THE PREDICTION ERROR OF THE CHAIN LADDER METHOD APPLIED TO CORRELATED RUN-OFF TRIANGLES

BY

CHRISTIAN BRAUN

ABSTRACT

It is shown how the distribution-free method of Mack (1993) can be extended in order to estimate the prediction error of the Chain Ladder method for a portfolio of several correlated run-off triangles.

KEYWORDS

Chain Ladder, Prediction Error, Correlation of Run-offs, Segmented Portfolio.

1. INTRODUCTION

In Mack (1993), a distribution-free method was developed in order to estimate the prediction error of Chain Ladder reserve estimates. For claims reserving purposes, an insurance company usually subdivides its portfolio into several subportfolios such that the development behavior of each subportfolio can be assumed to be homogeneous. Then, for each subportfolio, the Chain Ladder method can be applied in order to estimate the appropriate claims reserves and their prediction error.

But what is finally needed, are the claims reserves for the whole portfolio of the insurance company and their prediction error. Whereas the estimates of the claims reserves of each subportfolio can simply be added together in order to arrive at an estimate for the claims reserves of the whole portfolio, this is only the case for the prediction variances if the subportfolios can be assumed to be independent. But in long tail business, the development of different subportfolios is influenced to a substantial degree by the development behavior of bodily injury claims (medical and nursing costs). Even after correcting the data for the claims inflation, further direct and indirect sources for correlations between run-offs of a portfolio exist (see e.g. Houltram (2003)). Therefore, subportfolios in general can not be assumed to be independent. Then, the question arises how the prediction error of the aggregated portfolio can be arrived at.

In this situation, applying the Chain Ladder method to the overall triangle and taking the prediction error from this calculation is not a good solution

ASTIN BULLETIN, Vol. 34, No. 2, 2004, pp. 399-423

because already the reserve estimates obtained in this way will not be identical to the aggregation of the reserve estimates of the individual subportfolios, see e.g. Ajne (1994). Moreover, the aggregation of run-off triangles with different development patterns is like mixing apples and oranges and will normally lead to invalid results.

Therefore in this paper, a new, more sensible approach is developed. We assume that the correlation between two run-off triangles finds its manifestation in a fixed correlation coefficient between the individual development factors of the two corresponding development periods of the triangles. This correlation coefficient may depend on the development period, but not on the accident year. This assumption fits very well to the basic assumption behind the Chain Ladder method that the individual development factors of each development period fluctuate randomly around a fixed, but unknown age-to-age factor.

In actuarial practice, this approach enables the actuary to set up a range and a prudential margin for the reserves of the whole portfolio as required e.g. by several national accounting standards. The reserving bounds described in this paper are solely based on stochastic assumptions and on the observed data and not on assumed correlations between lines of business – as often done – which do not refer to the peculiarities of the underlying portfolio.

Due to the increasing importance of stochastic methods in claims reserving the prediction error of the reserves of a portfolio was subject of several publications, recently. In none of these papers, the author is aware of, the correlation between segments is defined such rigorously as it is done here. In Brehm (2002) for example, the correlation of the reserve distributions of the segments is simply set equal to the correlation of the separated calendar period inflation parameter estimate. Furthermore, Brehm does not use the Chain Ladder method for the ultimate projection.

In our approach the prediction error for the reserve estimate of a portfolio of correlated segments is based on a stochastic model. In a simulation based approach, Kirschner (2002) extended the bootstrapping technique for estimating the reserve variability of a single segment to a whole portfolio. This technique produces samples of the portfolio, but it is not clear what statistical properties these samples have actually and which correlations of the original segments are grasped in the samples at all. Aside, the bootstrapping technique assumes independent increments in the segments which does not fit with the Chain Ladder assumptions.

The paper is organized as follows: Section 2.1 gives the basic notations and repeats the recursive formulae for the prediction error of a single accident year for one triangle. From this, the prediction error of the total claims amount of all accident years is derived in section 2.2. In section 3, a second run-off triangle is introduced as well as the decisive assumption on the correlation between both triangles. In section 4, the recursive formulae for the prediction error of the sum of the two triangles are derived. In section 5, a numerical example is given including the derivation of a range for the best estimate of the portfolio reserve. In the final section 6, some remarks regarding the impact of claims inflation on the correlation of run-offs are made and properties of a simplified model are presented.

2. The prediction error for one run-off triangle

2.1. The prediction error of the ultimate claims amount of one accident year

Let $C_{ik} > 0$ be the cumulative claims amount of accident year i, $1 \leq i \leq n$, after k years of development, $1 \leq k \leq n$, for a certain subportfolio. The amounts C_{ik} with $i + k \leq n + 1$ are observable and we are interested in predicting the amounts C_{in} for $i = 2, 3, \ldots, n$. The Chain Ladder method does this recursively by

$$\hat{C}_{ik} = \hat{C}_{i,k-1} \cdot \hat{f}_k \tag{1}$$

with starting value $\hat{C}_{i,n+1-i} = C_{i,n+1-i}$ and age-to-age factor

$$\hat{f}_k = \frac{\sum_{i=1}^{n+1-k} C_{ik}}{C_{<,k-1}} = \sum_{i=1}^{n+1-k} \frac{C_{i,k-1}}{C_{<,k-1}} \cdot F_{ik} \tag{2}$$

which is a weighted average of individual development factors

$$F_{ik} := \frac{C_{ik}}{C_{i,k-1}} \text{ with } C_{<,k-1} := \sum_{i=1}^{n+1-k} C_{i,k-1}.$$

In the following we consider numerous conditional expectation values and variances. To avoid there lengthy expressions we introduce some notation. The condition "T_k" means that all variables $\{C_{ij} \mid 1 \leq i \leq n,\ 1 \leq j \leq k, i + j \leq n + 1\}$ of the run-off triangle up to and including development year k are given. Especially, the condition "T_n" indicates that the whole triangle is given. Furthermore, we use T_{ik} when the variables $\{C_{ij} \mid 1 \leq j \leq k\}$ are given.

On the basis of the stochastic assumptions (see Mack (1993) and (1999), where the further results of this section can be found, too)

$$\mathrm{E}\left(F_{ik} \middle| T_{i,k-1}\right) = f_k, \tag{3}$$

$$\mathrm{Var}\left(F_{ik} \middle| T_{i,k-1}\right) = \frac{\sigma_k^2}{C_{i,k-1}}, \tag{4}$$

for all $1 \leq i \leq n$ and $2 \leq k \leq n$ where f_k and σ_k^2 are unknown parameters, the estimation procedure (1) and (2) can be shown to be reasonable and conditionally unbiased, i.e. $\mathrm{E}(\hat{f}_k \mid T_{k-1}) = f_k$ and $\mathrm{E}(\hat{C}_{in} \mid T_{n+1-i}) = C_{i,n+1-\mathrm{i}} f_{n+2-i} \cdot \ldots \cdot f_n = \mathrm{E}(C_{in} \mid T_{n+1-i})$, if the accident years are independent. The assumptions (3) and (4) together with the assumption of the independence of the accident years are the basis for all considerations in this paper and are used without mentioning explicitly each time.

The prediction error $\mathrm{mse}(\hat{C}_{in})$ for the ultimate claims amount of an accident year is defined as

$$\mathrm{mse}(\hat{C}_{in}) := \mathrm{E}((C_{in} - \hat{C}_{in})^2 \mid T_n)$$

https://doi.org/10.2143/AST.34.2.505150 Published online by Cambridge University Press

because for reserving purposes only the future variability given the observable data is of interest. This can be written in the form

$$\mathrm{mse}(\hat{C}_{in}) = \mathrm{Var}(C_{in}|T_{n+1-i}) + (\mathrm{E}(\hat{C}_{in}|T_{n+1-i}) - \hat{C}_{in})^2$$

which for estimation purposes is approximated by

$$\mathrm{mse}(\hat{C}_{in}) \approx \mathrm{Var}(C_{in}|T_{n+1-i}) + \mathrm{Var}(\hat{C}_{in}|T_{n+1-i}). \tag{5}$$

In (5) $\mathrm{Var}(C_{in}|T_{n+1-i})$ is called the random error and $\mathrm{Var}(\hat{C}_{in}|T_{n+1-i})$ the estimation error. To keep the notation as simple as possible we omit from now on the conditions in the expectations. So, whenever for $i+k>n+1$ expectations like $\mathrm{E}(C_{ik})$, $\mathrm{E}(\hat{C}_{ik})$ and variances like $\mathrm{Var}(C_{ik})$ or $\mathrm{Var}(\hat{C}_{ik})$ are considered, in the strict sense $\mathrm{E}(C_{ik}|T_{n+1-i})$, $\mathrm{E}(\hat{C}_{ik}|T_{n+1-i})$, $\mathrm{Var}(C_{ik}|T_{n+1-i})$ and $\mathrm{Var}(\hat{C}_{ik}|T_{n+1-i})$ are meant. The exact formulations of the following derivations can be found in Mack (1993).

Now, we deduce recursions for the random error and for the estimation error. For this purpose, the equations (3) and (4) are used in the form

$$\mathrm{E}(C_{ik}|T_{i,k-1}) = C_{i,k-1}f_k,$$

$$\mathrm{Var}(C_{ik}|T_{i,k-1}) = C_{i,k-1}\sigma_k^2.$$

Then we have for $i+k>n+1$

$$\begin{aligned}\mathrm{Var}(C_{ik}) &= \mathrm{E}(\mathrm{Var}(C_{ik}|T_{i,k-1})) + \mathrm{Var}(\mathrm{E}(C_{ik}|T_{i,k-1}))\\ &= \mathrm{E}(C_{i,k-1})\sigma_k^2 + \mathrm{Var}(C_{i,k-1})f_k^2.\end{aligned}$$

This yields for the estimator $\widehat{\mathrm{Var}}(C_{in})$ of the random error $\mathrm{Var}(C_{in})$ of the ultimate claims amount the recursion

$$\widehat{\mathrm{Var}}(C_{ik}) = \widehat{\mathrm{Var}}(C_{i,k-1}) \cdot \hat{f}_k^2 + \hat{C}_{i,k-1}\hat{\sigma}_k^2 \tag{6}$$

with the starting value

$$\widehat{\mathrm{Var}}(C_{i,n+1-i}) = 0$$

as $C_{i,n+1-i}$ is already known. An unbiased estimator of $\hat{\sigma}_k^2$ is given by

$$\hat{\sigma}_k^2 = \frac{1}{n-k}\sum_{i=1}^{n+1-k} C_{i,k-1}(F_{ik} - \hat{f}_k)^2. \tag{7}$$

Similarly, $\hat{C}_{ik} = \hat{C}_{i,k-1}\hat{f}_k$ yields

$$\begin{aligned}\mathrm{Var}(\hat{C}_{ik}) &= \mathrm{E}(\mathrm{Var}(\hat{C}_{i,k-1}\hat{f}_k|T_{k-1})) + \mathrm{Var}(\mathrm{E}(\hat{C}_{i,k-1}\hat{f}_k|T_{k-1}))\\ &= \mathrm{E}(\hat{C}_{i,k-1}^2\mathrm{Var}(\hat{f}_k|T_{k-1})) + \mathrm{Var}(\hat{C}_{i,k-1})f_k^2.\end{aligned}$$

From this the following recursion for the estimator $\widehat{\mathrm{Var}}(\hat{C}_{in})$ of the estimation error $\mathrm{Var}(\hat{C}_{in})$ of the ultimate claims estimate $\hat{C}_{in}$ can be deduced:

$$\widehat{\mathrm{Var}}(\hat{C}_{ik}) = \widehat{\mathrm{Var}}(\hat{C}_{i,k-1})\hat{f}_k^2 + \hat{C}_{i,k-1}^2 \cdot \frac{\hat{\sigma}_k^2}{C_{<,\,k-1}} \tag{8}$$

because

$$\mathrm{Var}(\hat{f}_k \,|\, T_{k-1}) = \frac{\sigma_k^2}{C_{<,\,k-1}}. \tag{9}$$

The starting value for this recursion is

$$\widehat{\mathrm{Var}}(\hat{C}_{i,n+1-i}) = 0$$

because $\hat{C}_{i,n+1-i}$ is already observed. This yields the joint recursion for the estimate of the prediction error:

$$\widehat{\mathrm{mse}}(\hat{C}_{ik}) = \widehat{\mathrm{mse}}(\hat{C}_{i,k-1}) \cdot \hat{f}_k^2 + \hat{C}_{i,k-1}^2 \left(\frac{\hat{\sigma}_k^2}{\hat{C}_{i,\,k-1}} + \frac{\hat{\sigma}_k^2}{C_{<,\,k-1}} \right). \tag{10}$$

2.2. The prediction error of the total ultimate claims amount of one run-off triangle

Annual reports of insurance companies usually disclose estimates only for reserves and claims amounts for all accident years together. To estimate a range of those aggregated amounts, we have to consider the estimation error and prediction error for all accident years together.

C_{1n} is already known and no estimate is necessary. Therefore the first accident year adds nothing to the random error and the estimation error for the whole run-off. Taking this into account, the prediction error $\mathrm{mse}(\sum_{i=2}^{n} \hat{C}_{in})$ for all accident years is defined as

$$\mathrm{mse}\left(\sum_{i=2}^{n} \hat{C}_{in}\right) := \mathrm{E}\left(\left(\sum_{i=2}^{n} (C_{in} - \hat{C}_{in}) \right)^2 \middle| T_n \right).$$

We have (Mack (1993))

$$\begin{aligned}
\mathrm{mse}\left(\sum_{i=2}^{n} \hat{C}_{in}\right) &= \mathrm{Var}\left(\sum_{i=2}^{n} C_{in} \middle| T_n \right) + \left(\sum_{i=2}^{n} \left(\mathrm{E}\left(\hat{C}_{in} \middle| T_{n+1-i}\right) - \hat{C}_{in} \right) \right)^2 \\
&= \mathrm{Var}\left(\sum_{i=2}^{n} C_{in} \middle| T_n \right) + \sum_{i=2}^{n} \left(\mathrm{E}\left(\hat{C}_{in} \middle| T_{n+1-i}\right) - \hat{C}_{in} \right)^2 \\
&\quad + 2 \sum_{2 \le i < j \le n} \left(\mathrm{E}\left(\hat{C}_{in} \middle| T_{n+1-i}\right) - \hat{C}_{in} \right)\left(\mathrm{E}\left(\hat{C}_{jn} \middle| T_{n+1-j}\right) - \hat{C}_{jn} \right)
\end{aligned}$$

$$\approx \operatorname{Var}\left(\sum_{i=2}^{n} C_{in} \middle| T_n\right) + \sum_{i=2}^{n} \operatorname{Var}\left(\hat{C}_{in} \middle| T_{n+1-i}\right)$$
$$+ 2 \sum_{2 \le i < j \le n} \operatorname{Cov}\left(\hat{C}_{in}, \hat{C}_{jn} \middle| T_{n+1-i}\right)$$

The random error of the total ultimate loss amount is $\operatorname{Var}(\sum_{i=2}^{n} C_{in} | T_n)$. The estimation error $\operatorname{Var}(\sum_{i=2}^{n} \hat{C}_{in})$ of the ultimate claims amount of all accident years together is

$$\operatorname{Var}\left(\sum_{i=2}^{n} \hat{C}_{in}\right) := \sum_{i=2}^{n} \operatorname{Var}\left(\hat{C}_{in} \middle| T_{n+1-i}\right) + \sum_{2 \le i < j \le n} 2 \operatorname{Cov}\left(\hat{C}_{in}, \hat{C}_{jn} \middle| T_{n+1-i}\right). \tag{11}$$

It is important to note, that $\operatorname{Var}(\sum_{i=1}^{n} \hat{C}_{in})$ is only a notation for the right-hand-side in (11) and that it is not a variance since the right-hand-side of the definition (11) can not be rewritten as one single conditional variance due to the different conditions of the variances and covariances in the sum. This yields the following approximation for $\operatorname{mse}(\sum_{i=1}^{n} \hat{C}_{in})$ (which is analogous to (5)):

$$\operatorname{mse}\left(\sum_{i=1}^{n} \hat{C}_{in}\right) \approx \operatorname{Var}\left(\sum_{i=1}^{n} C_{in} \middle| T_n\right) + \operatorname{Var}\left(\sum_{i=1}^{n} \hat{C}_{in}\right).$$

Again, we omit the condition for simplicity. The random error $\operatorname{Var}(\sum_{i=2}^{n} C_{in})$ fulfills due to the independence of the accident years (which here implies that the variables C_{in}, $i = 1, \ldots, n$ are conditionally uncorrelated, Mack (2002), p. 255) the equation

$$\operatorname{Var}\left(\sum_{i=2}^{n} C_{in}\right) = \sum_{i=2}^{n} \operatorname{Var}\left(C_{in}\right). \tag{12}$$

Of course (12) can be generalized to

$$\operatorname{Var}\left(\sum_{i=n+2-k}^{n} C_{ik}\right) = \sum_{i=n+2-k}^{n} \operatorname{Var}\left(C_{ik}\right). \tag{13}$$

(13) and the recursion (6) for the random error of one accident year yield the recursion

$$\widehat{\operatorname{Var}}\left(\sum_{i=n+2-k}^{n} C_{ik}\right) = \widehat{\operatorname{Var}}\left(\sum_{i=n+3-k}^{n} C_{i,k-1}\right) \hat{f}_k^2 + \hat{C}_{\ge,k-1} \hat{\sigma}_k^2,$$

with

$$\hat{C}_{\ge,k-1} := \sum_{i=n+2-k}^{n} \hat{C}_{i,k-1}. \tag{14}$$

https://doi.org/10.2143/AST.34.2.505150 Published online by Cambridge University Press

Note, $\hat{C}_{\geq,k-1}$ is the sum of the estimated claims amounts of development period $k-1$ plus the known amount $C_{n+2-k,k-1}$ of the actual calendar year. This recursion starts with $k=2$ since for the first development year all claims amounts C_{i1}, $1 \leq i \leq n$, are already known. Here and in the following we use the convention that an empty summation is equal to 0.

For the estimation error $\mathrm{Var}(\sum_{i=2}^{n}\hat{C}_{in})$ such a simple relation as (12) does not hold since all correlations between the ultimate claims amount estimates of different accident years have to be considered. A recursion for $\widehat{\mathrm{Cov}}(\hat{C}_{in},\hat{C}_{jn})$ can be achieved by (with $k>n+1-i$ and $i<j$)

$$\begin{aligned}\mathrm{Cov}(\hat{C}_{ik},\hat{C}_{jk}) &= \mathrm{E}(\mathrm{Cov}(\hat{C}_{i,k-1}\hat{f}_k,\hat{C}_{j,k-1}\hat{f}_k \mid T_{k-1})) + \\ &\quad + \mathrm{Cov}(\mathrm{E}(\hat{C}_{i,k-1}\hat{f}_k \mid T_{k-1}),\, \mathrm{E}(\hat{C}_{j,k-1}\hat{f}_k \mid T_{k-1})) \\ &= \mathrm{E}(\hat{C}_{i,k-1}\hat{C}_{j,k-1}\mathrm{Var}(\hat{f}_k \mid T_{k-1})) + \mathrm{Cov}(\hat{C}_{i,k-1},\hat{C}_{j,k-1})f_k^2 \end{aligned} \tag{15}$$

and using (9)

$$\widehat{\mathrm{Cov}}(\hat{C}_{ik},\hat{C}_{jk}) = \widehat{\mathrm{Cov}}(\hat{C}_{i,k-1},\hat{C}_{j,k-1})\hat{f}_k^2 + \hat{C}_{i,k-1}\,\hat{C}_{j,k-1}\frac{\hat{\sigma}_k^2}{C_{<,\,k-1}} \tag{16}$$

starting with $\widehat{\mathrm{Cov}}(\hat{C}_{i,n+1-i},\hat{C}_{j,n+1-i}) = 0$ since $i<j$ and $C_{i,n+1-i}$ is known. (16) and (8) yield the following recursion for the estimation error:

$$\widehat{\mathrm{Var}}\left(\sum_{i=n+2-k}^{n}\hat{C}_{ik}\right) = \widehat{\mathrm{Var}}\left(\sum_{i=n+3-k}^{n}\hat{C}_{i,k-1}\right)\hat{f}_k^2 + (\hat{C}_{\geq,k-1})^2\frac{\hat{\sigma}_k^2}{C_{<,\,k-1}}.$$

For the same reason as before, this recursion starts with $k=2$.

The recursions for the random error $\mathrm{Var}(\sum_{i=2}^{n}C_{in})$ and the estimation error $\mathrm{Var}(\sum_{i=2}^{n}\hat{C}_{in})$ yield the recursion for the prediction error $\mathrm{mse}(\sum_{i=2}^{n}C_{in})$ of the total claims amounts for all accident years:

$$\begin{aligned}\widehat{\mathrm{mse}}\left(\sum_{i=n+2-k}^{n}\hat{C}_{ik}\right) &= \widehat{\mathrm{mse}}\left(\sum_{i=n+3-k}^{n}\hat{C}_{i,k-1}\right)\hat{f}_k^2 \\ &\quad + \left(\hat{C}_{\geq,\,k-1}\right)^2\left(\frac{\hat{\sigma}_k^2}{\hat{C}_{\geq,\,k-1}} + \frac{\hat{\sigma}_k^2}{C_{<,\,k-1}}\right).\end{aligned} \tag{17}$$

The recursion starts with $k=2$. Using (4) and (9) it can be shown that (17) is the same recursion as the one already given in Mack (1999) for the prediction error. Structure of recursion (17) is the same as in (10). The only difference between the two recursions are the estimated claims amounts $\hat{C}_{\geq,k-1}$ instead of the claims amount $\hat{C}_{i,k-1}$ for one accident year in (10).

The prediction error $\mathrm{mse}(\sum_{i=1}^{n}\hat{C}_{in})$ gives the mean squared deviation between the estimated ultimate claims amount $\sum_{i=1}^{n}\hat{C}_{in}$ and the true ultimate claims amount $\sum_{i=1}^{n}C_{in}$. The estimation error $\mathrm{Var}(\sum_{i=1}^{n}\hat{C}_{in})$ gives the mean squared deviation between the estimated ultimate claims amount $\sum_{i=1}^{n}\hat{C}_{in}$ and the expected ultimate claims amount $E(\sum_{i=1}^{n}C_{in}) = E(\sum_{i=1}^{n}\hat{C}_{in})$. Whereas the prediction error has to be used for the variability loading for a loss portfolio transfer, it is the estimation error which has to be used when assessing a confidence interval (range) around $\sum_{i=1}^{n}\hat{C}_{in}$ for the best estimate $E(\sum_{i=1}^{n}C_{in})$ of $\sum_{i=1}^{n}C_{in}$.

3. A Chain ladder-type model for the correlation between two run-off triangles

Now assume we have another subportfolio with cumulative run-off data $\{D_{ik}\}$ in addition to the data $\{C_{ik}\}$ of section 2. Considering that, we modify the condition "$T_{i,k-1}$". In the following "$T_{i,k-1}$" means, both sets of observable variables $\{C_{ij} \mid 1 \le j \le k-1\}$ and $\{D_{ij} \mid 1 \le j \le k-1\}$ are given. Moreover, we assume (3), (4), to hold for this "$T_{i,k-1}$".

Note, in this case (3) and (4) with the "$T_{i,k-1}$" as introduced in Section 2.1 still hold, being just a consequence of the new assumption, i.e. we have by using the notation C_{k-1} for the set $\{C_{ij} \mid 1 \le j \le k-1\}$ and D_{k-1} for $\{D_{ij} \mid 1 \le j \le k-1\}$

$$E(F_{ik} \mid C_{k-1}) = E(E(F_{ik} \mid C_{k-1}, D_{k-1}) \mid C_{k-1}) = f_k, \tag{18}$$

$$\begin{aligned} Var(F_{ik} \mid C_{k-1}) &= E(Var(F_{ik} \mid C_{k-1}, D_{k-1}) \mid C_{k-1}) \\ &\quad + Var(E(F_{ik} \mid C_{k-1}, D_{k-1}) \mid C_{k-1}) \\ &= \frac{\hat{\sigma}_k^2}{C_{i,k-1}}. \end{aligned} \tag{19}$$

Aside, (18) and (19) justify actuarial practice using the Chain Ladder method for a subportfolio without considering in addition the observables of all other segments of the portfolio.

For the subportfolio with cumulative run-off data $\{D_{ik}\}$ we denote with g_k and τ_k^2 its Chain-Ladder parameters corresponding to f_k and σ_k^2, respectively. The stochastic assumptions are

$$E(G_{ik} \mid T_{i,k-1}) = g_k \tag{20}$$

$$Var(G_{ik} \mid T_{i,k-1}) = \frac{\tau_k^2}{D_{i,k-1}}. \tag{21}$$

Again, the accident years $i = 1,\ldots, n$ are assumed to be independent.

We have the following estimators

$$\hat{g}_k = \frac{\sum_{i=1}^{n+1-k} D_{ik}}{D_{<,k-1}} \tag{22}$$

$$\hat{\tau}_k^2 = \frac{1}{n-k}\sum_{i=1}^{n-k+1} D_{i,k-1}(G_{ik} - \hat{g}_k)^2 \tag{23}$$

https://doi.org/10.2143/AST.34.2.505150 Published online by Cambridge University Press

with

$$G_{ik} := \frac{D_{ik}}{D_{i,k-1}},$$

$$D_{<,k-1} := \sum_{i=1}^{n+1-k} D_{i,k-1}.$$

For each of the data sets $\{C_{ik}\}$ and $\{D_{ik}\}$ the stochastic model for the Chain Ladder consists of an own submodel for each development period k, $2 \leq k \leq n$. In order to arrive at formulae for expectation and variance of the ultimate claims D_{in} in terms of the observable amounts $\{D_{ik},\ i+k \leq n+1\}$, the submodels are simply chained together.

Therefore it seems natural to restrict any assumptions regarding the correlation between the arrays $\{C_{ik}, 1 \leq i,k \leq n\}$ and $\{D_{ik}, 1 \leq i,k \leq n\}$ to each of the pairwise corresponding development years $k, 2 \leq k \leq n$, if we want to stay within the chain ladder world. In this sense, the natural generalization of (4) and (21) is the assumption

$$\mathrm{Cov}\left(F_{ik}, G_{ik} \middle| T_{i,k-1}\right) = \frac{\rho_k}{\sqrt{C_{i,k-1} D_{i,k-1}}} \tag{24}$$

which is equivalent to assuming that the correlation coefficient between the individual development factors F_{ik} and G_{ik}

$$\mathrm{Corr}\left(F_{ik}, G_{ik} \middle| T_{i,k-1}\right) := \frac{\mathrm{Cov}\left(F_{ik}, G_{ik} \middle| T_{i,k-1}\right)}{\sqrt{\mathrm{Var}\left(F_{ik} \middle| T_{i,k-1}\right) \cdot \mathrm{Var}\left(G_{ik} \middle| T_{i,k-1}\right)}} = \frac{\rho_k}{\sigma_k \tau_k}$$

is constant for k fixed.

Of course, different accident years of the portfolio consisting of the run-off data sets $\{C_{ik}\}$ and $\{D_{ik}\}$ are assumed to be independent. Then we have

$$\mathrm{Cov}(C_{ik}, D_{jk} \,|\, T_{k-1}) = 0 \ \text{for} \ i \neq j,$$

since

$$\begin{aligned} \mathrm{E}(C_{ik} \cdot D_{jk} \,|\, T_{k-1}) &= \mathrm{E}(\mathrm{E}(C_{ik} \cdot D_{jk} \,|\, T_{k-1}, T_{ik}) \,|\, T_{k-1}) \\ &= \mathrm{E}(C_{ik}\mathrm{E}(D_{jk} \,|\, T_{k-1}, T_{ik}) \,|\, T_{k-1}) \\ &= \mathrm{E}(C_{ik} \,|\, T_{k-1})\mathrm{E}(D_{jk} \,|\, T_{k-1}). \end{aligned} \tag{25}$$

(25) also holds for F_{ik} and G_{jk} instead of C_{ik} and D_{jk}. This shows

$$\mathrm{Cov}(F_{ik}, G_{jk} \,|\, T_{k-1}) = 0 \ \text{for} \ i \neq j.$$

https://doi.org/10.2143/AST.34.2.505150 Published online by Cambridge University Press

In analogy of the estimation of σ_k^2 and τ_k^2, the new parameter ρ_k can be estimated by

$$\hat{\rho}_k = \frac{1}{n-k-1+w_k^2} \sum_{i=1}^{n+1-k} \sqrt{C_{i,k-1} D_{i,k-1}} \, (F_{ik} - \hat{f}_k)(G_{ik} - \hat{g}_k) \tag{26}$$

with

$$w_k^2 := \frac{\left(\sum_{i=1}^{n+1-k} \sqrt{C_{i,k-1} D_{i,k-1}}\right)^2}{C_{<,k-1} \cdot D_{<,k-1}}.$$

The factor $\frac{1}{n-k-1+w_k^2}$ instead of $\frac{1}{n-k}$ as for $\hat{\sigma}_k^2$ and $\hat{\tau}_k^2$ ensures that the estimator $\hat{\rho}_k$ for ρ_k is unbiased. Note, that w_k^2 is positive and ≤ 1 (Cauchy-Schwarz inequality).

4. Estimation of the prediction error of the sum of two run-off triangles

First of all, we have to define the prediction error $\text{mse}(\hat{C}_{in} + \hat{D}_{in})$ for the ultimate claims amount of an accident year of the portfolio. It is defined analogously as for one run-off:

$$\text{mse}(\hat{C}_{in} + \hat{D}_{in}) := \text{E}((C_{in} + D_{in} - (\hat{C}_{in} + \hat{D}_{in}))^2 \,|\, T_n).$$

This can be approximated by

$$\text{mse}(\hat{C}_{in} + \hat{D}_{in}) \approx \text{Var}(C_{in} + D_{in} \,|\, T_{n+1-i}) + \text{Var}(\hat{C}_{in} + \hat{D}_{in} \,|\, T_{n+1-i}).$$

Here, $\text{Var}(C_{in} + D_{in} \,|\, T_{n+1-i})$ is the random error and $\text{Var}(\hat{C}_{in} + \hat{D}_{in} \,|\, T_{n+1-i})$ is the estimation error. Again, we omit these conditions in the following.

Based on the assumption (24) which can be rewritten as

$$\text{Cov}(C_{ik}, D_{ik} \,|\, T_{i,k-1}) = \sqrt{C_{i,k-1} D_{i,k-1}} \, \rho_k,$$

we now can calculate the random error $\text{Var}(C_{in} + D_{in})$ and the estimation error $\text{Var}(\hat{C}_{in} + \hat{D}_{in})$ of the combined triangle $\{C_{ik} + D_{ik} \,|\, i + k \leq n+1\}$. We have

$$\text{Var}(C_{in} + D_{in}) = \text{Var}(C_{in}) + 2\text{Cov}(C_{in}, D_{in}) + \text{Var}(D_{in})$$

and therefore, in addition to the recursions considered before, we need only a recursion for $\text{Cov}(C_{in}, D_{in})$, too. From

$$\begin{aligned} \text{Cov}(C_{ik}, D_{ik}) &= \text{E}(\text{Cov}(C_{ik}, D_{ik} \,|\, T_{i,k-1})) \\ &\quad + \text{Cov}(\text{E}(C_{ik} \,|\, T_{i,k-1}), \text{E}(D_{ik} \,|\, T_{i,k-1})) \\ &= \text{E}(\sqrt{C_{i,k-1} D_{i,k-1}})\rho_k + \text{Cov}(C_{i,k-1}, D_{i,k-1}) f_k g_k \end{aligned}$$

we deduce the recursion (for $i+k>n+1$)

$$\widehat{\mathrm{Cov}}(C_{ik},D_{ik}) = \widehat{\mathrm{Cov}}\left(C_{i,k-1},D_{i,k-1}\right)\hat{f}_k\hat{g}_k + \sqrt{\hat{C}_{i,k-1},\hat{D}_{i,k-1}}\,\hat{\rho}_k \tag{27}$$

for the estimated covariance between C_{ik} and D_{ik}. The starting value is

$$\widehat{\mathrm{Cov}}(C_{i,n+1-i},D_{i,n+1-i}) = 0$$

as both variables have already been observed. Similarly, for the estimation error we have

$$\mathrm{Var}(\hat{C}_{in}+\hat{D}_{in}) = \mathrm{Var}(\hat{C}_{in}) + 2\mathrm{Cov}(\hat{C}_{in},\hat{D}_{in}) + \mathrm{Var}(\hat{D}_{in})$$

and

$$\begin{aligned}\mathrm{Cov}(\hat{C}_{ik},\hat{D}_{ik}) &= \mathrm{E}(\mathrm{Cov}(\hat{C}_{i,k-1}\hat{f}_k,\hat{D}_{i,k-1}\hat{g}_k\,|\,T_{k-1}))\\ &\quad + \mathrm{Cov}(\mathrm{E}(\hat{C}_{i,k-1}\hat{f}_k\,|\,T_{k-1}),\mathrm{E}(\hat{D}_{i,k-1}\hat{g}_k\,|\,T_{k-1}))\\ &= \mathrm{E}(\hat{C}_{i,k-1}\hat{D}_{i,k-1}\mathrm{Cov}(\hat{f}_k,\hat{g}_k\,|\,T_{k-1}))\\ &\quad + \mathrm{Cov}(\hat{C}_{i,k-1},\hat{D}_{i,k-1})f_kg_k\end{aligned}$$

as well as

$$\begin{aligned}\mathrm{Cov}\left(\hat{f}_k,\hat{g}_k\,\middle|\,T_{k-1}\right) &= \mathrm{Cov}\left(\sum_{j=1}^{n+1-k}\frac{C_{j,k-1}}{C_{<,k-1}}F_{jk},\sum_{j=1}^{n+1-k}\frac{D_{j,k-1}}{D_{<,k-1}}G_{jk}\,\middle|\,T_{k-1}\right)\\ &= \sum_{j=1}^{n+1-k}\frac{C_{j,k-1}}{C_{<,k-1}}\frac{D_{j,k-1}}{D_{<,k-1}}\mathrm{Cov}\left(F_{jk},G_{jk}\,\middle|\,T_{k-1}\right)\\ &= \sum_{j=1}^{n+1-k}\frac{\sqrt{C_{j,k-1}D_{j,k-1}}}{C_{<,k-1}D_{<,k-1}}\rho_k\end{aligned} \tag{28}$$

Taken together, we have the recursion

$$\begin{aligned}\widehat{\mathrm{Cov}}(\hat{C}_{ik},\hat{D}_{ik}) &= \widehat{\mathrm{Cov}}(\hat{C}_{i,k-1},\hat{D}_{i,k-1})\cdot\hat{f}_k\hat{g}_k\\ &\quad + \frac{\hat{C}_{i,k-1}\hat{D}_{i,k-1}}{C_{<,k-1}\cdot D_{<,k-1}}\hat{\rho}_k\sum_{j=1}^{n+1-k}\sqrt{C_{j,k-1}D_{j,k-1}}\end{aligned} \tag{29}$$

with starting value

$$\widehat{\mathrm{Cov}}(\hat{C}_{i,n+1-i},\hat{D}_{i,n+1-i}) = 0.$$

This completes the derivation of formulae for the random error, for the estimation error and taken together for the prediction error for the ultimate claims amount of one accident year in a portfolio consisting of two correlated sub-portfolios.

For actuarial evaluation of the liabilities of a whole portfolio and their potential adverse development the errors of the ultimate claims amount for all accident years of the portfolio are important quantities. The prediction error of the total ultimate claims amount $\sum_{i=2}^{n}(\hat{C}_{in} + \hat{D}_{in})$ is

$$
\begin{aligned}
&\operatorname{mse}\left(\sum_{i=2}^{n}(\hat{C}_{in} + \hat{D}_{in})\right) := \mathrm{E}\left(\left(\sum_{i=2}^{n}\left(C_{in} + D_{in} - (\hat{C}_{in} + \hat{D}_{in})\right)\right)^2 \middle| T_n\right) \\
&= \operatorname{Var}\left(\sum_{i=2}^{n}(C_{in} + D_{in}) \middle| T_n\right) \\
&\quad + \left(\sum_{i=2}^{n}\left(\mathrm{E}\left(\hat{C}_{in} + \hat{D}_{in} \middle| T_{n+1-i}\right) - (\hat{C}_{in} + \hat{D}_{in})\right)\right)^2 \\
&= \operatorname{Var}\left(\sum_{i=2}^{n}(C_{in} + D_{in}) \middle| T_n\right) \\
&\quad + \left(\sum_{i=2}^{n}\left(\mathrm{E}\left(\hat{C}_{in} \middle| T_{n+1-i}\right) - \hat{C}_{in}\right) + \sum_{i=2}^{n}\left(\mathrm{E}\left(\hat{D}_{in} \middle| T_{n+1-i}\right) - \hat{D}_{in}\right)\right)^2 \\
&\approx \operatorname{Var}\left(\sum_{i=2}^{n}(C_{in} + D_{in}) \middle| T_n\right) \\
&\quad + \operatorname{Var}\left(\sum_{i=2}^{n}\hat{C}_{in}\right) + \operatorname{Var}\left(\sum_{i=2}^{n}\hat{D}_{in}\right) + \sum_{1 \le i,j \le n} 2\operatorname{Cov}\left(\hat{C}_{in}, \hat{D}_{jn} \middle| T_{n+1-min(i,j)}\right),
\end{aligned}
$$

where $min(i,j)$ denotes the Minimum of i and j. The first term is the random error, the last three together are the estimation error. Note, here we used the notation $\operatorname{Var}\left(\sum_{i=2}^{n}\hat{C}_{in}\right)$ and $\operatorname{Var}\left(\sum_{i=2}^{n}\hat{D}_{in}\right)$ as introduced in section 2.2.

The random error $\operatorname{Var}\left(\sum_{i=2}^{n}(\hat{C}_{in} + \hat{D}_{in})\right)$ – omitting conditions – can be written as

$$
\begin{aligned}
&\operatorname{Var}\left(\sum_{i=2}^{n}(C_{in} + D_{in})\right) \\
&= \operatorname{Var}\left(\sum_{i=2}^{n}C_{in}\right) + 2\operatorname{Cov}\left(\sum_{i=2}^{n}C_{in}, \sum_{i=2}^{n}D_{in}\right) + \operatorname{Var}\left(\sum_{i=2}^{n}D_{in}\right)
\end{aligned}
$$

For the random errors $\operatorname{Var}\left(\sum_{i=2}^{n}C_{in}\right)$ and $\operatorname{Var}\left(\sum_{i=2}^{n}D_{in}\right)$ we have already derived recursions in section 2. Therefore, only a recursion for the covariance of $\sum_{i=2}^{n}C_{in}$ and $\sum_{i=2}^{n}D_{in}$ is needed. Due to the independence of the accident years – which implies that the variables C_{in} and D_{jn} with $i,j = 1,\dots,n, i \neq j$ are conditionally uncorrelated – we have

$$
\operatorname{Cov}\left(\sum_{i=2}^{n}C_{in}, \sum_{i=2}^{n}D_{in}\right) = \sum_{i=2}^{n}\operatorname{Cov}(C_{in}, D_{in}).
$$

https://doi.org/10.2143/AST.34.2.505150 Published online by Cambridge University Press

Using the recursions for $\widehat{\mathrm{Cov}}(C_{ik}, D_{ik})$, $2 \le i \le n$ yields the recursion

$$\widehat{\mathrm{Cov}}\left(\sum_{i=n+2-k}^{n} C_{ik}, \sum_{i=n+2-k}^{n} D_{ik}\right) = \widehat{\mathrm{Cov}}\left(\sum_{i=n+3-k}^{n} C_{i,k-1}, \sum_{i=n+3-k}^{n} D_{i,k-1}\right)\hat{f}_k\hat{g}_k + \hat{\rho}_k \sum_{i=n+2-k}^{n} \sqrt{\hat{C}_{i,k-1}, \hat{D}_{i,k-1}} \tag{30}$$

starting with $k = 2$ since for the first development year all C_{i1} and D_{i1} are known.

For the covariances $\mathrm{Cov}(\hat{C}_{in}, \hat{D}_{jn})$ in the estimation error we proceed as in (15) and for (27). This leads to the recursion

$$\widehat{\mathrm{Cov}}(\hat{C}_{ik}, \hat{D}_{jk}) = \widehat{\mathrm{Cov}}(\hat{C}_{i,k-1}, \hat{D}_{j,k-1})\,\hat{f}_k\hat{g}_k + \frac{C_{i,k-1}D_{j,k-1}}{C_{<,k-1}\cdot D_{<,k-1}}\hat{\rho}_k \sum_{m=1}^{n+1-k}\sqrt{C_{m,k-1}\cdot D_{m,k-1}} \tag{31}$$

with starting value $k = n + 1 - min(i, j)$. Recursion (29) is a special case of (31). The recursion for $\sum_{i,j}\widehat{\mathrm{Cov}}(\hat{C}_{in}, \hat{D}_{jn})$ is then

$$\sum_{i,j=n+2-k}^{n}\widehat{\mathrm{Cov}}(\hat{C}_{ik}, \hat{D}_{jk}) = \sum_{i,j=n+3-k}^{n}\widehat{\mathrm{Cov}}(\hat{C}_{i,k-1}, \hat{D}_{j,k-1})\,\hat{f}_k\hat{g}_k + \hat{C}_{\ge,k-1}\hat{D}_{\ge,k-1}\,\hat{\rho}_k\frac{\sum_{i=1}^{n+1-k}\sqrt{C_{i,k-1}\cdot D_{i,k-1}}}{C_{<,k-1}\cdot D_{<,k-1}} \tag{32}$$

starting with $k = 2$ (cf. definition of $\hat{C}_{\ge,k-1}$ in (14)). This recursion completes the derivation of the recursions for the estimation error and the prediction error for the ultimate claims amounts estimates of the sum of two correlated subportfolios. The extension to more than two subportfolios is obvious.

5. Numerical example

In our numerical example we use data published by the Reinsurance Association of America (RAA) in their historical loss development study (RAA (2001)). Cumulative incurred losses $\{C_{ik}\}$ of General Liability (GL) reinsurance business are given in Table 1. Table 2 contains the corresponding data $\{D_{ik}\}$ for Auto Liability (AL) reinsurance business. For details see RAA (2001). For a demonstration of our approach with these runoffs we assume that the claims development comprised in each of these triangles is homogeneous so that we can limit our analysis to the two given triangles and we have not to perform any analysis of subtriangles. Moreover, we assume for simplicity that the development stops after the fourteenth year for both run-offs. Therefore we dispense with any extrapolation beyond the fourteenth development year.

https://doi.org/10.2143/AST.34.2.505150 Published online by Cambridge University Press

TABLE 1

GENERAL LIABILITY RUN-OFF

AY/DY	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1987	59.966	163.152	254.512	349.524	433.265	475.778	513.660	520.309	527.978	539.039	537.301	540.873	547.696	549.589
1988	49.685	153.344	272.936	383.349	458.791	503.358	532.615	551.437	555.792	556.671	560.844	563.571	562.795	
1989	51.914	170.048	319.204	425.029	503.999	544.769	559.475	577.425	588.342	590.985	601.296	602.710		
1990	84.937	273.183	407.318	547.288	621.738	687.139	736.304	757.440	758.036	782.084	784.632			
1991	98.921	278.329	448.530	561.691	641.332	721.696	742.110	752.434	768.638	768.373				
1992	71.708	245.587	416.882	560.958	654.652	726.813	768.358	793.603	811.100					
1993	92.350	285.507	466.214	620.030	741.226	827.979	873.526	896.728						
1994	95.731	313.144	553.702	755.978	857.859	962.825	1.022.241							
1995	97.518	343.218	575.441	769.017	934.103	1.019.303								
1996	173.686	459.416	722.336	955.335	1.141.750									
1997	139.821	436.958	809.926	1.174.196										
1998	154.965	528.080	1.032.684											
1999	196.124	772.971												
2000	204.325													

TABLE 2

AUTO LIABILITY RUN-OFF

AY/DY	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1987	114.423	247.961	312.982	344.340	371.479	371.102	380.991	385.468	385.152	392.260	391.225	391.328	391.537	391.428
1988	152.296	305.175	376.613	418.299	440.308	465.623	473.584	478.427	478.314	479.907	480.755	485.138	483.974	
1989	144.325	307.244	413.609	464.041	519.265	527.216	535.450	536.859	538.920	539.589	539.765	540.742		
1990	145.904	307.636	387.094	433.736	463.120	478.931	482.529	488.056	485.572	486.034	485.016			
1991	170.333	341.501	434.102	470.329	482.201	500.961	504.141	507.679	508.627	507.752				
1992	189.643	361.123	446.857	508.083	526.562	540.118	547.641	549.605	549.693					
1993	179.022	396.224	497.304	553.487	581.849	611.640	622.884	635.452						
1994	205.908	416.047	520.444	565.721	600.609	630.802	648.365							
1995	210.951	426.429	525.047	587.893	640.328	663.152								
1996	213.426	509.222	649.433	731.692	790.901									
1997	249.508	580.010	722.136	844.159										
1998	258.425	686.012	915.109											
1999	368.762	909.066												
2000	394.997													

TABLE 3

DEVELOPMENT FACTORS AND PARAMETER ESTIMATES

DY	2	3	4	5	6	7	8	9	10	11	12	13	14
$\hat{f}_k$	3,235	1,720	1,354	1,179	1,106	1,055	1,026	1,014	1,012	1,006	1,005	1,005	1,003
$\hat{s}_k^2$	17.642,53	7.027,84	1.432,51	685,21	144,32	209,99	50,81	52,03	136,96	43,45	2,66	54,03	2,66
$\hat{g}_k$	2,226	1,269	1,120	1,067	1,035	1,017	1,010	1,000	1,004	0,999	1,004	0,999	1,000
$\hat{\tau}_k^2$	11.104,38	607,07	321,80	363,48	156,37	30,81	20,41	4,52	26,45	1,95	10,31	1,86	0,34
$\hat{w}_k^2$	0,988	0,995	0,995	0,996	0,996	0,996	0,996	0,995	0,995	0,994	0,998	0,998	1,000
$\hat{\rho}_k$	3.434,41	1.022,71	463,29	222,82	73,14	36,25	–5,53	12,30	20,26	6,33	–0,02	10,04	–
$\hat{\rho}_k/(\hat{s}_k\hat{\tau}_k)$	0,245	0,495	0,682	0,446	0,487	0,451	–0,172	0,802	0,337	0,687	–0,004	1,001	–

The Chain-Ladder method yields the development factors $\hat{f}_k$ (for the GL run-off) and $\hat{g}_k$ (for AL run-off) and the parameter estimates $\hat{\sigma}_k$ (GL run-off) and $\hat{\tau}_k$ (AL run-off) as given in Table 3. The parameters σ_{14} and τ_{14} which can not be estimated via (7) and (23) since there is only one individual development factor in each run-off for the fourteenth development year, are selected as

$$\hat{\sigma}_{14}^2 = min(\hat{\sigma}_{13}^4 / \hat{\sigma}_{12}^2, \hat{\sigma}_{12}^2)$$

(see Mack (1993)) and $\hat{\tau}_{14}^2$ analogous. The parameters w_k^2 in the row 5 of Table 3 show that w_k^2 is approximately 1 for all development years in this example i.e. we have $\frac{1}{n-k-1+w_k^2} \approx \frac{1}{n-k}$. Rows 6 and 7 of Table 3 contain the estimate for ρ_k and for the correlation coefficient $\rho_k/(\sigma_k\tau_k)$. For development years 8 and 12 $\hat{\rho}_k$ is negative. This should not be overstated since the estimate of the covariance parameter ρ_k is based here only on seven and three observations, respectively and has no substantial contribution to the total errors due to the small ρ_k in the later development periods. ρ_k decays rapidly with respect to k, as it is usually the case for σ_k^2 and τ_k^2 (and also for f_k and g_k). Row 7 shows $\rho_k/(\sigma_k\tau_k)$ which gives the correlation coefficient of the individual development factors. It can be seen, that it is quite stable in the first seven development years.

Table 4 shows for each accident year i the estimated reserve $\hat{C}_{in} - C_{i,n+1-i}$ for GL run-off and the estimated reserve $\hat{D}_{in} - D_{i,n+1-i}$ for AL run-off and the sum of these two reserves ("Portfolio"). In the last column of Table 4 the estimated reserve is given when aggregating first both data triangles to one single triangle

TABLE 4

ESTIMATED IBNR RESERVES

Accident Year	GL run-off (A)	AL run-off (B)	Portfolio (A)+(B)	Overall Calculation
1987	0	0	**0**	0
1988	1.945	–135	**1.810**	1.988
1989	5.394	–740	**4.655**	5.117
1990	10.616	1.211	**11.827**	11.083
1991	15.220	992	**16.212**	15.344
1992	25.988	3.132	**29.120**	28.010
1993	42.133	3.661	**45.793**	44.553
1994	75.959	10.045	**86.004**	81.339
1995	135.599	21.567	**157.165**	149.553
1996	289.659	54.642	**344.301**	329.840
1997	561.237	118.575	**679.812**	644.927
1998	1.033.307	254.151	**1.287.458**	1.230.370
1999	1.887.590	565.448	**2.453.038**	2.331.408
2000	2.070.616	1.031.063	**3.101.679**	3.080.525
All years	6.155.261	2.063.612	**8.218.874**	7.954.058

and then estimating the reserve with the Chain-Ladder method. This (nonsense) calculation is only done for comparison purposes and is denoted "overall calculation" in the following and in the tables. The example shows that the overall calculation leads to another result which can be considered as unusable here since run-offs with different development patterns were added together. The reserve is about 265 Mio. lower than the one by separate calculation of the GL and AL reserves. To evaluate this difference we have to consider the variability in our estimates.

Tables 5-7 show the square roots of the random error, the estimation error and the prediction error, respectively for GL run-off in column 1 and AL run-off in column 2. The column "Portfolio" of these tables shows the corresponding figures for the whole portfolio consisting of the GL and AL subportfolios, computed with our method as described in section 4 taking into account the correlation between the individual development factors. Column 3a gives the implied average coefficent of correlation, i.e. the solution $\rho(X,Y)$ of the equation

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\rho(X,Y)\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)} \tag{33}$$

where X and Y are the reserves of the GL and AL run-off, and $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$ and $\mathrm{Var}(X+Y)$ are the squares of corresponding errors from columns (1)-(3). Columns 4 to 6 show the results of the calculation (33) but assuming a positive correlation of +1, no correlation and a negative correlation –1 between the corresponding individual development factors of all columns of the GL and AL run-off. In column 7 the roots of the errors are given for the overall calculation. The errors for the reserve for "Portfolio" of each accident year and all accident years together are between the ones assuming no correlation and a correlation equal to 1. Note that, the overall calculation yields for the accident year 1988 and 1989 errors which are larger than the corresponding error of the portfolio under the assumption of a complete positive correlation between both run-offs. This is a further hint that the overall calculation is not suited for the estimation of portfolio reserves and its range.

As discussed in subsection 2.2 we have to use the prediction error when assessing a range for the reserve of the portfolio. Assuming a log-normal distribution for the reserve a range for the reserve of all accident years of the portfolio can be calculated. For this, the mean of the distribution is set equal to the estimated reserve (see table 4) and the variance equal to the prediction error (see table 7 for the square root of the prediction error). Using the interval containing 90% probability around the mean with 45% probability on each side as range for the reserve, leads to a lower bound of 7.459.480 and an upper bound of 9.157.228. This range can be interpreted as follows. Under our model assumptions and the distribution assumption for the portfolio reserve the reserve which is finally needed for the complete development of the accident years 1987 to 2000 of the portfolio, is with 90% probability in this range. Of course, this ultimately necessary amount is not known until these accident years of the portfolio are fully developed, while this range can be computed by now.

When assessing a range for the best estimate of the reserve instead of the reserve itself, we have to use the estimation error instead of the prediction

https://doi.org/10.2143/AST.34.2.505150 Published online by Cambridge University Press

TABLE 5

SQUARE ROOT OF RANDOM ERROR

Accident Year	GL run-off (1)	AL run-off (2)	**Portfolio** (3)	Implied Portfolio Corr. (3a)	Assumed Portfolio Correlation			Overall Calculation (7)
					Corr. = 1 (4)	Corr. = 0 (5)	Corr. = –1 (6)	
1987	0	0	**0**		0	0	0	0
1988	1.224	404	**1.289**	0,000	1.628	1.289	820	2.536
1989	5.866	1.091	**6.863**	0,899	6.956	5.966	4.775	7.551
1990	6.864	2.461	**8.102**	0,369	9.325	7.292	4.403	8.453
1991	8.984	2.708	**10.428**	0,425	11.693	9.384	6.276	10.435
1992	14.204	4.750	**16.561**	0,370	18.954	14.977	9.454	16.665
1993	16.613	5.384	**19.522**	0,426	21.997	17.464	11.228	19.610
1994	19.488	6.577	**22.296**	0,289	26.064	20.567	12.911	21.971
1995	25.425	8.127	**29.240**	0,345	33.552	26.692	17.297	28.869
1996	31.823	14.609	**39.561**	0,365	46.432	35.016	17.215	40.512
1997	49.924	24.366	**63.867**	0,408	74.290	55.553	25.558	62.871
1998	78.731	33.227	**100.074**	0,518	111.959	85.456	45.504	98.979
1999	172.409	47.888	**199.643**	0,475	220.297	178.936	124.521	174.416
2000	261.006	117.293	**316.077**	0,294	378.298	286.149	143.713	304.289
All years	330.485	134.676	**397.054**	0,340	465.161	356.872	195.809	375.000

TABLE 6

SQUARE ROOT OF ESTIMATION ERROR

Accident Year	GL run-off (1)	AL run-off (2)	**Portfolio (3)**	Implied Portfolio Corr. (3a)	Assumed Portfolio Correlation			Overall Calculation (7)
					Corr. = 1 (4)	Corr. = 0 (5)	Corr. = –1 (6)	
1987	0	0	**0**		0	0	0	0
1988	1.241	449	**1.320**	-0,000	1.690	1.320	792	2.677
1989	4.436	934	**5.217**	0,805	5.370	4.533	3.502	6.119
1990	5.885	1.556	**6.701**	0,428	7.441	6.088	4.330	7.055
1991	6.656	1.708	**7.591**	0,457	8.364	6.872	4.948	7.834
1992	8.936	2.606	**10.265**	0,402	11.542	9.308	6.330	10.490
1993	10.570	3.115	**12.246**	0,433	13.685	11.019	7.455	12.538
1994	12.852	3.570	**14.506**	0,354	16.422	13.339	9.282	14.328
1995	15.129	4.144	**17.113**	0,373	19.272	15.686	10.985	16.799
1996	19.822	6.980	**23.300**	0,366	26.802	21.015	12.841	23.310
1997	28.775	11.021	**34.597**	0,390	39.797	30.814	17.754	33.519
1998	42.538	15.668	**51.888**	0,478	58.206	45.332	26.870	50.392
1999	87.209	23.624	**100.331**	0,462	110.832	90.352	63.585	87.217
2000	109.281	47.678	**131.984**	0,307	156.959	119.229	61.604	127.127
All years	270.843	91.594	**318.600**	0,398	362.437	285.911	179.249	304.841

TABLE 7

SQUARE ROOT OF PREDICTION ERROR

Accident Year	GL run-off (1)	AL run-off (2)	**Portfolio** (3)	Implied Portfolio Corr. (3a)	Assumed Portfolio Correlation Corr. = 1 (4)	Corr. = 0 (5)	Corr. = –1 (6)	Overall Calculation (7)
1987	0	0	**0**		0	0	0	0
1988	1.743	604	**1.845**	-0,000	2.347	1.845	1.139	3.688
1989	7.354	1.436	**8.621**	0,861	8.790	7.493	5.918	9.720
1990	9.042	2.912	**10.514**	0,386	11.953	9.499	6.130	11.010
1991	11.181	3.202	**12.898**	0,434	14.383	11.631	7.980	13.049
1992	16.781	5.418	**19.484**	0,378	22.199	17.634	11.363	19.692
1993	19.690	6.221	**23.045**	0,427	25.911	20.650	13.470	23.275
1994	23.344	7.483	**26.600**	0,305	30.827	24.514	15.861	26.230
1995	29.585	9.123	**33.880**	0,351	38.708	30.960	20.463	33.401
1996	37.492	16.191	**45.913**	0,363	53.682	40.838	21.301	46.739
1997	57.623	26.742	**72.636**	0,402	84.365	63.526	30.881	71.248
1998	89.488	36.736	**112.727**	0,509	126.224	96.735	52.752	111.068
1999	193.210	53.398	**223.436**	0,472	246.608	200.453	139.812	195.007
2000	282.960	126.613	**342.526**	0,296	409.573	309.995	156.347	329.777
All years	427.289	162.872	**509.075**	0,360	590.161	457.278	264.417	483.274

error. Assuming also a log-normal distribution for the best estimate of the reserve, but with mean and standard deviation according to table 4 and 6 a reasonable range for the best estimate for all accident years of our portfolio consisting of the GL and the AL runoff can be calculated. The mean is equal to the estimated reserve and the variance equal to the estimation error. We use the interval containing 50% probability around the mean with 25% probability on each side as range. This is a fair compromise between a non-informative 99%-range and the straight point estimate which would not contain the true expected reserve with 100% probability. This 50%-range leads to a lower bound of 8.008.292 and an upper bound of 8.438.171. Within this range, each amount can be taken as best estimate. This range for the best estimate of the reserve is much smaller than the range for the reserve itself as given above. The reserve estimate of the overall calculation (see table 4) is outside the range for the best estimate, since it is below the lower bound. This shows again that the overall calculation is not reasonable.

6. Final remarks

Correlations between run-off triangles are often attributed to the claims inflation affecting all or most of the segments of a portfolio in a similar way. For this reason, it may seem obvious to derive the correlation between the reserves from the correlation between the estimated inflation rates in the run-offs. But, since the inflation affects the diagonals in the run-offs, the basic Chain Ladder model assumption of independence of the accident years is violated. Therefore, calculating reserve ranges by using calendar year based correlations (Brehm (2002)) in conjunction with reserves estimated with the Chain Ladder method is inadvisable. In principle, all calendar year based dependencies should be removed from the run-offs, before the reserves are calculated with the Chain Ladder method. Since the inflation influences mainly payments and less incurred figures, applying the Chain Ladder method can be done for incurred run-offs with less problems.

Furthermore, the inflation rate of a calendar year does not affect the accident years of a run-off in the same way, since the payments are for different types of claims due to their different development periods. For instance, considering a fixed calendar year in a general liability portfolio, in earlier development years mainly property damages are paid while for later development years payments of bodily injury claims dominate. Thus, a run-off does not have a uniform calendar year inflation rate for all accident years, from which the correlation of the run-off triangles could be meaningful derived.

Our approach comes up with an individual correlation coefficent $\rho_k/(\sigma_k\tau_k)$ for each development period k. In contrast to this, some other approaches express the correlation between two run-offs by a single number, e.g. by a single overall correlation coefficient. If one likes to do this with our approach – even though it is not in line with the stochastic Chain-Ladder model which consists of own parameters f_k, σ_k for each development period k, $2 \leq k \leq n$ – one can simply set in the basic assumption (24) for the covariance in section 3

$\rho_k = \psi\sigma_k\tau_k$ with σ_k and τ_k as before and – now by using data from all development periods – estimate ψ by the weighted average of $\hat{\rho}_k/(\hat{\sigma}_k\hat{\tau}_k)$, i.e.

$$\hat{\psi} = \frac{1}{\sum_{k=2}^{n-1} v_k} \sum_{k=2}^{n-1} v_k \frac{\hat{\rho}_k}{\hat{\sigma}_k \hat{\tau}_k}.$$

with $v_k := n-k-1+w_k^2$ and $\hat{\rho}_k$, $\hat{\sigma}_k$ and $\hat{\tau}_k$ as given in sections 2 and 3. This simplified model implies a constant correlation coefficient ψ for all development years, i.e.

$$\mathrm{Corr}(F_{ik}, G_{ik} \mid T_{i,k-1}) = \psi \tag{34}$$

and using (9) and (28) yields

$$\mathrm{Corr}(\hat{f}_k, \hat{g}_k \mid T_{i,k-1}) = \psi \sum_{j=1}^{n+1-k} \frac{\sqrt{C_{j,k-1} D_{j,k-1}}}{\sqrt{C_{<,k-1} D_{<,k-1}}}.$$

The last equation shows that the correlation of $\hat{f}_k$ and $\hat{g}_k$ depends on the development period k even though $\hat{f}_k$ and $\hat{g}_k$ are weighted averages of individual development factors F_{ik} and G_{ik} (see (2)) whose correlation (34) is independent of k.

For the rest of this section we consider the case of a non-negative ψ. It results from the Cauchy-Schwarz inequality

$$\mathrm{Corr}(\hat{f}_k, \hat{g}_k \mid T_{i,k-1}) \le \psi.$$

Set $\hat{\rho}_k = \hat{\psi}\hat{\sigma}_k\hat{\tau}_k$ in the covariance estimates (27), (29), (30) and (32) of section 4. Using

$$\widehat{\mathrm{Corr}}(C_{in}, D_{in}) := \frac{\widehat{\mathrm{Cov}}(C_{in}, D_{in})}{\sqrt{\widehat{\mathrm{Var}}(C_{in})\,\widehat{\mathrm{Var}}(D_{in})}} \tag{35}$$

as an estimate for the correlation of the ultimate claims amounts C_{in} and D_{in} it can be shown via the explicit formulas for $\widehat{\mathrm{Cov}}(C_{in}, D_{in})$, $\widehat{\mathrm{Var}}(C_{in})$ (cf. Mack (2002), p. 252) and $\widehat{\mathrm{Var}}(D_{in})$ instead of the recursive formulas (6) and (27) that

$$\widehat{\mathrm{Corr}}(C_{in}, D_{in}) \le \hat{\psi}. \tag{36}$$

Defining the correlation estimates $\widehat{\mathrm{Corr}}(\hat{C}_{in}, \hat{D}_{in})$, $\widehat{\mathrm{Corr}}\left(\sum_{i=2}^{n} C_{in}, \sum_{i=2}^{n} D_{in}\right)$ and $\widehat{\mathrm{Corr}}\left(\sum_{i=2}^{n} \hat{C}_{in}, \sum_{i=2}^{n} \hat{D}_{in}\right)$ analogously to (35), it can also be shown

$$\widehat{\mathrm{Corr}}(\hat{C}_{in}, \hat{D}_{in}) \le \hat{\psi}, \tag{37}$$

$$\widehat{\mathrm{Corr}}\left(\sum_{i=2}^{n} C_{in}, \sum_{i=2}^{n} D_{in}\right) \le \hat{\psi}, \tag{38}$$

$$\widehat{\mathrm{Corr}}\left(\sum_{i=2}^{n}\hat{C}_{in}, \sum_{i=2}^{n}\hat{D}_{in}\right) \leq \hat{\psi}. \tag{39}$$

The estimated correlations (36) and (37) depend on the accident year i and the correlations (36)-(39) are different in general, but uniformly bounded by $\hat{\psi}$. (37) shows, the correlation of the developments of run-offs is underestimated by using the correlation of the ultimate estimates.

It can be easily seen by using the definition (33) and the identity

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + 2\mathrm{Cov}(X,Y) + \mathrm{Var}(Y)$$

that the correlation estimates on the left hand side of the inequalities (36)-(39) are the implied coefficient of correlations for the considered random variables, e.g.

$$\widehat{\mathrm{Corr}}(C_{in}, D_{in}) = \rho(C_{in}, D_{in}).$$

Furthermore, calculating the implied coefficient of correlation $\rho(\hat{C}_{in}, \hat{D}_{in})$ for the prediction error $\mathrm{mse}(\hat{C}_{in} + \hat{D}_{in})$ it can also be shown that it is different from $\hat{\psi}$ generally and

$$\rho(\hat{C}_{in}, \hat{D}_{in}) \leq \hat{\psi}$$

indicating that our estimated correlation coefficient $\hat{\psi}$ is at least as high as the implied average one. This holds not only for each accident year i, but also for all accident years together. To summarize, the implied coefficient of correlation underestimates the correlation of the run-offs, independent of whether it is calculated for the random error, the estimation error or the prediction error and whether it is calculated for a single accident year or for all accident years together.

Acknowledgements

The author would like to thank Dr. Thomas Mack for valuable comments and suggestions to improve the manuscript and Dr. Gerhard Quarg for clarifying discussions regarding conditional expectations.

References

Ajne, B. (1994) Additivity of Chain-Ladder Projections. *ASTIN Bulletin* **24**, 311-318.

Brehm, P. (2002) Correlation and the Aggregation of Unpaid Loss Distributions, *CAS Forum Fall 2002*, 1-24.

Houltram, A. (2003) Reserving Judgement – Considerations Relevant to Insurance Liability Assessment Under GPS210, XIV General Insurance Seminar, Institute of Actuaries of Australia.

Kirschner, G., Kerley, C. and Isaacs, B. (2002) Two Approaches to Calculating Correlated Reserve Indications Across Multiple Lines of Business, *CAS Forum Fall 2002*, 211-246.

MACK, T. (1993) Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates. *ASTIN Bulletin* **23**, 213-225.

MACK, T. (2002) Schadenversicherungsmathematik, 2nd Edition, Verlag Versicherungswirtschaft e.V., Karlsruhe.

MACK, T. (1999) The Standard Error of Chain Ladder Reserve Estimates: Recursive Calculation and Inclusion of a Tail Factor. *ASTIN Bulletin* **29**, 361-366.

REINSURANCE ASSOCIATION OF AMERICA (2001) *Historical Loss Development Study*, Washington.

DR. CHRISTIAN BRAUN
Munich Reinsurance Company
80791 Munich
Germany
E-mail: cbraun@munichre.com

https://doi.org/10.2143/AST.34.2.505150 Published online by Cambridge University Press