THE MEAN SQUARE ERROR OF PREDICTION IN THE CHAIN LADDER RESERVING METHOD – A COMMENT

BY

THOMAS MACK, GERHARD QUARG AND CHRISTIAN BRAUN

ABSTRACT

We discuss some questionable points of the approach taken in the paper by Buchwalder, Bühlmann, Merz and Wüthrich and come to the conclusion that this approach does not yield an improvement of Mack's original formula. The main reason is that the new approach disregards the negative correlation of the squares of the development factors. The same applies to the formula by Murphy (PCAS 1994).

1. INTRODUCTION

In their paper, Buchwalder, Bühlmann, Merz and Wüthrich (= BBMW) use a time series model to estimate the mean square error of prediction (= msep) for the chain ladder method. The msep consists of two components, the process variance and the estimation error. Whereas the formula for the process variance is identical to the one derived by Mack in [2], the estimation error according to the new formula is always greater than Mack's result. Therefore, we consider in this comment the time series model and the approach to the estimation error only. We use the same notation as BBMW and assume that the reader knows the time series approach by BBMW (section 2.3) and especially the approach to the estimation error (section 4.1.2). Any numbers of sections, equations or references relate to the original paper by BBMW.

2. THE TIME SERIES MODEL

A. The time series model is only a special case of Mack's model:

The reason is that BBMW assume the residuals $\varepsilon_{k,j}$ of the time series model to be independent within the same accident year whereas Mack does not. This is mentioned by BBMW at the end of section 2.3: "(T1) and (T2) imply Mack's conditions". Although this is not a fundamental difference, we mention it here in order to have a full comparison between Mack's original approach and the

Astin Bulletin 36(2), 543-552. doi: 10.2143/AST.36.2.2017937

BBMW approach. Moreover, this difference may have some practical relevance as an independence assumption is always a crucial point in claims reserving.

B. The time series model does not fit the chain ladder claims process:

This is touched by BBMW in the third remark of section 2.3: "Theoretically, our process could have negative values for the cumulative payments $X_{k,j}$. To avoid this problem, we could reformulate the definition of $\varepsilon_{k,j}$ such that its distribution is conditionally, given $X_{k,j-1}$, centered with variance 1 and such that $X_{k,j}$ is positive." But this reformulation proposed by BBMW is not a solution because then the residuals $\varepsilon_{k,j}$ are not independent any more: The distribution of $\varepsilon_{k,j}$ depends upon $X_{k,j-1}$ which itself depends upon $\varepsilon_{k,j-1}$. This point is not negligeable because later on, this independence is used.

A necessary condition for positivity would be the restriction that the distributions of the $\varepsilon_{k,j}$ are limited from below at negative amounts $u_{k,j}$ in such a way that $f_{j-1} X_{k,j-1} + \sigma_{j-1} \sqrt{X_{k,j-1}}\, u_{k,j} > 0$ holds for all k and j. But it seems almost impossible to define $u_{k,j}$ independently from $X_{k,j-1}$ such that the condition above holds. Probably, the distribution of $\varepsilon_{k,j-1}$ has to be limited from above, too, in order to secure that the factor $\sqrt{X_{k,j-1}}$ does not become too large in case of $\varepsilon_{k,j} < 0$. The crucial point is the dependency of the standard deviation upon the previous cumulative claims amount which is not known in advance. As a consequence, it is misleading that BBMW write at the beginning of section 2.3: "This route seems very promising as it … also defines how to simulate … triangles." In reality, using a reasonable distribution for $\varepsilon_{k,j}$ it seems almost impossible to guarantee positivity, i.e. the simulations may yield some negative amounts $X_{k,j}$. Even if it was possible to formulate the conditions for positivity in an appropriate way, it will probably be difficult to check in practice whether these conditions are fulfilled.

Conclusion. The independence assumption which is the only additional feature of the time series model as compared to Mack's model causes difficulties which indicate that this model does not fit the chain ladder claims process perfectly. At least, it will be very difficult to check if the assumptions are fulfilled in practice.

3. The three approaches to the estimation error

In section 4.1.2, BBMW give three approaches to the estimation error of which the third one is their own and the second one is said to be "similar to the one used in Mack [2]". This last statement is a misinterpretation of Mack's approach. Therefore, we give another description of the three approaches.

The basic formula for the estimation error is

$$\left(\widehat{X}_{k,j} - E\left(X_{k,j} \,\middle|\, \mathcal{D}_K\right)\right)^2 = X_{k,K-k+1}^2 \cdot \left(\hat{f}_{K-k+1} \cdot \ldots \cdot \hat{f}_{j-1} - f_{K-k+1} \cdot \ldots \cdot f_{j-1}\right)^2. \quad (4.11)$$

As this formula contains unknown parameters, it has to be approximated. The three approaches are based on (4.11) and two different algebraic reformulations of (4.11). BBMW give the following reformulation directly in (4.11) which we call (4.11a):

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

Mack [2] used the following more complicated but still exact reformulation

$$\begin{aligned}\left(\widehat{X}_{k,j} - E\left(X_{k,j}\,\middle|\,\mathcal{D}_K\right)\right)^2 &= X_{k,K-k+1}^2 \cdot \left(\hat{f}_{K-k+1} \cdot \ldots \cdot \hat{f}_{j-1} - f_{K-k+1} \cdot \ldots \cdot f_{j-1}\right)^2 \\ &= X_{k,K-k+1}^2 \cdot \left(\sum_{l=K-k+1}^{j-1} T_l\right)^2 = X_{k,K-k+1}^2 \cdot \left(\sum_{l=K-k+1}^{j-1} T_l^2 + 2\sum_{m<l} T_m T_l\right)\end{aligned} \tag{4.11b}$$

with $T_l = \hat{f}_{K-k+1} \cdot \ldots \cdot \hat{f}_{l-1}\left(f_l - \hat{f}_l\right) f_{l+1} \cdot \ldots \cdot f_{j-1}$.

In BBMW's approach 1 one would like to approximate (4.11) by

$$\begin{aligned}&X_{k,K-k+1}^2 \cdot E\left(\left(\hat{f}_{K-k+1} \cdot \ldots \cdot \hat{f}_{j-1} - f_{K-k+1} \cdot \ldots \cdot f_{j-1}\right)^2 \,\middle|\, \mathcal{B}_{K-k+1}\right) = \\ &\quad = X_{k,K-k+1}^2 \cdot \operatorname{Var}\left(\hat{f}_{K-k+1} \cdot \ldots \cdot \hat{f}_{j-1} \,\middle|\, \mathcal{B}_{K-k+1}\right) \\ &\quad = X_{k,K-k+1}^2 \cdot \left(E\left(\hat{f}_{K-k+1}^2 \cdot \ldots \cdot \hat{f}_{j-1}^2 \,\middle|\, \mathcal{B}_{K-k+1}\right) - f_{K-k+1}^2 \cdot \ldots \cdot f_{j-1}^2\right).\end{aligned}$$

But this cannot be calculated further because the squares $\hat{f}_l^2$ are (conditionally) negatively correlated with an unknown covariance matrix, see the theorem below. (Note that the $\hat{f}_l$ – not being squared – are uncorrelated, see Theorem 2 in Mack [2].)

Therefore Mack (see [2], page 219) used reformulation (4.11b) and approximated it as follows:

$$\begin{aligned}&X_{k,K-k+1}^2 \cdot \left(\sum_{l=K-k+1}^{j-1} T_l^2 + 2\sum_{m<l} T_m T_l\right) \\ &\quad \approx X_{k,K-k+1}^2 \cdot \left(\sum_{l=K-k+1}^{j-1} E\left(T_l^2 \,\middle|\, \mathcal{B}_l\right) + 2\sum_{m<l} E\left(T_m T_l \,\middle|\, \mathcal{B}_l\right)\right) \\ &\quad = X_{k,K-k+1}^2 \cdot \left(\sum_{l=K-k+1}^{j-1} E\left(T_l^2 \,\middle|\, \mathcal{B}_l\right)\right).\end{aligned}$$

This is very different from BBMW's approach 2 (called "similar to the one used in Mack [2]") because it does not average over one single σ-field $\mathcal{B}_l$ as BBMW write but over all relevant σ-fields $\mathcal{B}_{K-k+1}$, …, $\mathcal{B}_{j-1}$ as is the case in BBMW's own approach 3.

Regarding their approach 3, BBMW do not fully state how they arrived from (4.11) via (4.17) to (4.18). But apparently they started with reformulation (4.11a) and approximated it as follows:

$$X_{k,K-k+1}^2 \cdot \left(\prod_{l=K-k+1}^{j-1} \hat{f}_l^2 + \prod_{l=K-k+1}^{j-1} \hat{f}_l^2 - 2 \cdot \prod_{l=K-k+1}^{j-1} \hat{f}_l f_l \right) \approx \qquad (*)$$

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

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

Here, BBMW could have directly inserted

$$E\left(\hat{f}_l^2 \middle| \mathcal{B}_l\right) = f_l^2 + \sigma_l^2 / S_l$$

and then were arrived at their result (4.20) without using the time series model. The problem here is that the $\hat{f}_l^2$ are negatively correlated (see below) and therefore $\prod_l E\left(\hat{f}_l^2 \middle| \mathcal{B}_l\right)$ overestimates $\prod_l \hat{f}_l^2$ on average. Therefore, BBMW chose a more complicated approach with the intention to reach a better approximation. But this approach is not less questionable (see next section) and does give the same result (4.20).

4. The BBMW approach with resampling

In order to make the approximate equation (*) being exact, it is necessary and sufficient that

$$E\left(\prod_{l=K-k+1}^{j-1} \hat{f}_l^2 \middle| C \right) = \prod_{l=K-k+1}^{j-1} E\left(\hat{f}_l^2 \middle| C\right)$$

holds for an appropriate condition C. Ideally, one would like to have $C = \mathcal{D}_K$ because one is interested only in the variability given the data already known. But then

$$E\left(\hat{f}_l^2 \middle| \mathcal{D}_K\right) = \hat{f}_l^2$$

and one is back at formula (4.11) without having gained anything. Therefore, BBMW generate new development factors $\hat{f}_l^{\mathcal{D}_K}$, see equation (4.19), and then approximate (4.11) by

$$E\left(X_{k,K-k+1}^2 \cdot \left(\hat{f}_{K-k+1}^{\mathcal{D}_K} \cdot \ldots \cdot \hat{f}_{j-1}^{\mathcal{D}_K} - f_{K-k+1} \cdot \ldots \cdot f_{j-1} \right)^2 \middle| \mathcal{D}_K \right) \qquad (4.20)$$

– which can be calculated exactly – suggesting that this was a better approximation than Mack's approach. Later on, BBMW show that their result is always larger than Mack's.

The squares $\left(\hat{f}_l^{\mathcal{D}_K}\right)^2$ of the new factors are independent by their construction but they are not development factors of any run-off triangle because (see (4.19)) they use the "old" claims amounts $X_{k,j-1}$ as starting point in their denominator and not the claims amounts $Z_{k,j-1}$ generated in the previous resampling step. Thus, the newly generated set $\left\{ \hat{f}_1^{\mathcal{D}_K}, \hat{f}_2^{\mathcal{D}_K}, \ldots \right\}$ can never be the set of chain ladder development factors of a single run-off triangle. Moreover, these new factors disregard by their construction the fact that the squares $\hat{f}_l^2$ of "regular" development factors are conditionally negatively correlated, see the theorem below. This latter fact is relevant because the further calculations in (4.20) use that the $\left(\hat{f}_l^{\mathcal{D}_K}\right)^2$ are conditionally not correlated.

As shown at the end of section 3 above, the BBMW result could also have been obtained by approximating $\prod_l \hat{f}_l^2$ by $\prod_l E\left(\hat{f}_l^2 \middle| \mathcal{B}_l\right)$ which essentially assumes the $\hat{f}_l^2$ being uncorrelated. As they are in fact negatively correlated, the BBMW formula leads to an overestimation of the estimation error (4.11). This explains why the BBMW result is always larger than Mack's.

As communicated by Gary Venter, the BBMW formula for the estimation error gives numerically identical results to the corresponding formula by Murphy [4], see the comment by BBMW at the end of section 4.2. Therefore, also Murphy's formula overestimates the estimation error. Again, the reason is that Murphy explicitely assumes the independence of the various columnwise estimates.

Conclusion. As the newly generated development factors $\hat{f}_l^{\mathcal{D}_K}$ have relevant properties which regular development factors $\hat{f}_l$ don't have, it is very unclear whether formula (4.20) has any relevance for the estimation error in the original triangle. The resulting formula rather overstates the estimation error.

5. Theorem on negative correlation in the chain ladder model

Intuitively, the negative correlation of the $\hat{f}_l^2$ is obvious for the following reason: After an above-average $\hat{f}_{l-1}^2$ and thus also $\hat{f}_{l-1}$, the resulting $X_{k,l}$ are relatively large which due to the variance assumption (M4') (see BBMW's Remarks 2.1) entails a below-average variance of the next factor $\hat{f}_l$, i.e. a below-average second moment (as the first moment remains unchanged). In short: If $\hat{f}_{l-1}^2$ is above average, $\hat{f}_l^2$ is rather below average.

Theorem. In the chain ladder model (M1)-(M4), the squares $\hat{f}_{I-1}^2, \hat{f}_I^2$ of two successive development factors are negatively correlated, i.e. we have $\mathrm{Cov}(\hat{f}_{I-1}^2, \hat{f}_I^2 \,|\, \mathcal{B}_{I-1}) < 0$. This is not fulfilled for the newly resampled estimates $\hat{f}_j^{\mathcal{D}_K}$ for which we have $\mathrm{Cov}\big((\hat{f}_{I-1}^{D_K})^2, (\hat{f}_I^{D_K})^2 \,|\, \mathcal{B}_{I-1}\big) = 0$.

Proof: To fix ideas, we first prove the proposition for the individual development factors $X_{k,I}/X_{k,I-1}$, $X_{k,I+1}/X_{k,I}$ of any accident year k.

$$
\begin{aligned}
\mathrm{Cov}\left(\left(\frac{X_{k,I}}{X_{k,I-1}}\right)^2, \left(\frac{X_{k,I+1}}{X_{k,I}}\right)^2 \middle| \mathcal{B}_{I-1}\right) &= \mathrm{Cov}\left(\left(\frac{X_{k,I}}{X_{k,I-1}}\right)^2, E\left(\left(\frac{X_{k,I+1}}{X_{k,I}}\right)^2 \middle| \mathcal{B}_I\right) \middle| \mathcal{B}_{I-1}\right) \\
&= \frac{1}{X_{k,I-1}^2} \mathrm{Cov}\left(X_{k,I}^2, \mathrm{Var}\left(\frac{X_{k,I+1}}{X_{k,I}} \middle| \mathcal{B}_I\right) + \left(E\left(\frac{X_{k,I+1}}{X_{k,I}} \middle| \mathcal{B}_I\right)\right)^2 \middle| \mathcal{B}_{I-1}\right) \\
&= \frac{1}{X_{k,I-1}^2} \mathrm{Cov}\left(X_{k,I}^2, \frac{\sigma_I^2}{X_{k,I}} + f_I^2 \middle| \mathcal{B}_{I-1}\right) = \frac{\sigma_I^2}{X_{k,I-1}^2} \mathrm{Cov}\left(X_{k,I}^2, \frac{1}{X_{k,I}} \middle| \mathcal{B}_{I-1}\right) \\
&= \frac{\sigma_I^2}{X_{k,I-1}^2} \left(E\left(X_{k,I} \middle| \mathcal{B}_{I-1}\right) - E\left(X_{k,I}^2 \middle| \mathcal{B}_{I-1}\right) \cdot E\left(X_{k,I}^{-1} \middle| \mathcal{B}_{I-1}\right)\right) \\
&\le \frac{\sigma_I^2}{X_{k,I-1}^2} \left(E\left(X_{k,I} \middle| \mathcal{B}_{I-1}\right) - E\left(X_{k,I}^2 \middle| \mathcal{B}_{I-1}\right) \Big/ E\left(X_{k,I} \middle| \mathcal{B}_{I-1}\right)\right) \\
&= -\frac{\sigma_I^2}{X_{k,I-1}^2} \cdot \frac{\mathrm{Var}\left(X_{k,I} \middle| \mathcal{B}_{I-1}\right)}{E\left(X_{k,I} \middle| \mathcal{B}_{I-1}\right)} = -\frac{\sigma_I^2}{X_{k,I-1}^2} \cdot \frac{\sigma_{I-1}^2}{f_{I-1}} \\
&< 0
\end{aligned}
$$

where we have used Jensen's inequality for the "≤".

Now, we show the theorem for the (original) chain ladder development factors

$$
\hat{f}_I = \frac{\sum_{k=1}^{K-I} X_{k,I+1}}{\sum_{k=1}^{K-I} X_{k,I}} \quad \text{and} \quad \hat{f}_{I-1} = \frac{\sum_{k=1}^{K-I+1} X_{k,I}}{\sum_{k=1}^{K-I+1} X_{k,I-1}}.
$$

We shall use the following facts:

$$
E(\hat{f}_I \,|\, \mathcal{B}_I) = f_I \text{ and } E(\hat{f}_I^2 \,|\, \mathcal{B}_I) = f_I^2 + \mathrm{Var}(\hat{f}_I \,|\, \mathcal{B}_I) = f_I^2 + \sigma_I^2/S_I \text{ with } S_I := \sum_{k=1}^{K-I} X_{k,I}.
$$

With the latter notation we can write $\hat{f}_{I-1} = (S_I + X_{K-I+1,I})/S_{I-1}$.

Note further that, given $\mathcal{B}_l$, $\hat{f}_{l-1}$ is constant. Thus we have

$$\begin{aligned}
\mathrm{Cov}\Big(\hat{f}_{l-1}^2, \hat{f}_l^2 \Big| \mathcal{B}_{l-1}\Big) &= \mathrm{Cov}\Big(\hat{f}_{l-1}^2, E\Big(\hat{f}_l^2 \Big| \mathcal{B}_l\Big) \Big| \mathcal{B}_{l-1}\Big) \\
&= \mathrm{Cov}\Big(\hat{f}_{l-1}^2, \hat{f}_l^2 + \sigma_l^2 / S_l \Big| \mathcal{B}_{l-1}\Big) \\
&= \frac{\sigma_l^2}{S_{l-1}^2} \cdot \mathrm{Cov}\Big(\big(S_l + X_{K-l+1,l}\big)^2, S_l^{-1} \Big| \mathcal{B}_{l-1}\Big) \\
&= \frac{\sigma_l^2}{S_{l-1}^2}\Big(\mathrm{Cov}\Big(S_l^2, S_l^{-1} \Big| \mathcal{B}_{l-1}\Big) + 2\mathrm{Cov}\Big(X_{K-l+1,l} S_l, S_l^{-1} \Big| \mathcal{B}_{l-1}\Big) \\
&\quad + \mathrm{Cov}\Big(X_{K-l+1,l}^2, S_l^{-1} \Big| \mathcal{B}_{l-1}\Big)\Big)
\end{aligned}$$

Thus, we have only to show that all these covariances are ≤ 0 and one < 0. We first show

$$\begin{aligned}
&\mathrm{Cov}\big(X_{K-l+1,l}, S_l \big| \mathcal{B}_{l-1}\big) = \\
&E\big(X_{K-l+1,l} \cdot S_l \big| \mathcal{B}_{l-1}\big) - E\big(X_{K-l+1,l} \big| \mathcal{B}_{l-1}\big) \cdot E\big(S_l \big| \mathcal{B}_{l-1}\big) = 0.
\end{aligned}$$

This follows from the independence of the accident years (see model assumption (M1)), more precisely from the facts that $X_{K-l+1,l}$ is independent from S_l and that $\mathcal{B}_{l-1}$ can be split into two independent sub-algebras: $\mathcal{B}_{l-1} = \sigma(\mathcal{B}_{l-1}^-, \mathcal{B}_{l-1}^+)$ with

$$\mathcal{B}_{l-1}^- := \sigma\big(\{X_{k,j} \in \mathcal{D}_K \big| k \le K-l,\ j \le l-1\}\big)$$

and

$$\mathcal{B}_{l-1}^+ := \sigma\big(\{X_{k,j} \in \mathcal{D}_K \big| k > K-l,\ j \le l-1\}\big)$$

This yields

$$\begin{aligned}
E\big(X_{K-l+1,l} \cdot S_l \big| \mathcal{B}_{l-1}\big) &= E\Big(X_{K-l+1,l} \cdot E\big(S_l \big| \mathcal{B}_{l-1}, X_{K-l+1,l}\big) \Big| \mathcal{B}_{l-1}\Big) \\
&= E\Big(X_{K-l+1,l} \cdot E\Big(S_l \Big| \mathcal{B}_{l-1}^-\Big) \Big| \mathcal{B}_{l-1}\Big) \\
&= E\big(X_{K-l+1,l} \big| \mathcal{B}_{l-1}\big) \cdot E\big(S_l \big| \mathcal{B}_{l-1}\big).
\end{aligned}$$

Analogously one proves $\mathrm{Cov}\Big(X_{K-l+1,l}^2, S_l^{-1} \Big| \mathcal{B}_{l-1}\Big) = 0$.

From $\mathrm{Cov}\big(X_{K-l+1,l}, S_l \big| \mathcal{B}_{l-1}\big) = 0$ we infer,

$$\begin{aligned}
&\mathrm{Cov}\Big(X_{K-l+1,l} S_l, S_l^{-1} \Big| \mathcal{B}_{l-1}\Big) = \\
&E\big(X_{K-l+1,l} \big| \mathcal{B}_{l-1}\big) - E\big(X_{K-l+1,l} \big| \mathcal{B}_{l-1}\big) E\big(S_l \big| \mathcal{B}_{l-1}\big) E\Big(S_l^{-1} \Big| \mathcal{B}_{l-1}\Big) \le 0,
\end{aligned}$$

according to Jensen's inequality which we also use in the final step

$$\begin{aligned}&\mathrm{Cov}\left(S_l^2, S_l^{-1}\middle|\mathcal{B}_{l-1}\right) = E\left(S_l\middle|\mathcal{B}_{l-1}\right) - E\left(S_l^2\middle|\mathcal{B}_{l-1}\right)E\left(S_l^{-1}\middle|\mathcal{B}_{l-1}\right)\\&\leq -\mathrm{Var}\left(S_l\middle|\mathcal{B}_{l-1}\right)/E\left(S_l\middle|\mathcal{B}_{l-1}\right) < 0.\end{aligned}$$

Altogether, we have shown the main statement $\mathrm{Cov}\left(\hat{f}_{l-1}^2, \hat{f}_l^2\middle|\mathcal{B}_{l-1}\right) < 0$. More precisely, we have shown

$$\mathrm{Cov}\left(\hat{f}_{l-1}^2, \hat{f}_l^2\middle|\mathcal{B}_{l-1}\right) \leq -\frac{\sigma_l^2}{S_{l-1}^2}\cdot\frac{\mathrm{Var}\left(S_l\middle|\mathcal{B}_l\right)}{E\left(S_l\middle|\mathcal{B}_l\right)} = -\frac{\sigma_{l-1}^2\sigma_l^2}{S_{l-1}^2 f_{l-1}}.$$

This completes the proof as the other assertion is obvious:

$$\begin{aligned}&\mathrm{Cov}\left(\left(\hat{f}_{l-1}^{\mathcal{D}_K}\right)^2, \left(\hat{f}_l^{\mathcal{D}_K}\right)^2\middle|\mathcal{B}_{l-1}\right) =\\&\quad= E\left(\mathrm{Cov}\left(\left(\hat{f}_{l-1}^{\mathcal{D}_K}\right)^2, \left(\hat{f}_l^{\mathcal{D}_K}\right)^2\middle|\mathcal{D}_K\right)\middle|\mathcal{B}_{l-1}\right)\\&\qquad+\mathrm{Cov}\left(E\left(\left(\hat{f}_{l-1}^{\mathcal{D}_K}\right)^2\middle|\mathcal{D}_K\right), E\left(\left(\hat{f}_l^{\mathcal{D}_K}\right)^2\middle|\mathcal{D}_K\right)\middle|\mathcal{B}_{l-1}\right)\\&\quad= 0\end{aligned}$$

because from statements 1) and 3) shortly before (4.20) in the BBMW paper we see that $\mathrm{Cov}\left(\left(\hat{f}_{l-1}^{\mathcal{D}_K}\right)^2, \left(\hat{f}_l^{\mathcal{D}_K}\right)^2\middle|\mathcal{D}_K\right) = 0$ and $E\left(\left(\hat{f}_{l-1}^{\mathcal{D}_K}\right)^2\middle|\mathcal{D}_K\right) = f_{l-1}^2 + \frac{\sigma_{l-1}^2}{S_{l-1}}$ is constant given $\mathcal{B}_{l-1}$. □

6. Deeper insight in the case of two development factors

In the simplest case with only two factors $\hat{f}_1, \hat{f}_2$ we can calculate some relevant quantities exactly. E.g. we have the following formula for the average estimation error

$$\begin{aligned}&E\left(\left(\hat{f}_1\hat{f}_2 - f_1 f_2\right)^2\middle|\mathcal{B}_1\right) = \mathrm{Var}\left(\hat{f}_1\hat{f}_2\middle|\mathcal{B}_1\right) =\\&\quad= f_1^2\mathrm{Var}\left(\hat{f}_2\middle|\mathcal{B}_1\right) + \mathrm{Var}\left(\hat{f}_1\middle|\mathcal{B}_1\right)f_2^2\\&\qquad+\mathrm{Var}\left(\hat{f}_1\middle|\mathcal{B}_1\right)\mathrm{Var}\left(\hat{f}_2\middle|\mathcal{B}_1\right) + \mathrm{Cov}\left(\hat{f}_1^2, \hat{f}_2^2\middle|\mathcal{B}_1\right).\end{aligned}$$

This follows directly from the general formula

$$\begin{aligned}\operatorname{Var}(XY) &= E(X^2Y^2) - (E(XY))^2 \\ &= E(X^2)\,E(Y^2) + \operatorname{Cov}(X^2,Y^2) - (E(X)\,E(Y) + \operatorname{Cov}(X,Y))^2 \\ &= (E(X))^2 \operatorname{Var}(Y) + \operatorname{Var}(X)\,(E(Y))^2 + \operatorname{Var}(X)\operatorname{Var}(Y) + \operatorname{Cov}(X^2,Y^2) \\ &\quad - 2\,E(X)\operatorname{Cov}(X,Y)\,E(Y) - (\operatorname{Cov}(X,Y))^2\end{aligned}$$

and from $E(\hat{f}_1 \,|\, \mathcal{B}_1) = f_l$ as well as from $\operatorname{Cov}(\hat{f}_1 \hat{f}_2 \,|\, \mathcal{B}_1) = 0$ which latter two facts are consequences of theorem 2 of Mack [2].

On the other hand, the formulae of Mack ([2], p. 219) and of BBMW (cf. (4.20)) for the conditional estimation error of $\hat{f}_1 \hat{f}_2$ are

$$\begin{aligned}CEE_{\text{Mack}} &= f_1^2 \operatorname{Var}\left(\hat{f}_2 \,\middle|\, \mathcal{B}_2\right) + \operatorname{Var}\left(\hat{f}_1 \,\middle|\, \mathcal{B}_1\right) f_2^2, \\ CEE_{\text{BBMW}} &= CEE_{\text{Mack}} + \operatorname{Var}\left(\hat{f}_1 \,\middle|\, \mathcal{B}_1\right) \operatorname{Var}\left(\hat{f}_2 \,\middle|\, \mathcal{B}_2\right).\end{aligned}$$

Whereas the average estimation error $\operatorname{Var}(\hat{f}_1 \hat{f}_2 \,|\, \mathcal{B}_1)$ depends upon $\mathcal{B}_1$, the *CEE*-formulae depend upon $\mathcal{B}_2$, too, and, using $E(\operatorname{Var}(\hat{f}_2 \,|\, \mathcal{B}_2) \,|\, \mathcal{B}_1) = \operatorname{Var}(\hat{f}_2 \,|\, \mathcal{B}_1)$, we have as average

$$E\left(CEE_{\text{BBMW}} \,|\, \mathcal{B}_1\right) = \operatorname{Var}\left(\hat{f}_1\, \hat{f}_2 \,\middle|\, \mathcal{B}_1\right) - \operatorname{Cov}\left(\hat{f}_1^2, \hat{f}_2^2 \,\middle|\, \mathcal{B}_1\right).$$

As $\operatorname{Cov}(\hat{f}_1^2, \hat{f}_2^2 \,|\, \mathcal{B}_1) < 0$ (see the theorem above), this shows again that CEE_{BBMW} on average overestimates the estimation error.

On the other hand, we have

$$\begin{aligned}E\left(CEE_{\text{Mack}} \,|\, \mathcal{B}_1\right) &= \operatorname{Var}\left(\hat{f}_1\, \hat{f}_2 \,\middle|\, \mathcal{B}_1\right) - \\ &\left(\operatorname{Var}\left(\hat{f}_1 \,\middle|\, \mathcal{B}_1\right) \operatorname{Var}\left(\hat{f}_2 \,\middle|\, \mathcal{B}_1\right) + \operatorname{Cov}\left(\hat{f}_1^2, \hat{f}_2^2 \,\middle|\, \mathcal{B}_1\right)\right)\end{aligned}$$

where the correction terms $\operatorname{Var}(\hat{f}_1 \,|\, \mathcal{B}_1) \operatorname{Var}(\hat{f}_2 \,|\, \mathcal{B}_1)$ and $\operatorname{Cov}(\hat{f}_1^2, \hat{f}_2^2 \,|\, \mathcal{B}_1)$ are of similar size but opposite sign and thus almost cancel out.

For the purpose of this comparison, we should take into account that the factors $\hat{f}_l$ are in practice mostly between 1 and 2 and their standard error is < 1. Thus variances and covariances are of even smaller size and so on. This means that $\operatorname{Var}(\hat{f}_1 \,|\, \mathcal{B}_1) \operatorname{Var}(\hat{f}_2 \,|\, \mathcal{B}_1)$ and $\operatorname{Cov}(\hat{f}_1^2, \hat{f}_2^2 \,|\, \mathcal{B}_1)$ are mostly negligeable as compared to $\hat{f}_1^2 \operatorname{Var}(\hat{f}_2 \,|\, \mathcal{B}_2)$ or $\operatorname{Var}(\hat{f}_1 \,|\, \mathcal{B}_1)\, \hat{f}_2^2$.

Finally, we give a small numerical example for the statements of this section where we can generate all possible triangles according to the time series model of BBMW. Given are the first column $X_{11} = 110$, $X_{21} = 100$, $X_{31} = 90$ and the true parameters $f_1 = 1.15$, $f_2 = 2.4$, $\sigma_1 = 10$, $\sigma_2 = 8$. As distribution for $\varepsilon_{k,j}$ we choose the simplest distribution possible, i.e. each $\varepsilon_{k,j}$ can only take on the values +1 and –1 with 50% probability each. With these parameters we

have exactly $2^5 = 32$ different sets of data $\{X_{12}, X_{13}, X_{22}, X_{23}, X_{32}\}$ using the time series

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

For each of these trapezoids, we can calculate the parameter estimates $\hat{f}_1, \hat{f}_2$, $\hat{\sigma}_1^2, \hat{\sigma}_2^2$, the true value of the estimation error $(\hat{f}_1\hat{f}_2 - f_1 f_2)^2$ and its estimators

$$\widehat{CEE}_{\text{Mack}} = \hat{f}_1^2\, \hat{\sigma}_2^2 / S_2 + \hat{f}_2^2\, \hat{\sigma}_1^2 / S_1,$$

$$\widehat{CEE}_{\text{BBMW}} = \widehat{CEE}_{\text{Mack}} + \frac{\hat{\sigma}_1^2\, \hat{\sigma}_2^2}{S_1\, S_2}.$$

In 22 of all 32 trapezoids, $\widehat{CEE}_{\text{Mack}}$ is closer to the true estimation error than $\widehat{CEE}_{\text{BBMW}}$. Averaged over all 32 trapezoids, we obtain

$$E\left(\left(\hat{f}_1\, \hat{f}_2 - f_1\, f_2\right)^2 \middle| \mathcal{B}_1\right) = 2.37,$$

$$E\left(\widehat{CEE}_{\text{Mack}} \middle| \mathcal{B}_1\right) = 2.53,$$

$$E\left(\widehat{CEE}_{\text{BBMW}} \middle| \mathcal{B}_1\right) = 2.70.$$

This confirms the theoretical findings of this section.

Conclusion. On average, the BBMW formula for the estimation error overestimates the true estimation error whereas Mack's formula is only very slightly less than BBMW and therefore on average yields a better approximation to the estimation error.

7. Overall conclusion

The approach proposed by BBMW does not lead to an improvement of Mack's formula but rather overstates the estimation error as the corresponding formula by Murphy [4] does. The reason is that the squares of the development factors are negatively correlated. This fact was not known before and was discovered (essentially by Gerhard Quarg) when analysing the BBMW approach. Therefore we are grateful to BBMW for their stimulating approach which helped to clarify that Mack's approximate formula for the estimation error is difficult to be improved.

Reference

Buchwalder, M., Bühlmann, H., Merz, M. and Wüthrich M.V. (2006) The Mean Square Error of Prediction in the Chain Ladder Reserving Method (Mack and Murphy Revisited), *ASTIN Bulletin*, **36(2)**, 521-542.

THE MEAN SQUARE ERROR OF PREDICTION IN THE CHAIN LADDER RESERVING METHOD – FINAL REMARK

BY

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

It seems to us that the discussion with Thomas Mack, Gerhard Quarg and Christian Braun illustrates how important it is to clearly define the understanding of the estimation error: Basically the estimation error derives from the fact that the true parameters of the underlying stochastic model are unknown and need to be estimated from the observed data. But the outcome of these observed figures in the claims trapezoid could also have been different. How to interpret this "being different"?

In our paper we have described three possible answers to this question (approaches in formulae (4.15)-(4.17)), namely complete resampling and conditional resampling as the extreme cases plus a compromise between these extremes. We have chosen to write our paper on the understanding of conditional resampling because we wanted our measure of estimation error to depend on the actual observed data which lead us to work with the conditional probability distributions given the observed trapezoid.

The question of the "right form of resampling" is indeed a fundamental one. It is not a chain ladder specific question, but needs to be answered for all methods of stochastic claims reserving, in particular for all approaches using bootstrapping techniques. In the latter case the bootstrapping technique directly follows from this understanding.

Finally, we acknowledge that the discussions with Thomas Mack, Gerhard Quarg and Christian Braun have been a great stimulus to us, firstly to better understand our own ideas and secondly to learn that other interpretations of the estimation error may also be important. We hope to continue learning from continuing discussions with them on a private basis.

We would also like to thank another anonymous referee and the ASTIN Editors for all the work and effort that went into having our paper published.

Astin Bulletin 36(2), 553. doi: 10.2143/AST.36.2.2017938