

THE MEAN SQUARE ERROR OF PREDICTION
IN THE CHAIN LADDER RESERVING METHOD
(MACK AND MURPHY REVISITED)

BY

MARKUS BUCHWALDER, HANS BÜHLMANN, MICHAEL MERZ
AND MARIO V. WÜTHRICH

ABSTRACT

We revisit the famous Mack formula [2], which gives an estimate for the mean square error of prediction MSEP of the chain ladder claims reserving method: We define a time series model for the chain ladder method. In this time series framework we give an approach for the estimation of the conditional MSEP. It turns out that our approach leads to results that differ from the Mack formula. But we also see that our derivation leads to the same formulas for the MSEP estimate as the ones given in Murphy [4]. We discuss the differences and similarities of these derivations.

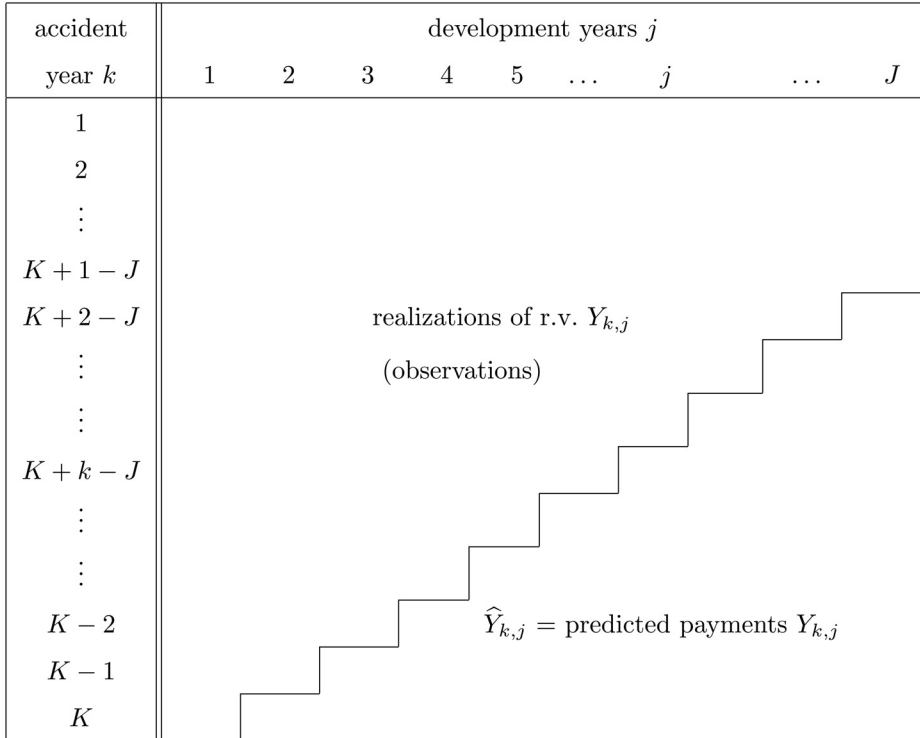
1. MOTIVATION

In this article we revisit the famous Mack formula [2] to estimate the mean square error of prediction MSEP of the chain ladder method. The chain ladder method has to be understood as an algorithm to estimate claims reserves (under certain homogeneity assumptions). Mack [2] was the first who gave a probabilistic background to the chain ladder method which allowed for the derivation of MSEP estimates for the claims reserves. We formulate a time series model for the chain ladder method. Within this time series framework we discuss different approaches for the estimation of the MSEP (see Subsection 4.1.2). These different approaches contain conditional and unconditional views of the problem. In the present work we concentrate on the conditional view, i.e. we give formulas and estimates for the MSEP in the conditional approach. Our (conditional) approach leads to the same MSEP estimates as an other derivation given by Murphy [4] (in Subsection 4.2 we discuss and compare our derivation to Murphy's derivation). Moreover, we see that the Mack formula [2] is a linear approximation to our result.

We want to emphasize that different approaches to the estimation of the MSEP are acceptable, if they are correct in a mathematical sense. They describe the problem from different angles (conditional or unconditional probability laws). The expert in practice should decide which is the appropriate view for his problem.

1.1. The development trapezoid

Our claims data have the following structure:



In this trapezoid the variables

- $k, \quad 1 \leq k \leq K,$ refer to accident years (rows)
- $j, \quad 1 \leq j \leq J,$ refer to development years (columns).

Usually we have observations in $\mathcal{D}_K := \{Y_{k,j} | 1 \leq k \leq K \text{ and } 1 \leq j \leq \min\{J, K - k + 1\}\}$ and we need to estimate the distributions in \mathcal{D}_K^c . For $K > J$ the set \mathcal{D}_K yields a trapezoid and for $K = J$ we get a triangle of the observations.

The entries $Y_{k,j}$ (as well as their partial sums $X_{k,j}$) may be interpreted

- a) on a paid claims basis,
- b) on a occurred claims basis.

Henceforth,

$$X_{k,j} := \sum_{l=1}^j Y_{k,l}, \tag{1.1}$$

denotes the cumulative payments of accident year k up to development period j , or the claims incurred for accident year k reported up to reporting year j , respectively.

	$Y_{k,j}$	$X_{k,j}$
Interpretation a)	Non-cumulative payments for accident year k in development year j	Cumulative payments for accident year k up to development year j
Interpretation b)	Change of incurred loss of accident year k in development year j	Incurred loss of accident year k up to development year j

For our exposition we use in the sequel interpretation a). The reader can easily translate into interpretation b).

Whereas $X_{k,j} (1 \leq j \leq J)$ refers to cumulative payments with ultimate loss $X_{k,J}$ for accident year k , it is sometimes useful to consider the incremental payments in accounting years (diagonals):

$$Y_t := \sum_{\substack{k+j=t+1 \\ 1 \leq j \leq J}} Y_{k,j} \quad (\text{sum over } t\text{-diagonal}) \tag{1.2}$$

stands then for the payments in the accounting year $t (1 \leq t \leq K + J - 1)$.

The Problems:

We have observed \mathcal{D}_K , or all t -diagonals of Y_t with $1 \leq t \leq K$, respectively. Intuitively this means that we are at time K .

- Estimate the reserves $R_k = X_{k,J} - X_{k,K-k+1}$ for all accident years k .
- Predict Y_t for all future t -diagonals for $K < t \leq K + J - 1$.

The estimation of the claims reserves R_k is the classical (actuarial) reserving problem studied in every non-life insurance company. The prediction of Y_t for future diagonals yields the cashflow associated with the reserving problem, which is especially helpful for economic considerations of the reserves and the valuation portfolio constructions (see [1]).

Remarks 1.1.

- For each future t -diagonal we have to predict in formula (1.2) not only Y_t but each of its summands $Y_{k,j}$ with $k + j = t + 1$.
- This allows – by linearity – also the prediction of all missing $X_{k,j} (k + j > K + 1)$, in particular the ultimate $X_{k,J}$ and the reserve R_k .

Notation

- We use $\hat{}$ for both “prediction” and “estimation” to simplify the notation. The interpretation of its meaning should be clear from the context.
- We use capital letters $Y_{k,j}$ (and $X_{k,j}$) to denote both the random variable as well as its realisations. The meaning of the symbol must be interpreted from the context.

2. THE CHAIN LADDER RESERVING MODEL AND THE CORRESPONDING CANONICAL MODEL

The chain ladder method is based on cumulative payments $X_{k,j}$. From the cumulative estimates we can easily derive estimates for $Y_{k,j}$, as we see below.

2.1. Chain ladder method

The algorithmic definition of the chain-ladder method reads as follows

1. There are constants $f_l (l = 1, \dots, J-1)$ such that for all accident years k the future cumulative payments $X_{k,j} (j > k^*)$ are estimated by

$$\widehat{X}_{k,j} := X_{k,k^*} \cdot f_{k^*} \cdot f_{k^*+1} \cdot \dots \cdot f_{j-1}, \quad (2.1)$$

where $k^* = k^*(k) = K - k + 1$ is the last observation for accident year k on the K -diagonal.

2. The chain ladder factors f_l (age-to-age factors) are estimated by

$$\hat{f}_l := \frac{\sum_{k=1}^{K-l} X_{k,l+1}}{S_l}, \quad (2.2)$$

where

$$S_l := \sum_{k=1}^{K-l} X_{k,l}. \quad (2.3)$$

The actuarial literature often explains the chain ladder method (as above) purely as a computational algorithm and leaves the question open which probabilistic model would lead to this procedure.

It is Mack's merit [2] that he was the first who gave an answer to this question.

Remark. Formula (2.2) gives one possibility for the estimation of the chain ladder factors f_l . Of course there are other possibilities to estimate f_l . We do not further discuss these in this paper (see also Remarks 2.1).

2.2. Mack’s conditions

- (M1) Accident years k are independent.
- (M2) The $X_{k,j}$ ($j = 1, 2, \dots, J$) form a Markov chain for each k ($1 \leq k \leq K$).
- (M3) $E[X_{k,j} | X_{k,j-1}] = f_{j-1} \cdot X_{k,j-1}$.

It is quite natural (but not imperative) to add a condition about conditional variances

(M4) $\text{Var}[X_{k,j} | X_{k,j-1}] = \sigma_{j-1}^2 \cdot X_{k,j-1}$.

Remarks 2.1.

- One could also write

(M4’) $\text{Var}\left[\frac{X_{k,j}}{X_{k,j-1}} \middle| X_{k,j-1}\right] = \frac{\sigma_{j-1}^2}{X_{k,j-1}}$.

Formula (M4’) is known as the Bühlmann-Straub condition in credibility theory. One may therefore interpret $X_{k,j-1}$ as (conditional) volume measure.

- Assumption (M4) is not imperative and could be generalized to other powers $X_{k,j-1}^\gamma$ (see [3]). However, if we choose $\gamma = 1$ as in (M4), then the estimator \hat{f}_l (formula (2.2)) has minimal variance, given S_l (see [2] or [5]). If $\gamma \neq 1$ one uses other estimates for f_l than (2.2).
- It is also worth to mention that the chain ladder estimate is only based on the information in the upper triangle. Of course, in practice, we have additional (external) information. In this case our estimate should not only be based on \mathcal{D}_K .

It is easy to check that under conditions (M1)-(M3) we have for $j > k^*$ (see Mack [2])

$$E[\widehat{X}_{k,j} | \mathcal{D}_K] = f_{k^*} \cdot f_{k^*+1} \cdot \dots \cdot f_{j-1} \cdot X_{k,k^*} = E[X_{k,j} | X_{k,k^*}],$$

and

$$E[X_{k,k^*} \cdot \hat{f}_{k^*} \cdot \dots \cdot \hat{f}_{j-1} | X_{k,k^*}] = E[X_{k,j} | X_{k,k^*}].$$

(CL)

This means that we have unbiased estimators for the conditional expectations.

2.3. The chain ladder consistent time series model

We choose here a different route. We specify a time series model which yields the chain ladder method. This route seems very promising as it – besides allowing to calculate means and variances – also defines how to simulate development trapezoids (or triangles).

Different accident years are independent, and for $1 \leq j \leq J$ and $0 \leq k \leq K$ we have

$$X_{k,j} = f_{j-1} \cdot X_{k,j-1} + \sigma_{j-1} \cdot \sqrt{X_{k,j-1}} \cdot \varepsilon_{k,j}, \tag{T1}$$

with $\sigma_{j-1} > 0$ and $\varepsilon_{k,j}$ independent with $E[\varepsilon_{k,j}] = 0$ and $E[\varepsilon_{k,j}^2] = 1$. $(T2)$

Remarks 2.2

- This time series model coincides with the model studied in Murphy [4], Model IV (weighted average development factor model). We discuss Murphy’s results below.
- $X_{k,0}$ ($1 \leq k \leq K$) can be interpreted as the (deterministic) premium P_k earned for accident year k . In an enlarged trapezoid this allows the simulation of the first year development figures.
- Theoretically, our process could have negative values for the cumulative payments $X_{k,j}$. To avoid this problem, we could reformulate the definition of $\varepsilon_{k,j}$ such that its distribution is conditionally, given $X_{k,j-1}$, centered with variance 1 and such that $X_{k,j}$ is positive.
- The time series model is particularly useful for the derivation of the (conditional) estimation error. It reflects the mechanism of generating sets of “other possible” observations: Having an observation $X_{k,j-1}$, (T1) tells us how to model/generate the chain-ladder factors \hat{f}_{j-1} (see also (T1’) and (4.19), below).

It is easy to check that (T1) and (T2) imply again that the chain ladder prediction formula (CL) holds. Also one sees that (T1) and (T2) imply Mack’s conditions (M1)-(M4).

3. ESTIMATION OF MODEL PARAMETERS

3.1. Estimators for f_j and σ_j^2

Our time series model specified by assumptions (T1) and (T2) has as parameters

$$\begin{aligned} &f_1, f_2, \dots, f_{J-1}, \\ &\sigma_1^2, \sigma_2^2, \dots, \sigma_{J-1}^2, \end{aligned} \tag{3.1}$$

which we need to estimate. To estimate the chain ladder factors f_{j-1} we use the standard formulae (2.2)-(2.3). The appropriate estimator for the variance σ_{j-1}^2 is given by ($1 \leq j \leq J-1$)

$$\hat{\sigma}_{j-1}^2 := \frac{1}{K-j} \cdot \sum_{k=1}^{K-j+1} X_{k,j-1} \cdot \left(\frac{X_{k,j}}{X_{k,j-1}} - \hat{f}_{j-1} \right)^2. \tag{3.2}$$

TABLE 1
PARAMETERS VERSUS OBSERVATIONS FOR THE TRIANGLE

J	Parameters	Observations
3	4	6
4	6	10
5	8	15
6	10	21
\vdots	\vdots	\vdots
10	18	55

Remarks 3.1.

- Estimator (2.2) is minimum variance among linear estimators.
- Estimator (3.2) is minimum variance for normal $\varepsilon_{k,j}$.
- If we have enough data (i.e. $K > J$), we are able to estimate σ_{J-1}^2 with (3.2), otherwise there are several possibilities to estimate σ_{J-1}^2 (see e.g. [2]).

3.2. Statistical meaning of the estimators \hat{f}_j and $\tilde{\sigma}_j^2$

Generally speaking our model has quite an excessive number of parameters $(2J - 2)$ and our estimation resides on a quite small number of observations $J \cdot (\frac{2 \cdot K - J + 1}{2})$, in the “worst case” of the triangles $\frac{J \cdot (J + 1)}{2}$.

This seems to indicate that the traditional chain ladder does “overparameterize”. One obvious remedy is to put $\sigma_{l-1}^2 = \sigma^2$ for all l . One might even look for a “smoothing procedure” for the age-to-age factors e.g. a parametric curve for the f_j depending on fewer parameters. We would also like to mention that the chain ladder method makes rather strong homogeneity assumptions across all accident years. Practical data do often have outliers, trends, cycles, etc. which need to be smoothed.

The statistical questions about overparameterization and about checking the real data against the chain-ladder axioms go beyond the scope of this paper and are not further pursued here. We therefore address these challenges as open problems to the reader.

4. MEAN SQUARE ERROR OF PREDICTION

At time $t = K$ we have realization for the cumulative payments $X_{k,j}$ given by the upper trapezoid $\mathcal{D}_K = \{X_{k,j} | 1 \leq k \leq K \text{ and } 1 \leq j \leq \min\{J, K - k + 1\}\}$. Moreover, given \mathcal{D}_K , the chain ladder estimators

$$\widehat{X}_{k,j} = X_{k,k^*} \cdot \widehat{f}_{k^*} \cdot \widehat{f}_{k^*+1} \cdot \dots \cdot \widehat{f}_{j-1} \tag{4.1}$$

denote the estimators of the future cumulative payments $X_{k,j}$ ($1 \leq k \leq K$ and $K - k + 1 < j \leq J$). Formula (T1) implies for $X_{k,j}$ with $K - k + 1 < j \leq J$

$$E[X_{k,j} | \mathcal{D}_K] = X_{k,k^*} \cdot f_{k^*} \cdot f_{k^*+1} \cdot \dots \cdot f_{j-1}. \tag{4.2}$$

4.1. Prediction error for single accident years

As usual the prediction error of $\widehat{X}_{k,j}$ can be decomposed into two parts: a) stochastic error (process variance) and b) estimation error (c.f. Mack [2], Section 3):

$$E\left[\left(\widehat{X}_{k,j} - X_{k,j}\right)^2 \middle| \mathcal{D}_K\right] = \underbrace{E\left[\left(X_{k,j} - E[X_{k,j} | \mathcal{D}_K]\right)^2 \middle| \mathcal{D}_K\right]}_{\text{process variance}} + \underbrace{\left(\widehat{X}_{k,j} - E[X_{k,j} | \mathcal{D}_K]\right)^2}_{\text{estimation error}}. \tag{4.3}$$

The process variance $\text{Var}(X_{k,j} | \mathcal{D}_K) = E[(X_{k,j} - E[X_{k,j} | \mathcal{D}_K])^2 | \mathcal{D}_K]$ originates from the stochastic movement of the process, whereas the estimation error reflects the uncertainty in the estimation of the expectation (chain ladder parameters). We derive estimates for both the process variance and the estimation error. For the estimates of the estimation error there are different approaches. These are described in Subsection 4.1.2, below. The reader who wants to find quickly the relevant formulae for the different errors should use Table 2 for the corresponding numbers of formulae:

TABLE 2
MEAN SQUARE ERROR OF PREDICTION FOR DIFFERENT CONSIDERATIONS

	X	Y
Single accident years	Result 4.1	Result 4.4
Ultimate of aggregated accident years	Result 4.2	
Accounting years	Result 4.3	Result 4.4

For the definition of the symbols $\Gamma_{k,j}^2$ and $\Delta_{k,j}^2$ in this formulae see (4.8) and (4.21) respectively.

4.1.1. (Conditional) Process variance

Using (4.2) we obtain for the (conditional) process variance in formula (4.3): Choose $1 \leq k \leq K$ and $K - k + 1 < j \leq J$

$$E[(X_{k,j} - E[X_{k,j} | \mathcal{D}_K])^2 | \mathcal{D}_K] = E(X_{k,j}^2 | \mathcal{D}_K) - X_{k,K-k+1}^2 \cdot f_{K-k+1}^2 \cdot \dots \cdot f_{j-1}^2. \tag{4.4}$$

Moreover, using (T1) and (T2) we have

$$\begin{aligned} E(X_{k,j}^2 | \mathcal{D}_K) &= f_{j-1}^2 \cdot E(X_{k,j-1}^2 | \mathcal{D}_K) + \sigma_{j-1}^2 \cdot E(X_{k,j-1} \cdot \varepsilon_{k,j}^2 | \mathcal{D}_K) \\ &\quad \vdots \\ &= f_{j-1}^2 \cdot f_{j-2}^2 \cdot \dots \cdot f_{K-k+1}^2 \cdot X_{k,K-k+1}^2 \\ &\quad + \sigma_{j-1}^2 \cdot f_{j-2} \cdot f_{j-3} \cdot \dots \cdot f_{K-k+1} \cdot X_{k,K-k+1} \\ &\quad + f_{j-1}^2 \cdot \sigma_{j-2}^2 \cdot f_{j-3} \cdot \dots \cdot f_{K-k+1} \cdot X_{k,K-k+1} \\ &\quad + \dots + f_{j-1}^2 \cdot f_{j-2}^2 \cdot \dots \cdot f_{K-k+2}^2 \cdot \sigma_{K-k+1}^2 \cdot X_{k,K-k+1}. \end{aligned} \tag{4.5}$$

Hence, from (4.5) we obtain for the (conditional) process variance:

$$\begin{aligned} \text{Var}(X_{k,j} | \mathcal{D}_K) &= E[(X_{k,j} - E[X_{k,j} | \mathcal{D}_K])^2 | \mathcal{D}_K] \\ &= X_{k,K-k+1} \cdot \sum_{l=K-k+1}^{j-1} \sigma_l^2 \cdot \sum_{m=l+1}^{j-1} f_m^2 \cdot \sum_{m=K-k+1}^{l-1} f_m \end{aligned} \tag{4.6}$$

Finally, using (CL) the (conditional) process variance can be rewritten in the form:

$$E[(X_{k,j} - E[X_{k,j} | \mathcal{D}_K])^2 | \mathcal{D}_K] = (E[X_{k,j} | \mathcal{D}_K])^2 \cdot \sum_{l=K-k+1}^{j-1} \frac{\sigma_l^2}{E[X_{k,l} | \mathcal{D}_K] \cdot f_l^2}. \tag{4.7}$$

Replacing $E[X_{k,j} | \mathcal{D}_K]$, $E[X_{k,l} | \mathcal{D}_K]$ and the parameters σ_l^2 , f_l by their estimators leads to the following estimator of the (conditional) process variance

$$\widehat{E}[(X_{k,j} - E[X_{k,j} | \mathcal{D}_K])^2 | \mathcal{D}_K] := \underbrace{\left(\widehat{X}_{k,j}\right)^2 \cdot \sum_{l=K-k+1}^{j-1} \frac{\widehat{\sigma}_l^2}{\widehat{X}_{k,l} \cdot \widehat{f}_l^2}}_{=: \Gamma_{k,j}^2} \tag{4.8}$$

$\Gamma_{k,j}^2$ can be rewritten in a recursive form:

$$\Gamma_{k,j}^2 = \Gamma_{k,j-1}^2 \cdot \widehat{f}_{j-1}^2 + \widehat{\sigma}_{j-1}^2 \cdot \widehat{X}_{k,j-1}. \tag{4.9}$$

Using (4.8), we obtain for the non-cumulative payments $Y_{k,j}$ the following estimator of the (conditional) process variance

$$\begin{aligned} \widehat{E}[(Y_{k,j} - E[Y_{k,j} | \mathcal{D}_K])^2 | \mathcal{D}_K] &:= \Gamma_{k,j}^2 + \Gamma_{k,j-1}^2 \cdot (1 - 2 \cdot \widehat{f}_{j-1}) \\ &= \Gamma_{k,j-1}^2 \cdot (\widehat{f}_{j-1} - 1)^2 + \widehat{\sigma}_{j-1}^2 \cdot \widehat{X}_{k,j-1}. \end{aligned} \tag{4.10}$$

4.1.2. (Conditional) Estimation error

Using (4.1) and (4.2) leads to the following formula for the estimation error:

$$\begin{aligned}
 (\widehat{X}_{k,j} - E[X_{k,j} | \mathcal{D}_K])^2 &= X_{k,K-k+1}^2 \cdot (\widehat{f}_{K-k+1} \cdot \dots \cdot \widehat{f}_{j-1} - f_{K-k+1} \cdot \dots \cdot f_{j-1})^2 \\
 &= X_{k,K-k+1}^2 \cdot \left(\prod_{l=K-k+1}^{j-1} \widehat{f}_l^2 + \prod_{l=K-k+1}^{j-1} f_l^2 - 2 \cdot \prod_{l=K-k+1}^{j-1} \widehat{f}_l \cdot f_l \right). \tag{4.11}
 \end{aligned}$$

The realizations of the estimators $\widehat{f}_{K-k+1}, \dots, \widehat{f}_{j-1}$ are known at time $t = K$, and the “true” chain ladder factors $f_{K-k+1}, \dots, f_{j-1}$ are unknown (and have to be estimated). In order to determine the (conditional) estimation error we need to determine the volatilities of \widehat{f}_j around its true values f_j . These volatilities measure the quality of the estimates \widehat{f}_j . We determine the volatilities with the help of resampled observations for \widehat{f}_j . There are different approaches to resample our time series, conditional ones and unconditional ones. For $j \in \{1, \dots, J\}$ we define the σ -fields

$$\mathcal{B}_j = \sigma(\{X_{k,l} \in \mathcal{D}_K \mid l \leq j\}). \tag{4.12}$$

Hence, $\{\mathcal{B}_j\}_{j=1, \dots, J}$ defines a filtration and \widehat{f}_j is \mathcal{B}_{j+1} -measurable for all $j = 1, \dots, J-1$. If we fix accident year k for the moment, then in order to determine the volatility in the estimates we need to resample (see right-hand side of (4.11))

$$\widehat{f}_{K-k+1}^2 \cdot \dots \cdot \widehat{f}_{j-1}^2. \tag{4.13}$$

Define the upper right corner of the observations \mathcal{D}_K (with respect to development year $K - k + 2$)

$$\mathcal{D}_K^O = \mathcal{D}_K \cap \{X_{i,j}; j > K - k + 1, i \geq 1\} \tag{4.14}$$

Approach 1 (Complete resampling in \mathcal{D}_K^O). Calculate

$$E[\widehat{f}_{K-k+1}^2 \cdot \dots \cdot \widehat{f}_{j-1}^2 \mid \mathcal{B}_{K-k+1}]. \tag{4.15}$$

This is the complete averaging over the multidimensional distribution after time $K - k + 1$. The observed realizations in \mathcal{D}_K^O do not play any role in this consideration. Therefore we call this the unconditional version.

Approach 2. Calculate

$$\widehat{f}_{K-k+1}^2 \cdot \dots \cdot \widehat{f}_{j-2}^2 \cdot E[\widehat{f}_{j-1}^2 \mid \mathcal{B}_{j-1}]. \tag{4.16}$$

The averaging is only partly done. In this approach one has also the possibility to choose the position at which the averaging should be done. This approach is in some sense similar to the one used in Mack [2] (see also Section 4.2).

Approach 3 (Conditional resampling in \mathcal{D}_K^O).

$$E[\hat{f}_{K-k+1}^2 | \mathcal{B}_{K-k+1}] \cdot E[\hat{f}_{K-k+2}^2 | \mathcal{B}_{K-k+2}] \cdot \dots \cdot E[\hat{f}_{j-1}^2 | \mathcal{B}_{j-1}]. \tag{4.17}$$

The averaging is only done over the conditional distributions. Thereby the observed realizations in \mathcal{D}_K^O have a direct influence on the estimate since they are used in the conditions. Therefore we call this the conditional version. From a numerical point of view it is important to note that Approach 3 allows for a multiplicative structure of the measure of volatility.

The question, which approach should be chosen, is not a mathematical one. It depends on the circumstances of the questions which approach should be used for a specific practical problem. Henceforth, only the practitioner can choose the appropriate approach for his problems and questions.

In the present work we focus on Approach 3, the conditional version. Hence (4.11) is estimated by (see also (4.20), below)

$$X_{k, K-k+1}^2 \cdot \left(\prod_{l=K-k+1}^{j-1} E[\hat{f}_l^2 | \mathcal{B}_l] - \prod_{l=K-k+1}^{j-1} f_l^2 \right). \tag{4.18}$$

We therefore resample the observations $\hat{f}_{K-k+1}, \dots, \hat{f}_{j-1}$, given the upper trapezoid \mathcal{D}_K (see Remarks 2.2). This means that for the determination of an estimator for the conditional estimation error we have to take into account that, given the upper trapezoid \mathcal{D}_K , the observation for \hat{f}_j could have been different from the observed values \hat{f}_j . To regard this source of uncertainty, we generate for $1 \leq j \leq J$ and $0 \leq k \leq K$, given \mathcal{D}_K , a set of “new” observations by the formula

$$Z_{k,j} = f_{j-1} \cdot X_{k,j-1} + \sigma_{j-1} \cdot \sqrt{X_{k,j-1}} \cdot \tilde{\varepsilon}_{k,j}, \tag{T1'}$$

$$\text{with } \sigma_{j-1} > 0 \text{ and } \varepsilon_{k,j}, \tilde{\varepsilon}_{k,j} \text{ independent and identically distributed} \tag{T2'}$$

(cf. (T1) and (T2)). Observe that, given $X_{k,j-1}$, $Z_{k,j} \stackrel{(d)}{=} X_{k,j}$.

From formula (T1') and (2.2) we obtain the following representation for the resampled estimates of the development factors

$$\hat{f}_{j-1}^{\mathcal{D}_K} = \frac{\sum_{k=1}^{K-j+1} Z_{k,j}}{\sum_{k=1}^{K-j+1} X_{k,j-1}} = f_{j-1} + \frac{\sigma_{j-1}}{S_{j-1}} \cdot \sum_{k=1}^{K-j+1} \sqrt{X_{k,j-1}} \cdot \tilde{\varepsilon}_{k,j} \quad (1 \leq j \leq J) \tag{4.19}$$

with

$$S_{j-1} = \sum_{k=1}^{K-j+1} X_{k,j-1}.$$

Observe that, given \mathcal{B}_{j-1} , we have $\hat{f}_{j-1} \stackrel{(d)}{=} \hat{f}_{j-1}^{\mathcal{D}_K}$.

Unlike the observations $\{X_{k,j} | 1 \leq k \leq K \text{ and } 1 \leq j \leq \min\{J, K-k+1\}\}$ the observations $\{Z_{k,j} | 1 \leq k \leq K \text{ and } 1 \leq j \leq \min\{J, K-k+1\}\}$ and also the resampled estimates $\hat{f}_0^{\mathcal{D}_K}, \dots, \hat{f}_{j-1}^{\mathcal{D}_K}$ are random variables given the upper trapezoid \mathcal{D}_K . Furthermore the observations $X_{k,j}$ and the random variables $\tilde{\varepsilon}_{k,j}$ are unconditionally independent. From (4.19), (T2') and (T2) we see that

- 1) the estimators $\hat{f}_0^{\mathcal{D}_K}, \dots, \hat{f}_{j-1}^{\mathcal{D}_K}$ are conditionally independent w.r.t. \mathcal{D}_K ,
- 2) $E[\hat{f}_{j-1}^{\mathcal{D}_K} | \mathcal{D}_K] = f_{j-1}$ for $1 \leq j \leq J$ and
- 3) $E[(\hat{f}_{j-1}^{\mathcal{D}_K})^2 | \mathcal{D}_K] = f_{j-1}^2 + \frac{\sigma_{j-1}^2}{S_{j-1}}$ for $1 \leq j \leq J$.

Hence in Approach 3 (using 1)-3) from above) we can explicitly calculate the conditional estimation error, i.e. the right-hand side of (4.11) is estimated by

$$\begin{aligned}
 & E \left[X_{k, K-k+1}^2 \cdot \left(\hat{f}_{K-k+1}^{\mathcal{D}_K} \cdot \dots \cdot \hat{f}_{j-1}^{\mathcal{D}_K} - f_{K-k+1} \cdot \dots \cdot f_{j-1} \right)^2 \middle| \mathcal{D}_K \right] \\
 &= X_{k, K-k+1}^2 \cdot \left(\prod_{l=K-k+1}^{j-1} E \left[\left(\hat{f}_l^{\mathcal{D}_K} \right)^2 \middle| \mathcal{D}_K \right] - \prod_{l=K-k+1}^{j-1} f_l^2 \right) \tag{4.20} \\
 &= X_{k, K-k+1}^2 \cdot \left(\prod_{l=K-k+1}^{j-1} \left(f_l^2 + \frac{\sigma_l^2}{S_l} \right) - \prod_{l=K-k+1}^{j-1} f_l^2 \right).
 \end{aligned}$$

Next, we replace the parameters $\sigma_{K-k+1}^2, \dots, \sigma_{j-1}^2$ and $f_{K-k+1}, \dots, f_{j-1}$ by their estimators, and we obtain the following estimator for the (conditional) estimation error

$$\begin{aligned}
 & \hat{E}_{\mathcal{D}_K} \left[\left(\hat{X}_{k,j} - E[X_{k,j} | \mathcal{D}_K] \right)^2 \right] \\
 &:= \hat{E} \left[X_{k, K-k+1}^2 \cdot \left(\hat{f}_{K-k+1}^{\mathcal{D}_K} \cdot \dots \cdot \hat{f}_{j-1}^{\mathcal{D}_K} - f_{K-k+1} \cdot \dots \cdot f_{j-1} \right)^2 \middle| \mathcal{D}_K \right] \tag{4.21} \\
 &:= X_{k, K-k+1}^2 \cdot \underbrace{\left(\prod_{l=K-k+1}^{j-1} \left(\hat{f}_l^2 + \frac{\hat{\sigma}_l^2}{S_l} \right) - \prod_{l=K-k+1}^{j-1} \hat{f}_l^2 \right)}_{=: \Delta_{k,j}^2}.
 \end{aligned}$$

$\Delta_{k,j}^2$ can be rewritten in a recursive form:

$$\begin{aligned}
 \Delta_{k,j}^2 &= \Delta_{k,j-1}^2 \cdot \hat{f}_{j-1}^2 + \frac{\hat{\sigma}_{j-1}^2}{S_{j-1}} \cdot \prod_{l=K-k+1}^{j-2} \left(\hat{f}_l^2 + \frac{\hat{\sigma}_l^2}{S_l} \right) \\
 &= \Delta_{k,j-1}^2 \cdot \left(\hat{f}_{j-1}^2 + \frac{\hat{\sigma}_{j-1}^2}{S_{j-1}} \right) + \frac{\hat{\sigma}_{j-1}^2}{S_{j-1}} \cdot \prod_{l=K-k+1}^{j-2} \hat{f}_l^2. \tag{4.22}
 \end{aligned}$$

Analogously we define for non-cumulative payments the following estimator:

$$\widehat{Y}_{k,j} := \widehat{X}_{k,j} - \widehat{X}_{k,j-1}. \tag{4.23}$$

Hence the (conditional) estimation error is estimated by (in the 3rd step we use (4.22))

$$\begin{aligned} & \widehat{E}_{\mathcal{D}_K} \left[\left(\widehat{Y}_{k,j} - E[Y_{k,j} | \mathcal{D}_K] \right)^2 \right] \\ & := \widehat{E} \left[X_{k, K-k+1}^2 \cdot \left(\prod_{l=K-k+1}^{j-2} \widehat{f}_l^{\mathcal{D}_K} \cdot (\widehat{f}_{j-1}^{\mathcal{D}_K} - 1) - \prod_{l=K-k+1}^{j-2} f_l \cdot (f_{j-1} - 1) \right)^2 \middle| \mathcal{D}_K \right] \\ & := X_{k, K-k+1}^2 \cdot \left(\Delta_{k,j-1}^2 \cdot (\widehat{f}_{j-1} - 1)^2 + \prod_{l=K-k+1}^{j-2} \left(\widehat{f}_l^2 + \frac{\widehat{\sigma}_l^2}{S_l} \right) \cdot \frac{\widehat{\sigma}_{j-1}^2}{S_{j-1}} \right) \tag{4.24} \\ & = X_{k, K-k+1}^2 \cdot \left(\Delta_{k,j-1}^2 \cdot (\widehat{f}_{j-1} - 1)^2 + \Delta_{k,j}^2 - \Delta_{k,j-1}^2 \cdot \widehat{f}_{j-1} \right) \\ & = X_{k, K-k+1}^2 \cdot \left(\Delta_{k,j}^2 - (2\widehat{f}_{j-1} - 1) \cdot \Delta_{k,j-1}^2 \right). \end{aligned}$$

Remark: The (conditional) estimation error in the cumulative case and the incremental case have exactly the same structure (compare 3rd line of (4.24) with (4.22)).

4.2. Result and comparison to the Mack [2] formula and the result by Murphy [4]

From our results we obtain the following estimator for the (conditional) prediction error of a single accident year (cf. (4.8) and (4.21)):

Result 4.1. (MSEP for single accident years) *Under Assumptions (T1), (T2), (T1'), (T2') for the estimation of the process variance and the estimation error, respectively, we have the following estimator for the (conditional) mean square prediction error (4.3) of a single accident year*

$$\begin{aligned} & \widehat{E} \left[\left(\widehat{X}_{k,j} - X_{k,j} \right)^2 \middle| \mathcal{D}_K \right] := \Gamma_{k,j}^2 + X_{k, K-k+1}^2 \cdot \Delta_{k,j}^2 \\ & = \underbrace{\left(\widehat{X}_{k,j} \right)^2 \cdot \sum_{l=K-k+1}^{j-1} \frac{\widehat{\sigma}_l^2}{\widehat{X}_{k,l} \cdot \widehat{f}_l^2}}_{\text{process variance}} + \underbrace{X_{k, K-k+1}^2 \cdot \left(\prod_{l=K-k+1}^{j-1} \left(\widehat{f}_l^2 + \frac{\widehat{\sigma}_l^2}{S_l} \right) - \prod_{l=K-k+1}^{j-1} \widehat{f}_l^2 \right)}_{\text{estimation error}}. \tag{4.25} \end{aligned}$$

The recursive form of the estimator of the prediction error is given by (cf. (4.9) and (4.22))

$$\begin{aligned} \widehat{E} \left[\left(\widehat{X}_{k,j} - X_{k,j} \right)^2 \middle| \mathcal{D}_K \right] &:= \Gamma_{k,j-1}^2 \cdot \widehat{f}_{j-1}^2 + \widehat{\sigma}_{j-1}^2 \cdot \widehat{X}_{k,j-1} \\ &+ X_{k,K-k+1}^2 \cdot \left(\Delta_{k,j-1}^2 \cdot \widehat{f}_{j-1}^2 + \prod_{l=K-k+1}^{j-2} \left(\widehat{f}_l^2 + \frac{\widehat{\sigma}_l^2}{S_l} \right) \cdot \frac{\widehat{\sigma}_{j-1}^2}{S_{j-1}} \right). \end{aligned} \tag{4.26}$$

$\Gamma_{k,j-1}^2$ and $\Delta_{k,j-1}^2$ were defined in (4.8) and (4.21) respectively.

Comparison to the Mack formula [2].

Formula (4.25) looks very similar to the Mack formula [2]. But, at the first sight surprisingly, the formula for the estimation error given in Mack [2] differs from our formula (4.22). In [2] the recursive formula for $\Delta_{k,j}^2$ only considers the following terms

$$\Delta_{k,j}^2 = \Delta_{k,j-1}^2 \cdot \widehat{f}_{j-1}^2 + \prod_{l=K-k+1}^{j-2} \widehat{f}_l^2 \cdot \frac{\widehat{\sigma}_{j-1}^2}{S_{j-1}}, \tag{4.27}$$

whereas we consider for the second term the following expression (cf. (4.22))

$$\prod_{l=K-k+1}^{j-2} \left(\widehat{f}_l^2 + \frac{\widehat{\sigma}_l^2}{S_l} \right) \cdot \frac{\widehat{\sigma}_{j-1}^2}{S_{j-1}}.$$

The difference comes from the fact that Mack calculates the estimation error for a version of Approach 2. Mack [2] sums up the following expressions (see [2], p. 219)

$$\begin{aligned} E \left[\left(\widehat{f}_{K+1-k} \cdots \widehat{f}_{j-1} \cdot (f_j - \widehat{f}_j) \cdot f_{j+1} \cdots f_{j-1} \right)^2 \middle| \mathcal{B}_j \right] &= \\ \widehat{f}_{K+1-k}^2 \cdots \widehat{f}_{j-1}^2 \cdot \frac{\sigma_j^2}{S_j} \cdot f_{j+1}^2 \cdots f_{j-1}^2, \end{aligned} \tag{4.28}$$

for $K + 1 - k \leq j \leq J - 1$. Formula (4.28) is similar to Approach 2 (see (4.16)) if we choose smartly the position j at which we want to average.

In fact, it can be shown that the Mack formula is a linear approximation to our result:

$$\begin{aligned} \widehat{E} \left[X_{k,K-k+1}^2 \cdot \left(\widehat{f}_{K-k+1}^{\mathcal{D}_K} \cdots \widehat{f}_{j-1}^{\mathcal{D}_K} - f_{K-k+1} \cdots f_{j-1} \right)^2 \middle| \mathcal{D}_K \right] & \tag{4.29} \\ = X_{k,K-k+1}^2 \cdot \left(\prod_{l=K-k+1}^{j-1} \left(\widehat{f}_l^2 + \frac{\widehat{\sigma}_l^2}{S_l} \right) - \prod_{l=K-k+1}^{j-1} \widehat{f}_l^2 \right) \end{aligned}$$

$$\begin{aligned}
 &= X_{k, K-k+1}^2 \cdot \prod_{l=K-k+1}^{j-1} \hat{f}_l^2 \cdot \underbrace{\left(\prod_{l=K-k+1}^{j-1} \left(1 + \frac{\hat{\sigma}_l^2}{S_l \cdot \hat{f}_l^2} \right) - 1 \right)}_{\approx \sum_{l=K-k+1}^{j-1} \frac{\hat{\sigma}_l^2 / \hat{f}_l^2}{S_l}} \\
 &\approx X_{k, K-k+1}^2 \cdot \prod_{l=K-k+1}^{j-1} \hat{f}_l^2 \cdot \sum_{l=K-k+1}^{j-1} \frac{\hat{\sigma}_l^2 / \hat{f}_l^2}{S_l},
 \end{aligned}$$

which is the Mack term for the estimation error (cf. [2], page 219).

In the practical examples that we have looked at the numerical differences resulting from our formula presented in (4.22) and the Mack formula is rather small (see Table 5).

Comparison to the result of Murphy [4].

We are grateful to one of the referees of our paper for having pointed out that our numerical results for the ultimate losses coincide with those obtained on the basis of the weighted average development (WAD) factor model in Murphy [4], Model IV.

If one looks at this paper [4] one finds indeed that the WAD model coincides with ours. It is also the chain ladder model stated in time series language. To obtain the crucial recursive formula for the estimation error (Theorem 3 in Appendix C of [4]) Murphy assumes independence of estimates of the chain ladder factors. This assumption is inconsistent with the model assumptions. One can easily see that chain ladder factors are uncorrelated (see Mack [2], Theorem 2). But other direct calculations show that the squares of chain ladder estimates are negatively correlated.

The point is that Murphy by his assumptions gets a multiplicative structure of the measure of volatility. In this paper we get the multiplicative structure by the choice of Approach 3 for the measure of the (conditional) volatility of the chain ladder estimate (see discussion in Section 4.1.2). Henceforth, since in both estimates one uses a multiplicative structure it turns out that our recursive estimate (4.22) is exactly the estimate presented in Theorem 3 of Murphy [4].

4.3. Aggregation of ultimate loss over different accident years

Consider two different accident years $k < l$. From our assumptions we know that the ultimate losses $X_{k,J}$ and $X_{l,J}$ are independent. Nevertheless we have to be careful if we aggregate $\hat{X}_{k,J}$ and $\hat{X}_{l,J}$. The estimators are no longer independent since they use the same observations for estimating the age-to-age factors f_j .

$$\begin{aligned}
 &E[(\hat{X}_{k,J} + \hat{X}_{l,J} - (X_{k,J} + X_{l,J}))^2 | \mathcal{D}_K] \\
 &= \text{Var}(X_{k,J} + X_{l,J} | \mathcal{D}_K) + (\hat{X}_{k,J} + \hat{X}_{l,J} - E[X_{k,J} + X_{l,J} | \mathcal{D}_K])^2.
 \end{aligned} \tag{4.30}$$

Using the independence of the different accident years, we obtain for the first term

$$\text{Var}(X_{k,J} + X_{l,J} | \mathcal{D}_K) = \text{Var}(X_{k,J} | \mathcal{D}_K) + \text{Var}(X_{l,J} | \mathcal{D}_K), \tag{4.31}$$

whereas for the second term we obtain

$$\begin{aligned} & (\widehat{X}_{k,J} + \widehat{X}_{l,J} - E[X_{k,J} + X_{l,J} | \mathcal{D}_K])^2 \\ &= (\widehat{X}_{k,J} - E[X_{k,J} | \mathcal{D}_K])^2 + (\widehat{X}_{l,J} - E[X_{l,J} | \mathcal{D}_K])^2 \\ & \quad + 2 \cdot (\widehat{X}_{k,J} - E[X_{k,J} | \mathcal{D}_K]) \cdot (\widehat{X}_{l,J} - E[X_{l,J} | \mathcal{D}_K]). \end{aligned} \tag{4.32}$$

Hence we have the following decomposition for the (conditional) prediction error of the sum of two accident years

$$\begin{aligned} & E[(\widehat{X}_{k,J} + \widehat{X}_{l,J} - (X_{k,J} + X_{l,J}))^2 | \mathcal{D}_K] \\ &= E[(\widehat{X}_{k,J} - X_{k,J})^2 | \mathcal{D}_K] + E[(\widehat{X}_{l,J} - X_{l,J})^2 | \mathcal{D}_K] \\ & \quad + 2 \cdot (\widehat{X}_{k,J} - E[X_{k,J} | \mathcal{D}_K]) \cdot (\widehat{X}_{l,J} - E[X_{l,J} | \mathcal{D}_K]). \end{aligned} \tag{4.33}$$

In addition to the (conditional) mean square error of prediction of single accident years, we need to determine the volatilities of \hat{f}_j for the cross-products terms, which gives an additional term in the estimation errors for aggregated accident years.

$$\begin{aligned} & (\widehat{X}_{k,J} - E[X_{k,J} | \mathcal{D}_K]) \cdot (\widehat{X}_{l,J} - E[X_{l,J} | \mathcal{D}_K]) \\ &= X_{k,K-k+1} \cdot (\hat{f}_{K-k+1} \cdot \dots \cdot \hat{f}_{J-1} - f_{K-k+1} \cdot \dots \cdot f_{J-1}) \\ & \quad \cdot X_{l,K-l+1} \cdot (\hat{f}_{K-l+1} \cdot \dots \cdot \hat{f}_{J-1} - f_{K-l+1} \cdot \dots \cdot f_{J-1}). \end{aligned} \tag{4.34}$$

As in (4.20) (Approach 3) this term is (conditionally) estimated by

$$\begin{aligned} & E \left[X_{k,K-k+1} X_{l,K-l+1} \left(\prod_{j=K-k+1}^{J-1} \hat{f}_j^{\mathcal{D}_K} - \prod_{j=K-k+1}^{J-1} f_j \right) \left(\prod_{j=K-l+1}^{J-1} \hat{f}_j^{\mathcal{D}_K} - \prod_{j=K-l+1}^{J-1} f_j \right) \middle| \mathcal{D}_K \right] \\ &= X_{k,K-k+1} \cdot X_{l,K-l+1} \cdot f_{K-l+1} \cdot \dots \cdot f_{K-k} \cdot \left(\prod_{j=K-k+1}^{J-1} E \left[(\hat{f}_j^{\mathcal{D}_K})^2 \middle| \mathcal{D}_K \right] - \prod_{j=K-k+1}^{J-1} f_j^2 \right) \\ &= X_{k,K-k+1} \cdot E[X_{l,K-k+1} | X_{l,K-l+1}] \cdot \left(\prod_{j=K-k+1}^{J-1} \left(f_j^2 + \frac{\sigma_j^2}{S_j} \right) - \prod_{j=K-k+1}^{J-1} f_j^2 \right). \end{aligned} \tag{4.35}$$

But then the estimation of the covariance term is straightforward from the estimate of a single accident year.

Result 4.2. (MSEP for aggregated accident years). *Under Assumptions (T1), (T2), (T1'), (T2') for the estimation of the process variance and the estimation error respectively, we have the following estimator for the (conditional) mean square prediction error of the ultimate loss of aggregated accident years*

$$\begin{aligned} & \widehat{E} \left[\left(\sum_{k=K+2-J}^K \widehat{X}_{k,J} - \sum_{k=K+2-J}^K X_{k,J} \right)^2 \middle| \mathcal{D}_K \right] \\ & := \sum_{k=K+2-J}^K \widehat{E} \left[\left(\widehat{X}_{k,J} - X_{k,J} \right)^2 \middle| \mathcal{D}_K \right] + 2 \cdot \sum_{k+2-J \leq k < l \leq K} X_{k,K-k+1} \cdot \widehat{X}_{l,K-k+1} \cdot \Delta_{k,J}^2. \end{aligned} \tag{4.36}$$

Remarks:

1) The last term (covariance terms) from the result above can be rewritten as

$$2 \cdot \sum_{k+2-J \leq k < l \leq K} \frac{\widehat{X}_{l,K-k+1}}{X_{k,K-k+1}} \cdot X_{k,K-k+1}^2 \cdot \Delta_{k,J}^2, \tag{4.37}$$

where $X_{k,K-k+1}^2 \cdot \Delta_{k,J}^2$ is the (conditional) estimation error of the single accident year k (see (4.21)). This may be helpful in the implementation since it leads to matrix multiplications.

2) Note that in this section we have presented the aggregation of only two ultimate losses $X_{k,J}$ and $X_{l,J}$ which leads to one covariance term. If we aggregate over all accident years, of course, we obtain the sum over all covariance terms $l \neq k$.

4.4. Prediction error for accounting years

In the sequel X_t denotes the total of the cumulative payments originating from all accident years up to accounting year t ($1 \leq t \leq K + J - 1$). Formally this corresponds to the sum of the $X_{k,j}$ over the t -diagonal

$$X_t := \sum_{\substack{k+j=t+1 \\ 1 \leq j \leq J}} X_{k,j}. \tag{4.38}$$

We are at time K , i.e. the total payments X_t are observed for $t \leq K$. Conveniently, for accident year k we denote the last observation by X_{k,k^*} with $k + k^* = K + 1$. We want to predict $X_{K+\tau}$ for $1 \leq \tau \leq J - 1$ and the corresponding estimator is given by

$$\widehat{X}_{K+\tau} := \sum_{\substack{k+j=K+\tau+1 \\ t+1 \leq j \leq J}} \widehat{X}_{k,j}. \tag{4.39}$$

Sometimes it is easier to write this sum as

$$\widehat{X}_{K+\tau} = \sum_{k=K+\tau+1-J}^K \widehat{X}_{k,k^*+\tau} = \sum_{k^*=1}^{J-\tau} \widehat{X}_{k,k^*+\tau}. \tag{4.40}$$

The (conditional) prediction error of $\widehat{X}_{K+\tau}$ can be decomposed as follows (see (4.3))

$$\begin{aligned} & E \left[\left(\widehat{X}_{K+\tau} - X_{K+\tau} \right)^2 \middle| \mathcal{D}_K \right] \\ &= \underbrace{\text{Var}(X_{K+\tau} | \mathcal{D}_K)}_{\text{process variance}} + \underbrace{\left(\widehat{X}_{K+\tau} - E[X_{K+\tau} | \mathcal{D}_K] \right)^2}_{\text{estimation error}}. \end{aligned} \tag{4.41}$$

Using (4.39), (4.40) and the independence of the accident years we obtain for the process variance of $\widehat{X}_{K+\tau}$

$$\text{Var}(X_{K+\tau} | \mathcal{D}_K) = \sum_{k=K+\tau+1-J}^K \text{Var}(X_{k,k^*+\tau} | \mathcal{D}_K). \tag{4.42}$$

Finally, using (4.8) we obtain as estimator for the (conditional) process variance of $\widehat{X}_{K+\tau}$

$$\widehat{E} \left[\left(X_{K+\tau} - E[X_{K+\tau} | \mathcal{D}_K] \right)^2 \middle| \mathcal{D}_K \right] := \sum_{\substack{k+j=K+\tau+1 \\ \tau+1 \leq j \leq J}} \Gamma_{k,j}^2. \tag{4.43}$$

For the estimation error we proceed as in Subsection 4.1.2 (Approach 3). We average

$$\left(\widehat{X}_{K+\tau} - E[X_{K+\tau} | \mathcal{D}_K] \right)^2$$

over the resampled chain ladder factors $\widehat{f}_{k^*}^{\mathcal{D}_K}, \dots, \widehat{f}_{k^*+\tau-1}^{\mathcal{D}_K}$, given the upper trapezoid \mathcal{D}_K , to determine the (conditional) estimation error of $\widehat{X}_{K+\tau}$. We define

$$\widehat{g}_{k^*,\tau}^{\mathcal{D}_K} := \widehat{f}_{k^*}^{\mathcal{D}_K} \cdot \dots \cdot \widehat{f}_{k^*+\tau-1}^{\mathcal{D}_K} \quad \text{and} \quad g_{k^*,\tau} := f_{k^*} \cdot \dots \cdot f_{k^*+\tau-1}. \tag{4.44}$$

Observe that

$$\begin{aligned} & \left(\widehat{X}_{K+\tau} - E[X_{K+\tau} | \mathcal{D}_K] \right)^2 = \left(\sum_{k^*=1}^{J-\tau} \left(\widehat{X}_{k,k^*+\tau} - E[X_{k,k^*+\tau} | \mathcal{D}_K] \right) \right)^2 \\ &= \left(\sum_{k^*=1}^{J-\tau} X_{k,k^*} \cdot \left(\widehat{f}_{k^*} \cdot \dots \cdot \widehat{f}_{k^*+\tau-1} - f_{k^*} \cdot \dots \cdot f_{k^*+\tau-1} \right) \right)^2. \end{aligned} \tag{4.45}$$

Hence we calculate the average over the (conditionally) resampled observations as follows

$$\begin{aligned}
 & E \left[\left(\sum_{k^*=1}^{J-\tau} X_{k,k^*} \cdot \left(\hat{f}_{k^*}^{\mathcal{D}_K} \cdot \dots \cdot \hat{f}_{k^*+\tau-1}^{\mathcal{D}_K} - f_{k^*} \cdot \dots \cdot f_{k^*+\tau-1} \right) \right)^2 \middle| \mathcal{D}_K \right] \\
 &= E \left[\sum_{k^*,l^*=1}^{J-\tau} X_{k,k^*} \cdot X_{l,l^*} \cdot \left(\hat{g}_{k^*,\tau}^{\mathcal{D}_K} - g_{k^*,\tau} \right) \cdot \left(\hat{g}_{l^*,\tau}^{\mathcal{D}_K} - g_{l^*,\tau} \right) \middle| \mathcal{D}_K \right].
 \end{aligned} \tag{4.46}$$

Next, we use $\Delta_{k,k^*+\tau}^2$ (cf. (4.21)) and replace $f_{k^*}, \dots, f_{l^*-\tau}, f_{k^*+\tau}, \dots, f_{l^*+\tau-1}$ by their estimators. Thus we obtain the following (conditional) estimator for the estimation error:

$$\begin{aligned}
 & \hat{E}_{\mathcal{D}_K} \left[\left(\hat{X}_{K+\tau} - E[X_{K+\tau} | \mathcal{D}_K] \right)^2 \right] \\
 &:= \sum_{k^*=1}^{J-\tau} X_{k,k^*}^2 \cdot \Delta_{k,k^*+\tau}^2 + 2 \cdot \sum_{k^*=1}^{J-\tau} \sum_{\substack{l^* > k^* \\ \tau+1 \leq j \leq J}}^{\min\{k^*+\tau-1, J-\tau\}} \hat{X}_{k,l^*} \cdot X_{l,l^*} \cdot \Delta_{l,k^*+\tau}^2 \cdot \prod_{m=k^*+\tau}^{l^*+\tau-1} \hat{f}_m.
 \end{aligned} \tag{4.47}$$

From our results (4.43) and (4.47) we obtain for the (conditional) prediction error of an accounting year:

Result 4.3. (MSEP for accounting years, cumulative payments) *For the estimator of the (conditional) prediction error (4.41) we have*

$$\begin{aligned}
 & \hat{E} \left[\left(\hat{X}_{K+\tau} - X_{K+\tau} \right)^2 \middle| \mathcal{D}_K \right] := \sum_{\substack{k+j=K+\tau+1 \\ \tau+1 \leq j \leq J}} \hat{E} \left[\left(\hat{X}_{k,j} - X_{k,j} \right)^2 \middle| \mathcal{D}_K \right] \\
 &+ 2 \cdot \sum_{k^*=1}^{J-\tau} \sum_{\substack{l^* > k^* \\ \tau+1 \leq j \leq J}}^{\min\{k^*+\tau-1, J-\tau\}} \hat{X}_{k,l^*} \cdot X_{l,l^*} \cdot \Delta_{l,k^*+\tau}^2 \cdot \prod_{m=k^*+\tau}^{l^*+\tau-1} \hat{f}_m,
 \end{aligned} \tag{4.48}$$

Analogously for the estimated non-cumulative payments in accounting year $K + \tau$ we have

$$\hat{Y}_{K+\tau} = \sum_{\substack{k+j=K+1+\tau \\ \tau+1 \leq j \leq J}} \hat{Y}_{k,j} = \sum_{k^*=1}^{J-\tau} \left(\hat{X}_{k,k^*+\tau} - \hat{X}_{k,k^*+\tau-1} \right). \tag{4.49}$$

The corresponding (conditional) prediction error can again be decomposed:

$$E \left[\left(\hat{Y}_{K+\tau} - Y_{K+\tau} \right)^2 \middle| \mathcal{D}_K \right] = \underbrace{\text{Var}(Y_{K+\tau} | \mathcal{D}_K)}_{\text{process variance}} + \underbrace{\left(\hat{Y}_{K+\tau} - E[Y_{K+\tau} | \mathcal{D}_K] \right)^2}_{\text{estimation error}}. \tag{4.50}$$

Using (4.49), (4.10) and the independence of the accident years we obtain for the (conditional) process variance of $\widehat{Y}_{K+\tau}$ the estimator

$$\begin{aligned} \widehat{E} \left[(Y_{K+\tau} - E[Y_{K+\tau} | \mathcal{D}_K])^2 \middle| \mathcal{D}_K \right] & \\ & := \sum_{\substack{k+j=K+1+\tau \\ \tau+1 \leq j \leq J}} \left(\Gamma_{k,j}^2 + \Gamma_{k,j-1}^2 \cdot (1 - 2 \cdot \widehat{f}_{j-1}) \right). \end{aligned} \tag{4.51}$$

Using (4.49) and (4.44) we get for the estimation error of $\widehat{Y}_{K+\tau}$ (Approach 3)

$$\begin{aligned} & E \left[\left(\sum_{k^*=1}^{J-\tau} X_{k,k^*} \cdot \left(\widehat{g}_{k^*,\tau-1}^{\mathcal{D}_K} \cdot \left(\widehat{f}_{k^*+\tau-1}^{\mathcal{D}_K} - 1 \right) - g_{k^*,\tau-1} \cdot (f_{k^*+\tau-1} - 1) \right) \right)^2 \middle| \mathcal{D}_K \right] \\ & = \sum_{k^*=1}^{J-\tau} X_{k,k^*}^2 \cdot E \left[\left(\widehat{g}_{k^*,\tau-1}^{\mathcal{D}_K} \cdot \left(\widehat{f}_{k^*+\tau-1}^{\mathcal{D}_K} - 1 \right) - g_{k^*,\tau-1} \cdot (f_{k^*+\tau-1} - 1) \right)^2 \middle| \mathcal{D}_K \right] \\ & + 2 \cdot \sum_{k^*=1}^{J-\tau} \min \{ k^* + \tau - 1, J - \tau \} \sum_{l^* > k^*} E \left[X_{k,k^*} \cdot X_{l,l^*} \cdot \left(\left(\widehat{g}_{k^*,\tau}^{\mathcal{D}_K} - \widehat{g}_{k^*,\tau-1}^{\mathcal{D}_K} \right) - (g_{k^*,\tau} - g_{k^*,\tau-1}) \right) \right. \\ & \quad \left. \cdot \left(\left(\widehat{g}_{l^*,\tau}^{\mathcal{D}_K} - \widehat{g}_{l^*,\tau-1}^{\mathcal{D}_K} \right) - (g_{l^*,\tau} - g_{l^*,\tau-1}) \right) \middle| \mathcal{D}_K \right]. \end{aligned} \tag{4.52}$$

As before, we obtain in addition to the terms of a single accident years the covariance terms, which we need to calculated separately. If we proceed as in (4.47) we obtain for the single covariance terms in (4.52) the following estimate ($l^* > k^*$)

$$\Lambda_{k,l} := \widehat{X}_{k,l^*} \cdot X_{l,l^*} \cdot [\Delta_{l,k^*+\tau}^2 - \Delta_{l,k^*+\tau-1}^2 \cdot \widehat{f}_{k^*+\tau-1}] \cdot \prod_{m=k^*+\tau}^{l^*+\tau-2} \widehat{f}_m \cdot (\widehat{f}_{l^*+\tau-1} - 1). \tag{4.53}$$

So, if we collect all our terms, we obtain

Result 4.4 (MSEP for accounting years, incremental payments) *For the (conditional) prediction error (4.50) we have the estimator*

$$\begin{aligned} E \left[(\widehat{Y}_{K+\tau} - Y_{K+\tau})^2 \middle| \mathcal{D}_K \right] & := \sum_{\substack{k+j=K+1+\tau \\ \tau+1 \leq j \leq J}} \widehat{E} \left[(\widehat{Y}_{k,j} - Y_{k,j})^2 \middle| \mathcal{D}_K \right] \\ & + 2 \cdot \sum_{k^*=1}^{J-\tau} \min \{ k^* + \tau - 1, J - \tau \} \sum_{l^* > k^*} \Lambda_{k,l}, \end{aligned} \tag{4.54}$$

where (see (4.10) and (4.24)) **(MSEP for single accident year, incremental payments)**

$$\widehat{E}\left[\left(\widehat{Y}_{k,j} - Y_{k,j}\right)^2 \middle| \mathcal{D}_K\right] := \Gamma_{k,j}^2 + \Gamma_{k,j-1}^2 \cdot \left(1 - 2 \cdot \widehat{f}_{j-1}\right) + X_{k,K-k+1}^2 \cdot \left(\Delta_{k,j}^2 - \left(2\widehat{f}_{j-1} - 1\right) \cdot \Delta_{k,j-1}^2\right). \tag{4.55}$$

5. EXAMPLE

We use the first run off triangle of cumulative payments given in Mack [2] (cf. Table 1 in [2]).

TABLE 3
RUN-OFF-TRIANGLE (CUMULATIVE PAYMENTS), SOURCE [2].

AY <i>k</i>	Development period <i>j</i>									
	1	2	3	4	5	6	7	8	9	10
1	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
2	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	
3	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315		
4	310608	1418858	2195047	3757447	4029929	4381982	4588268			
5	443160	1136350	2128333	2897821	3402672	3873311				
6	396132	1333217	2180715	2985752	3691712					
7	440832	1288463	2419861	3483130						
8	359480	1421128	2864498							
9	376686	1363294								
10	344014									

The estimators \widehat{f}_j of the chain ladder factors (cf. Table 4) show, that the example chosen is a relatively simple run off triangle.

TABLE 4
ESTIMATORS OF THE CHAIN LADDER FACTORS \widehat{f}_j .

\widehat{f}_j								
1	2	3	4	5	6	7	8	9
3.490607	1.747333	1.457413	1.173852	1.103824	1.086269	1.053874	1.076555	1.017725

For the estimation of $\sigma_{\mathcal{D}_K}^2$ we use the formula given in [2], before Theorem 3. Hence, using the estimators \widehat{f}_j and $\widehat{\sigma}_j^2$ we find an estimate (and the corresponding error) for the aggregate reserve of all accident years $\sum_k R_k$.

In Table 5 the process standard deviation, (estimation error)^{1/2}, MSPE and standard error of prediction for the aggregated ultimate loss over (all) different accident years are given. We see that the results from Mack's formula and our formula presented in (4.22) are nearly the same. But since the Mack formula for the estimation error is a linear bound from below to our formula it is also clear that our results are slightly higher compared to Mack's results.

Other examples that we have examined (long and short term business) confirm the findings in this example.

TABLE 5
MSEP AND STANDARD ERROR OF PREDICTION FOR AGGREGATED ACCIDENT YEARS.

	reserves $\Sigma_k \hat{R}_k$	process std.dev.	$\sqrt{\text{estim error}}$	MSEP	pred. std. error
Mack	18'680'856	1'878'292	1'568'532	5'988'273'257'923	2'447'095
BBMW	18'680'856	1'878'292	1'569'349	5'990'835'395'887	2'447'618
difference	0	0	817	2'562'137'964	523

REFERENCES

- [1] BUCHWALDER, M., BÜHLMANN H., MERZ, M. and WÜTHRICH, M.V. (2005) *Legal valuation portfolio in non-life insurance*. Conference paper presented at the 36th International ASTIN Colloquium, 4-7 September 2005, ETH Zürich. available under www.astin2005.ch
- [2] MACK, T. (1993) Distribution-free calculation of the standard error of Chain Ladder reserve estimates. *Astin Bulletin* **23(2)**, 213-225.
- [3] MACK, T. (1999) The standard error of Chain Ladder reserve estimates: recursive calculation and inclusion of a tail factor. *Astin Bulletin* **29(2)**, 361-366.
- [4] MURPHY, D.M. (1994) Unbiased loss development factors. *Proc. CAS*, **LXXXI**, 154-222.
- [5] SCHMIDT, K.D. and SCHNAUS, A. (1996). An extension of Mack's model for the chain ladder method. *Astin Bulletin* **26(2)**, 247-262.

MARKUS BUCHWALDER

Bâloise, Aeschengraben 21, CH-4002 Basel, Switzerland

Hans BÜHLMANN

ETH Zürich, Department of Mathematics, CH-8092 Zürich, Switzerland

MICHAEL MERZ

University Tübingen, Faculty of Economics, D-72074 Tübingen, Germany

MARIO V. WÜTHRICH

ETH Zürich, Department of Mathematics, CH-8092 Zürich, Switzerland