). Choose log-link $g(x) = \log x$ and assume the following: There is given a fixed covariance matrix Σ of the form Equation (8) having positive definite Schur complements $S_{[J]}^P$ and $S_{[J]}^I$.

- Conditionally, given parameter vector $\Theta = (\Theta_0^P, \dots, \Theta_J^P, \Theta_0^I, \dots, \Theta_J^I)'$, the random vectors Ξ_i are independent for different accident years $i = 0, \dots, J$ with

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

- The parameter vector Θ has prior distribution

$$\Theta \sim \mathcal{N}(\theta, T),$$

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 $i = 0, \dots, J$ to one random vector

$$\Xi = (\Xi'_0, \dots, \Xi'_J)',$$

which has conditional distribution

$$\Xi | \Theta \sim \mathcal{N}(B\Theta, \Sigma^+),$$

for an appropriate matrix $B \in \mathbb{R}^{2(J+1)^2 \times 2(J+1)}$ and covariance matrix $\Sigma^+ = \text{diag}(\Sigma, \dots, \Sigma)$. The following lemma is crucial, we refer to Corollary 4.3 in [6].

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

$$\zeta = \begin{pmatrix} \Xi \\ \Theta \end{pmatrix} \sim \mathcal{N} \left(\mu = \begin{pmatrix} B\theta \\ \theta \end{pmatrix}, S = \begin{pmatrix} \Sigma^+ + BTB' & BT \\ TB' & T \end{pmatrix} \right).$$

An easy consequence of Lemma 5 is the following marginal distribution

$$\Xi \sim \mathcal{N}(B\theta, \Sigma^+ + BTB').$$

This shows that, in the Bayesian multivariate normal model with Gaussian priors, we can completely “integrate out” the hierarchy of parameters Θ . However, we keep the hierarchy of parameters in order to obtain Bayesian parameter estimates for Θ .

Denote the dimension of ζ 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 ζ

$$\zeta \mapsto (\zeta_t, \zeta_v) = (P_t \zeta, P_v \zeta). \tag{16}$$

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

$$S_t = \text{Cov}(\zeta_t) = P_t S P_t', \quad S_v = \text{Cov}(\zeta_v) = P_v S P_v', \quad S'_{v,t} = S_{t,v} = \text{Cov}(\zeta_t, \zeta_v) = P_t S P_v'.$$

The projections in Equation (16) only describe a permutation of the components of ζ . In complete analogy to Lemma 1 we have the following lemma.

Lemma 6. Under Assumption 5, the random vector $\zeta_v |_{\{\zeta_t\}}$ has a multivariate Gaussian distribution with the first two conditional moments given by

$$\begin{aligned} \mu_v^{\text{post}} &= \mathbb{E}[\zeta_v | \zeta_t] = \mu_v + S_{v,t} (S_t)^{-1} (\zeta_t - \mu_t), \\ S_v^{\text{post}} &= \text{Cov}(\zeta_v | \zeta_t) = S_v - S_{v,t} (S_t)^{-1} S_{t,v}. \end{aligned}$$

This lemma allows us to estimate the parameters and calculate the predictions at time J , conditionally given observations

$$\begin{aligned} \mathcal{D}_J^P &= \{P_{i,j}; 0 \leq i \leq J, 0 \leq j \leq J; i + j \leq J\}, \\ \mathcal{D}_J^I &= \{I_{i,j}; 0 \leq i \leq J, 0 \leq j \leq J; i + j \leq J\}, \\ \mathcal{D}_J &= \mathcal{D}_J^P \cup \mathcal{D}_J^I. \end{aligned}$$

Choose $t = |\mathcal{D}_J|$ and $v = n - t$ and denote by P_t the projection of ζ 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 6 allows us to calculate the posterior distribution of ζ_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

$$P_\Theta \in \mathbb{R}^{2(J+1) \times v} \text{ with } P_\Theta \zeta_v = \Theta.$$

This projection extracts the parameter vector Θ from the unobserved components ζ_v .

Corollary 3 (parameter estimation). Under Assumption 5, the Bayesian estimator for the parameter vector Θ is at time J given by

$$\theta^{\text{post}} = \mathbb{E}[\Theta | \mathcal{D}_J] = P_\Theta \mu_v^{\text{post}}.$$

This can now be compared to the individual estimates

$$\theta^{(*),\text{post}} = \mathbb{E}[\Theta | \mathcal{D}_J^*], \tag{17}$$

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 total claim amount of accident year i , we have two different possibilities, either we use the predictor of cumulative payments $P_{i,J}$ or the one of incurred claims $I_{i,J}$. Naturally, these two predictors differ and the Munich chain-ladder method exactly aims at diminishing this difference by including the incurred-paid and paid-incurred ratios, see Remark 2 and [1]. Choose the log-link $g(x) = \log x$, then we calculate for $i = 1, \dots, J$ the best predictors

$$\mathbb{E}[P_{i,J} | \mathcal{D}_J] = P_{i,J-i} \mathbb{E} \left[\exp \left\{ \sum_{l=J-i+1}^J \xi_{i,l}^P \right\} \middle| \mathcal{D}_J \right],$$

and

$$\mathbb{E}[I_{i,J} | \mathcal{D}_J] = I_{i,J-i} \mathbb{E} \left[\exp \left\{ \sum_{l=J-i+1}^J \xi_{i,l}^I \right\} \middle| \mathcal{D}_J \right].$$

Assume again that ζ_t exactly corresponds to the observations in \mathcal{D}_J . Then we define for $i = 1, \dots, J$ and $*$ $\in \{P, I\}$ the linear maps

$$G_i^* \in \mathbb{R}^{1 \times v} \quad \text{with} \quad G_i^* \zeta_v = \sum_{l=J-i+1}^J \zeta_{i,l}^*.$$

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

Theorem 4 (modified Munich chain-ladder (mMCL) predictors). *Under Assumption 5, the Bayesian predictors for the total claim amount of accident year $i = 1, \dots, J$ at time J are*

$$\widehat{P}_{i,J}^{\text{mMCL}} = \mathbb{E} [P_{i,J} | \mathcal{D}_J] = P_{i,J-i} \exp \left\{ G_i^P \mu_v^{\text{post}} + G_i^P S_v^{\text{post}} (G_i^P)' / 2 \right\},$$

and

$$\widehat{I}_{i,J}^{\text{mMCL}} = \mathbb{E} [I_{i,J} | \mathcal{D}_J] = I_{i,J-i} \exp \left\{ G_i^I \mu_v^{\text{post}} + G_i^I S_v^{\text{post}} (G_i^I)' / 2 \right\}.$$

The conditional mean-square error of prediction is given by

$$\begin{aligned} \text{mse}_{\sum_{i=1}^J P_{i,J} | \mathcal{D}_J} \left(\sum_{i=1}^J \widehat{P}_{i,J}^{\text{mMCL}} \right) &= \text{Var} \left(\sum_{i=1}^J P_{i,J} \middle| \mathcal{D}_J \right) \\ &= \sum_{i,k=1}^J \widehat{P}_{i,J}^{\text{mMCL}} \widehat{P}_{k,J}^{\text{mMCL}} \left(\exp \left\{ G_i^P S_v^{\text{post}} (G_k^P)' \right\} - 1 \right), \end{aligned}$$

and analogously for incurred claims $\text{mse}_{\sum_{i=1}^J I_{i,J} | \mathcal{D}_J} \left(\sum_{i=1}^J \widehat{I}_{i,J}^{\text{mMCL}} \right)$.

This can now again be compared to the individual predictors

$$\widehat{P}_{i,J}^{\text{HCL}} = \mathbb{E} [P_{i,J} | \mathcal{D}_J^P] \quad \text{and} \quad \widehat{I}_{i,J}^{\text{HCL}} = \mathbb{E} [I_{i,J} | \mathcal{D}_J^I], \tag{18}$$

and the corresponding conditional mean-square errors of prediction. Note that these individual predictors correspond to the predictors in the model of Hertig [7] under Gaussian prior assumptions for the (unknown) mean parameters. Predictors and prediction uncertainty of Equation (18) can (easily) be obtained from Theorem 4 using the particular choice $A = 0$ in Σ .

Before we give a numerical example, we briefly describe these predictors. The likelihood function of Assumption 5 is given by

$$\begin{aligned} \mathcal{L}(\Xi, \Theta) &= \frac{1}{(2\pi)^{2(J+1)^2/2} \det(\Sigma^+)^{1/2}} \exp \left\{ -\frac{1}{2} (\Xi - \mathbf{B}\Theta)' (\Sigma^+)^{-1} (\Xi - \mathbf{B}\Theta) \right\} \\ &\times \frac{1}{(2\pi)^{2(J+1)/2} \det(T)^{1/2}} \exp \left\{ -\frac{1}{2} (\Theta - \theta)' T^{-1} (\Theta - \theta) \right\}. \end{aligned}$$

Under the additional assumption of diagonal matrices

$$\begin{aligned} \Sigma &= \text{diag} \left((s_0^P)^2, \dots, (s_{J+1}^P)^2, (s_0^I)^2, \dots, (s_{J+1}^I)^2 \right), \\ T &= \text{diag} \left((t_0^P)^2, \dots, (t_{J+1}^P)^2, (t_0^I)^2, \dots, (t_{J+1}^I)^2 \right), \end{aligned} \tag{19}$$

we obtain log-likelihood (we drop all normalizing constants)

$$\log \mathcal{L}(\Xi, \Theta) \propto -\frac{1}{2} \sum_{j=0}^J \left[\sum_{i=0}^J \frac{(\xi_{i,j}^P - \Theta_j^P)^2}{(s_j^P)^2} + \frac{(\Theta_j^P - \theta_j^P)^2}{(t_j^P)^2} + \sum_{i=0}^J \frac{(\xi_{i,j}^I - \Theta_j^I)^2}{(s_j^I)^2} + \frac{(\Theta_j^I - \theta_j^I)^2}{(t_j^I)^2} \right].$$

From this, we see that the Bayesian estimators of the parameters are for $j = 0, \dots, J$ and $* \in \{P, I\}$ under Equation (19) given by, see also Corollary 3,

$$\mathbb{E} [\Theta_j^* | \mathcal{D}_J] = z_j^* \widehat{\Theta}_j^* + (1 - z_j^*) \theta_j^*,$$

with prior mean θ_j^* , and empirical mean $\widehat{\Theta}_j^*$ and credibility weight z_j^* given by

$$\widehat{\Theta}_j^* = \frac{1}{I - j + 1} \sum_{i=0}^{J-i} \xi_{i,j}^* \quad \text{and} \quad z_j^* = \frac{J - i + 1}{J - i + 1 + (\sigma_j^*/t_j^*)^2}.$$

If we now let the prior information become non-informative, *i.e.*, $t_j^* \rightarrow \infty$, we obtain estimate

$$\lim_{t_j^* \rightarrow \infty} \mathbb{E} [\Theta_j^* | \mathcal{D}_J] = \widehat{\Theta}_j^*, \tag{20}$$

and posterior variances $(\sigma_j^*)^2 / (J - i + 1)$. In view of Theorem 4, this provides under Equation (19) and in the non-informative prior limit

$$\lim_{t_j^P \rightarrow \infty} \widehat{P}_{i,J}^{\text{mMCL}} = P_{i,J-i} \prod_{j=J-i+1}^J \exp \left\{ \widehat{\Theta}_j^P + \frac{(\sigma_j^P)^2}{2} \left(1 + \frac{1}{J - i + 1} \right) \right\} = P_{i,J-i} \prod_{j=J-i}^{J-1} \widehat{f}_j^P, \tag{21}$$

where the latter identity defines the chain-ladder parameter estimates \widehat{f}_j^P for our model. This is exactly the chain-ladder predictor obtained in Hertig’s log-normal chain-ladder model, see formula (5.9) in [7]. The corresponding result also holds true for incurred claims under Equation (19).

As was investigated by Quarg and Mack [1], see also Remark 2 above and Figure 1 below, we expect positive dependence between cumulative payment residuals and incurred-paid ratios (and between incurred claims residuals and paid-incurred ratios). This will be reflected by a covariance matrix choice Σ that does not have diagonal form Equation (19) but a general off-diagonal matrix A in Equation (8) such that the Schur complements are positive definite (see Assumption 5). In this case, the best predictors are provided by Theorem 4. They do not have a simple form (though their calculation is straightforward using matrix algebra). We will compare these predictors to the Munich chain-ladder predictors Equation (15) which are non-optimal in our context (see Theorem 3).

7. Example

We provide an explicit example which is based on the original data of Quarg and Mack [1], the data is provided in the appendix. We calculate for this data set Hertig’s chain-ladder (HCL) reserves according to Equations (18) and (21), the reserves in the modified Munich chain-ladder (mMCL) method of Theorem 4 and the (non-optimal) log-normal Munich chain-ladder (LN-MCL) reserves Equation (15) (according to Assumption 4). These reserves are based on the Bayesian multivariate log-normal framework of Assumption 5 with log-link $g(x) = \log x$. For comparison purposes, we also

provide the classical chain-ladder (CL) reserves together with the Quarg and Mack Munich chain-ladder reserves (QM–MCL); these two latter methods differ from our results because of the different variance assumption in Assumption 1. In order to have comparability between the different approaches, we choose non-informative priors $t_j^P, t_j^I \rightarrow \infty$ in the former Bayesian methods, see also Equations (20) and (19).

First, we need to estimate the parameters in the log-normal model of Assumption 5. For s_j^P and s_j^I , we choose the sample standard deviations of the observed log-link ratios $\xi_{i,j}^P$ and $\xi_{i,j}^I, i + j \leq J$, with the usual exponential extrapolation for the last period $j = 6$. Using these sample estimators, we calculate the posterior means $\theta_j^{(P),\text{post}}$ and $\theta_j^{(I),\text{post}}$ using Corollary 3 under choice $A = 0$. In the non-informative prior limit, these posterior means are given by Equation (21) (and similarly for incurred claims). This then allows one to calculate Hertig’s chain-ladder parameters \hat{f}_j^P and \hat{f}_j^I , see Equation (21). These parameters are provided in Table 1. Note that these chain-ladder factors differ from the ones in the classical chain-ladder model because of the different variance assumptions.

Table 1. Sample standard deviations s_j^P and s_j^I ; posterior means $\theta_j^{(P),\text{post}}$ and $\theta_j^{(I),\text{post}}$ obtained from Corollary 3, see also Equation (21); and Hertig’s chain-ladder estimates \hat{f}_j^P and \hat{f}_j^I according to Equation (21).

a.y. <i>i/d.y. j</i>	0	1	2	3	4	5	6
$\theta_j^{(P),\text{post}}$	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
\hat{f}_j^P	1,573	2.5376	1.1296	1.0301	1.0219	1.0208	1.0138
$\theta_j^{(I),\text{post}}$	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
\hat{f}_j^I	2,963	1.6959	1.0148	1.0004	1.0116	0.9912	0.9963

Using these parameters, we calculate the HCL reserves (prediction minus the last observed cumulative payments $P_{i,J-i}$ at time J), and for comparison purposes we provide the classical CL reserves. These results are provided in Table 2. The main observation is that there are quite substantial differences between the HCL reserves from cumulative payments of 6,205 and the HCL reserves from incurred claims of 7,730, see Table 2. This also holds true for the classical CL reserves 5,938 versus 7,503. This gap mainly comes from the last accident year $i = 6$ because incurred claims observation $I_{6,0}$ is comparably high. We also note that the HCL reserves are more conservative than the classical CL ones. This mainly comes from the variance correction that enters the mean of log-normal random variables, see Equation (21).

To bridge this gap between the cumulative payments and the incurred claims methods we study the other reserving methods. We start with the LN–MCL method under the log-normal assumptions of Assumption 4. First we determine the correlation parameters. We use the estimators of Section 3.1.2 in [1] with changed variance functions. This provides estimates $\hat{\lambda}^P = 49\%$ and $\hat{\lambda}^I = 45\%$. Note that this exactly corresponds to the positive linear dependence illustrated in Figure 1; Quarg and Mack [1] obtain under their (changed) variance assumption 64% and 44%, respectively, which is in line with our findings. Using these estimates we can then calculate the reserves in our LN–MCL method and in

Quarg-Mack’s QM–MCL method. The results are provided in Table 2. We observe that the gap between the cumulative payments reserves of 6,729 and the incurred claims reserves of 7,140 becomes more narrow due to the correction factors. The same holds true for QM–MCL with reserves 6,847 and 7,120, respectively. Moreover, both models LN–MCL and QM–MCL provide rather close results, though their model assumptions differ in the variance assumption.

Table 2. Resulting reserves from the Hertig’s chain-ladder (HCL) method based on paid and incurred; from the log-normal Munich chain-ladder (LN–MCL) method based on paid and incurred; from the modified Munich chain-ladder (mMCL) paid method; the classical chain-ladder (CL) method based on paid and incurred (inc.); and the Quarg and Mack Munich chain-ladder (QM–MCL) method paid and incurred.

a.y. <i>i</i>	HCL		LN-MCL		mMCL	CL		QM-MCL	
	paid	inc.	paid	inc.	paid	paid	inc.	paid	inc.
1	32	97	35	95	16	32	97	35	96
2	157	92	92	147	115	158	88	103	135
3	337	286	262	346	375	332	276	269	326
4	416	201	289	330	382	408	191	289	302
5	925	459	656	688	906	924	466	646	655
6	4,339	6,594	5,395	5,534	5,130	4,084	6,385	5,505	5,606
total	6'205	7'730	6'729	7'140	6'924	5'938	7'503	6'847	7'120

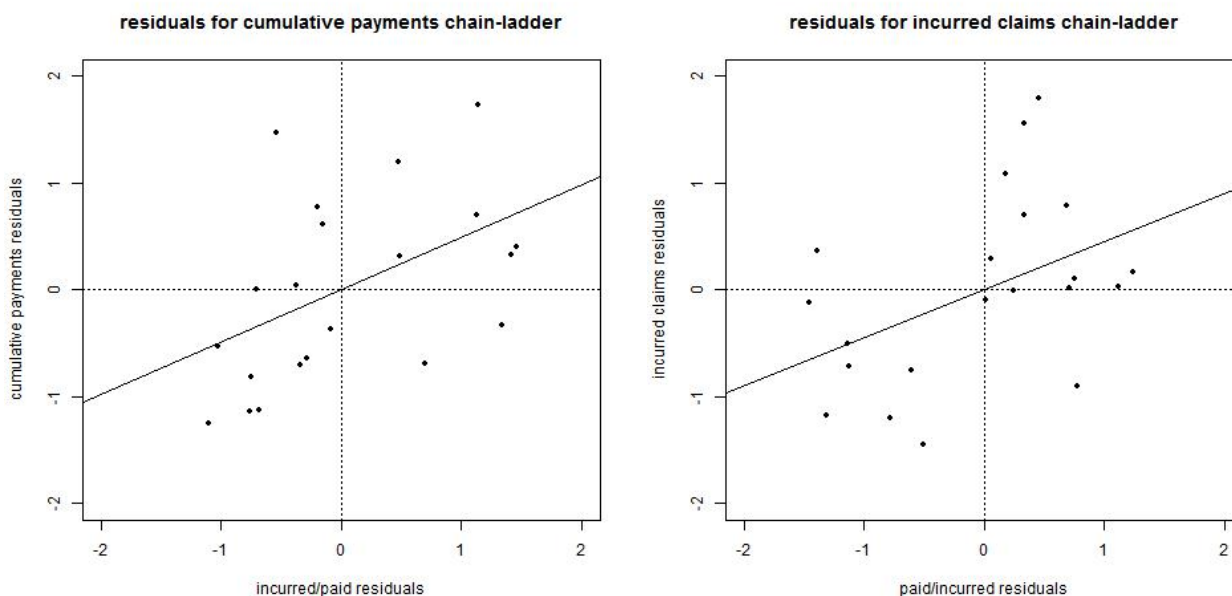


Figure 1. (lhs) Incurred-paid residuals obtained from $Q_{i,j}^{-1} = I_{i,j}/P_{i,j}$, see Remark 2, versus claims payments residuals obtained from $P_{i,j+1}$, straight line has slope $\hat{\lambda}^P = 49\%$; (rhs) paid-incurred residuals obtained from $Q_{i,j} = P_{i,j}/I_{i,j}$ versus incurred claims residuals obtained from $I_{i,j+1}$, straight line has slope $\hat{\lambda}^I = 45\%$.

Finally, we study the modified Munich chain-ladder method mMCL of Assumption 5, see Theorem 4. We therefore need to specify the off-diagonal matrix $A = (a_{k,l})_{0 \leq k, l \leq J}$, see Equation (8). A first idea to calibrate this matrix A is to use correlation estimate $\hat{\lambda}^P = 49\%$ from the LN-MCL method. A crude approximation using Theorem 3 provides

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

From this, we see that in our numerical example we need comparatively high correlations, for instance, $\text{Corr}(\xi_{i,j+1}^P, \xi_{i,k}^I) \geq 40\%$ would be in line with $\hat{\lambda}^P = 49\%$. The difficulty with this choice is that the resulting matrix Σ of type Equation (8) is not positive definite. Therefore, we need to choose smaller correlations. We do the following choice for all $i, j \geq 0$

$$\text{Corr}(\xi_{i,j+m}^P, \xi_{i,j}^I) = \begin{cases} 40\% & \text{for } m = 1, \\ 30\% & \text{for } m = 2, \\ 20\% & \text{for } m = 3, \end{cases} \tag{22}$$

and 0% otherwise. This provides a positive definite choice for Σ of type Equation (8) in our example. This choice means that we can learn from incurred claims observations $\xi_{i,j}^I$ (which relate to residuals $\varepsilon_{i,j}^{I|P}$) for cumulative payments observations $\xi_{i,j+m}^P$ with development lags $m = 1, 2, 3$, but no other conclusions can be drawn from other observations. Note that, in this example, we only use correlation choices Equation (22), but *no* similar choice for $\text{Corr}(\xi_{i,j+m}^I, \xi_{i,j}^P)$ is done. The reason is that if we choose positive correlations for the latter, in general, Σ is not positive definite. This shows that requirement Equation (8) is rather restrictive and we expect that data usually does not satisfy Assumption 1, because both plots in Figure 1 show a positive slope.

The resulting mMCL reserves

$$\hat{R}_i^{\text{mMCL}} = \hat{P}_{i,J}^{\text{mMCL}} - P_{i,J-i},$$

according to Theorem 4, are provided in Table 2. Correlation choice Equation (22) means that we learn from incurred claims, which are above average for accident year $i = 6$. This substantially increases the mMCL reserves based on cumulative payments to 6'924. Note that we do not provide the values for incurred claims: positive definiteness of Σ restricts $\text{Corr}(\xi_{i,j+m}^I, \xi_{i,j}^P) = 0$ (under Equation (22)) which implies that we obtain almost identical values to the HCL incurred reserves for $\hat{I}_{i,J}^{\text{mMCL}} - P_{i,J-i}$.

Finally, we analyze the prediction uncertainty measured by the square-rooted conditional mean square error of prediction. The results are provided in Table 3.

The prediction uncertainties of the HCL reserves and of the mMCL reserves were calculated according to Theorem 4. For the former (HCL reserves), we simply need to set $A = 0$. We see that the uncertainties in the modified version for cumulative payments are reduced from 1'249 to 1'208 because correlations Equation (22) imply that we can learn from incurred claims for cumulative payments. For incurred claims, they remain (almost) invariant because of choices $a_{k,l} = 0$ for $k < l$.

We can now calculate the prediction uncertainty for the LN–MCL method (which is still an open problem). Within Assumption 5, we know that the mMCL predictor is optimal, therefore, we obtain prediction uncertainty for the LN–MCL method

$$\text{mse}_{\text{P}_{\sum_i P_{i,J}|\mathcal{D}_J}} \left(\sum_i \hat{P}_{i,J}^{\text{MCL}} \right) = \text{mse}_{\text{P}_{\sum_i P_{i,J}|\mathcal{D}_J}} \left(\sum_i \hat{P}_{i,J}^{\text{mMCL}} \right) + \left(\sum_i \hat{P}_{i,J}^{\text{MCL}} - \sum_i \hat{P}_{i,J}^{\text{mMCL}} \right)^2, \quad (23)$$

and similarly for incurred claims. The second term in Equation (23) is the approximation error because the LN–MCL predictor is non-optimal within Assumption 5.

Table 3. Resulting reserves and square-rooted conditional mean square error of prediction of the different chain-ladder methods. * is calculated from Equation (23).

	Reserves	msep ^{1/2}
Hertig’s chain-ladder HCL paid	6,205	1,249
Hertig’s chain-ladder HCL incurred	7,730	1,565
log-normal Munich chain-ladder LN–MCL paid	6,729	1,224*
log-normal Munich chain-ladder LN–MCL incurred	7,140	1,673*
modified Munich chain-ladder mMCL paid	6,924	1,208
modified Munich chain-ladder mMCL incurred	7,730	1,565
classical chain-ladder CL paid	5,938	994
classical chain-ladder CL incurred	7,503	995
Quarg-Mack Munich chain-ladder QM–MCL paid	6,847	n/a
Quarg-Mack Munich chain-ladder QM–MCL incurred	7,120	n/a

To resume, the modified Munich chain-ladder method for cumulative payments and under assumption Equation (8) provides in our example, claims reserves that are between the HCL paid and the HCL incurred reserves (as requested). Moreover, it provides the smallest prediction uncertainty among the methods based on multivariate normal distributions. This is because, in contrast to the HCL paid and HCL incurred methods, it simultaneously considers the entire information \mathcal{D}_J , and because there is no bias (approximation error) compared LN–MCL paid and LN–MCL incurred. These conclusions are always based on the validity of Assumption 5 which is the weakness of the method because real data typically requires different covariance matrix choices than Equation (8).

8. Conclusions

We have studied the Munich chain-ladder axioms of Quarg and Mack [1] under the moment assumptions of Assumption 1. In a multivariate log-normal modeling framework, this provides rather restrictive covariance matrix Σ requirements, see Equation (8), so that Assumption 1 is simultaneously fulfilled for cumulative payments and incurred claims. For instance, a reasonable choice of Σ for the data of Quarg and Mack [1] will differ from structure Equation (8), see Section 7 where a simultaneous

choice of Equation (22) for cumulative payments and a similar choice for incurred claims would lead to a covariance matrix Σ that is not positive definite.

If Equation (8) holds, then there exists a consistent Munich chain-ladder framework, see Assumption 3, for which we can analyze claims reserves and their prediction uncertainty, see Theorem 4. Moreover, the Munich chain-ladder predictor is non-optimal in this framework, see Theorem 2, and the approximation error is provided in Theorem 3.

Author Contributions

Both authors have contributed to this document to a similar extent.

Conflicts of Interest

The authors declare no conflict of interest.

Appendix

A. Data of Quarg and Mack [1]

Table A1. Observed cumulative payments $P_{i,j}, i + j \leq 6$, source Quarg and Mack [1].

a.y. <i>i</i> /d.y. <i>j</i>	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						

Table A2. Observed incurred claims $I_{i,j}, i + j \leq 6$, source Quarg and Mack [1].

a.y. <i>i</i> /d.y. <i>j</i>	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						

B. Inverse Matrix $\Sigma_{[1]}$

Consider the matrix

$$\Sigma_{[1]} = \left(\begin{array}{cc|cc} (s_0^P)^2 & 0 & a_{0,0} & a_{0,1} \\ 0 & (s_1^P)^2 & a_{1,0} & a_{1,1} \\ \hline a_{0,0} & a_{1,0} & (s_0^I)^2 & 0 \\ a_{0,1} & a_{1,1} & 0 & (s_1^I)^2 \end{array} \right).$$

Set

$$\begin{aligned} b_{0,0} &= (s_0^I)^2 - a_{0,0}^2/(s_0^P)^2 - a_{1,0}^2/(s_1^P)^2, \\ b_{1,1} &= (s_1^I)^2 - a_{0,1}^2/(s_0^P)^2 - a_{1,1}^2/(s_1^P)^2, \\ b_{0,1} &= -a_{0,0}a_{0,1}/(s_0^P)^2 - a_{1,1}a_{1,0}/(s_1^P)^2. \end{aligned}$$

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

$$(\Sigma_{[1]})^{-1} = \left(\begin{array}{cc|cc} \frac{1}{(s_0^P)^2} + \frac{b_{0,0}a_{0,1}^2 - 2b_{0,1}a_{0,0}a_{0,1} + b_{1,1}a_{0,0}^2}{(b_{0,0}b_{1,1} + b_{0,1}^2)(s_0^P)^4} & \frac{b_{0,0}a_{0,1}a_{1,1} - b_{0,1}(a_{0,0}a_{1,1} + a_{1,0}a_{0,1}) + b_{1,1}a_{1,0}a_{0,0}}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_0^P)^2(s_1^P)^2} & \frac{b_{0,1}a_{0,1} - b_{1,1}a_{0,0}}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_0^P)^2} & \frac{-b_{0,0}a_{0,1} + b_{0,1}a_{0,0}}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_0^P)^2} \\ \frac{b_{0,0}a_{0,1}a_{1,1} - b_{0,1}(a_{0,0}a_{1,1} + a_{1,0}a_{0,1}) + b_{1,1}a_{1,0}a_{0,0}}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_0^P)^2(s_1^P)^2} & \frac{1}{(s_1^P)^2} + \frac{b_{0,0}a_{1,1}^2 - 2b_{0,1}a_{1,0}a_{1,1} + b_{1,1}a_{1,0}^2}{(b_{0,0}b_{1,1} + b_{0,1}^2)(s_1^P)^4} & \frac{b_{0,1}a_{1,1} - b_{1,1}a_{1,0}}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_1^P)^2} & \frac{-b_{0,0}a_{1,1} + b_{0,1}a_{1,0}}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_1^P)^2} \\ \hline \frac{b_{0,1}a_{0,1} - b_{1,1}a_{0,0}}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_0^P)^2} & \frac{b_{0,1}a_{1,1} - b_{1,1}a_{1,0}}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_1^P)^2} & \frac{b_{1,1}}{b_{0,0}b_{1,1} - b_{0,1}^2} & \frac{-b_{0,1}}{b_{0,0}b_{1,1} - b_{0,1}^2} \\ \frac{-b_{0,0}a_{0,1} + b_{0,1}a_{0,0}}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_0^P)^2} & \frac{-b_{0,0}a_{1,1} + b_{0,1}a_{1,0}}{(b_{0,0}b_{1,1} - b_{0,1}^2)(s_1^P)^2} & \frac{-b_{0,1}}{b_{0,0}b_{1,1} - b_{0,1}^2} & \frac{b_{0,0}}{b_{0,0}b_{1,1} - b_{0,1}^2} \end{array} \right).$$

References

1. Quarg, G.; Mack, T. Munich chain ladder. *Blätter DGVFM* **2004**, *XXVI*, 597–630.
2. Liu, H.; Verrall, R. Bootstrap estimation of the predictive distributions of reserves using paid and incurred claims. *Variance* **2010**, *4/2*, 121–135.
3. Merz, M.; Wüthrich, M.V. A credibility approach to the Munich chain-ladder method. *Blätter DGVFM* **2006**, *XXVII*, 619–628.
4. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*, 2nd ed.; Prentice-Hall: Upper Saddle River, NJ, USA, 1988.
5. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
6. Wüthrich, M.V.; Merz, M. Stochastic claims reserving manual: Advances in dynamic modeling. Available online: <http://papers.ssrn.com/sol3/papers.cfm?abstract-id=2649057> (accessed on 21 December 2015).
7. Hertig, J. A statistical approach to the IBNR-reserves in marine insurance. *ASTIN Bull.* **1985**, *15/2*, 171–183.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).