CHAIN LADDER AND ERROR PROPAGATION

BY

ANCUS RÖHR

ABSTRACT

We show how estimators for the chain ladder prediction error in Mack's (1993) distribution-free stochastic model can be derived using the error propagation formula. Our method allows for the treatment of the general case of the prediction error of the loss development result between two arbitrary future horizons. In the well-known special cases considered previously by Mack (1993) and Merz and Wüthrich (2008), our estimators coincide with theirs. However, the algebraic form in which we cast them is new, considerably more compact and more intuitive to understand. For example, in the classical case treated by Mack (1993), we show that the mean squared prediction error divided by the squared estimated ultimate loss can be written as $\sum_j \hat{u}_j^2 \hat{q}_j$, where $\hat{u}_j$ measures the (relative) uncertainty around the jth development factor and $\hat{q}_j$ the proportion of the estimated ultimate loss that it affects. The error propagation method also provides a natural split into process error and parameter error. Our proofs identify and exploit symmetries of "chain ladder processes" in a novel way. For the sake of wider practical applicability of the formulae derived, we allow for incomplete historical data and the exclusion of outliers in the triangles.

KEYWORDS

Chain ladder, mean squared prediction error, Mack error formula, claims development result, solvency.

1. INTRODUCTION

The idea of this paper is to derive approximations for the chain ladder prediction error in Mack's (1993) distribution-free stochastic model using the error propagation formula. This approach seems to have been given little attention in the literature — in Mack (2008), it is mentioned in passing in the introduction, and in Matitschka (2010) it is applied in an analysis of the prediction error within the overdispersed Poisson model. Gisler (2013) analyzed the chain ladder error using Taylor approximation in 1992, but never published his results. In our view,

Astin Bulletin 46(2), 293-330. doi: 10.1017/asb.2016.9

the error propagation approach has advantages in that it is very straightforward and gives a transparent derivation.

The error propagation principle (see Ku, 1966) linearizes a random variable which is a function of other random variables around a chosen point using first-order Taylor approximation, and then reduces the stochastic analysis of quadratic deviations of the function values to an analysis of the (co-)variances of the variables occurring in the Taylor series. In our context, the latter variables are the link ratios from the chain ladder triangle, and the function of the link ratios is the ultimate loss.

The approximation of the mean squared prediction error for the ultimate loss which we find agrees with Mack's result (Mack, 1993). However, our method is different from Mack's, and we also present the result in a new form. Using the quantities $\hat{u}_j$ and $\hat{q}_j$ defined in Section 2.7, of which $\hat{u}_j$ measures the (relative) accuracy of the development factor estimate $\hat{f}_j$ and $\hat{q}_j$ the proportion of the predicted ultimate loss that $\hat{f}_j$ affects, our estimator for the (relative) mean squared prediction error for the ultimate loss, i.e. the mean squared prediction error divided by the squared expected ultimate loss, takes on the very compact and intuitive form

$$\sum_j \hat{u}_j^2 \hat{q}_j,$$

where the sum runs over all development periods j. In our Main Result 5.3, we show that this formula generalizes to

$$\sum_j \hat{u}_j^2 \left(\frac{1}{1-\tilde{q}_j} - \frac{1}{1-\breve{q}_j} \right) \Big/ \frac{1}{1-\hat{q}_j}$$

for the relative mean squared prediction error of the claims development result between two arbitrary future horizons, the first one represented by the coefficients $\tilde{q}_j$ and the second one by the coefficients $\breve{q}_j$ (all defined in analogy to the $\hat{q}_j$). This new result is a generalization of the case of the one-year development horizon — which goes back to Merz and Wüthrich, e.g. (Merz and Wüthrich, 2008) — and apparently agrees (as numerical examples show) with recent results obtained by a different method by Merz and Wüthrich (2014). The general case is of interest in risk margin calculations for solvency purposes.

Result 5.3 also provides a split of the total mean squared error of prediction into process and parameter error parts.

Applying these formulae in practice, one often encounters situations where the historical data is incomplete or there are outliers in the data. For example, the upper left corner of some incurred claims triangle may not be available due to a change in claims reserving practice. Practitioners might be tempted to account for this by adapting the estimators $\hat{u}_j$ in the above formulae. However, more changes are required, as can be seen from Result 5.3.

The paper is organized as follows. In Section 2, we review the notation and basic facts about the chain ladder model. Much of this is standard material;

non-standard items include the definition of a "chain ladder process" and even a "non-homogeneous" variant of it, the emphasis on underlying symmetries of such processes, the aforementioned systematic treatment of incomplete historical triangle data and the possibility of allowing "ragged" triangles. Section 3 explains how to apply the error propagation formula to calculate the prediction error for the ultimate loss. In Section 4, we recover Mack's error formula (Mack, 1993), applying our method from Section 3. In Section 5, we extend our analysis to the claims development result up to an arbitrary loss development horizon and further between two future development horizons. In Section 6, we provide a calculated example. In the appendix, we explicitly verify the agreement of our formulae with previously published formulae for the ultimate development and the case of the one-year development horizon.

2. The chain ladder method

2.1. General notation

Throughout this article, let $E[X]$ and $V[X]$ denote the expectation and variance of a random variable X, respectively. Conditional expectation and variance of X, given the σ-algebra $\sigma[Y]$ generated by another random variable Y, are denoted $E[X|Y]$ and $V[X|Y]$, respectively.

The shorthand $\mathbf{1}_Q$ is used as an indicator function with value 1 when the condition Q is true, and 0 otherwise. For example, $\mathbf{1}_{i<j}$ is 1 if $i < j$, and 0 otherwise.

Partial differentiation is denoted $\partial_x f$ rather than the more customary $\frac{\partial f}{\partial x}$.

2.2. Chain ladder processes

In this section, we recall the definition and basic properties of Mack's distribution-free stochastic model (Mack, 1993). All of the material is standard (Mack, 1993; Wüthrich and Merz, 2008), except that we dwell a little longer than usual on the stochastic process aspects underlying the model. The discussion of loss development triangles with their double indices is deferred to Section 2.4; here, we consider the development of a "single accident year", for example, for which we need one development time index j only. A single accident year would be an example of a "homogeneous chain ladder process", as we will call it. But Mack's axioms allow to accommodate additional terms ("A_j" in the following definition) which serve to describe processes more general than that. These give rise to "non-homogeneous" chain ladder processes. We will encounter such a process in Section 4.1 — it arises from a loss development triangle by aggregating the future development column by column. In this case, development is relevant only conditionally, given the historical triangle $\mathcal{D}$, which is why we include the σ-algebra $\mathcal{A}$ (think of $\mathcal{A} = \sigma[\mathcal{D}]$) in the definition below.

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

Because of this terminology and the slight generalization over standard treatments, we spell out the adapted, but otherwise well-known proofs.

Definition 2.1. *Let* $\{X_j\}_{j\geq 0}$ *be a discrete time, real-valued stochastic process, and* $\mathcal{A}$ *a* σ*-algebra such that* $\sigma[X_0] \subseteq \mathcal{A}$. *Furthermore, denote by*

$$\mathcal{X}_j := \text{the } \sigma\text{-algebra generated by } \mathcal{A},\, X_0,\, X_1, \ldots,\, X_j.$$

Then we call $\{X_j\}_{j\geq 0}$ *a* chain ladder process (given $\mathcal{A}$)*, if each* X_j *is positive and if for each* $j > 0$*, there exist positive real numbers* f_j*,* ϕ_j *and an* $\mathcal{A}$*-measurable real random variable* A_j *such that*

$$E\left[X_j|\mathcal{X}_{j-1}\right] = f_j X_{j-1} + A_j,$$
$$V\left[X_j|\mathcal{X}_{j-1}\right] = \phi_j X_{j-1}.$$

The parameter f_j *is called the* development factor *with index* j *and the quotient*

$$F_j := \frac{X_j - A_j}{X_{j-1}}$$

the link ratio *with index* j*. We say that the chain ladder process is* homogeneous*, if* $\sigma[X_0] = \mathcal{A}$ *and* $A_j = 0$ *for all* $j > 0$*. Finally, for notational convenience, we define an* A_j *also for* $j = 0$ *by setting* $A_0 := X_0$.

Note that even in the case of a homogeneous process, where $\mathcal{A} = \sigma[X_0]$, conditioning with respect to $\mathcal{A}$ makes sense, since X_0 represents the starting point of the recursive formulae for the expectation values and variances of subsequent elements. Without knowing the value of X_0, little can be inferred about the future development. Also, the random variable X_0 itself is not required to have finite moments, whereas for the conditional expectations and variances of the X_j, $j > 0$, the finiteness is an assumption implicit in the formulae.

Link ratios are themselves random variables, but the f_j and ϕ_j are parameters of the chain ladder process which need to be estimated from data.

The following proposition summarizes the well-known basic properties of the distribution-free model (see Mack, 1993).

Proposition 2.2. *For every integer* $j > 0$*, define the map* α_j *by*

$$\alpha_j\,[x] := f_j x + A_j.$$

Then for any integers $j \geq k > 0$*,*

$$\begin{aligned} E\left[X_j|\mathcal{X}_{k-1}\right] &= f_j E[X_{j-1}|\mathcal{X}_{k-1}] + A_j \\ &= (\alpha_j \circ \alpha_{j-1} \circ \ldots \circ \alpha_k)\,[X_{k-1}] \qquad (1) \\ &= (\alpha_j \circ \alpha_{j-1} \circ \ldots \circ \alpha_k)\,[0] + f_j \cdot \ldots \cdot f_k \cdot X_{k-1} \end{aligned}$$

$$V\left[X_j|\mathcal{X}_{k-1}\right] = \phi_j E[X_{j-1}|\mathcal{X}_{k-1}] + f_j^2\, V\left[X_{j-1}|\mathcal{X}_{k-1}\right], \qquad (2)$$

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

where the circle denotes map composition;

$$E\left[F_j|\mathcal{X}_{k-1}\right] = f_j = E[F_j] \tag{3}$$

$$V\left[F_j|\mathcal{X}_{k-1}\right] = \phi_j E[\frac{1}{X_{j-1}}|\mathcal{X}_{k-1}] \geq \frac{\phi_j}{E[X_{j-1}|\mathcal{X}_{k-1}]} \tag{4}$$

and for any integers $i > j \geq k > 0$,

$$E\left[F_i F_j|\mathcal{X}_{k-1}\right] = f_i f_j = E[F_i F_j] \tag{5}$$

which means that F_i and F_j are uncorrelated. Finally, all statements remain true if we replace $\mathcal{X}_{k-1}$ by $\sigma[X_{k-1}, \mathcal{A}]$.

Proof. We prove the first statement by induction on j, the starting point $j = k$ being identical to the condition from the definition of a chain ladder process. Assuming the statement to be true up to index $j-1$, we have

$$\begin{aligned} E\left[X_j|\mathcal{X}_{k-1}\right] &= E\left[E\left[X_j|\mathcal{X}_{j-1}\right]|\mathcal{X}_{k-1}\right] = E\left[f_j X_{j-1} + A_j|\mathcal{X}_{k-1}\right] \\ &= f_j E\left[X_{j-1}|\mathcal{X}_{k-1}\right] + A_j \\ &= f_j \left((\alpha_{j-1} \circ \ldots \circ \alpha_k)[0] + f_{j-1} \cdot \ldots \cdot f_k \cdot X_{k-1}\right) + A_j \\ &= (\alpha_j \circ \ldots \circ \alpha_k)[0] + f_j \cdot \ldots \cdot f_k \cdot X_{k-1} \end{aligned}$$

as claimed. For the second statement, look at

$$\begin{aligned} V\left[X_j|\mathcal{X}_{k-1}\right] &= E\left[V\left[X_j|\mathcal{X}_{j-1}\right]|\mathcal{X}_{k-1}\right] + V\left[E\left[X_j|\mathcal{X}_{j-1}\right]|\mathcal{X}_{k-1}\right] \\ &= E\left[\phi_j X_{j-1}|\mathcal{X}_{k-1}\right] + V\left[f_j X_{j-1} + A_j|\mathcal{X}_{k-1}\right] \\ &= \phi_j E[X_{j-1}|\mathcal{X}_{k-1}] + f_j^2 V\left[X_{j-1}|\mathcal{X}_{k-1}\right]. \end{aligned}$$

To prove the third claim, consider

$$\begin{aligned} E\left[F_j|\mathcal{X}_{k-1}\right] &= E\left[E\left[F_j|\mathcal{X}_{j-1}\right]|\mathcal{X}_{k-1}\right] \\ &= E\left[E\left[\frac{X_j - A_j}{X_{j-1}}|\mathcal{X}_{j-1}\right]|\mathcal{X}_{k-1}\right] = E\left[f_j|\mathcal{X}_{k-1}\right] = f_j. \end{aligned}$$

For the fourth statement, we have

$$\begin{aligned} V\left[F_j|\mathcal{X}_{k-1}\right] &= E\left[V\left[F_j|\mathcal{X}_{j-1}\right]|\mathcal{X}_{k-1}\right] + V\left[E\left[F_j|\mathcal{X}_{j-1}\right]|\mathcal{X}_{k-1}\right] \\ &= E\left[\left.\frac{\phi_j X_{j-1}}{X_{j-1}^2}\right| \mathcal{X}_{k-1}\right] + V\left[f_j|\mathcal{X}_{k-1}\right] \\ &= \phi_j E\left[\left.\frac{1}{X_{j-1}}\right| \mathcal{X}_{k-1}\right] \geq \frac{\phi_j}{E\left[X_{j-1}|\mathcal{X}_{k-1}\right]}, \end{aligned}$$

where at the end we have used Jensen's inequality. As to the fifth claim, we have

$$E\left[F_iF_j|\mathcal{X}_{k-1}\right] = E[E[F_iF_j|\mathcal{X}_{i-1}]|\mathcal{X}_{k-1}] = E[F_jE[F_i|\mathcal{X}_{i-1}]|\mathcal{X}_{k-1}]$$
$$= E[F_jf_i|\mathcal{X}_{k-1}] = f_iE[F_j|\mathcal{X}_{k-1}] = f_if_j$$

as stated. Finally, our last claim follows by replacing all $\mathcal{X}_{k-1}$ by $\sigma[X_{k-1}, \mathcal{A}]$ (but keeping the $\mathcal{X}_{j-1}$ and $\mathcal{X}_{i-1}$) in the above proofs, using again induction and what we have proved already. ■

The inequality in (4) is sharp — in fact, we will use it as an approximation later on.

2.3. Aggregating chain ladder processes

Loss development triangles contain chain ladder processes which — by assumption — have identical sets of parameters. Before studying such triangles, we will show that any finite number of such (independent) processes may be combined into a new chain ladder process by summation. This fact will be used later on. We only need it for homogeneous chain ladder processes. We also look at what happens if we aggregate a homogeneous chain ladder process in the development direction by combining several steps into one.

Throughout this section, let $\{X_j\}_{j\geq 0}$ be a homogeneous chain ladder process with parameters f_j, ϕ_j.

Proposition 2.3. *Let $\{X'_j\}_{j\geq 0}$ be another homogeneous chain ladder process with parameters f'_j, ϕ'_j. Assume $\{X_j\}_{j\geq 0}$ and $\{X'_j\}_{j\geq 0}$ to be independent, and assume $f_j = f'_j$ and $\phi_j = \phi'_j$ for all j. Then the sum $\{X_j + X'_j\}_{j\geq 0}$ is a homogeneous chain ladder process with the same parameters f_j, ϕ_j.*

Proof. Denoting by $\mathcal{X}'_j$ the σ-algebra generated by $X'_0, X'_1, \ldots, X'_j$ and by $\mathcal{Y}_j$ the σ-algebra generated by the random variables $X_0 + X'_0, \ldots, X_j + X'_j$, we have

$$\begin{aligned}
E[X_j + X'_j|\mathcal{Y}_{j-1}] &= E[E[X_j + X'_j|\mathcal{X}_{j-1}, \mathcal{X}'_{j-1}]|\mathcal{Y}_{j-1}]\\
&= E[f_jX_{j-1} + f'_jX'_{j-1}|\mathcal{Y}_{j-1}]\\
&= E[f_j(X_{j-1} + X'_{j-1})|\mathcal{Y}_{j-1}]\\
&= f_j(X_{j-1} + X'_{j-1})
\end{aligned}$$

using the assumption $f_j = f'_j$ and the independence assumption. Furthermore,

$$\begin{aligned}
V[X_j + X'_j|\mathcal{Y}_{j-1}] = {}& E[V[X_j + X'_j|\mathcal{X}_{j-1}, \mathcal{X}'_{j-1}]|\mathcal{Y}_{j-1}]\\
&+ V[E[X_j + X'_j|\mathcal{X}_{j-1}, \mathcal{X}'_{j-1}]|\mathcal{Y}_{j-1}],
\end{aligned}$$

where the second term on the right-hand side vanishes by what we have just proved about the expectation value. By the independence assumption,

$$V[X_j + X'_j|\mathcal{X}_{j-1}, \mathcal{X}'_{j-1}] = V[X_j|\mathcal{X}_{j-1}] + V[X'_j|\mathcal{X}'_{j-1}] = \phi_j X_{j-1} + \phi'_j X'_{j-1},$$

and our claim now follows from the assumption $\phi_j = \phi'_j$. ■

There is a second way to "aggregate" a chain ladder process, namely, by combining several development steps into one. We will look into this now.

Proposition 2.4. *For any* $0 < k \leq j$*, set*

$$\phi_{k,j} := \frac{\phi_k}{f_k^2} f_k f_{k+1} \cdot \ldots \cdot f_j, \tag{6}$$

such that in particular $\phi_{j,j} = \phi_j / f_j$*. Then for* $j \geq k > 0$*,*

$$V[X_j|\mathcal{X}_{k-1}] = (\phi_{j,j} + \phi_{j-1,j} + \cdots + \phi_{k,j}) E[X_j|\mathcal{X}_{k-1}]. \tag{7}$$

Furthermore, if $j_0 < j_1 < j_2 < \ldots$ *forms a sub-sequence of the non-negative integers, then the stochastic process* $\{X'_k\}_{k\geq 0}$*, where* $X'_k := X_{j_k}$*, is again a homogeneous chain ladder process.*

Proof. For the first claim, we proceed by induction on j, starting with $j = k$, in which case the assertion follows directly from the defining properties of a chain ladder process. Assuming our claim to hold for $j - 1$, we get from (2), (1) and our homogeneity assumption

$$\begin{aligned} V[X_j|\mathcal{X}_{k-1}] &= \phi_j E[X_{j-1}|\mathcal{X}_{k-1}] + f_j^2 V\left[X_{j-1}|\mathcal{X}_{k-1}\right] \\ &= \left(\phi_j + f_j^2(\phi_{j-1,j-1} + \cdots + \phi_{k,j-1})\right) E[X_{j-1}|\mathcal{X}_{k-1}] \\ &= \left(f_j\phi_{j,j} + f_j(\phi_{j-1,j} + \cdots + \phi_{k,j})\right) E[X_{j-1}|\mathcal{X}_{k-1}] \\ &= \left(\phi_{j,j} + \phi_{j-1,j} + \cdots + \phi_{k,j}\right) E[X_j|\mathcal{X}_{k-1}], \end{aligned}$$

which completes the induction step and proves the first claim. From it, we conclude that if we jump from X_{k-1} directly to X_j, we have

$$E[X_j|\mathcal{X}_{k-1}] = f_j f_{j-1} \cdot \ldots \cdot f_k X_{k-1}, \tag{8}$$

$$V[X_j|\mathcal{X}_{k-1}] = (\phi_{j,j} + \phi_{j-1,j} + \cdots + \phi_{k,j}) f_j f_{j-1} \cdot \ldots \cdot f_k X_{k-1}, \tag{9}$$

which are the defining properties of a homogeneous chain ladder process. This proves our second claim. ■

In Section 5.3, we will meet the $\phi_{k,j}$ again (see Remark 5.7).

Remark 2.5. Note that the model assumption on the variance, $V[X_j|\mathcal{X}_{j-1}] = \phi_j X_{j-1}$, plays a crucial role in this section. Alternative chain ladder models have been studied which specify $V[X_j|\mathcal{X}_{j-1}] = \phi_j$ or $V[X_j|\mathcal{X}_{j-1}] = \phi_j X_{j-1}^2$. For these, the results of this section do not hold.

2.4. Loss development triangles

Our treatment of loss development triangles differs in two respects from standard treatments (e.g. Mack, 1993; Wüthrich and Merz, 2008): we allow for missing data and also for flexibility with respect to the time granularity. Instead of accident years, we speak of loss portfolios, and instead of development years, we consider development periods or steps, which do not have to be equidistant. Given the facts presented in the previous section, Propositions 2.3 and 2.4, this is a natural thing to do.

Consider a collection of pairwise disjoint claims portfolios identified by an index $i \in \mathcal{I}$, where $\mathcal{I}$ is a finite non-empty set. We assume that each portfolio is subject to a development described by a homogeneous chain ladder process. Let $C_{i,j} > 0$ denote the value of the process of the portfolio with index i at development index j. This could be claim payments, claim numbers, or incurred losses, for example. The values could be incremental or cumulative, although in practice the interesting applications will more likely be of the cumulative type, and we will make this assumption throughout the paper.

We assume that the portfolios are comparable in that their chain ladder development parameters f_j, ϕ_j are identical, i.e., do not depend on i. We also assume the chain ladder processes to be independent, and that for each portfolio at least one of the development states is known, i.e. has been observed.

Our notation describing the data is summarized in the following:

Definition 2.6. *Assuming each set $\{j \mid C_{i,j}$ is known$\}$, for $i \in \mathcal{I}$, to be non-empty, we define*

$$\begin{aligned}
j_i &:= \max\left\{j \mid C_{i,j} \textit{ is known}\right\}, \quad \textit{for each} \quad i \in \mathcal{I};\\
J &:= \max\{j_i \mid i \in \mathcal{I}\};\\
\mathcal{I}_j &:= \left\{i \in \mathcal{I} \mid C_{i,j} \textit{ is known}\right\}, \quad \textit{for each non-negative integer } j;\\
\mathcal{I}_j^* &:= \{i \in \mathcal{I} \mid j_i < j\}, \quad \textit{for each non-negative integer } j;\\
f_{i,j} &:= C_{i,j}/C_{i,j-1}, \quad \textit{for each } (i,j) \textit{ such that } i \in \mathcal{I}_{j-1} \cap \mathcal{I}_j;\\
F_{i,j} &:= C_{i,j}/C_{i,j-1}, \quad \textit{for each } (i,j) \textit{ such that } i \in \mathcal{I}_j^*;\\
\mathcal{D} &:= \{C_{i,j} \mid i \in \mathcal{I}_j\}.
\end{aligned}$$

Thus, j_i is the development index of the latest known development state of portfolio i, and the maximum of all j_i is denoted J. The development up to the index J will be referred to as the *ultimate development* or *development up to the ultimate horizon*, although this does not require that all claims have to be finally settled at that point.

The states $C_{i,j}$ with $j \leq j_i$ will be called the *historical*, the others the *future* development states (the future states residing in the lower right part of Figure 1 — the column-wise grouping indicated by the dashed lines will be explained in Section 4.1). We allow for some of the historical states to be unknown

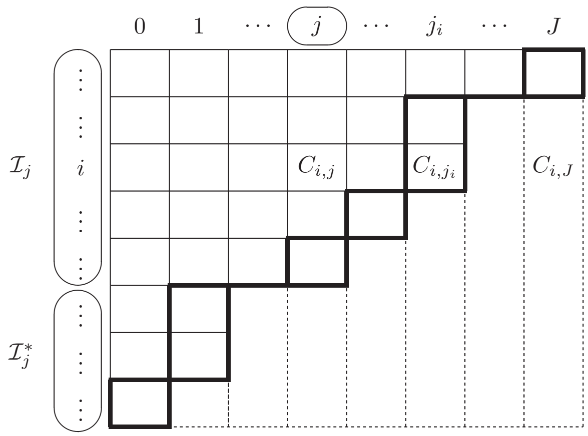

FIGURE 1: A loss development triangle.

or missing (more on this in Section 2.5). The set $\mathcal{I}_j$ represents the set of all portfolios with known development at development index j, whereas $\mathcal{I}_j^*$ is the set of all portfolios for which the state at development index j is a future development state. In particular, by definition of J, we have $\mathcal{I}_j = \emptyset$ and $\mathcal{I}_j^* = \mathcal{I}$ for $j > J$. The future development state $C_{i,j}$ is a random variable from today's point of view. The set $\mathcal{D}$ is the set of those states which have already been observed; its subset

$$\{C_{i,j_i} | i \in \mathcal{I}\} \subseteq \mathcal{D}$$

which represents the set of "latest" known states for our loss portfolios, will be referred to as *"today's diagonal"* (the elements surrounded by thick lines in Figure 1).

The link ratios $f_{i,j}$ with $j \leq j_i$ will be called *historical link ratios*, the link ratios $F_{i,j}$ with $j > j_i$ *future link ratios*.

Figure 1 illustrates these definitions. The triangular shape is the general case, since we may always order the loss portfolios by decreasing j_i.

For $0 \leq j \leq \min\{j_i \,|\, i \in \mathcal{I}\}$, we have $\mathcal{I}_j^* = \emptyset$ by definition. We make the additional assumption that, in fact, $\min\{j_i \,|\, i \in \mathcal{I}\} = 0$, such that

$$\mathcal{I}_j^* = \emptyset \Leftrightarrow j = 0. \tag{10}$$

This is no loss of generality, as the values $C_{i,j}$ with $j < \min\{j_i \,|\, i \in \mathcal{I}\}$ would play no role in our analysis, and instead of summing over all $j = 1$ to J, we would frequently have to sum from $j = \min\{j_i \,|\, i \in \mathcal{I}\} + 1$ to J instead.

We close this section by introducing a practical summation convention:

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

Summation Convention 2.7. *For any subset $\mathcal{H} \subseteq \mathcal{I}$, we define*

$$C_{\mathcal{H},j} := \sum_{i \in \mathcal{H}} C_{i,j},$$

where the sum is understood to be 0 *if* $\mathcal{H} = \emptyset$ *and to take precedence over further algebraic transformation such as squaring. Dropping indices altogether will mean summation over all* $i \in \mathcal{I}$ *at the ultimate horizon, i.e.*

$$C := C_{\mathcal{I},J}.$$

We will also apply this convention to the predictors $\hat{C}_{i,j}$ *to be defined later on.*

2.5. Non-standard-shaped and incomplete triangles

The notation introduced in the previous section allows for the treatment of missing data, which we explain in more detail in this section.

The triangle in Figure 1 is somewhat "ragged" in comparison to a standard accident year/development year triangle, because two elements on the (geometric) diagonal are missing (third and last but one column in Figure 1). We also allow for missing data to occur "inside" the historical triangle, as we did not require $\mathcal{I}_j \cup \mathcal{I}_j^*$ to be equal to $\mathcal{I}$. The states at development step j belonging to the portfolios enumerated by the set $\mathcal{I} \setminus (\mathcal{I}_j \cup \mathcal{I}_j^*)$ are historical, but their value is not known — hence this data is missing.

Missing data are a fairly common situation. For example, a change in individual claims reserving practice may render some old part of the triangle useless for the analysis of incurred development. Or, while payments may be available on a granular level for various types of claims payments, case reserves may only be available at that level from a certain calendar date onward — again a situation, where for the purpose of the analysis of incurred loss development some old part of the historical data must be considered unknown. Then there is the case of an obviously false data recording for a single individual claim that occurs in one development period and is reversed in the next. If this cannot be rectified in the data preparation step, the state affected must be declared unknown. Finally, for paid data analysis, some long tail lines of business have zero payments during the initial phases of development. Strictly speaking, this would fall outside the scope of our models, since we required all random variables to take on strictly positive values, but a practical way to deal with this situation would be to treat the zero values as unknown data.

Note, however, that Definition 2.6 does imply $\mathcal{I}_J \cup \mathcal{I}_J^* = \mathcal{I}$, because if $i \in \mathcal{I}$, then either $J > j_i$, in which case $i \in \mathcal{I}_J^*$, or $J = j_i$, in which case $C_{i,J}$ is known and hence $i \in \mathcal{I}_J$. In other words, any missing state at index J is considered a future state, and hence there are no missing historical states at development index J.

While the values of both $C_{i,j-1}$ and $C_{i,j}$ must be known to calculate the historical link ratio $f_{i,j}$, it is not unusual in practice to discard some of the available link ratios when estimating the development factors. To accommodate the deletion of "outliers" among the link ratios, or the selection of only the "latest 5 diagonals", we assume that for each $j \in \{1, \ldots, J\}$, a non-empty set

$$\mathcal{H}_j \subseteq \mathcal{I}_{j-1} \cap \mathcal{I}_j$$

has been specified that reflects the selection made for the purpose of estimating f_j and ϕ_j by the unbiased estimators (see Mack, 1993)

$$\hat{f}_j := \frac{C_{\mathcal{H}_j,j}}{C_{\mathcal{H}_j,j-1}} = \frac{\sum_{i\in\mathcal{H}_j} C_{i,j-1} f_{i,j}}{\sum_{i\in\mathcal{H}_j} C_{i,j-1}} \quad \text{and} \quad \hat{\phi}_j := \frac{\sum_{i\in\mathcal{H}_j} C_{i,j-1}\left(f_{i,j} - \hat{f}_j\right)^2}{-1 + \sum_{i\in\mathcal{H}_j} 1}.$$

While incomplete data are important in practical applications, the theory and the formulae are somewhat simpler if completeness may be assumed. Therefore, we make the following definition:

Definition 2.8. *If* $\mathcal{I}_j \cup \mathcal{I}_j^* = \mathcal{I}$ *for all* j, *and* $\mathcal{H}_j = \mathcal{I}_{j-1} \cap \mathcal{I}_j$ *for all* $j > 0$, *then we say that* the historical triangle is complete. *In this case, we also have* $\mathcal{I} = \mathcal{I}_0 \supseteq \mathcal{I}_1 \supseteq \ldots \supseteq \mathcal{I}_J$.

The last statement needs verification: the sets $\mathcal{I}_j^*$ always form an ascending chain, being pre-images of an ascending chain of sets of integers, and starting with $\mathcal{I}_0^* = \emptyset$ by (10). Hence, if $\mathcal{I}_j \cup \mathcal{I}_j^* = \mathcal{I}$ for all j, then the $\mathcal{I}_j$ form a descending chain, starting with $\mathcal{I}_0 = \mathcal{I}$ (note that $\mathcal{I}_j$ and $\mathcal{I}_j^*$ are disjoint by definition).

Our treatment of incompleteness is different from that presented by Dahms (2008). For example, in the cases mentioned, where we have a cumulative incurred value that is classified as valid, it must reflect the complete payment history plus the current value of case reserves. Dahms, by contrast, allows for part of the payment history to be missing, too.

Remark 2.9. We explained the "non-standard" shape of the triangle depicted above by missing data "on the diagonal". This gave rise to a "non-standard" or "ragged" shape of the triangle. Note, however, that Definition 2.6 also covers non-standard shapes that may arise although no data is missing at all. For example, assume that we observe similar development behavior in two lines of business and have reasons to believe that they could be modeled using the same chain ladder parameters. Assume further that for some reason we need to keep the results for the two lines of business separate. Then instead of just adding two "standard accident year/development year" triangles, we might, for each accident year, look at two loss portfolios and their development separately. Arranged by increasing degree of development, this would result in a non-standard triangle shape, although no data is missing. Doing so allows to keep the results of the analysis separate (see Result 5.3), while the joint analysis helps to reduce the parameter error. Furthermore, it provides a way to estimate ϕ_J, which otherwise can only be estimated by extrapolation.

Remark 2.10. One should exercise care when discarding link ratios, i.e. when removing elements from the sets $\mathcal{H}_j$. Possible reasons for discarding an observed link ratio include the following: (a) the loss development triangle element forming the numerator of the link ratio is unreliable; (b) the denominator is unreliable; or (c) the development step between denominator and the numerator is considered not to have followed the stochastic laws assumed to be underlying the overall chain ladder model. Unreliable data in this sense may be caused, for example, by data entry errors in the claims system that are reversed in the next development period. In cases (a) and (b), one usually must discard two successive link ratios, because an unreliable element in the loss development triangle in general affects two successive link ratios. If a diagonal element is unreliable, then only one link ratio is affected, but in this case not only the link ratio, but the whole diagonal element should be discarded, because the predicted future development would rely on a value that already has been identified as unreliable.

Case (c) is obviously very subjective, but it may arise naturally as follows: suppose we have a loss portfolio of accident years $i = 0, 1, \ldots, J$ with development years $j = 0, 1, \ldots, J$, whose development is known for all accounting years $k = i + j = 0, \ldots, J$. Next, suppose we have a second triangle from the same line of business that we would like to add to the first triangle. Suppose, however, that in the second triangle, the cumulative data only starts in a specific accounting year $k \leq J$, i.e. that its entries in accounting years less than k are missing. Then we may add the two triangles, and all elements of the merged triangle may be called reliable, but the jump at the accounting year k diagonal clearly has nothing to do with the stochastic development in the model, but is predominantly caused by the additional volume of the second triangle. So in this case, the link ratios arising between diagonal $k - 1$ and k of the merged triangle are justifiably discarded.

Certain other schemes of discarding link ratios, on the other hand, for example the method of discarding the highest and lowest values at each development step, will introduce a bias, and are outside the scope of this paper.

2.6. Chain ladder predictors

Given the latest known cumulative values C_{i,j_i}, the chain ladder predictors $\hat{C}_{i,j}$ of future values $C_{i,j}$ are given by

$$\hat{C}_{i,j} := C_{i,j_i} \cdot \hat{f}_{j_i+1} \cdot \ldots \cdot \hat{f}_j \tag{11}$$

(see Mack, 1993). Setting $\hat{C}_{i,j} := C_{i,j}$ for historical (i, j) will permit us to apply our summation convention 2.7 to the $\hat{C}_{i,j}$ and in particular allow us to abbreviate the predictor for the total ultimate loss C by $\hat{C}$.

2.7. Accuracy and influence of development factors

As mentioned in the introduction, we will formulate our results in terms of certain quantities $\hat{u}_j$ and $\hat{q}_j$. These are statistics of the historical data, but are usually not found in standard treatments of the chain ladder method. This section defines and interprets them.

We start with the conditional coefficient of variation of the link ratio F_j of a chain ladder process, i.e. the quantity

$$\sqrt{\frac{V[F_j|\mathcal{X}_{j-1}]}{(E[F_j|\mathcal{X}_{j-1}])^2}} = \sqrt{\frac{V[\frac{X_j-A_j}{X_{j-1}}|\mathcal{X}_{j-1}]}{(E[\frac{X_j-A_j}{X_{j-1}}|\mathcal{X}_{j-1}])^2}} = \sqrt{\frac{\phi_j}{f_j^2 X_{j-1}}}. \tag{12}$$

If we calculate this quantity for the process of portfolio i from our loss development triangle of Section 2.4, we get $\sqrt{\phi_j/(f_j^2 C_{i,j-1})}$; and for the chain ladder process $\{C_{\mathcal{H}_j,k}\}_{k\geq 0}$ obtained by aggregating, in accordance with Proposition 2.3, the portfolios enumerated by the set $\mathcal{H}_j$, we get $\sqrt{\phi_j/(f_j^2 C_{\mathcal{H}_j,j-1})}$. Replacing unknown quantities by their estimators, we arrive at the following estimators for the conditional coefficients of variation of these two processes:

$$\hat{u}_{i,j} := \sqrt{\frac{\hat{\phi}_j}{\hat{f}_j \hat{C}_{i,j}}} \quad \text{for } j > j_i \text{, and} \quad \hat{u}_j := \sqrt{\frac{\hat{\phi}_j}{\hat{f}_j C_{\mathcal{H}_j,j}}} \quad \text{for } j > 0, \tag{13}$$

where we have used that $\hat{f}_j = C_{\mathcal{H}_j,j}/C_{\mathcal{H}_j,j-1}$ by definition of $\hat{f}_j$, and $\hat{f}_j \hat{C}_{i,j-1} = \hat{C}_{i,j}$ by (11).

Being an estimator for the coefficient of variation of $\hat{f}_j$ as a link ratio of the aggregate chain ladder process, $\hat{u}_j$ is also a measure of the relative accuracy of $\hat{f}_j$ as an estimator for f_j.

The next quantity we introduce measures the influence that any given development factor has on the ultimate loss C. For example, one might ask, if we misestimate f_j by a few percent, by how many percent will that change our estimated ultimate loss? The quantity we are about to define measures the quotient of these percentages.

Lemma 2.11. *For $j \in \{0, 1, \ldots, J\}$, let*

$$q_j := \frac{E[C_{\mathcal{I}_j^*,J}|\mathcal{D}]}{E[C|\mathcal{D}]} \quad \textit{and} \quad \hat{q}_j := \frac{\hat{C}_{\mathcal{I}_j^*,J}}{\hat{C}}.$$

Then $q_0 = \hat{q}_0 = 0$, and for $j > 0$, the quantity $\hat{q}_j$ is obtained from q_j by replacing each f_j with its estimator $\hat{f}_j$, and

$$q_j = \partial_{\log[f_j]} \log[E[C|\mathcal{D}]].$$

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

Proof. The assertion about $\hat{q}_j$ follows from (11) and (1). By the latter,

$$\partial_{f_j} E[C|\mathcal{D}] = \partial_{f_j} \sum_{i\in\mathcal{I}} \left(C_{i,j_i} \prod_{k=j_i+1}^{J} f_k \right) = \sum_{i\in\mathcal{I}} \left(\frac{C_{i,j_i} \mathbf{1}_{j>j_i}}{f_j} \prod_{k=j_i+1}^{J} f_k \right).$$

Now $j > j_i$ is equivalent to $i \in \mathcal{I}_j^*$, hence

$$\begin{aligned} \partial_{\log[f_j]} \log[E[C|\mathcal{D}]] &= \frac{f_j \partial_{f_j} E[C|\mathcal{D}]}{E[C|\mathcal{D}]} = \frac{1}{E[C|\mathcal{D}]} \sum_{i\in\mathcal{I}_j^*} \left(C_{i,j_i} \prod_{k=j_i+1}^{J} f_k \right) \\ &= \frac{E[C_{\mathcal{I}_j^*,J}|\mathcal{D}]}{E[C|\mathcal{D}]} \end{aligned}$$

as claimed. Finally, $q_0 = \hat{q}_0 = 0$ because $\mathcal{I}_0^* = \emptyset$ by (10). ■

Thus $\hat{q}_j$ estimates the influence that f_j has on the ultimate loss — in one interpretation as a ratio of log changes, and in another as the proportion of the ultimate loss that is actually affected by the development factor. In particular, $0 \le \hat{q}_j < 1$.

3. Error propagation calculation of the MSEP

In this section, we recall the usual definition of the mean squared prediction error and apply the error propagation principle to approximate the mean squared prediction error for a class of random variables and associated predictors arising naturally in our chain ladder context.

3.1. Mean squared error of prediction

Let us briefly recall the definition of the mean squared error of prediction (e.g. Wüthrich and Merz, 2008). The difference between the true (but unknown) ultimate loss C and its predictor $\hat{C}$, given $\mathcal{D}$, is called the (conditional) *prediction error*, and its squared expectation the (conditional) *mean squared error of prediction* ("MSEP"):

$$\begin{aligned} \mathrm{msep}_{C-\hat{C}} &:= E[(C-\hat{C})^2|\mathcal{D}] \\ &= E[(C - E[C|\mathcal{D}] + E[C|\mathcal{D}] - \hat{C})^2|\mathcal{D}] \\ &= E[(C - E[C|\mathcal{D}])^2|\mathcal{D}] + (E[C|\mathcal{D}] - \hat{C})^2 \\ &= V[C|\mathcal{D}] + (E[C|\mathcal{D}] - \hat{C})^2 \\ &=: \mathrm{msep}_{C-\hat{C}}^{\mathrm{proc}} + \mathrm{msep}_{C-\hat{C}}^{\mathrm{parm}}. \end{aligned} \tag{14}$$

The first summand in the fourth line is called the *mean squared process error* (or *process variance*) and the second the squared estimation (or parameter) error. In the following section, we will estimate these error terms using the error propagation principle.

3.2. Applying the error propagation principle

Formula (14) applies in a context much more general than our chain ladder context. But within the latter, the quantities occurring have a number of special properties which allow us to apply the error propagation principle. We will investigate this in the current section. For a general discussion of error propagation and its applications, see e.g. Ku (1966).

The special thing about C, which by definition equals

$$C = \sum_{i \in \mathcal{I}} C_{i,j_i} \prod_{j=j_i+1}^{J} F_{i,j},$$

is that it is a function of random variables — the future link ratios $F_{i,j}$ — whose first and second moments, including their correlations, can be readily estimated, as we will show in this section (following Mack, 1993). Therefore, approximating C by a first-order Taylor series in the $F_{i,j}$ will allow us to derive estimators for quantities such as the MSEP. This is the basic idea behind the approach taken in this section.

This approach can be applied to C, and to other functions of the future link ratios. Such more general functions will be needed in Section 5. So let us consider any random variable G which is a rational function, also denoted by G, of the future link ratios, with coefficients that depend on the known values from the historical triangle $\mathcal{D}$. We might even consider a more general class than that of rational functions over the real numbers, but we restrict to this class in order to fix ideas and because it will serve our purposes. We assume that the first two moments of G exist, conditionally given $\mathcal{D}$.

The first thing we will do is to substitute the $F_{i,j}$ in the algebraic expression for G by a new set of variables related to the two sources of uncertainty, the process and parameter error. We will proceed in two steps. First, we apply the substitution

$$F_{i,j} =: \xi_{i,j} + f_j, \tag{15}$$

introducing the new variables $\xi_{i,j}$ (one for each future link ratio). The parameters f_j are deterministic real numbers; however, their values are unknown to us, and therefore the f_j are indeterminates in the algebraic expression resulting after the substitution (15) in G. Therefore, the second substitution we will apply, for $1 \leq j \leq J$,

$$f_j =: \eta_j + \hat{f}_j, \tag{16}$$

is well defined, introduces the new variables η_j, and together with the first substitution converts G into a function of the variables $\xi_{i,j}$ and η_j. Together, (15)

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

and (16) amount to the combined substitution

$$F_{i,j} = \xi_{i,j} + \eta_j + \hat{f}_j \tag{17}$$

for all future $F_{i,j}$, i.e. for all (i, j) such that $i \in \mathcal{I}_j^*$.

Evaluating G at the point where $F_{i,j} = \hat{f}_j$ for all future link ratios provides a predictor for the random variable G which we denote by $\hat{G}$, in analogy to the special case of $\hat{C}$ defined above:

$$\hat{G} := G_{|\hat{f}}.$$

We will systematically use the subscript notations

$$(\ldots)_{|\hat{f}} \quad , \quad (\ldots)_{|f} \quad \text{and} \quad (\ldots)_{|0}$$

when we interpret the subscripted expression as a function of the $F_{i,j}$ and evaluate it at the point $F_{i,j} = \hat{f}_j$, when we interpret the subscripted expression as a function of the $F_{i,j}$ and evaluate it at the point $F_{i,j} = f_j$, and when we interpret the subscripted expression as a function of the $\xi_{i,j}$ and η_j (after the substitution (17)) and evaluate it at the point $\xi_{i,j} = \eta_j = 0$, respectively. For example, we have $\hat{G} = G_{|\hat{f}} = G_{|0}$.

Taking $\hat{G}$ as a predictor for G, our goal is to find the prediction error $\text{msep}_{G-\hat{G}}$. Regarding G as a function of the $\xi_{i,j}$ and η_j, the MSEP is now seen to come from two kinds of sources: the deviations of the variables $\xi_{i,j}$ from 0, which stem from the deviations of the $F_{i,j}$ from their expectation values f_j, and the deviations of the η_j from 0, which are attributable to the amount by which we misestimate the true f_j by $\hat{f}_j$. In fact, we may expect these deviations to be in the vicinity of 0, since $E[\xi_{i,j}|\mathcal{D}] = E[F_{i,j}|\mathcal{D}] - f_j = 0$ by the assumed independence of loss portfolios and (3), and since $\hat{f}_j$ is an unbiased estimator of f_j (see Mack, 1993). Therefore, we may expect the first order Taylor series of G about $\xi_{i,j} = \eta_j = 0$,

$$L_{\hat{f}}[G] := \hat{G} + \sum_{j=1}^{J} \sum_{i \in \mathcal{I}_j^*} (\partial_{\xi_{i,j}} G)_{|0}\, \xi_{i,j} + \sum_{j=1}^{J} (\partial_{\eta_j} G)_{|0}\, \eta_j. \tag{18}$$

to be an approximation of the random variable G (assuming differentiability at 0).

Result 3.1 (Preliminary Approximations). *In this situation, the following statements hold true:*

1. *The (conditional) MSEP of the prediction of G by $\hat{G}$ is approximated by*

$$msep_{G-\hat{G}} \approx msep_{L_{\hat{f}}[G]-\hat{G}} \tag{19}$$

$$\approx \sum_{j=1}^{J} \sum_{i \in \mathcal{I}_j^*} (\partial_{F_{i,j}} G)^2_{|\hat{f}} \, V[F_{i,j}|\mathcal{D}] + \sum_{j=1}^{J} \left(\sum_{i \in \mathcal{I}_j^*} (\partial_{F_{i,j}} G)_{|\hat{f}} \right)^2 V[\hat{f}_j \mid C_{\mathcal{H}_j, j-1}]. \tag{20}$$

2. *The first summand in the last expression is an approximation for* $msep^{proc}_{G-\hat{G}}$ *and the second an approximation for* $msep^{parm}_{G-\hat{G}}$.
3. *The variances occurring in these formulae may be estimated as follows:*

$$V\left[F_{i,j} \middle| \mathcal{D} \right] \approx \hat{f}_j^2 \hat{u}_{i,j}^2, \tag{21}$$

$$V[\hat{f}_j | C_{\mathcal{H}_j, j-1}] \approx \hat{f}_j^2 \hat{u}_j^2. \tag{22}$$

Derivation. Taking $L_{\hat{f}}[G]$ as an approximation for G given $\mathcal{D}$ justifies the approximation (19). We now claim that

$$\text{msep}_{L_{\hat{f}}[G]-\hat{G}} = \sum_{j=1}^{J} \sum_{i \in \mathcal{I}_j^*} (\partial_{\xi_{i,j}} G)^2_{|0} \, V[F_{i,j}|\mathcal{D}] + \left(\sum_{j=1}^{J} (\partial_{\eta_j} G)_{|0} \, \eta_j \right)^2. \tag{23}$$

To see this, we square the right-hand side of (18) and take expectations $E[.|\mathcal{D}]$; the rest then boils down to an analysis of cross-terms, which we will now carry out, following Mack (1993). First, $E[\xi_{i,j}|\mathcal{D}] = 0$ by (3), and consequently $E[\xi_{i,j}\xi_{i',j'}|\mathcal{D}] = 0$ if either $i \neq i'$ (by the assumed independence of loss portfolios) or $j \neq j'$ (by (5)); and $E[\xi_{i,j}^2|\mathcal{D}] = V[\xi_{i,j}|\mathcal{D}] = V[F_{i,j}|\mathcal{D}]$. Next, we note that conditionally, given $\mathcal{D}$, the η_j are not random at all, being the difference between a fixed (if unknown) parameter f_j and a statistic $\hat{f}_j$ that has been calculated from observed data. Therefore $E[\xi_{i,j}\eta_k|\mathcal{D}] = \eta_k E[\xi_{i,j}|\mathcal{D}] = 0$. Putting this together, we arrive at (23).

Our next claim is

$$\left(\sum_{j=1}^{J} (\partial_{\eta_j} G)_{|0} \, \eta_j \right)^2 \approx \sum_{j=1}^{J} (\partial_{\eta_j} G)^2_{|0} \, V[\hat{f}_j \mid C_{\mathcal{H}_j, j-1}], \tag{24}$$

which again can be justified by an analysis of cross terms, still following Mack (1993): In order to estimate $\eta_j \eta_k$ — assuming $j \leq k$ without loss of generality — we approximate its unknown value by an average value:

$$\eta_j \eta_k \approx E[\eta_j \eta_k | \{C_{i,j-1} | i \in \mathcal{H}_j\}], \tag{25}$$

where on the right-hand side we interpret the $\hat{f}_j$ and $\hat{f}_k$ occurring in $\eta_j \eta_k = (f_j - \hat{f}_j)(f_k - \hat{f}_k)$ as random variables, and not as observed statistics. Now if

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

$j < k$, then this becomes

$$\begin{aligned} E[\eta_j \eta_k | \{C_{i,j-1} | i \in \mathcal{H}_j\}] &= E[E[\eta_j \eta_k | \{C_{i,\ell-1} | i \in \mathcal{H}_\ell, \ell \leq k\}] | \{C_{i,j-1} | i \in \mathcal{H}_j\}] \\ &= E[\eta_j E[\eta_k | \{C_{i,\ell-1} | i \in \mathcal{H}_\ell, \ell \leq k\}] | \{C_{i,j-1} | i \in \mathcal{H}_j\}] \\ &= 0 \end{aligned}$$

since $E[\eta_k | \{C_{i,\ell-1} | i \in \mathcal{H}_\ell, \ell \leq k\}] = 0$ by our chain ladder assumptions and the assumed independence of the loss portfolios. For $j = k$, the right-hand side of (25) becomes $V[\hat{f}_j | C_{\mathcal{H}_j, j-1}]$, and we arrive at the approximation (24). Finally, by the chain rule of differentiation, $(\partial_{\eta_j} G)_{|0} = \sum_{i \in \mathcal{I}_j^*} (\partial_{F_{i,j}} G)_{|\hat{f}}$. Putting this together, we have justified the approximation (20).

To prove statement 2), we first consider the case where

$$E[G|\mathcal{D}] = G_{|f}. \tag{26}$$

For example, $G = C$ is such a function, by (1). Under this assumption, we have, by definition of the process and parameter error (see (14)),

$$\text{msep}^{\text{proc}}_{G-\hat{G}} = E[(G - E[G|\mathcal{D}])^2 | \mathcal{D}] = E[(G - G_{|f})^2 | \mathcal{D}]$$

and

$$\text{msep}^{\text{parm}}_{G-\hat{G}} = (E[G|\mathcal{D}] - \hat{G})^2 = (G_{|f} - G_{|\hat{f}})^2$$

respectively. Like above, we use first-order Taylor expansion. For the process error, using the same arguments about the cross terms already presented above, this leads to

$$\text{msep}^{\text{proc}}_{G-\hat{G}} \approx \sum_{j=1}^{J} \sum_{i \in \mathcal{I}_j^*} (\partial_{F_{i,j}} G)^2_{|f} V[F_{i,j} | \mathcal{D}], \tag{27}$$

but $(\partial_{F_{i,j}} G)_{|\hat{f}}$ is an estimator for $(\partial_{F_{i,j}} G)_{|f}$, which proves the first claim in statement 2) for the special type of G fulfilling (26). As for the second claim, regarding the parameter error, this is equally straightforward, considering that $\sum (\partial_{\eta_j} G)_{|0} \eta_j$ is also the linear term of the Taylor expansion of the function $G_{|f}$ of the f_j at the point $\eta_j = 0$ for all j.

Besides C, another example of a function satisfying the condition (26) is $L_{\hat{f}}[G]$. Indeed, since $E[\xi_{i,j} | \mathcal{D}] = 0$,

$$E[L_{\hat{f}}[G] | \mathcal{D}] = \hat{G} + \sum_{j=1}^{J} (\partial_{\eta_j} G)_{|0} \eta_j,$$

and the same result is obtained by substituting each occurrence of $\xi_{i,j} = F_{i,j} - f_j$ by 0 in (18). Therefore, by what we have proved of statement 2) already, the two summands of (24) approximate $\text{msep}^{\text{proc}}_{L_{\hat{f}}[G]-\hat{G}}$ and $\text{msep}^{\text{parm}}_{L_{\hat{f}}[G]-\hat{G}}$, respectively.

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

Hence, to the extent that we accept the random variable $L_{\hat{f}}[G]$ as an approximation of G, we may also consider the summands of (24) to represent approximations of $\mathrm{msep}^{\mathrm{proc}}_{G-\hat{G}}$ and $\mathrm{msep}^{\mathrm{parm}}_{G-\hat{G}}$, respectively. This proves the statement 2).

For statement 3), note first that due to the assumed independence of the loss portfolios, we have $V[F_{i,j}|\mathcal{D}] = V[F_{i,j}|C_{i,0}, \ldots, C_{i,j_i}]$. Next, remember that in Section 2.7 we found that $\hat{u}_{i,j}$ and $\hat{u}_j$ estimate the (conditional) coefficient of variation of the link ratio at development index j of the chain ladder processes $\{C_{i,k}\}_{k\geq 0}$ and $\{C_{\mathcal{H}_j,k}\}_{k\geq 0}$, respectively. Since the expectation value of the link ratio is estimated in each case by $\hat{f}_j$, we get the approximations (21) and (22). ■

Remark 3.2. For a discussion of the split into process and parameter error, see also Merz and Wüthrich (2008) and, in the context of a Bayesian chain ladder model, Bühlmann *et al.* (2009), Remarks 4.2 and 4.8.

4. The MSEP for the ultimate loss

In Section 4 and Section 5, we will apply our Preliminary Approximations 3.1 to derive approximations for $\mathrm{msep}_{G-\hat{G}}$ for various functions G of the future link ratios. The case considered in the present Section 4 is the case $G = C$, i. e. the development to the ultimate horizon. Later on, in Section 5, this will be generalized to the case of development to any intermediate horizon, and even to the development between two future horizons. Correspondingly, the results of Section 5 contain the results obtained in the present Section 4 as a special case. However, some proofs and insights do not carry over to the more general case of Section 5 and are therefore presented here in Section 4: first of all, the use of non-homogeneous chain ladder processes provides straightforward insight into the fact that the process error of the process(es) underlying the historical data has the effect of an "additional process error" when considering the future development (see Remark 4.2). Secondly, the split into process and parameter error may be accomplished for the case of development to the ultimate horizon by direct Taylor approximation of the defining terms because condition (26) holds for the ultimate loss C, while in the more general context of Section 5, this will no longer be possible. On the other hand, the detailed evaluation of partial derivatives presented in Section 5.2 is not necessary for the purposes of Section 4.

4.1. The chain ladder process of future development

The future part of the development of our given loss portfolios can be aggregated to a single, non-homogeneous chain ladder process. At development index j, we simply sum over all future states $C_{i,j}$ at this index (in Figure 1, this corresponds to the dashed rectangle in column j), and as A_j we take the sum of the elements on today's diagonal at development step j (the rectangles with thick

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

solid lines in Figure 1). More formally, the process

$$\{C_{\mathcal{I}^*_{j+1},j}\}_{j\geq 0} \tag{28}$$

is a (non-homogeneous) chain ladder process (given $\mathcal{A} := \sigma[\mathcal{D}]$) with parameters f_j, ϕ_j, $A_j = C_{\mathcal{I}^*_{j+1}\setminus\mathcal{I}^*_j,j}$. To see this, we check the condition on the expectation value (leaving the rest to the reader):

$$\begin{aligned} E[C_{\mathcal{I}^*_{j+1},j}|C_{\mathcal{I}^*_j,j-1},\ldots,C_{\mathcal{I}^*_1,0},\mathcal{D}] &= E[C_{\mathcal{I}^*_{j+1},j} - A_j|C_{\mathcal{I}^*_j,j-1},\ldots,C_{\mathcal{I}^*_1,0},\mathcal{D}] + A_j \\ &= E[C_{\mathcal{I}^*_j,j}|C_{\mathcal{I}^*_j,j-1},\ldots,C_{\mathcal{I}^*_1,0},\mathcal{D}] + A_j \\ &= f_j C_{\mathcal{I}^*_j,j-1} + A_j. \end{aligned}$$

In the third equation, we have used our assumption of independence of loss portfolios, which allows us to ignore the loss portfolios not represented in $\mathcal{I}^*_j$ in the conditioning, as well as Proposition 2.3, which tells us that $\{C_{\mathcal{I}^*_j,k}\}_{k\geq 0}$ is a homogeneous chain ladder process with parameters f_k and ϕ_k. Furthermore, at $j = 0$, the value of the process (28) is positive by (10). Let

$$F_j := \frac{C_{\mathcal{I}^*_{j+1},j} - C_{\mathcal{I}^*_{j+1}\setminus\mathcal{I}^*_j,j}}{C_{\mathcal{I}^*_j,j-1}}$$

be the link ratio of the process (28), for $j \geq 1$. Note that the value of the process (28) at development index $j = J$ is just the ultimate loss C.

Result 4.1 (preliminary approximations — ultimate horizon). *In this situation, we have*

$$msep_{C-\hat{C}} \approx \hat{C}^2 \sum_{j=1}^{J} \hat{q}_j^2 \frac{V[F_j|\mathcal{D}] + V[\hat{f}_j|C_{\mathcal{H}_j,j-1}]}{\hat{f}_j^2}, \tag{29}$$

$$msep^{proc}_{C-\hat{C}} \approx (E[C|\mathcal{D}])^2 \sum_{j=1}^{J} q_j^2 \frac{V[F_j|\mathcal{D}]}{f_j^2}, \tag{30}$$

$$msep^{parm}_{C-\hat{C}} \approx \hat{C}^2 \sum_{j=1}^{J} \hat{q}_j^2 \frac{V[\hat{f}_j|C_{\mathcal{H}_j,j-1}]}{\hat{f}_j^2}. \tag{31}$$

Derivation. This will follow quite straightforwardly from our Preliminary Approximations 3.1. Note first that Result 3.1 may be adapted easily to the situation we consider here, where instead of a two-dimensional array of link ratios $F_{i,j}$, we have the one-dimensional array of the F_j. Effectively, statements (1) and (2) of 3.1 hold with the sums over the index i replaced by a single item (involving F_j instead of $F_{i,j}$). Instead of the substitution (15), we now use $F_j = \xi_j + f_j$, with new coordinates ξ_j. To derive (29), we therefore refer to (20), and all that remains to be shown is that $(\partial_{F_j} C)_{|\hat{f}} = \hat{C}\hat{q}_j/\hat{f}_j$. But by (1), Lemma 2.11 and (11)

$$(\partial_{F_j} C)_{|f_j} = \partial_{f_j} E[C|\mathcal{D}] = \frac{E[C|\mathcal{D}]}{f_j}\partial_{\log[f_j]}\log[E[C|\mathcal{D}]] = \frac{E[C|\mathcal{D}]}{f_j}q_j$$

and hence

$$(\partial_{F_j} C)_{|\hat{f}_j} = \frac{\hat{C}}{\hat{f}_j}\hat{q}_j,$$

which proves our claim about $(\partial_{F_j} C)_{|\hat{f}_j}$. The second statement of Result 3.1 now implies (30) and (31). ■

Remark 4.2. Note here that the parameter error of the future development stems from the process error of the historical data, as becomes apparent from (31) (remember $\hat{f}_j$ is a link ratio of the chain ladder process $\{C_{\mathcal{H}_j,k}\}_{k\geq 0}$), and that its influence may be interpreted as just some "additional process uncertainty" of the non-homogeneous chain ladder process describing the future development, as becomes clear from (29) and (30). This insight is the reason why we introduced the concept of a non-homogeneous chain ladder process.

Remark 4.3. In (30), we might have gone one step further and might have replaced $E[C|\mathcal{D}]$, q_j and f_j by $\hat{C}$, $\hat{q}_j$ and $\hat{f}_j$, respectively, but we need the formula as it stands to build on it in Remark 4.7.

4.2. Mack's formula revisited

Equipped with the results of the last section, we now proceed to derive a simple error formula when the historical triangle is complete, and we will show that it coincides with Mack's well-known formula (Mack, 1993).

The following lemma will be crucial for achieving our goal.

Lemma 4.4. *Suppose the historical triangle is complete, and consider any* $j \in \{1, \dots, J\}$ *and any set* $\mathcal{H}$ *such that* $\mathcal{I}_j \subseteq \mathcal{H} \subseteq \mathcal{I}$. *Then* $\hat{C}_{\mathcal{H},j} = \hat{C}_{\mathcal{H},j-1}\hat{f}_j$.

Proof. $\hat{C}_{\mathcal{H},j} = \hat{C}_{\mathcal{I}_j,j} + \hat{C}_{\mathcal{H}\setminus\mathcal{I}_j,j}$, since $\mathcal{H}$ is a superset of $\mathcal{I}_j$. By completeness (see Definition 2.8) and since $j > 0$, $\hat{C}_{\mathcal{I}_j,j} = \hat{C}_{\mathcal{I}_{j-1}\cap\mathcal{I}_j,j} = \hat{C}_{\mathcal{H}_j,j}$. By definition of $\hat{f}_j$, we have $\hat{C}_{\mathcal{H}_j,j} = \hat{f}_j\hat{C}_{\mathcal{H}_j,j-1}$. Finally, all summands of $C_{\mathcal{H}\setminus\mathcal{I}_j,j}$ are future development states, hence $\hat{C}_{\mathcal{H}\setminus\mathcal{I}_j,j} = \hat{C}_{\mathcal{H}\setminus\mathcal{I}_j,j-1}\hat{f}_j$. Putting this together proves our claim. ■

We can now proceed to formulate our first main result:

Result 4.5 (Mack's formula revisited). *If the historical triangle is complete, then the (conditional) mean squared error of prediction* $\mathrm{msep}_{C-\hat{C}}$ *of the ultimate loss*

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

C by the predictor $\hat{C}$ can be approximated as

$$\frac{msep_{C-\hat{C}}}{\hat{C}^2} \approx \sum_{j=1}^{J} \hat{u}_j^2 \hat{q}_j = \sum_{j=1}^{J} \hat{u}_j^2 (1-\hat{q}_j)\,\hat{q}_j + \sum_{j=1}^{J} \hat{u}_j^2 \hat{q}_j^2$$

and the decomposition at the end gives the split into process and parameter error. Furthermore, the approximation for $msep_{C-\hat{C}}$ given here coincides with that given by Mack (Mack, 1993, Corollary to Theorem 3).

Derivation. Using (22) in (31), we infer

$$\mathrm{msep}^{\mathrm{parm}}_{C-\hat{C}} \approx \hat{C}^2 \sum_{j=1}^{J} \hat{q}_j^2 \frac{\hat{f}_j^2 \hat{u}_j^2}{\hat{f}_j^2} = \hat{C}^2 \sum_{j=1}^{J} \hat{u}_j^2 \hat{q}_j^2.$$

By our completeness assumption, repeated application of Lemma 4.4, and Lemma 2.11, we see that

$$\frac{C_{\mathcal{H}_j,j-1}}{\hat{C}_{\mathcal{I}_j^*,j-1}} = \frac{\hat{C}_{\mathcal{H}_j,J}}{\hat{C}_{\mathcal{I}_j^*,J}} = \frac{\hat{C}_{\mathcal{I}_j,J}}{\hat{C}_{\mathcal{I}_j^*,J}} = \frac{1-\hat{q}_j}{\hat{q}_j}.$$

Using this, the definitions of $\hat{f}_j$ and $\hat{u}_j$, and (4), we can approximate and estimate

$$V[F_j|\mathcal{D}] \approx \frac{\phi_j}{E[C_{\mathcal{I}_j^*,j-1}|\mathcal{D}]} \approx \frac{\hat{\phi}_j}{\hat{C}_{\mathcal{I}_j^*,j-1}} = \frac{\hat{\phi}_j}{C_{\mathcal{H}_j,j-1}} \frac{1-\hat{q}_j}{\hat{q}_j} = \hat{f}_j^2 \hat{u}_j^2 \frac{1-\hat{q}_j}{\hat{q}_j}.$$

Plugging this into (30) and replacing $E[C|\mathcal{D}]$, q_j and f_j by their estimators $\hat{C}$, $\hat{q}_j$ and $\hat{f}_j$ yields

$$\mathrm{msep}^{\mathrm{proc}}_{C-\hat{C}} \approx \hat{C}^2 \sum_{j=1}^{J} \hat{q}_j^2 \frac{\hat{f}_j^2 \hat{u}_j^2 (1-\hat{q}_j)/\hat{q}_j}{\hat{f}_j^2} = \hat{C}^2 \sum_{j=1}^{J} \hat{u}_j^2 (1-\hat{q}_j)\hat{q}_j.$$

Now that we have found approximations for the right-hand sides of (30) and (31), we see that their sum serves as an approximation for the right-hand side of (29), and all claims of Result 4.5 are thus proved — except the identity of our approximation of $\mathrm{msep}_{C-\hat{C}}$ with Mack's version, for which we refer to the appendix. ■

Remark 4.6. Loosely speaking, Result 4.5 states that the (relative) mean squared error is the sum of all products "relative (squared) accuracy" $\hat{u}_j^2$ times "influence" $\hat{q}_j$ of the development factors (cf. Section 2.7). See also Remark 5.7 for yet another way to present these formulae.

Remark 4.7. We close this section with a remark on the Taylor approximation error. For the process error part, Taylor approximation gave (30), and Jensen's inequality (4) allowed us to approximate $V[F_j|\mathcal{D}]$ by $\phi_j/E[C_{\mathcal{I}_j^*,j-1}|\mathcal{D}]$ in the derivation of Result 4.5, which in combination results in the approximation

$$\mathrm{msep}^{\mathrm{proc}}_{C-\hat{C}} \approx (E[C|\mathcal{D}])^2 \sum_{j=1}^{J} q_j^2 \frac{\phi_j}{f_j^2 E[C_{\mathcal{I}_j^*,j-1}|\mathcal{D}]}$$

which was used in the proof of Result 4.5. The point we would like to make is that this is actually an exact equation. To see this, we need to recall Mack's derivation of the process error from Mack (1993) (adapted to our non-homogeneous chain ladder process): note first that, by definition,

$$\mathrm{msep}^{\mathrm{proc}}_{C-\hat{C}} = V[C|\mathcal{D}].$$

Next, by (1),

$$E[C|\mathcal{D}] = (\alpha_J \circ \ldots \circ \alpha_j)\,[0] + f_J \cdot \ldots \cdot f_j \cdot E[C_{\mathcal{I}_j^*,j-1}|\mathcal{D}],$$

where both the first summand and $E[C_{\mathcal{I}_j^*,j-1}|\mathcal{D}]$, when written as a function of the A_k and f_k, are independent of the particular development factor f_j; hence

$$q_j = \partial_{\log[f_j]} \log[E[C|\mathcal{D}]] = \frac{f_j \partial_{f_j} E[C|\mathcal{D}]}{E[C|\mathcal{D}]} = \frac{f_J \cdot \ldots \cdot f_j \cdot E[C_{\mathcal{I}_j^*,j-1}|\mathcal{D}]}{E[C|\mathcal{D}]}.$$

Now by (1) and (2),

$$\begin{aligned} f_J^2 \cdot f_{J-1}^2 \cdot \ldots \cdot f_{j+1}^2 V[C_{\mathcal{I}_{j+1}^*,j}|\mathcal{D}] &= f_J^2 \cdot f_{J-1}^2 \cdot \ldots \cdot f_{j+1}^2 V[C_{\mathcal{I}_j^*,j}|\mathcal{D}] \\ &= f_J^2 \cdot \ldots \cdot f_{j+1}^2 \phi_j E[C_{\mathcal{I}_j^*,j-1}|\mathcal{D}] \\ &\quad + f_J^2 \cdot \ldots \cdot f_{j+1}^2 \cdot f_j^2 V[C_{\mathcal{I}_j^*,j-1}|\mathcal{D}] \\ &= (E[C|\mathcal{D}])^2 q_j^2 \frac{\phi_j}{f_j^2 E[C_{\mathcal{I}_j^*,j-1}|\mathcal{D}]} \\ &\quad + f_J^2 \cdot \ldots \cdot f_j^2 V[C_{\mathcal{I}_j^*,j-1}|\mathcal{D}], \end{aligned}$$

from which our claim follows by (descending) induction on j (starting with $j = J$, and interpreting the empty product $f_J^2 \cdot f_{J-1}^2 \cdot \ldots \cdot f_{J+1}^2$ as 1).

5. The MSEP under partial loss development

While in Section 4 we applied the results of Section 3.2 to the analysis of the mean squared error of prediction of the ultimate loss, we extend our analysis in this section to the case of partial development between today and some future

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

development horizon, and also to today's expectation of the claims development result between two future horizons.

The key observation in Section 5.1 is that future chain ladder predictors $\tilde{C}$ for the ultimate loss are rational functions of the future link ratios $F_{i,j}$; not surprisingly, our current predictor $\hat{C}$ is recovered in all of these predictors upon replacing the random variables $F_{i,j}$ by today's estimator $\hat{f}_j$. After a careful analysis of the partial derivatives in Section 5.2, we may then use the Preliminary Approximations 3.1 and derive an approximate MSEP for the prediction of $\tilde{C}$ by $\hat{C}$. But using today's predictor $\hat{C}$ as a predictor for any future $\tilde{C}$ is the same as using 0 as today's predictor of the claims development result between any two development horizons, which leads us to formulate our Main Result 5.3 for this general case.

5.1. Predicting future chain ladder estimates

Our first step will be to analyze how the chain ladder predictor of the ultimate loss changes once more of the development becomes known. The notation used to deal with the extended information is the following:

Definition 5.1. *Suppose we are given a mapping $i \mapsto \tilde{j}_i$ defined for all $i \in \mathcal{I}$ such that $\tilde{j}_i$ is an integer and $j_i \le \tilde{j}_i \le J$. We then set, for each non-negative integer j,*

$$\tilde{\mathcal{I}}_j := \mathcal{I}_j \cup \{i \in \mathcal{I} \mid j_i < j \le \tilde{j}_i\},$$

$$\tilde{\mathcal{I}}^*_j := \mathcal{I}^*_j \setminus \{i \in \mathcal{I} \mid j_i < j \le \tilde{j}_i\} = \{i \in \mathcal{I} \mid \tilde{j}_i < j\},$$

in analogy to $\mathcal{I}_j$ and $\mathcal{I}^*_j$, and for each $j \in \{1, \ldots, J\}$,

$$\tilde{\mathcal{H}}_j := \mathcal{H}_j \cup \{i \in \mathcal{I} \mid j_i < j \le \tilde{j}_i\} = \mathcal{H}_j \cup \mathcal{I}^*_j \setminus \tilde{\mathcal{I}}^*_j,$$

and we call the sets

$$\tilde{\mathcal{D}} := \{C_{i,j} | i \in \tilde{\mathcal{I}}_j\} \quad \text{and} \quad \{C_{i,\tilde{j}_i} | i \in \mathcal{I}\} \subseteq \tilde{\mathcal{D}}$$

of random variables a future development horizon *and its corresponding* future diagonal*, respectively.*

For example, development to the ultimate horizon would be described by setting $\tilde{j}_i := J$, while development over the next period (the "one-period horizon") would be described by $\tilde{j}_i := \min[J, j_i + 1]$.

After the additional development, the chain ladder predictors will become

$$\tilde{C}_{i,J} := C_{i,j_i} \left(\prod_{j=j_i+1}^{\tilde{j}_i} F_{i,j} \right) \left(\prod_{j=\tilde{j}_i+1}^{J} \tilde{f}_j \right)$$

where

$$\tilde{f}_j := \frac{C_{\tilde{\mathcal{H}}_j,j}}{C_{\tilde{\mathcal{H}}_j,j-1}} = \frac{C_{\mathcal{H}_j,j} + \sum_{i\in\mathcal{I}_j^*\setminus\tilde{\mathcal{I}}_j^*} C_{i,j_i} F_{i,j_i+1}\cdot\ldots\cdot F_{i,j-1}F_{i,j}}{C_{\mathcal{H}_j,j-1} + \sum_{i\in\mathcal{I}_j^*\setminus\tilde{\mathcal{I}}_j^*} C_{i,j_i} F_{i,j_i+1}\cdot\ldots\cdot F_{i,j-1}}.$$

Summing over all loss portfolios represented in $\mathcal{I}$, and using our shorthand summation notation 2.7 also for the $\tilde{C}_{i,j}$, our future predictor of the ultimate loss will become

$$\tilde{C} := \tilde{C}_{\mathcal{I},J}.$$

From today's point of view, $\tilde{C}$ is a random variable, the randomness stemming from the future link ratios $F_{i,j}$ up to the development horizon $\tilde{\mathcal{D}}$.

This function $\tilde{C}$ of the future link ratios is another example of the class of functions G discussed in Section 3.2. We have

$$\tilde{C}_{|\hat{f}} = \hat{C},$$

because $\tilde{f}_j$, as can be seen by substituting $C_{\mathcal{H}_j,j} = C_{\mathcal{H}_j,j-1}\hat{f}_j$ in the above fraction, is a weighted average of $\hat{f}_j$ and some future link ratios $F_{i,j}$ with the same index j (note, however, that in general $\tilde{C}_{|f} \neq E[\tilde{C}|\mathcal{D}]$). That way, $\hat{C}$ serves as today's predictor of any such future predictor $\tilde{C}$ of the ultimate loss.

We are now almost ready to apply Result 3.1 to calculate an MSEP for such future predictors as $\tilde{C}$. The only thing missing is a further evaluation of the sensitivities $\partial_{F_{i,j}}\tilde{C}$. We will address this in the next section.

5.2. Calculating the sensitivities

This section is devoted to the calculation of the partial derivatives $\partial_{F_{i,j}}\tilde{C}_{t,J}$.

Lemma 5.2. *For each $t \in \mathcal{I}$ and each future (i, j) (i.e. $i \in \mathcal{I}_j^*$),*

$$\left(\partial_{F_{i,j}}\tilde{C}_{t,J}\right)_{|\hat{f}} = \mathbf{1}_{i\in\mathcal{I}_j^*\setminus\tilde{\mathcal{I}}_j^*}\frac{\hat{C}_{t,J}}{\hat{f}_j}\left(\mathbf{1}_{t=i} + \frac{\hat{C}_{i,j}}{\hat{C}_{\tilde{\mathcal{H}}_j,j}}\mathbf{1}_{t\in\tilde{\mathcal{I}}_j^*}\right). \tag{32}$$

Proof. For any $k \in \{1, \ldots, J\}$, we have

$$\partial_{F_{i,j}}C_{t,k} = \partial_{F_{i,j}}C_{t,j_t}\cdot F_{t,j_t+1}\cdot\ldots\cdot F_{t,k} = \mathbf{1}_{i=t}\mathbf{1}_{j_i<j\le k}\frac{C_{i,k}}{F_{i,j}}.$$

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

Using this, let us calculate $\partial_{F_{i,j}} \tilde{f}_k$:

$$\begin{aligned}
\partial_{F_{i,j}} \tilde{f}_k &= \partial_{F_{i,j}} \frac{C_{\tilde{\mathcal{H}}_k,k}}{C_{\tilde{\mathcal{H}}_k,k-1}} \\
&= \frac{\mathbf{1}_{i\in\tilde{\mathcal{H}}_k}}{F_{i,j}} \frac{\mathbf{1}_{j_i<j\le k} C_{i,k} C_{\tilde{\mathcal{H}}_k,k-1} - C_{\tilde{\mathcal{H}}_k,k} \mathbf{1}_{j_i<j\le k-1} C_{i,k-1}}{C^2_{\tilde{\mathcal{H}}_k,k-1}} \\
&= \frac{\mathbf{1}_{i\in\tilde{\mathcal{H}}_k}}{F_{i,j}} \frac{\mathbf{1}_{j_i<j\le k} F_{i,k} C_{i,k-1} C_{\tilde{\mathcal{H}}_k,k-1} - \tilde{f}_k C_{\tilde{\mathcal{H}}_k,k-1} \mathbf{1}_{j_i<j\le k-1} C_{i,k-1}}{C^2_{\tilde{\mathcal{H}}_k,k-1}} \\
&= \frac{\mathbf{1}_{i\in\tilde{\mathcal{H}}_k} \mathbf{1}_{j_i<j}}{F_{i,j}} \left(\mathbf{1}_{j=k} \frac{C_{i,j}}{C_{\tilde{\mathcal{H}}_j,j-1}} + \mathbf{1}_{j\le k-1} \frac{(F_{i,k} - \tilde{f}_k) C_{i,k-1}}{C_{\tilde{\mathcal{H}}_k,k-1}} \right).
\end{aligned}$$

But replacing future link ratios with their estimators, both $F_{i,k}$ and $\tilde{f}_k$ turn into $\hat{f}_k$, which means that

$$\left(\partial_{F_{i,j}} \tilde{f}_k\right)_{|\hat{f}} = \mathbf{1}_{j=k} \mathbf{1}_{i\in\tilde{\mathcal{H}}_j} \mathbf{1}_{j_i<j} \frac{\hat{C}_{i,j}}{\hat{f}_j \hat{C}_{\tilde{\mathcal{H}}_j,j-1}} = \mathbf{1}_{j=k} \mathbf{1}_{i\in\mathcal{I}^*_j \setminus \tilde{\mathcal{I}}^*_j} \frac{\hat{C}_{i,j}}{\hat{C}_{\tilde{\mathcal{H}}_j,j}}.$$

With this we are now ready to prove the lemma:

$$\begin{aligned}
\left(\partial_{F_{i,j}} \tilde{C}_{t,J}\right)_{|\hat{f}} &= \left(\partial_{F_{i,j}} \left(C_{t,j_t} \left(\prod_{k=j_t+1}^{\tilde{j}_t} F_{t,k} \right) \left(\prod_{k=\tilde{j}_t+1}^{J} \tilde{f}_k \right) \right) \right)_{|\hat{f}} \\
&= \hat{C}_{t,J} \left(\sum_{k=j_t+1}^{\tilde{j}_t} \frac{\partial_{F_{i,j}} F_{t,k}}{F_{t,k}} + \sum_{k=\tilde{j}_t+1}^{J} \frac{\partial_{F_{i,j}} \tilde{f}_k}{\tilde{f}_k} \right)_{|\hat{f}} \\
&= \hat{C}_{t,J} \left(\sum_{k=j_t+1}^{\tilde{j}_t} \frac{\mathbf{1}_{i=t} \mathbf{1}_{j=k}}{\hat{f}_k} + \sum_{k=\tilde{j}_t+1}^{J} \frac{\mathbf{1}_{j=k} \mathbf{1}_{i\in\mathcal{I}^*_j \setminus \tilde{\mathcal{I}}^*_j} \hat{C}_{i,j}}{\hat{f}_k \hat{C}_{\tilde{\mathcal{H}}_j,j}} \right) \\
&= \hat{C}_{t,J} \left(\mathbf{1}_{i=t} \frac{\mathbf{1}_{j_i<j\le\tilde{j}_i}}{\hat{f}_j} + \frac{1}{\hat{f}_j} \frac{\hat{C}_{i,j}}{\hat{C}_{\tilde{\mathcal{H}}_j,j}} \mathbf{1}_{i\in\mathcal{I}^*_j \setminus \tilde{\mathcal{I}}^*_j} \mathbf{1}_{j>\tilde{j}_t} \right).
\end{aligned}$$

To see that the last expression equals the one stated in the lemma, note that the conditions $j_i < j \le \tilde{j}_i$ and $i \in \mathcal{I}^*_j \setminus \tilde{\mathcal{I}}^*_j$ are equivalent. Finally, $j > \tilde{j}_t$ is equivalent to $t \in \tilde{\mathcal{I}}^*_j$. ■

5.3. Error formulae

Equipped with our Preliminary Approximations 3.1 and the formula for $\partial_{F_{i,j}} \tilde{C}_{t,J}$ evaluated at $\hat{f}$ derived in the previous section, we are now almost ready to prove our Main Result 5.3. Our aim is to generalize Result 4.5 so that we can handle incomplete triangles, the development between two arbitrary horizons and also calculate prediction errors for subportfolios. To achieve all that, some additional notation needs to be introduced.

We start with the definition of a quantity that will capture the effect of the missing data and the discarded outliers: for $j \in \{1, \ldots, J\}$, let

$$\hat{r}_j := \frac{\hat{C}_{\mathcal{H}_j \cup \mathcal{I}_j^*, j}}{\hat{C} / (\hat{f}_{j+1} \cdot \ldots \cdot \hat{f}_J)} = \frac{\hat{C}_{\mathcal{H}_j \cup \mathcal{I}_j^*, j}}{\hat{C}\, \hat{C}_{i,j} / \hat{C}_{i,J}} = \frac{C_{\mathcal{H}_j, j}}{\hat{C}\, \hat{C}_{i,j} / \hat{C}_{i,J}} + \hat{q}_j \quad (\text{for } i \in \mathcal{I}_j^*).$$

The second equality holds since for any $i \in \mathcal{I}_j^*$, the quotient $\hat{C}_{i,J}/\hat{C}_{i,j}$ is just $\hat{f}_{j+1} \cdot \ldots \cdot \hat{f}_J$. If the triangle is complete, then $\mathcal{H}_j \cup \mathcal{I}_j^* = \mathcal{I}$, which together with Lemma 4.4 implies $\hat{r}_j = 1$. If the triangle is not complete, the ratios $\hat{r}_j$ may differ from 1, and they summarize all information we need about the deviation from completeness, as we will see shortly. Note that in any case, $\hat{r}_j > \hat{q}_j$.

Our next definition introduces a new set of influence factors,

$$\tilde{q}_j := \hat{C}_{\tilde{\mathcal{I}}_j^*, J} / \hat{C}, \qquad j \in \{0, 1, \ldots, J\},$$

which are based on $\tilde{\mathcal{I}}_j$ rather than $\mathcal{I}_j$. Note that while the $\tilde{C}_{i,j}$ are random variables, the $\tilde{q}_j$ are statistics of the historical data — we might have chosen a notation like $\hat{q}_{\mathcal{I}_j^*}$ and $\hat{q}_{\tilde{\mathcal{I}}_j^*}$, but chose $\hat{q}_j$ and $\tilde{q}_j$ for brevity.

Suppose that besides $\tilde{\mathcal{D}}$ we consider a second future horizon, for which we put the sign $\breve{}$ on top to distinguish the respective items — e.g. $\breve{\jmath}_i$ or $\breve{\mathcal{D}}$ or $\breve{C}_{i,J}$ (and we also apply our summation convention 2.7 to the latter). We demand that $\breve{\mathcal{D}}$ is beyond $\tilde{\mathcal{D}}$, i.e. that $\breve{\jmath}_i \geq \tilde{\jmath}_i$ for all $i \in \mathcal{I}$. We then have two future predictors, $\tilde{C}$ and $\breve{C}$, which are both random variables from today's point of view, and for both of which today's prediction is just $\hat{C}$. In other words, today's prediction of the loss development result between the two future horizons is 0. This number has its prediction error, for which we will provide an approximation in Result 5.3.

The prediction error for $\breve{C} - \tilde{C}$ refers to the whole loss portfolio, since $\breve{C}$ and $\tilde{C}$ do. In the literature, one also generally finds formulae that refer to the error related to a subportfolio only — usually a single accident year. We will also present such formulae, but take the more general point of view of considering the subportfolio identified by any given non-empty subset $\mathcal{H} \subseteq \mathcal{I}$. Thus, we will approximate the prediction error that 0 has as a predictor of the loss development result $\breve{C}_{\mathcal{H},J} - \tilde{C}_{\mathcal{H},J}$. To deal with subportfolios, we use the following notation (for $1 \le j \le J$):

$$\hat{q}_{\mathcal{H},j} := \hat{C}_{\mathcal{H}\cap\mathcal{I}_j^*,J}/\hat{C}, \qquad \tilde{q}_{\mathcal{H},j} := \hat{C}_{\mathcal{H}\cap\tilde{\mathcal{I}}_j^*,J}/\hat{C}, \qquad \check{q}_{\mathcal{H},j} := \hat{C}_{\mathcal{H}\cap\check{\mathcal{I}}_j^*,J}/\hat{C},$$

$$\hat{s}_{\mathcal{H},j} := 1 + \frac{\hat{q}_{\mathcal{H},j}}{\hat{r}_j - \hat{q}_j}, \qquad \tilde{s}_{\mathcal{H},j} := 1 + \frac{\tilde{q}_{\mathcal{H},j}}{\hat{r}_j - \tilde{q}_j}, \qquad \check{s}_{\mathcal{H},j} := 1 + \frac{\check{q}_{\mathcal{H},j}}{\hat{r}_j - \check{q}_j},$$

$$\hat{s}_j := \hat{s}_{\mathcal{I},j}, \qquad \tilde{s}_j := \tilde{s}_{\mathcal{I},j}, \qquad \check{s}_j := \check{s}_{\mathcal{I},j}.$$

Note that $\tilde{q}_{\mathcal{I},j} = \tilde{q}_j$ etc., and that for disjoint sets $\mathcal{H}, \mathcal{H}'$, we have

$$\tilde{q}_{\mathcal{H}\cup\mathcal{H}',j} = \tilde{q}_{\mathcal{H},j} + \tilde{q}_{\mathcal{H}',j} \quad \text{and} \quad \tilde{s}_{\mathcal{H}\cup\mathcal{H}',j} - 1 = (\tilde{s}_{\mathcal{H},j} - 1) + (\tilde{s}_{\mathcal{H}',j} - 1),$$

so the notation is compatible with our summation convention 2.7 (strictly so for the q's, and "with a twist" for the s's).

Let us apply our new notation to establish a useful relation between the $\hat{u}_{i,j}$ and $\hat{u}_j$, namely

$$\hat{u}_{i,j}^2 = \hat{u}_j^2 \frac{C_{\mathcal{H}_j,j}}{\hat{C}_{i,j}} = \hat{u}_j^2(\hat{r}_j - \hat{q}_j)\frac{\hat{C}}{\hat{C}_{i,J}} = \frac{\hat{u}_j^2 \hat{r}_j}{\hat{s}_j}\frac{\hat{C}}{\hat{C}_{i,J}}, \tag{33}$$

which holds for any $i \in \mathcal{I}_j^*$ in accordance with the definition of $\hat{r}_j$.

We are now ready to state our main result:

Main Result 5.3. *For $j \in \{1, \ldots, J\}$, let $\hat{U}_j := \hat{u}_j \hat{r}_j$, and let $\mathcal{H} \subseteq \mathcal{I}$ denote any non-empty subset of $\mathcal{I}$. Then the (conditional) mean squared prediction error of the development result* 0 *between the two horizons $\tilde{\mathcal{D}}$ and $\check{\mathcal{D}}$ for the portfolio $\mathcal{H}$,*

$$msep_{(\check{C}_{\mathcal{H},J} - \tilde{C}_{\mathcal{H},J})-0} := E[(\check{C}_{\mathcal{H},J} - \tilde{C}_{\mathcal{H},J})^2 | \mathcal{D}],$$

satisfies the approximate equation

$$\frac{\mathrm{msep}_{(\check{C}_{\mathcal{H},J} - \tilde{C}_{\mathcal{H},J})-0}}{\hat{C}^2} \approx \sum_{j=1}^{J} \hat{U}_j^2 \left(\frac{\frac{\tilde{s}_{\mathcal{H},j}}{\tilde{s}_j}(\tilde{s}_{\mathcal{H},j} - 1) - \frac{\check{s}_{\mathcal{H},j}}{\check{s}_j}(\check{s}_{\mathcal{H},j} - 1)}{\hat{s}_j} \right) \tag{34}$$

and the parameter error (divided by $\hat{C}^2$) can be approximated by

$$\frac{\mathrm{msep}^{\mathrm{parm}}_{(\check{C}_{\mathcal{H},J} - \tilde{C}_{\mathcal{H},J})-0}}{\hat{C}^2} \approx \sum_{j=1}^{J} \hat{U}_j^2 \left(\frac{\tilde{s}_{\mathcal{H},j} - \check{s}_{\mathcal{H},j}}{\hat{s}_j} \right)^2, \tag{35}$$

while the process error (divided by $\hat{C}^2$) is approximated by the difference of the right-hand sides of (34) *and* (35)*. If $\mathcal{H} = \mathcal{I}$, setting $\hat{Q}_j := (\tilde{s}_j - \check{s}_j)/\hat{s}_j$, this becomes*

$$\frac{\mathrm{msep}_{(\check{C} - \tilde{C})-0}}{\hat{C}^2} \approx \sum_{j=1}^{J} \hat{U}_j^2 \hat{Q}_j = \sum_{j=1}^{J} \hat{U}_j^2 (1 - \hat{Q}_j)\hat{Q}_j + \sum_{j=1}^{J} \hat{U}_j^2 \hat{Q}_j^2, \tag{36}$$

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

and the first sum in the last term approximates the process error and the second the parameter error (divided by $\hat{C}^2$).

Derivation. We use Lemma 5.2 and the fact that

$$\hat{r}_j = \frac{\hat{C}_{\mathcal{H}_j \cup \mathcal{I}_j^*, j}}{\hat{C}\,\hat{C}_{i,j}/\hat{C}_{i,J}} = \frac{\hat{C}_{\mathcal{H}_j \cup \mathcal{I}_j^* \setminus \tilde{\mathcal{I}}_j^*, j}}{\hat{C}\,\hat{C}_{i,j}/\hat{C}_{i,J}} + \tilde{q}_j$$

to conclude that for $i \in \mathcal{I}_j^*$,

$$\begin{aligned}\left(\partial_{F_{i,j}} \tilde{C}_{\mathcal{H},J}\right)_{|\hat{f}} &= \mathbf{1}_{i \in \mathcal{I}_j^* \setminus \tilde{\mathcal{I}}_j^*} \left(\mathbf{1}_{i \in \mathcal{H}} \frac{\hat{C}_{i,J}}{\hat{f}_j} + \frac{\hat{C}_{i,j}\hat{C}_{\mathcal{H} \cap \tilde{\mathcal{I}}_j^*, J}}{\hat{f}_j \hat{C}_{\mathcal{H}_j \cup \mathcal{I}_j^* \setminus \tilde{\mathcal{I}}_j^*, j}} \right) \\ &= \mathbf{1}_{i \in \mathcal{I}_j^* \setminus \tilde{\mathcal{I}}_j^*} \frac{\hat{C}_{i,J}}{\hat{f}_j} \left(\mathbf{1}_{i \in \mathcal{H}} + \frac{\tilde{q}_{\mathcal{H},j}}{\hat{r}_j - \tilde{q}_j} \right) = \mathbf{1}_{i \in \mathcal{I}_j^* \setminus \tilde{\mathcal{I}}_j^*} \frac{\hat{C}_{i,J}}{\hat{f}_j} \left(\tilde{s}_{\mathcal{H},j} - \mathbf{1}_{i \notin \mathcal{H}} \right),\end{aligned}$$

and we get an analogous expression for $(\partial_{F_{i,j}} \check{C}_{\mathcal{H},J})_{|\hat{f}}$.

We now apply Result 3.1 to the random variable $G = \check{C}_{\mathcal{H},J} - \tilde{C}_{\mathcal{H},J}$, for which $G_{|\hat{f}} = 0$. Together with the partial derivative just calculated, we obtain

$$\begin{aligned}\mathrm{msep}^{\mathrm{parm}}_{(\check{C}_{\mathcal{H},J} - \tilde{C}_{\mathcal{H},J})-0} &\approx \sum_{j=1}^{J} \left(\sum_{i \in \mathcal{I}_j^*} \left(\partial_{F_{i,j}} (\check{C}_{\mathcal{H},J} - \tilde{C}_{\mathcal{H},J}) \right)_{|\hat{f}} \right)^2 \hat{f}_j^2 \hat{u}_j^2 \\ &= \sum_{j=1}^{J} \left(\sum_{i \in \mathcal{I}_j^*} \frac{\hat{C}_{i,J}}{\hat{f}_j} \left(\mathbf{1}_{i \notin \check{\mathcal{I}}_j^*} \left(\check{s}_{\mathcal{H},j} - \mathbf{1}_{i \notin \mathcal{H}} \right) - \mathbf{1}_{i \notin \tilde{\mathcal{I}}_j^*} \left(\tilde{s}_{\mathcal{H},j} - \mathbf{1}_{i \notin \mathcal{H}} \right) \right) \right)^2 \hat{f}_j^2 \hat{u}_j^2 \\ &= \hat{C}^2 \sum_{j=1}^{J} \hat{U}_j^2 \left(\sum_{i \in \mathcal{I}_j^*} \frac{\hat{C}_{i,J}}{\hat{r}_j \hat{C}} \left(\mathbf{1}_{i \notin \check{\mathcal{I}}_j^*} \left(\check{s}_{\mathcal{H},j} - \mathbf{1}_{i \notin \mathcal{H}} \right) - \mathbf{1}_{i \notin \tilde{\mathcal{I}}_j^*} \left(\tilde{s}_{\mathcal{H},j} - \mathbf{1}_{i \notin \mathcal{H}} \right) \right) \right)^2\end{aligned}$$

Now

$$\sum_{i \in \mathcal{I}_j^*} \frac{\hat{C}_{i,J}}{\hat{r}_j \hat{C}} \mathbf{1}_{i \notin \check{\mathcal{I}}_j^*} = \sum_{i \in \mathcal{I}_j^*} \frac{\hat{C}_{i,J}}{\hat{r}_j \hat{C}} (1 - \mathbf{1}_{i \in \check{\mathcal{I}}_j^*}) = \frac{\hat{q}_j - \check{q}_j}{\hat{r}_j} = \frac{1}{\check{s}_j} - \frac{1}{\hat{s}_j}$$

and

$$\begin{aligned}\sum_{i \in \mathcal{I}_j^*} \frac{\hat{C}_{i,J}}{\hat{r}_j \hat{C}} \mathbf{1}_{i \notin \check{\mathcal{I}}_j^*} \mathbf{1}_{i \notin \mathcal{H}} &= \sum_{i \in \mathcal{I}_j^*} \frac{\hat{C}_{i,J}}{\hat{r}_j \hat{C}} (1 - \mathbf{1}_{i \in \check{\mathcal{I}}_j^*})(1 - \mathbf{1}_{i \in \mathcal{H}}) \\ &= \frac{\hat{q}_j - \check{q}_j}{\hat{r}_j} - \frac{\hat{q}_{\mathcal{H},j} - \check{q}_{\mathcal{H},j}}{\hat{r}_j} = \frac{\check{s}_{\mathcal{H},j}}{\check{s}_j} - \frac{\hat{s}_{\mathcal{H},j}}{\hat{s}_j}\end{aligned}$$

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

and similarly for the terms involving $\tilde{\mathcal{I}}_j$. With these formulae, we can continue our evaluation of the above approximation of the parameter error:

$$\begin{aligned}
\dots &= \hat{C}^2 \sum_{j=1}^{J} \hat{U}_j^2 \left(\frac{\breve{s}_{\mathcal{H},j}}{\breve{s}_j} - \frac{\breve{s}_{\mathcal{H},j}}{\hat{s}_j} - \frac{\hat{s}_{\mathcal{H},j}}{\breve{s}_j} + \frac{\hat{s}_{\mathcal{H},j}}{\hat{s}_j} - \frac{\tilde{s}_{\mathcal{H},j}}{\tilde{s}_j} + \frac{\tilde{s}_{\mathcal{H},j}}{\hat{s}_j} + \frac{\hat{s}_{\mathcal{H},j}}{\tilde{s}_j} - \frac{\hat{s}_{\mathcal{H},j}}{\hat{s}_j} \right)^2 \\
&= \hat{C}^2 \sum_{j=1}^{J} \hat{U}_j^2 \left(\frac{\tilde{s}_{\mathcal{H},j} - \breve{s}_{\mathcal{H},j}}{\hat{s}_j} \right)^2,
\end{aligned}$$

which proves (35).

For the process error, we invoke Result 3.1 and (33) to conclude

$$\begin{aligned}
\mathrm{msep}^{\mathrm{proc}}_{(\breve{C}_{\mathcal{H},J} - \tilde{C}_{\mathcal{H},J}) - 0} &\approx \sum_{j=1}^{J} \sum_{i \in \mathcal{I}_j^*} \left(\partial_{F_{i,j}} (\breve{C}_{\mathcal{H},J} - \tilde{C}_{\mathcal{H},J}) \right)^2_{|\hat{f}} \hat{f}_j^2 \frac{\hat{u}_j^2 \hat{r}_j}{\hat{s}_j} \frac{\hat{C}}{\hat{C}_{i,J}} \\
&= \hat{C}^2 \sum_{j=1}^{J} \frac{\hat{U}_j^2}{\hat{s}_j} \sum_{i \in \mathcal{I}_j^*} \frac{\hat{C}_{i,J}}{\hat{r}_j \hat{C}} \left(\mathbf{1}_{i \notin \breve{\mathcal{I}}_j^*} \left(\breve{s}_{\mathcal{H},j} - \mathbf{1}_{i \notin \mathcal{H}} \right) - \mathbf{1}_{i \notin \tilde{\mathcal{I}}_j^*} \left(\tilde{s}_{\mathcal{H},j} - \mathbf{1}_{i \notin \mathcal{H}} \right) \right)^2.
\end{aligned}$$

Let us analyze the sum over i: expanding the squares, we get three sums,

$$\sum_{i \in \mathcal{I}_j^*} \frac{\hat{C}_{i,J}}{\hat{r}_j \hat{C}} \mathbf{1}_{i \notin \breve{\mathcal{I}}_j^*} \left(\breve{s}_{\mathcal{H},j} - \mathbf{1}_{i \notin \mathcal{H}} \right)^2 = \left(\frac{1}{\breve{s}_j} - \frac{1}{\hat{s}_j} \right) \breve{s}_{\mathcal{H},j}^2 + \left(\frac{\breve{s}_{\mathcal{H},j}}{\breve{s}_j} - \frac{\hat{s}_{\mathcal{H},j}}{\hat{s}_j} \right) (1 - 2\breve{s}_{\mathcal{H},j}),$$

$$\sum_{i \in \mathcal{I}_j^*} \frac{\hat{C}_{i,J}}{\hat{r}_j \hat{C}} \mathbf{1}_{i \notin \tilde{\mathcal{I}}_j^*} \left(\tilde{s}_{\mathcal{H},j} - \mathbf{1}_{i \notin \mathcal{H}} \right)^2 = \left(\frac{1}{\tilde{s}_j} - \frac{1}{\hat{s}_j} \right) \tilde{s}_{\mathcal{H},j}^2 + \left(\frac{\tilde{s}_{\mathcal{H},j}}{\tilde{s}_j} - \frac{\hat{s}_{\mathcal{H},j}}{\hat{s}_j} \right) (1 - 2\tilde{s}_{\mathcal{H},j})$$

and

$$\begin{aligned}
&\sum_{i \in \mathcal{I}_j^*} \frac{\hat{C}_{i,J}}{\hat{r}_j \hat{C}} (-2) \mathbf{1}_{i \notin \breve{\mathcal{I}}_j^*} \mathbf{1}_{i \notin \tilde{\mathcal{I}}_j^*} \left(\breve{s}_{\mathcal{H},j} - \mathbf{1}_{i \notin \mathcal{H}} \right) \left(\tilde{s}_{\mathcal{H},j} - \mathbf{1}_{i \notin \mathcal{H}} \right) \\
&\quad = \sum_{i \in \mathcal{I}_j^*} \frac{\hat{C}_{i,J}}{\hat{r}_j \hat{C}} (-2) \mathbf{1}_{i \notin \tilde{\mathcal{I}}_j^*} \left(\breve{s}_{\mathcal{H},j} \tilde{s}_{\mathcal{H},j} + \mathbf{1}_{i \notin \mathcal{H}} \left(1 - \breve{s}_{\mathcal{H},j} - \tilde{s}_{\mathcal{H},j} \right) \right) \\
&\quad = -2 \left(\frac{1}{\tilde{s}_j} - \frac{1}{\hat{s}_j} \right) \breve{s}_{\mathcal{H},j} \tilde{s}_{\mathcal{H},j} - 2 \left(\frac{\tilde{s}_{\mathcal{H},j}}{\tilde{s}_j} - \frac{\hat{s}_{\mathcal{H},j}}{\hat{s}_j} \right) (1 - \breve{s}_{\mathcal{H},j} - \tilde{s}_{\mathcal{H},j}),
\end{aligned}$$

where we used our assumption that the horizon $\breve{\mathcal{D}}$ is beyond $\tilde{\mathcal{D}}$ when we concluded that $\mathbf{1}_{i \notin \breve{\mathcal{I}}_j^*} \mathbf{1}_{i \notin \tilde{\mathcal{I}}_j^*} = \mathbf{1}_{i \notin \tilde{\mathcal{I}}_j^*}$. We add these three sums by expanding the products and regrouping the resulting 20 terms and obtain

$$\sum_{i\in\mathcal{I}_j^*}\frac{\hat{C}_{i,J}}{\hat{r}_j\hat{C}}\left(\mathbf{1}_{i\notin\check{\mathcal{I}}_j^*}\left(\check{s}_{\mathcal{H},j}-\mathbf{1}_{i\notin\mathcal{H}}\right)-\mathbf{1}_{i\notin\tilde{\mathcal{I}}_j^*}\left(\tilde{s}_{\mathcal{H},j}-\mathbf{1}_{i\notin\mathcal{H}}\right)\right)^2$$

$$=\left(-\frac{1}{\hat{s}_j}-\frac{1}{\hat{s}_j}+\frac{2}{\hat{s}_j}\right)\hat{s}_{\mathcal{H},j}+\left(\frac{1}{\tilde{s}_j}-\frac{2}{\tilde{s}_j}\right)\tilde{s}_{\mathcal{H},j}+\left(\frac{1}{\check{s}_j}\right)\check{s}_{\mathcal{H},j}$$
$$+\left(\frac{1}{\tilde{s}_j}-\frac{1}{\hat{s}_j}-\frac{2}{\tilde{s}_j}+\frac{2}{\tilde{s}_j}\right)\tilde{s}^2_{\mathcal{H},j}+\left(-\frac{2}{\tilde{s}_j}+\frac{2}{\hat{s}_j}+\frac{2}{\tilde{s}_j}\right)\tilde{s}_{\mathcal{H},j}\check{s}_{\mathcal{H},j}$$
$$+\left(\frac{1}{\check{s}_j}-\frac{1}{\hat{s}_j}-\frac{2}{\check{s}_j}\right)\check{s}^2_{\mathcal{H},j}+\left(\frac{2}{\hat{s}_j}-\frac{2}{\hat{s}_j}\right)\hat{s}_{\mathcal{H},j}\tilde{s}_{\mathcal{H},j}+\left(\frac{2}{\hat{s}_j}-\frac{2}{\hat{s}_j}\right)\hat{s}_{\mathcal{H},j}\check{s}_{\mathcal{H},j}$$
$$=-\frac{\tilde{s}_{\mathcal{H},j}}{\tilde{s}_j}+\frac{\check{s}_{\mathcal{H},j}}{\check{s}_j}+\frac{\tilde{s}^2_{\mathcal{H},j}}{\tilde{s}_j}-\frac{\tilde{s}^2_{\mathcal{H},j}}{\hat{s}_j}+2\frac{\tilde{s}_{\mathcal{H},j}\check{s}_{\mathcal{H},j}}{\hat{s}_j}-\frac{\check{s}^2_{\mathcal{H},j}}{\check{s}_j}-\frac{\check{s}^2_{\mathcal{H},j}}{\hat{s}_j}$$
$$=\frac{\tilde{s}_{\mathcal{H},j}}{\tilde{s}_j}(\tilde{s}_{\mathcal{H},j}-1)-\frac{\check{s}_{\mathcal{H},j}}{\check{s}_j}(\check{s}_{\mathcal{H},j}-1)-\frac{(\tilde{s}_{\mathcal{H},j}-\check{s}_{\mathcal{H},j})^2}{\hat{s}_j}.$$

Combining this with the formula for the parameter error already established proves (34). Finally, if $\mathcal{H}=\mathcal{I}$, then the terms in brackets in (34) and (35) both specialize to $\hat{Q}_j$, and we get (36). ■

We discuss the Main Result in a series of remarks:

Remark 5.4. Merz and Wüthrich (2014) analyze $\text{msep}_{(\check{C}-\tilde{C})-0}$ by embedding the classical chain ladder model in a Bayesian chain ladder model and deriving approximations in the non-informative prior case. Numerical examples indicate that their results coincide exactly with our formula (36) above for the total MSEP. Merz and Wüthrich apply their results to address the issue of risk margin calculations for solvency purposes. Indeed, since

$$E[(\check{C}-\tilde{C})^2|\mathcal{D}]=E[E[(\check{C}-\tilde{C})^2|\tilde{\mathcal{D}}]|\mathcal{D}],$$

the formula (36) gives an approximation of today's expectation of the "future MSEP (at $\tilde{\mathcal{D}}$)" for the claims development result between $\tilde{\mathcal{D}}$ and $\check{\mathcal{D}}$. As such, it is important for the calculation of risk margins and cost-of-capital loadings. We refer to Salzmann and Wüthrich (2010) and Merz and Wüthrich (2014) for a detailed discussion.

Remark 5.5. For the development over the one-year horizon, our formula generalizes the formula given by Merz and Wüthrich (2008) and Bühlmann *et al.* (2009) — for a proof, see the appendix. It also specializes to Result 4.5 (case $\mathcal{H}=\mathcal{I}$, $\tilde{q}_j=\hat{q}_j$, $\check{q}_j=0$, $\hat{r}_j=1$ for all j).

Remark 5.6. The right-hand side of (34) depends on the two horizons $\tilde{\mathcal{D}}$ and $\check{\mathcal{D}}$ only via the difference $\tilde{s}_{\mathcal{H},j}(\tilde{s}_{\mathcal{H},j}-1)/\tilde{s}_j-\check{s}_{\mathcal{H},j}(\check{s}_{\mathcal{H},j}-1)/\check{s}_j$. From this, it is clear that our approximation for the total mean squared error of prediction is additive: dividing the development step between $\tilde{\mathcal{D}}$ and $\breve{\mathcal{D}}$ in two parts corresponds to a split of the approximate MSEP into two parts which add up to the approximate MSEP of the original step.

Note that in the Bayesian chain ladder model in Merz and Wüthrich (2008) mentioned above (see also Salzmann and Wüthrich, 2010), the predictors of the ultimate loss form a martingale, and in this case the additivity of the mean squared prediction error of the claims development result follows directly from the properties of a martingale. In our case, by contrast, the predictors do not form a martingale in general, and the additivity property is a non-trivial — and highly desirable — result.

Remark 5.7. Replacing the f_j and ϕ_j in (6) with their estimators, we obtain estimators $\hat{\phi}_{j,J}$ for the $\phi_{j,J}$. They are linked to the quantities defined in this section:

$$\frac{\hat{C}\hat{U}_j^2}{\hat{r}_j\hat{s}_j} = \frac{\hat{C}\hat{u}_j^2\hat{r}_j^2}{\hat{r}_j^2/(\hat{r}_j - \hat{q}_j)} = \frac{\hat{C}\hat{\phi}_j(\hat{r}_j - \hat{q}_j)}{C_{\mathcal{H}_j,j}\hat{f}_j} = \frac{\hat{C}\hat{\phi}_j}{C_{\mathcal{H}_j,j}\hat{f}_j}\frac{C_{\mathcal{H}_j,j}}{\hat{C}/(\hat{f}_{j+1}\cdot\ldots\cdot\hat{f}_J)} = \hat{\phi}_{j,J}.$$

Going back to Result 4.5, where $\mathcal{H} = \mathcal{I}$, $\hat{r}_j = 1$ and hence $1 - \hat{q}_j = 1/\hat{s}_j$, we see that we may write Mack's formula as

$$\mathrm{msep}_{C-\hat{C}} \approx \sum_{j=1}^{J}\hat{\phi}_{j,J}\hat{s}_j\hat{C}_{I_j^*,J} = \sum_{j=1}^{J}\hat{\phi}_{j,J}\hat{C}_{I_j^*,J} + \sum_{j=1}^{J}\hat{\phi}_{j,J}(\hat{s}_j - 1)\hat{C}_{I_j^*,J}.$$

Note the analogy between the estimated process error (first sum in last expression) and the right-hand side of (7). Note also that the parameter error estimator (second sum) is formally equivalent to some "additional process error", as already observed in Remark 4.2. Generalizing to the development between arbitrary future horizons, (34) tells us

$$\mathrm{msep}_{(\breve{C}-\tilde{C})-0} \approx \sum_{j=1}^{J}\hat{\phi}_{j,J}\left(\tilde{s}_j - \breve{s}_j\right)\hat{C} = \sum_{j=1}^{J}\hat{\phi}_{j,J}\left(\tilde{s}_j\hat{C}_{\tilde{I}_j^*,J} - \breve{s}_j\hat{C}_{\breve{I}_j^*,J}\right),$$

because $\tilde{s}_j - \breve{s}_j = \tilde{s}_j - 1 - (\breve{s}_j - 1) = \tilde{s}_j\tilde{q}_j - \breve{s}_j\breve{q}_j$. If we were to retrieve the process error part from this by replacing $\tilde{s}_j$ and $\breve{s}_j$ by 1 in the last sum, inspired by Mack's case, we would merely pick up the direct effect that the process volatility has on the portfolios in $\tilde{\mathcal{I}}_j^* \setminus \breve{\mathcal{I}}_j^*$. But in our approach, the volatility of these portfolios also affects the remaining ones via an update on the development factors, and we consider this as part of the process error. Indeed, formula (36) shows that in our approach, the way to split off the process error is to multiply $\tilde{s}_j$ and $\breve{s}_j$ both by $1 - \hat{Q}_j$. In Mack's case, both methods coincide.

Remark 5.8. The quantities $\hat{\phi}_{j,J}$ have properties similar to a cash flow pattern and hence may be thought of as some kind of "risk flow pattern". Indeed, by

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

Proposition 2.4, the total process variance $V[X_J|\mathcal{X}_0]$ is a sum of pattern values $\phi_{j,J}$ all multiplied by the same ultimate value $E[X_J|\mathcal{X}_0]$. And if we move to a coarser granularity in the development direction, the new pattern is obtained by grouping and summing the values of the original pattern, again by Proposition 2.4.

Hence, the value $\phi_{j,J}$ describes, in an invariant way, the "amount of risk" that stems from period j in the development, and allows to make comparisons between different time periods, or between different lines of business. One should make sure that the development after step J is negligible, because $\phi_{k,j} = \phi_{k,j-1} f_j$ for $j > k$.

At the other end, at $j = 0$, the first element and its associated uncertainty is outside the scope of our analysis. This is usually called the "premium risk". We can include it in our analysis by shifting the loss development one step to the right and choosing the premium as the starting point $C_{i,0}$. That way, we obtain an integrated view of premium and reserving risk.

Remark 5.9. Since $\hat{q}_j \geq \tilde{q}_j \geq \check{q}_j$ by our assumptions on the horizons, and since $\hat{r}_j > \hat{q}_j$, it follows that $\hat{s}_j \geq \tilde{s}_j \geq \check{s}_j$ and $0 \leq \hat{Q}_j \leq 1$; we also see that if the development time between $\tilde{\mathcal{D}}$ and $\check{\mathcal{D}}$ is short, then $\hat{Q}_j$ is small, and in this case the process error term dominates the parameter error term. Hence, over short development horizons, the process error dominates.

Remark 5.10. In the case of development from today to the ultimate horizon, $\hat{U}_j^2 \hat{Q}_j = \hat{u}_j^2 \hat{q}_j \hat{r}_j = (\hat{\phi}_j \hat{q}_j / \hat{f}_j) \cdot (\hat{r}_j / C_{\mathcal{H}_j,j})$. Here we see that if we randomly delete some historical data, the term $\hat{r}_j / C_{\mathcal{H}_j,j}$ increases, while $\hat{\phi}_j \hat{q}_j / \hat{f}_j$ on average stays the same. This illustrates the effect that using less data increases the uncertainty of prediction.

6. Example

The example we present in Table 1 is the first example analyzed by Mack (1993) and is based on data taken from Taylor and Ashe (1983). Below the (cumulative) triangle, the table shows the estimators $\hat{f}_j$, $\hat{\phi}_j$, $\hat{u}_j$ and $\hat{q}_j$. The parameter ϕ_9 cannot be estimated, for lack of data in development period 9. We follow Mack (1993) approach in extrapolating $\hat{\phi}_9 := \min[\hat{\phi}_8^2/\hat{\phi}_7, \hat{\phi}_7, \hat{\phi}_8]$ from the preceding values (in our case $\hat{\phi}_9 = \hat{\phi}_7$). Note also that all calculations are made to machine precision from the original triangle data and then rounded to the precision shown.

As an example on how to interpret the coefficients $\hat{u}_j$ and $\hat{q}_j$, consider the values $\hat{f}_1 = 3.491$ and $\hat{u}_1 = 6.29\%$. From these we infer that

$$f_1 \approx 3.491 \pm 6.29\% = 3.491 \pm 0.219.$$

The uncertainty about the true f_1 only affects a rather small part of the ultimate loss predictor, namely just $\hat{q}_1 = 9.37\%$, because $\hat{f}_1$ is only used to estimate the ultimate loss of the least developed accident period $i = 9$.

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

TABLE 1

EXAMPLE

i	$C_{i,0}$	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$	$C_{i,4}$	$C_{i,5}$	$C_{i,6}$	$C_{i,7}$	$C_{i,8}$	$C_{i,9}$
0	357,848	1,124,788	1,735,330	2,218,270	2,745,596	3,319,994	3,466,336	3,606,286	3,833,515	3,901,463
1	352,118	1,236,139	2,170,033	3,353,322	3,799,067	4,120,063	4,647,867	4,914,039	5,339,085	
2	290,507	1,292,306	2,218,525	3,235,179	3,985,995	4,132,918	4,628,910	4,909,315		
3	310,608	1,418,858	2,195,047	3,757,447	4,029,929	4,381,982	4,588,268			
4	443,160	1,136,350	2,128,333	2,897,821	3,402,672	3,873,311				
5	396,132	1,333,217	2,180,715	2,985,752	3,691,712					
6	440,832	1,288,463	2,419,861	3,483,130						
7	359,480	1,421,128	2,864,498							
8	376,686	1,363,294								
9	344,014									
$\hat{f}_j$		3.491	1.747	1.457	1.174	1.104	1.086	1.054	1.077	1.018
$\hat{\phi}_j$		160,280	37,737	41,965	15,183	13,731	8,186	447	1,147	447
$\hat{\phi}_{j,9}$		190,039	51,154	46,796	17,907	15,603	8,701	464	1,085	439
$\hat{u}_j$		6.29%	3.47%	3.62%	2.44%	2.50%	2.09%	0.56%	1.08%	1.06%
$\hat{q}_j$		9.37%	20.01%	32.80%	43.47%	53.11%	62.27%	72.26%	82.4%	92.64%
Horizon		+1	+2	+3	+4	+5	+6	+7	+8	+9
Process		8.68%	9.80%	10.12%	10.15%	10.11%	10.08%	10.07%	10.06%	10.05%
Parameter		3.92%	5.86%	7.11%	7.78%	8.15%	8.32%	8.36%	8.39%	8.40%
Total		9.52%	11.42%	12.36%	12.79%	12.99%	13.07%	13.09%	13.10%	13.10%

The lower part of the table exhibits the square roots of the process, parameter and total mean squared prediction errors as a percentage of the total outstanding development (i.e. the difference between the ultimate values and the latest known values on today's diagonal — in our case 18, 680, 856). The values are calculated for the development horizons "+k", where k ranges from 1 to 9; by this we mean $\tilde{j}_i := \min[J, j_i + k]$, while $J = 9$ and $j_i = J - i$, as is evident from the triangle. The square root of the total mean squared prediction error for the development up to the ultimate horizon is 2, 447, 095, whereas for the development up to the 1 year horizon it is 1, 778, 968. These values amount to 13.10% and 9.52% of the total outstanding development, as shown in the table.

The process error shows an interesting behavior. It is dominant for short development horizons as explained in Remark 5.9. This reflects the fact that any realization $F_{i,j}$ with $i \in \mathcal{I}_j^* \setminus \tilde{\mathcal{I}}_j^*$ not only affects the ultimate of the loss portfolio i, but also all ultimates influenced by the future development factor $\tilde{f}_j$, and these are hit in a fully correlated manner. Another observation about the process error in our example is that it is not monotonous, but reaches a maximum at the development horizon "+4". The total MSEP, by contrast, is always increasing, a consequence of the additivity property derived in Remark 5.6.

Acknowledgements

I am grateful to M. Wüthrich and A. Gisler for alerting me to the recent publication Merz and Wüthrich (2014), to A. Gisler, T. Mack and R. Dahms for valuable discussions on the subject of this paper, and to the editor and the referees for many corrections and useful suggestions.

References

Buchwalder, M., Bühlmann, H., Merz, M. and Wüthrich, M. (2006) The mean square error of prediction in the chain ladder reserving method (Mack and Murphy Revisited), *ASTIN Bulletin,* **36**(2), 521–542.

Bühlmann, H., De Felice, M., Gisler, A., Moriconi, F. and Wüthrich, M.V. (2009) Recursive credibility formula for chain ladder factors and the claims development result. *ASTIN Bulletin,* **39**(1), 275–306.

Dahms, R. (2008) A loss reserving method for incomplete data. *Mitteilungen-Bulletin SAV,* **1–2**, 127–148.

Gisler, A. (2013) *"Die rasante Entwicklung der Mathematik in der Versicherung: eine Zeitreise über die letzten 30 Jahre*". Presentation given at the annual conference of the Swiss Actuarial Association in Winterthur, 7 September 2013.

Ku, H.H. (1966) Notes on the use of propagation of error formulas. *Journal of Research of the National Bureau of Standards - C. Engineering and Instrumentation*, **70C**(4), 263–273.

Mack, T. (1993) Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin,* **23**(2), 213–225.

Mack, T. (2008) The prediction error of Bornhuetter/Ferguson. *ASTIN Bulletin,* **38**(1), 87–103.

Matitschka, H. (2010) Prognosefehler im Overdispersed Poisson Modell für Abwicklungsdreiecke. *Blätter DGVFM,* **31**, 291–306.

MERZ, M. and WÜTHRICH, M.V. (2008). Modelling the claims development result for solvency purposes. *CAS E-Forum Fall 2008*, 542–568.
MERZ, M. AND WÜTHRICH, M.V. (2014) Claims run-off uncertainty: The full picture, *http://ssrn.com/abstract=2524352*
SALZMANN, R. and WÜTHRICH, M.V. (2010) Cost-of-capital margin for a general insurance liability runoff, *ASTIN Bulletin,* **40/2**, 415–451.
TAYLOR, G.C. and ASHE, F.R. (1983) Second moments of estimates of outstanding claims. *Journal of Econometrics,* **23**, 37–61.
WÜTHRICH, M.V. and MERZ, M. (2008) *Stochastic Claims Reserving Methods in Non-Life Insurance.* New York: Wiley.

ANCUS RÖHR
Helvetia Insurance Switzerland
St. Alban-Anlage 26
4002 Basel, Switzerland
E-mail: ancus.roehr@helvetia.ch

APPENDIX

A. COMPARISON TO PREVIOUSLY PUBLISHED FORMULAE

We will first show that the formula from Result 4.5 agrees with the original formula given by Mack (1993). He considered "standard triangles" fulfilling (in our notation) $\mathcal{I} = \{0, 1, \ldots, J\}$, $j_i = J - i$, and $\mathcal{I}_j = \{0, 1, \ldots, J - j\}$. He also assumed the historical triangle to be complete. We also restrict our attention to the case of all loss portfolios combined, as in 4.5, and skip the case of a single accident period.

Mack studied the case of development to the ultimate horizon. He uses a different convention of labeling link ratios and development factors. While our link ratio equation is $C_{i,j} f_{i,j+1} = C_{i,j+1}$, Mack uses the notation $C_{i,j} f_{i,j} = C_{i,j+1}$. Therefore, converting Mack's formulae to our notation, the first step is to replace Mack's $\hat{f}_j$ by our $\hat{f}_{j+1}$. The same applies to the ϕ_j: Mack's $\hat{\sigma}_j^2$ must be replaced by our $\hat{\phi}_{j+1}$. Keeping this in mind, Mack's original formula (Mack (1993), Corollary to Theorem 3) for the mean squared prediction error in our notation reads (note that $\mathcal{I}_{j_i}^* = \{i + 1, \ldots, J\}$)

$$M^2 := \sum_{i \in \mathcal{I}_J^*} \left(\hat{C}_{i,J}^2 \sum_{j=j_i+1}^{J} \frac{\hat{\phi}_j}{\hat{f}_j^2} \left(\frac{1}{\hat{C}_{i,j-1}} + \frac{1}{C_{\mathcal{I}_j, j-1}} \right) + \hat{C}_{i,J} \hat{C}_{\mathcal{I}_{j_i}^*, J} \sum_{j=j_i+1}^{J} \frac{2\hat{\phi}_j / \hat{f}_j^2}{C_{\mathcal{I}_j, j-1}} \right).$$

Proposition A.1. *This formula coincides with our formula in Result 4.5, i.e.* $M^2 = \hat{C}^2 \sum_{j=1}^{J} \hat{u}_j^2 \hat{q}_j$.

Proof. Using the definitions of $\hat{u}_{i,j}$ and $\hat{u}_j$, we get

$$M^2 = \sum_{i \in \mathcal{I}_J^*} \left(\left(\hat{C}_{i,J}^2 \sum_{j=j_i+1}^{J} \left(\hat{u}_{i,j}^2 + \hat{u}_j^2 \right) \right) + \hat{C}_{i,J} \hat{C}_{\mathcal{I}_{j_i}^*, J} \sum_{j=j_i+1}^{J} 2\, \hat{u}_j^2 \right).$$

Using (33) as well as historical completeness, which allows us to replace $\hat{u}_{i,j}^2$ with $\hat{u}_j^2(1-\hat{q}_j)\hat{C}/\hat{C}_{i,J}$, yields

$$M^2 = \sum_{i\in\mathcal{I}_J^*}\left(\left(\hat{C}_{i,J}^2\sum_{j=j_i+1}^{J}\hat{u}_j^2\left(\frac{(1-\hat{q}_j)\hat{C}}{\hat{C}_{i,J}}+1\right)\right)+\hat{C}_{i,J}\hat{C}_{\mathcal{I}_{j_i}^*,J}\sum_{j=j_i+1}^{J}2\,\hat{u}_j^2\right)$$
$$=\sum_{i\in\mathcal{I}_J^*}\sum_{j=j_i+1}^{J}\left(\hat{u}_j^2\hat{C}_{i,J}(1-\hat{q}_j)+\hat{u}_j^2\hat{C}_{i,J}^2+2\,\hat{u}_j^2\hat{C}_{i,J}\hat{C}_{\mathcal{I}_{j_i}^*,J}\right).$$

The double sum runs exactly over all future (i, j). Changing the summation order, we get

$$M^2 = \sum_{j=1}^{J}\hat{u}_j^2\sum_{i\in\mathcal{I}_j^*}\left(\hat{C}_{i,J}\left(1-\hat{q}_j\right)\hat{C}+\hat{C}_{i,J}^2+2\,\hat{C}_{i,J}\hat{C}_{\mathcal{I}_{j_i}^*,J}\right)$$
$$=\sum_{j=1}^{J}\hat{u}_j^2\left(\hat{C}_{\mathcal{I}_j^*,J}(1-\hat{q}_j)\hat{C}+\hat{C}_{\mathcal{I}_j^*,J}^2\right)$$
$$=\hat{C}^2\sum_{j=1}^{J}\hat{u}_j^2(\hat{q}_j(1-\hat{q}_j)+\hat{q}_j^2)=\hat{C}^2\sum_{j=J_{i_1}+1}^{j_1}\hat{u}_j^2\hat{q}_j,$$

■

Bühlmann *et al.* (2009) considered the case of a one-period development and derived various formulae for the prediction error of the claims development result. We will show that their "linear approximation" (Bühlmann *et al.*, 2009, Result 4.10, which goes back to Merz and Wüthrich, 2008), is a special case of our formula (36). The "standard trapezoids" they considered are given by portfolios $i = 0, 1, \ldots, J+n$ for some integer $n \geq 0$, with known development of portfolio i from development index 0 up to $j_i = \min[J, J+n-i]$. Their Result 4.10 shows that the mean squared prediction error of the claims development result of all accident periods over a single calendar period can be approximated by

$$m^2 := \sum_{i\in\mathcal{I}_J^*}\left(C_{i,j_i}\tilde{\Gamma}_{j_i}^* + C_{i,j_i}^2\tilde{\Delta}_{j_i}^* + 2\,C_{i,j_i}\hat{C}_{\mathcal{I}_{j_i}^*,j_i}\left(\tilde{\Phi}_{j_i}^*+\tilde{\Delta}_{j_i}^*\right)\right),$$

where for $j < J$

$$\tilde{\Phi}_j^* := \frac{\hat{\phi}_{j+1}}{C_{\mathcal{I}_j,j}}\frac{\hat{C}^2}{\hat{C}_{\mathcal{I},j+1}^2} = \hat{u}_{j+1}^2\frac{C_{\mathcal{I}_{j+1},j+1}}{C_{\mathcal{I}_j,j+1}}\frac{\hat{C}^2}{\hat{C}_{\mathcal{I},j}^2} = \hat{u}_{j+1}^2\frac{1-\hat{q}_{j+1}}{1-\hat{q}_j}\frac{\hat{C}^2}{\hat{C}_{\mathcal{I},j}^2},$$

$$\tilde{\Gamma}_j^* := \tilde{\Phi}_j^* C_{\mathcal{I}_j,j}\left(1+\frac{C_{\mathcal{I}_j\setminus\mathcal{I}_{j+1},j}}{C_{\mathcal{I}_j,j}}\right) = \tilde{\Phi}_j^* C_{\mathcal{I}_j,j}\left(1+\frac{\hat{q}_{j+1}-\hat{q}_j}{1-\hat{q}_j}\right),$$

$$\tilde{\Delta}_j^* := \frac{\hat{C}^2}{\hat{C}_{\mathcal{I},j}^2}\sum_{k=j}^{J-1}\frac{C_{\mathcal{I}_k\setminus\mathcal{I}_{k+1},k}}{C_{\mathcal{I}_k,k}}\frac{\hat{\phi}_{k+1}/\hat{f}_{k+1}^2}{C_{\mathcal{I}_{k+1},k}} = \frac{\hat{C}^2}{\hat{C}_{\mathcal{I},j}^2}\sum_{k=j}^{J-1}\frac{\hat{q}_{k+1}-\hat{q}_k}{1-\hat{q}_k}\hat{u}_{k+1}^2.$$

To justify the transformations applied here to the original definitions, a few comments are in order. First note that, by the assumptions on the shape of the trapezoid, $\mathcal{I}_{j_i}\setminus\mathcal{I}_{j_i+1}=\{i\}$ for

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press

each $j_i < J$. Next, Bühlmann *et al.* use products like $\prod_{k=j+1}^{J} \hat{f}_k$, which we write as $\hat{C}/\hat{C}_{\mathcal{I},j}$, relying on Lemma 4.4 (which requires completeness). This lemma also links quotients of partial sums at the same development index to the $\hat{q}_j$, e.g. $\hat{C}_{\mathcal{I}_j,k}/\hat{C}_{\mathcal{I},k} = 1 - \hat{q}_j$ for any $j \leq k < J$, which we have used repeatedly above.

For the development between today and the one period horizon that we are considering here, we have $\tilde{q}_j = \hat{q}_j$ and $\check{q}_j = \hat{q}_{j-1}$. We may now prove:

Proposition A.2. *Result 4.10 in Bühlmann et al. (2009) is a special case of* (36)*:*

$$\frac{m^2}{\hat{C}^2} = \sum_{j=1}^{J} \hat{u}_j^2 \frac{\hat{q}_j - \hat{q}_{j-1}}{1 - \hat{q}_{j-1}} = \sum_{j=1}^{J} \hat{u}_j^2 \left(\frac{1}{1-\hat{q}_j} - \frac{1}{1-\hat{q}_{j-1}} \right) \Big/ \frac{1}{1-\hat{q}_j}.$$

Proof. Since $\mathcal{I}_{j_i} \setminus \mathcal{I}_{j_i+1} = \{i\}$, we have $C_{i,j_i}^2 + 2C_{i,j_i}\hat{C}_{\mathcal{I}_{j_i}^*,j_i} = \hat{C}^2_{\mathcal{I}^*_{j_i+1},j_i} - \hat{C}^2_{\mathcal{I}^*_{j_i},j_i}$, and we may write m^2 as a sum over j rather than i:

$$m^2 = \sum_{j=0}^{J-1} C_{\mathcal{I}_j \setminus \mathcal{I}_{j+1},j} (\tilde{\Gamma}_j^* + 2\,\hat{C}_{\mathcal{I}_j^*,j} \tilde{\Phi}_j^*) + (\hat{C}^2_{\mathcal{I}^*_{j+1},j} - \hat{C}^2_{\mathcal{I}^*_j,j}) \tilde{\Delta}_j^*.$$

Plugging in the above expressions for $\tilde{\Gamma}_j^*$, $\tilde{\Phi}_j^*$ and $\tilde{\Delta}_j^*$ and using also $C_{\mathcal{I}_j \setminus \mathcal{I}_{j+1},j} = (\hat{q}_{j+1} - \hat{q}_j)\,\hat{C}_{\mathcal{I},j}$, we get on the one hand

$$\begin{aligned}
&\sum_{j=0}^{J-1} C_{\mathcal{I}_j \setminus \mathcal{I}_{j+1},j} (\tilde{\Gamma}_j^* + 2\,\hat{C}_{\mathcal{I}_j^*,j} \tilde{\Phi}_j^*) \\
&\quad = \sum_{j=0}^{J-1} \frac{C_{\mathcal{I}_j \setminus \mathcal{I}_{j+1},j}}{\hat{C}_{\mathcal{I},j}} \left(\frac{C_{\mathcal{I}_j,j}}{\hat{C}_{\mathcal{I},j}} \left(1 + \frac{\hat{q}_{j+1} - \hat{q}_j}{1 - \hat{q}_j} \right) + 2\,\frac{C_{\mathcal{I}_j,j}}{\hat{C}_{\mathcal{I},j}} \right) \tilde{\Phi}_j^* \hat{C}^2_{\mathcal{I},j} \\
&\quad = \sum_{j=0}^{J-1} (\hat{q}_{j+1} - \hat{q}_j) \left((1 - \hat{q}_j) \left(1 + \frac{\hat{q}_{j+1} - \hat{q}_j}{1 - \hat{q}_j} \right) + 2\,\hat{q}_j \right) \hat{u}_{j+1}^2 \frac{1 - \hat{q}_{j+1}}{1 - \hat{q}_j} \hat{C}^2 \\
&\quad = \hat{C}^2 \sum_{j=0}^{J-1} (1 - \hat{q}_{j+1}^2)\, \hat{u}_{j+1}^2 \frac{\hat{q}_{j+1} - \hat{q}_j}{1 - \hat{q}_j},
\end{aligned}$$

while on the other hand

$$\begin{aligned}
\sum_{j=0}^{J-1} (\hat{C}^2_{\mathcal{I}^*_{j+1},j} - \hat{C}^2_{\mathcal{I}^*_j,j}) \tilde{\Delta}_j^* &= \hat{C}^2 \sum_{j=0}^{J-1} (\hat{q}_{j+1}^2 - \hat{q}_j^2) \sum_{k=j}^{J-1} \hat{u}_{k+1}^2 \frac{\hat{q}_{k+1} - \hat{q}_k}{1 - \hat{q}_k} \\
&= \hat{C}^2 \sum_{j=0}^{J-1} \hat{q}_{j+1}^2 \hat{u}_{j+1}^2 \frac{\hat{q}_{j+1} - \hat{q}_j}{1 - \hat{q}_j},
\end{aligned}$$

as the double sum has many canceling terms (note that $\hat{q}_0 = 0$). Putting this together proves the first stated equality, while the remaining one follows by simple algebraic transformations. ■

https://doi.org/10.1017/asb.2016.9 Published online by Cambridge University Press