A QUANTITATIVE STUDY OF CHAIN LADDER BASED PRICING APPROACHES FOR LONG-TAIL QUOTA SHARES

BY

ULRICH RIEGEL

ABSTRACT

Pricing approaches for long-tail quota shares are often based on the chain ladder method. Apart from IBNR calculation, common pricing methods require volume measures for accident years in the observation period, and for the quotation period. In practice, in most cases restated premiums are used as the volume measures. The prediction error of the chain ladder method is an important part of the prediction uncertainty of these pricing approaches. There are, however, two sources of uncertainty that are not addressed by the chain ladder model: the stochastic volatility of the claims in the first development year; and the *restatement uncertainty*, the risk that the restated premium is not a good volume measure. We extend Mack's chain ladder model to cover these two sources of uncertainty, and calculate the mean-squared error of chain ladder pricing approaches with arbitrary weights for the accident years in the observation period. Then we focus on the problem of finding optimal weights for the accident years. First, we assume that the parameters for restatement uncertainty are given, and provide recursion formulas to calculate approximately-optimal weights. Second, we describe a maximum likelihood approach that can be used to estimate the restatement uncertainty.

KEYWORDS

Chain ladder, reinsurance pricing, quota share, mean squared error.

1. INTRODUCTION

Despite a trend towards non-proportional reinsurance, long-tail quota shares remain extremely important for reinsurers. For instance, in 2012, German insurers ceded more than EUR 3 billion motor premium on a pro rata basis (see Haas (2013)). In view of large volumes and relatively small margins, pricing accuracy is fundamental in this segment.

Long-tail quota shares often have risk-mitigating features, such as sliding scale commissions or profit commissions. In these cases, a point estimate for

Astin Bulletin 45(2), 267-307. doi: 10.1017/asb.2015.2

the expected loss ratio is not sufficient for pricing — a stochastic loss model is required. Practitioners often use the method of moments to fit an aggregate loss model. In this context, it is very useful to have a formula for the standard error of the selected pricing approach, since it can be used to estimate the standard deviation of an aggregate distribution that includes parameter uncertainty.

There are two very common but rather different approaches for pricing long-tail quota shares: the *average loss ratio (ALR) method*, and the *roll forward (RF) method*. Both methods are based on a chain ladder calculation for the claims triangle.

The ALR method requires trending of losses and restatement of premiums in order to make figures comparable across years. The average ultimate loss ratio of the observation period, as predicted by the chain ladder, is used to forecast the ultimate loss ratio of the quotation year.

The RF method relies on the chain ladder prediction for the ultimate loss ratio of the most recent accident year of the observation period. RF parameters for the premium and expected loss are used to reflect the expected changes from the last accident year in the observation period to the quotation year.

The ALR method should be preferred for segments with large stochastic volatility (e.g. quota shares covering industrial liability). On the other hand, the RF method is the standard method for segments with low stochastic volatility of the claims burden, and relatively high uncertainty of restated premiums (e.g. large motor quota shares, where the loss ratios are mainly driven by the premium cycle).

The ALR and RF methods are two extreme cases. The ALR method employs the average over all accident years in the observation period, whereas the RF method only uses the chain ladder ultimate of the last accident year. An obvious first approach to generalization of both methods is to allow arbitrary weights for the accident years in the observation period. In practice, these weights are normally used to average the ultimate loss ratios resulting from the chain ladder calculation. There is, however, a second approach that should be considered: using the weights to average the loss ratios after the first development year, and then multiplying the resulting loss ratio by the product of the development factors from the chain ladder calculation (see Mack (1993)). This second approach is less intuitive, but we will see that it should be preferred since it employs estimators with smaller (conditional) variances.

These pricing methods are suited mainly for attritional losses, i.e. it makes sense to remove large losses before applying the methods. An approach to separate large and attritional losses in a consistent way has been introduced by Riegel (2014).

The standard chain ladder model describes the stochastic claims development, provided that the claims burden of the first development year is given. For predictions using the discussed pricing approaches, there are additional sources of uncertainty that need to be addressed. First is the stochastic volatility of the claims in the first development year. We extend the chain ladder model by assumptions for the first development year, and require that the claims burden has a compound Poisson distribution. The second source of uncertainty is that the restated premiums, used as estimators for the volume measures, deviate from the correct volume measures. We call this effect *restatement uncertainty* and handle it with a simple time series model. These extensions of the chain ladder model allow calculation of and comparison between the mean squared errors of the different pricing approaches, and estimation of approximately-optimal weights for the accident years in the observation period.

In Section 2, we introduce the mentioned extensions of Mack's chain ladder model. We assume that all data have been trended to the quotation year, and use an axiomatic definition of restatement uncertainty, to keep the model as simple as possible. In Section 3, we discuss certain aspects of claims trending and restatement uncertainty to provide a practical motivation for the corresponding model assumption. In Section 4, we define the pricing approaches in detail and apply them to an example based on data from a German motor quota share. The example is used in each of the subsequent sections to illustrate results. Section 5 is dedicated to the calculation of the mean squared errors of the considered pricing methods. Assuming that the parameters for restatement uncertainty are known, or at least estimated using expert judgement, we derive a recursion that allows calculation of approximately-optimal weights for the accident years in Section 6. In Section 7, we use a maximum likelihood approach to estimate the parameters for restatement uncertainty.

An Excel implementation of the discussed pricing methods, the calculation of the standard errors and the recursions for optimal weights is provided by the author on www.researchgate.net.

2. Stochastic model

In this section, we extend Mack's chain ladder model in such a way that it can be used to calculate the prediction error of the most common chain ladder based pricing approaches for long tail quota shares.

We consider the typical situation of a pricing actuary. Assume that a claims triangle is available for an observation period consisting of the accident years $i = 1, \ldots, n$. The task is to predict the ultimate loss ratio of the *quotation year* $q := n + 2$. For $i = 1, \ldots, q$ and $j = 1, \ldots, n$ let $C_{i,j}$ denote the cumulative claims of accident year i after j development years (payments or incurred amounts). We assume that all claims data have been trended to the calendar year q, i.e. that inflation has been removed. More details on trending are found in Section 3.

The observable claims data at the end of calendar year n (which is the data available to the pricing actuary) is described by the σ-algebra

$$\mathcal{D} := \sigma\{C_{i,j} \,|\, i + j \leq n + 1\}.$$

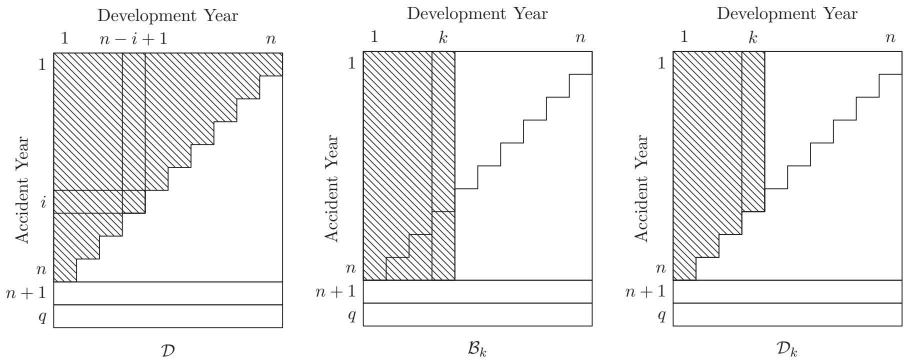

FIGURE 1: Illustration of the used σ-algebras.

Moreover, let $\mathcal{B}_k$ be the claims information given after k development years of each accident year, i.e.

$$\mathcal{B}_k := \sigma\{C_{i,j} \mid 1 \leq i \leq q,\ j \leq k\},$$

and let

$$\mathcal{D}_k := \mathcal{D} \cap \mathcal{B}_k.$$

The σ-algebras are illustrated in Figure 1.

In addition to the claims information, we assume that restated premiums $\widehat{v}_i$ have been calculated for the accident years $i = 1, \ldots, q$. The first three model assumptions combine the common chain ladder model with additive assumptions for the first development year.

Model Assumption 1. *The accident years are independent.*

Model Assumption 2. *There exist volume measures $v_1, \ldots, v_q$ and factors $f_0, \ldots, f_{n-1} \geq 0$ such that*

$$\mathrm{E}(C_{i,1}) = f_0 v_i \quad \textit{and} \quad \mathrm{E}(C_{i,j+1} \mid \mathcal{B}_j) = f_j C_{i,j},$$

for $i = 1, \ldots, q$ and $j = 1, \ldots, n-1$.

Model Assumption 3. *There are $\sigma_1^2, \ldots, \sigma_{n-1}^2 > 0$ such that*

$$\mathrm{Var}(C_{i,j+1} \mid \mathcal{B}_j) = \sigma_j^2 C_{i,j},$$

for $j \geq 1$. Moreover, there exist $x > 0$ and $c > 0$ such that $C_{1,1}, \ldots, C_{q,1}$ can be represented as collective risk models

$$C_{i,1} = \sum_{k=1}^{N_i} X_{i,k},$$

(i.e. $X_{i,1}, X_{i,2}, \ldots$ i.i.d. and independent of N_i) with Poisson distributed claims count N_i and $\mathrm{E}(X_{i,k}) = x$, $\mathrm{CV}(X_{i,k}) = c$.

Here, $\mathrm{CV}(X)$ denotes the coefficient of variation of a random variable X. An introduction to collective risk models is found in Klugmann *et al.* (2004). Eventually, we add a supplementary model assumption for restatement uncertainty. The practical background for this assumption is discussed in the next section.

Model Assumption 4. *The volume measure v_q equals $\widehat{v}_q$. The unknown volume measures $v_1, \ldots, v_{n+1}$ are estimated by the restated premiums $\widehat{v}_1, \ldots, \widehat{v}_{n+1}$ such that*

$$\widehat{v}_i = v_i(1 + e_i + \cdots + e_{n+1}),$$

with independent and normally distributed random errors $e_i \sim N(0, \varepsilon_i^2)$. Moreover, the random vector $(e_1, \ldots, e_{n+1})$ and the claims development $\sigma\{C_{i,j} \mid 1 \leq i \leq q,\ 1 \leq j \leq n\}$ are independent.

Remark 2.1. *Since the normal distribution is not bounded from below, Model Assumption 4 can yield negative $\widehat{v}_i$ in extreme cases. In practice, restated premiums are always positive of course. If the model is used for simulation, paths with $\widehat{v}_i$ below a certain threshold, say $0.05 \cdot v_i$, should therefore be ignored. Correspondingly, we will truncate the distribution of $\widehat{v}_i$ for the calculation of the prediction uncertainty (cf. Appendix C).*

Remark 2.2. *From Model Assumption 3, Wald's equation and the Blackwell–Girshick equation we have*

$$\mathrm{E}(C_{i,1}) = \mathrm{E}(N_i)\,\mathrm{E}(X_{i,1}) = \mathrm{E}(N_i)x,$$

$$\mathrm{Var}(C_{i,1}) = \mathrm{Var}(N_i)\,\mathrm{E}(X_{i,1})^2 + \mathrm{E}(N_i)\,\mathrm{Var}(X_{i,1}) = \mathrm{E}(N_i)(1 + c^2)x^2.$$

With $\sigma_0^2 := f_0(1 + c^2)x$ we obtain

$$\mathrm{Var}(C_{i,1}) = \mathrm{E}(C_{i,1})(1 + c^2)x = \sigma_0^2 v_i.$$

3. Comments on trending and restatement uncertainty

In the last section, we have assumed that claims data have been trended and that premiums have been restated to the calendar year q. In order to provide a practical background for Model Assumption 4, we describe the relevant aspects of the methods that are typically used in this respect. Restating premiums (also called *trending* or *on-leveling of premiums*) can be very complex. For a detailed discussion of this topic cf. Jones (2002).

Let $p_1^*, \ldots, p_n^*$ be the (untrended) premiums of the accident years in the observation period and let p_{n+1}^* and p_q^* be the estimated premiums for the accident years $n+1$ and $q = n+2$. Let $C_{i,j}^*$ denote the cumulative claims before trending and let I_i^C be the expected individual claims size in $C_{i,1}^*$. There are a number

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

of different methods for trending the claims $C^*_{i,j}$ to the calendar year q. The most popular ones are *calendar year trending* and *accident year trending*. We do not go into detail here, but the usual methods have in common that the claims $C^*_{i,1}$ of the first development year should be multiplied by I^C_q / I^C_i to obtain the trended claims $C_{i,1}$. Of course, the index I^C_i is not known. The pricing actuary uses estimates $\widehat{I}^C_i$ and calculates the trended claims $C_{i,1}$ as

$$C_{i,1} := \frac{\widehat{I}^C_q}{\widehat{I}^C_i} \cdot C^*_{i,1}.$$

Regarding the premiums, we proceed in two steps. First, we trend the premiums p^*_i in the same way as the claims $C^*_{i,1}$, i.e.

$$p_i := \frac{\widehat{I}^C_q}{\widehat{I}^C_i} \cdot p^*_i.$$

In the following, we assume that claims trending works perfectly. In particular we assume that $\widehat{I}^C_i = I^C_i$ and that the trended claims are inflation-free. This idealization is a common requirement for most IBNR methods (in particular for chain ladder). Note that this assumption is not more critical in our context than for normal chain ladder calculations since $C_{i,1}/p_i = C^*_{i,1}/p^*_i$ holds even if $\widehat{I}^C_i$ deviates from I^C_i.

In a second step, we restate the trended premiums p_i in order to correct for effects that influence the premium quality and thus the expected loss ratio. Examples are tariff changes, frequency trends and portfolio changes. We define the premium quality index I^Q_i by

$$I^Q_i := \frac{p_i}{\mathrm{E}(C_{i,1})}.$$

With

$$f_0 := \frac{E(C_{q,1})}{p_q} \quad \text{and} \quad v_i := \frac{I^Q_q}{I^Q_i} \cdot p_i,$$

we then have $v_q = p_q = p^*_q$ and

$$\mathrm{E}(C_{i,1}) = f_0 \cdot v_i,$$

for all i. The premium quality index I^Q_i is not known, of course. The pricing actuary has to rely on estimates $\widehat{I}^Q_i$ which are typically derived from market data and additional individual information regarding tariff and portfolio changes. Using this estimated index, we obtain the restated premiums

$$\breve{v}_i := \frac{\widehat{I}^Q_q}{\widehat{I}^Q_i} \cdot p_i,$$

that are used as estimators for the volume measures v_i. Assume that the annual changes of the premium quality index can only be estimated up to random errors e_i with $\mathrm{E}(e_i) = 0$, i.e.

$$\frac{\widehat{I}^Q_{i+1}}{\widehat{I}^Q_i} = \frac{I^Q_{i+1}}{I^Q_i}(1 + e_i).$$

For $i = 1, \ldots, n+1$ we then have

$$\breve{v}_i = \frac{\widehat{I}^Q_q}{\widehat{I}^Q_i} \cdot p_i = \frac{I^Q_q}{I^Q_i}(1 + e_i) \cdots (1 + e_{n+1}) p_i = v_i (1 + e_i) \cdots (1 + e_{n+1}).$$

Neglecting the error terms of higher order, we obtain

$$\breve{v}_i \approx v_i (1 + e_i + \cdots + e_{n+1}) =: \widehat{v}_i. \qquad (*)$$

With the additional assumption that $e_1, \ldots, e_{n+1}$ are independent and normally distributed with standard deviations $\varepsilon_1, \ldots, \varepsilon_{n+1}$ these considerations lead quite naturally to Model Assumption 4.

Note that $\mathrm{E}(\breve{v}_i) = \mathrm{E}(\widehat{v}_i) = v_i$. For the parameters used in practice, $\breve{v}_i$ and $\widehat{v}_i$ have similar distributions and the random vectors $(\breve{v}_1, \ldots, \breve{v}_{n+1})^t$ and $(\widehat{v}_1, \ldots, \widehat{v}_{n+1})^t$ have very similar covariance matrices (see Appendix A). Therefore, the approximation $(*)$ makes sense in our context. Figure 2 shows the realizations of $\breve{v}_1/v_1, \ldots, \breve{v}_{n+1}/v_{n+1}$ and $\widehat{v}_1/v_1, \ldots, \widehat{v}_{n+1}/v_{n+1}$ for 100 simulations of $e_1, \ldots, e_{n+1}$ with $n = 10$ and $\varepsilon_1 = \cdots = \varepsilon_{n+1} = 3\%$. The realizations from each simulation have been connected to a path. Apart from a few extreme cases (where $|\widehat{v}_1/v_1 - 1| > 0.2$) the paths for $\breve{v}_i$ and $\widehat{v}_i$ are quite similar.

In the next section, we introduce methods that provide predictions $\widehat{C}_{q,j}$ for the cumulative claims $C_{q,j}$ of the quotation year q. Note that the prediction $\widehat{C}_{q,j}$ has the monetary value of the calendar year q and has to be projected to the calendar year $q + j - 1$.

4. Chain ladder based pricing approaches for long-tail quota shares

In this section, we describe the pricing approaches in more detail. All of the considered pricing methods are based on a chain ladder calculation for the triangle $(C_{i,j})_{i+j \le n+1}$. For $j = 1, \ldots, n-1$ we use the $\mathcal{D}_j$-conditionally unbiased estimators

$$\widehat{f}_j := \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}},$$

for f_j. For $i = 1, \ldots, n$ and $j \le n - i + 1$ let $\widehat{C}_{i,j} := C_{i,j}$ and for $j = n - i + 2, \ldots, n$

$$\widehat{C}_{i,j} := C_{i,n-i+1} \widehat{f}_{n-i+1} \ldots \widehat{f}_{j-1}.$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

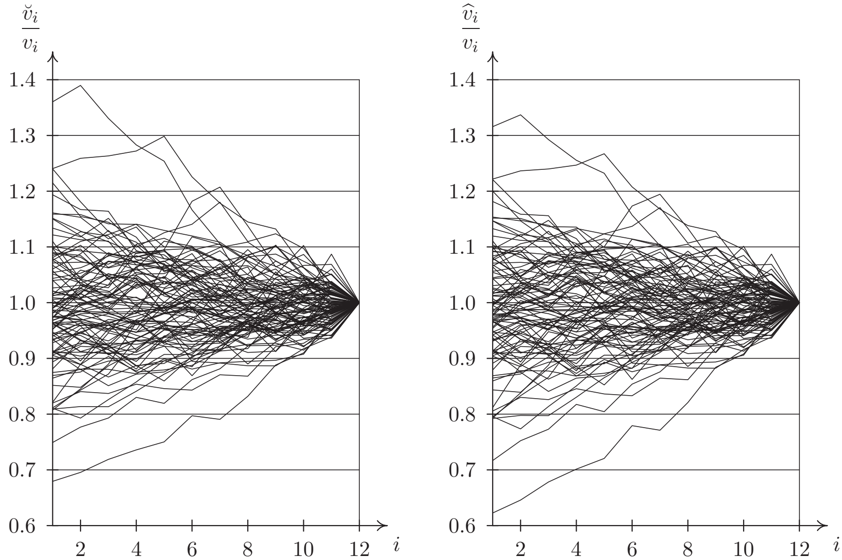

FIGURE 2: 100 simulated paths for $\breve{v}_i/v_i$ and $\widehat{v}_i/v_i$ in the case $n = 10$ and $\varepsilon_1 = \cdots = \varepsilon_{n+1} = 3\%$.

The ALR method uses the predictions

$$\widehat{C}^{\text{ALR}}_{q,j} := \frac{v_q}{\sum_{i=1}^n \widehat{v}_i} \sum_{i=1}^n \widehat{C}_{i,j},$$

i.e. the predicted loss ratio $\widehat{C}^{\text{ALR}}_{q,j}/v_q$ is a weighted average of the restated loss ratios $\widehat{C}_{i,j}/\widehat{v}_i$, where the restated premium $\widehat{v}_i$ is used as weight for the accident year i. The calculation of the predicted ultimate loss ratio $\widehat{C}^{\text{ALR}}_{q,n}/v_q$ is illustrated in Figure 3.

The RF method predicts the cumulative claims burdens $C_{q,j}$ by

$$\widehat{C}^{\text{RF}}_{q,j} := \frac{v_q}{\widehat{v}_n} \widehat{C}_{n,j},$$

i.e. the loss ratio $\widehat{C}_{n,j}/\widehat{v}_n$ of the last accident year in the observation period is used to estimate the corresponding loss ratio $\widehat{C}^{\text{RF}}_{q,j}/v_q$ of the quotation year. An illustration of the RF method is found in Figure 4

The ALR and the RF method are two extreme cases: the ALR method relies on the volume weighted ALR of the observation period, whereas the RF method only uses the last accident year of the observation period. These two approaches can be generalized by using arbitrary weights for the accident years in the observation period.

Let $\mathbf{w} = (w_1, \ldots, w_n)^t$ be a vector with non-negative components and $\sum_{i=1}^n w_i > 0$. The most popular and intuitive way to generalize the ALR and

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

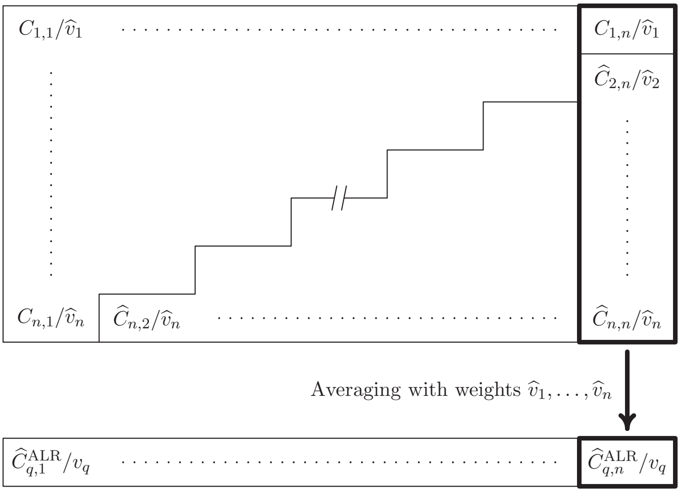

FIGURE 3: Illustration of the ALR method.

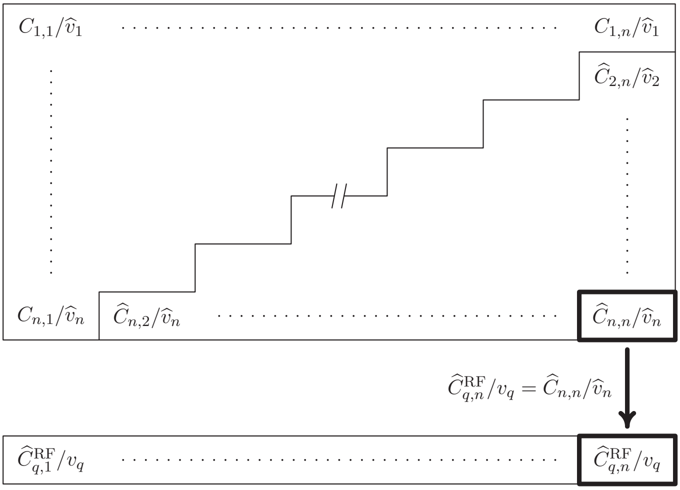

FIGURE 4: Illustration of the RF method.

RF methods is to predict the loss ratio $C_{q,j}/v_q$ by the weighted average of the loss ratios $\widehat{C}_{i,j}/\widehat{v}_i$ from chain ladder calculation, where the weight $w_i\widehat{v}_i$ is used for accident year i. Then the cumulative claims burden $C_{q,j}$ is predicted by

$$\widehat{C}^{u,\mathbf{w}}_{q,j} := \frac{v_q}{\sum_{i=1}^n w_i \widehat{v}_i} \sum_{i=1}^n w_i \widehat{C}_{i,j}.$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

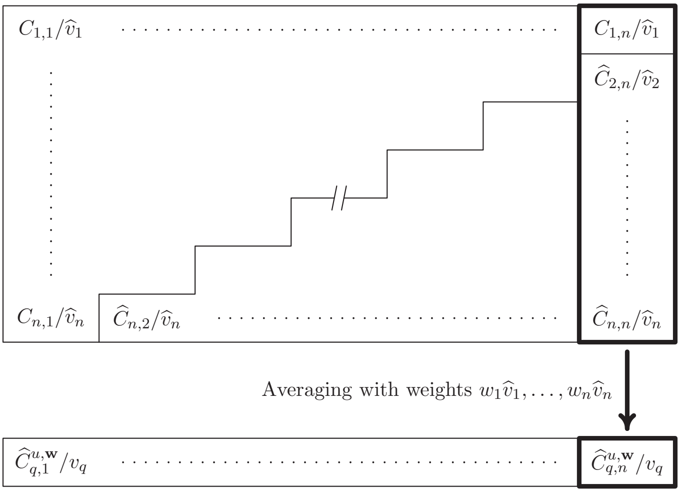

FIGURE 5: Calculation of $\widehat{C}^{u,\mathbf{w}}_{q,n}$.

We use the superscript "u" in the notation to indicate that the predictor $\widehat{C}^{u,\mathbf{w}}_{q,n}/v_q$ is obtained by averaging over the *ultimate* loss ratios $\widehat{C}_{i,n}/\widehat{v}_i$ of the chain ladder calculation. Figure 5 provides an illustration of this method.

There is, however, a second way to use arbitrary weights. We consider the estimator

$$\widehat{f}_0^{\mathbf{w}} := \frac{\sum_{i=1}^n w_i C_{i,1}}{\sum_{i=1}^n w_i \widehat{v}_i},$$

for the factor f_0. Note that $\widehat{f}_0^{\mathbf{w}}$ is the ALR after the first development year, where $w_i\widehat{v}_i$ is used as weight for accident year i. Then we define alternative predictors

$$\widehat{C}^{f,\mathbf{w}}_{q,j} := v_q \widehat{f}_0^{\mathbf{w}} \widehat{f}_1 \cdots \widehat{f}_{j-1},$$

for $j = 1, \ldots, n$. In this case, we use the superscript "f" to indicate that the predictor $\widehat{C}^{f,\mathbf{w}}_{q,n}/v_q$ is obtained by averaging over the loss ratios $\widehat{C}_{i,1}/\widehat{v}_i$ of the *first* development year before the factor $\widehat{f}_1 \cdots \widehat{f}_{n-1}$ is applied. Figure 6 illustrates the calculation of $\widehat{C}^{f,\mathbf{w}}_{q,n}$.

Note that the ALR method and the RF method are really special cases of these approaches: If $\mathbf{w} = \mathbf{w}^{\mathrm{ALR}} := \mathbb{1}_n := (1, \ldots, 1)^t$ then we have

$$\widehat{C}^{\mathrm{ALR}}_{q,j} = \widehat{C}^{f,\mathbf{w}}_{q,j} = \widehat{C}^{u,\mathbf{w}}_{q,j}.$$

In the case $\mathbf{w} = \mathbf{w}^{\mathrm{RF}} := (0, \ldots, 0, 1)^t$ we have

$$\widehat{C}^{\mathrm{RF}}_{q,j} = \widehat{C}^{f,\mathbf{w}}_{q,j} = \widehat{C}^{u,\mathbf{w}}_{q,j}.$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

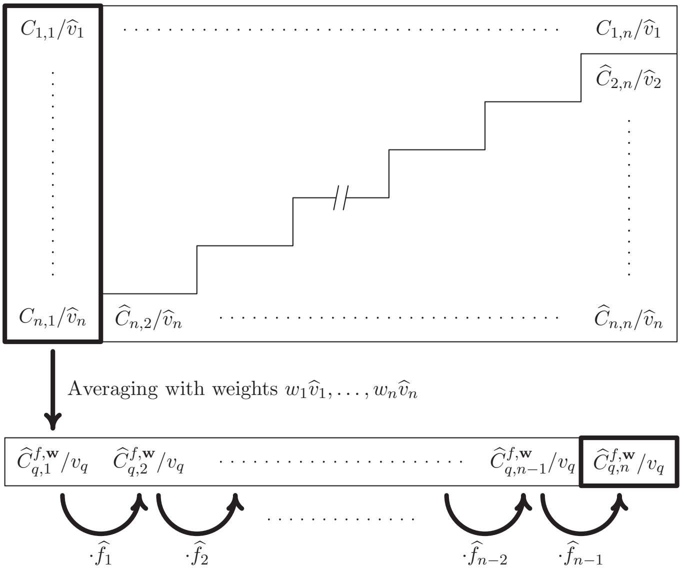

FIGURE 6: Calculation of $\widehat{C}^{f,\mathbf{w}}_{q,n}$.

Remark 4.1. *The predictions $\widehat{C}^{u,\mathbf{w}}_{q,k}$ can alternatively be written as*

$$\widehat{C}^{u,\mathbf{w}}_{q,k} = v_q \widehat{f}^{\mathbf{w}}_0 \widehat{f}^{\mathbf{w}}_1 \cdots \widehat{f}^{\mathbf{w}}_{k-1},$$

with the $\mathcal{D}_j$-conditionally unbiased estimators

$$\widehat{f}^{\mathbf{w}}_j := \frac{\sum_{i=1}^n w_i \widehat{C}_{i,j+1}}{\sum_{i=1}^n w_i \widehat{C}_{i,j}},$$

for f_j. A short calculation shows that

$$\widehat{f}^{\mathbf{w}}_j = \sum_{i=1}^{n-j} \frac{\lambda^{\mathbf{w}}_{i,j}}{\sum_{k=1}^{n-j} \lambda^{\mathbf{w}}_{k,j}} \cdot \frac{C_{i,j+1}}{C_{i,j}} \quad \text{with} \quad \lambda^{\mathbf{w}}_{i,j} := \left(w_i + \frac{\sum_{k=n-j+1}^n w_k \widehat{C}_{k,j}}{\sum_{k=1}^{n-j} C_{k,j}} \right) C_{i,j},$$

i.e. $\widehat{f}^{\mathbf{w}}_j$ is a convex combination of the $\mathcal{D}_j$-conditionally unbiased estimators $C_{i,j+1}/C_{i,j}$. It is well known that conditionally, given $\mathcal{D}_j$, $\widehat{f}_j$ has the smallest variance amongst all convex combinations of $C_{i,j+1}/C_{i,j}$ (see Mack (2002)). Therefore, we have

$$\mathrm{Var}(\widehat{f}^{\mathbf{w}}_j \mid \mathcal{D}_j) \geq \mathrm{Var}(\widehat{f}_j \mid \mathcal{D}_j),$$

for $j = 1, \ldots, n-1$. Based on this observation it is possible to show that $\widehat{C}^{f,\mathbf{w}}_{q,n}$ should be preferred to $\widehat{C}^{u,\mathbf{w}}_{q,n}$ if the Conditional Resampling Approach (cf. Wüthrich and Merz (2008)) is used to compare the prediction errors. A proof is available on request from the author. At first sight, this is surprising since the predictor $\widehat{C}^{f,\mathbf{w}}_{q,n}$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

is not based on the last known diagonal of the triangle. Here, we see a fundamental difference between pricing and reserving. In the chain ladder model the last diagonal is crucial for reserving since the reserve $\mathrm{E}(C_{i,n} - C_{i,n-i+1} \,|\, \mathcal{D}) = (f_{n-i+1} \dots f_{n-1} - 1)C_{i,n-i+1}$ *is a linear function of* $C_{i,n-i+1}$*. On the contrary, in pricing we want to predict the random variable* $C_{q,n}$ *which is independent of the last diagonal* $C_{1,n}, \dots, C_{n,1}$.

Although $\widehat{C}_{q,n}^{f,\mathbf{w}}$ *should be preferred, we calculate the standard error for both predictions in the next section since most actuaries intuitively rather use* $\widehat{C}_{q,n}^{u,\mathbf{w}}$ *in practice.*

For $j = 1, \dots, n-2$ we have the usual estimators

$$\widehat{\sigma}_j^2 := \frac{1}{n-j-1} \sum_{i=1}^{n-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \widehat{f}_j \right)^2,$$

(see Mack (1993)) and for $j = n-1$ we use the extrapolation

$$\widehat{\sigma}_{n-1}^2 := \min\left(\widehat{\sigma}_{n-2}^4 / \widehat{\sigma}_{n-3}^2, \widehat{\sigma}_{n-2}^2\right).$$

The average claims size x in the first development year and the corresponding coefficient of variation c cannot be estimated from the data described above. In segments like motor third party liability or motor hull, x and c are typically very similar for all portfolios in a market, i.e. estimators $\widehat{x}$ and $\widehat{c}$ can be obtained from individual claims information of an arbitrary portfolio. In other segments, where the market is less homogeneous, individual claims data for the respective portfolio is needed to estimate x and c. Given estimators $\widehat{x}$ and $\widehat{c}$ we use the estimator

$$\left(\widehat{\sigma}_0^{\mathbf{w}}\right)^2 := \widehat{f}_0^{\mathbf{w}}(1 + \widehat{c}^2)\widehat{x},$$

for σ_0^2.

Example 4.2. We will now apply the discussed methods to a numerical example which is based on a German MTPL portfolio. The observation period consists of $n = 8$ accident years and we predict the loss ratio of the quotation year $q = 10$. Table 1 contains the restated premiums and cumulative payments. All numbers are in 1,000 EUR. The result of the chain ladder calculation is provided in Table 2.

We apply the methods with the following vectors $\mathbf{w}$:

$$\mathbf{w}^{\mathrm{ALR}} = \begin{pmatrix} 1\\1\\1\\1\\1\\1\\1\\1 \end{pmatrix}, \quad \mathbf{w}^{\mathrm{RF}} = \begin{pmatrix} 0\\0\\0\\0\\0\\0\\0\\1 \end{pmatrix}, \quad \mathbf{w}^{4\mathrm{y}} := \begin{pmatrix} 0\\0\\0\\0\\1\\1\\1\\1 \end{pmatrix}, \quad \mathbf{w}^{\mathrm{inc}} := \begin{pmatrix} 0.125\\0.250\\0.375\\0.500\\0.625\\0.750\\0.875\\1.000 \end{pmatrix}.$$

TABLE 1

RESTATED PREMIUMS $\widehat{v}_i$ AND CLAIMS TRIANGLE $(C_{i,j})_{i+j\leq n+1}$.

i	$\widehat{v}_i$	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$	$C_{i,4}$	$C_{i,5}$	$C_{i,6}$	$C_{i,7}$	$C_{i,8}$
1	107,309	54,547	74,053	76,573	78,825	80,118	81,419	82,250	82,905
2	114,101	56,961	76,587	79,144	82,421	83,196	84,295	85,203	
3	115,849	56,885	73,400	74,865	76,723	77,543	77,737		
4	116,540	57,588	72,977	75,295	77,484	79,074			
5	114,632	55,629	69,854	72,425	73,274				
6	120,303	58,679	80,071	84,169					
7	117,410	61,948	80,254						
8	116,293	60,105							
9	117,000								
10	118,000								

TABLE 2

LOSS RATIOS $\widehat{C}_{i,j}/\widehat{v}_i$ RESULTING FROM THE CHAIN LADDER CALCULATION.

	j							
i	1	2	3	4	5	6	7	8
1	50.8%	69.0%	71.4%	73.5%	74.7%	75.9%	76.6%	77.3%
2	49.9%	67.1%	69.4%	72.2%	72.9%	73.9%	74.7%	75.3%
3	49.1%	63.4%	64.6%	66.2%	66.9%	67.1%	67.8%	68.3%
4	49.4%	62.6%	64.6%	66.5%	67.9%	68.6%	69.3%	69.9%
5	48.5%	60.9%	63.2%	63.9%	64.8%	65.5%	66.2%	66.7%
6	48.8%	66.6%	70.0%	71.9%	72.9%	73.7%	74.5%	75.1%
7	52.8%	68.4%	70.7%	72.7%	73.7%	74.5%	75.3%	75.9%
8	51.7%	67.7%	70.1%	72.0%	73.0%	73.8%	74.6%	75.2%

The vectors $\mathbf{w}^{\text{ALR}}$ and $\mathbf{w}^{\text{RF}}$ correspond to the ALR and RF method, respectively. With $\mathbf{w}^{4\text{y}}$ we average over the last four accident years and with $\mathbf{w}^{\text{inc}}$ we apply increasing weights for the accident years. The predicted loss ratios $\widehat{C}^{f,\mathbf{w}}_{q,j}/v_q$ and $\widehat{C}^{u,\mathbf{w}}_{q,j}/v_q$ are provided in Tables 3 and 4. We observe that $\widehat{C}^{f,\mathbf{w}}_{q,j} = \widehat{C}^{u,\mathbf{w}}_{q,j}$ for $\mathbf{w} = \mathbf{w}^{\text{ALR}}$ and $\mathbf{w} = \mathbf{w}^{\text{RF}}$. For $\mathbf{w} = \mathbf{w}^{4\text{y}}$ and $\mathbf{w} = \mathbf{w}^{\text{inc}}$ we have $\widehat{C}^{f,\mathbf{w}}_{q,j} \neq \widehat{C}^{u,\mathbf{w}}_{q,j}$ for $j \geq 2$.

5. CALCULATION OF THE MEAN SQUARED ERROR

Let $\widehat{C}^{*,\mathbf{w}}_{q,j}$ denote one of the predictions $\widehat{C}^{f,\mathbf{w}}_{q,j}$ and $\widehat{C}^{u,\mathbf{w}}_{q,j}$. Let $\mathcal{D}^{\mathbf{v}}$ be the σ-algebra generated by $\mathcal{D}$ and $\widehat{v}_1, \ldots, \widehat{v}_{n+1}$. In this section we study the mean squared error

$$\operatorname{mse}(\widehat{C}^{*,\mathbf{w}}_{q,n}) := \mathrm{E}\left[\left(C_{q,n} - \widehat{C}^{*,\mathbf{w}}_{q,n}\right)^2 \mid \mathcal{D}^{\mathbf{v}}\right],$$

TABLE 3

PREDICTED LOSS RATIOS $\widehat{C}_{q,i}^{f,\mathbf{w}}/v_q$.

$\mathbf{w}$	i = 1	2	3	4	5	6	7	8
$\mathbf{w}^{\mathrm{ALR}}$	50.12%	65.69%	67.97%	69.85%	70.84%	71.60%	72.35%	72.93%
$\mathbf{w}^{\mathrm{RF}}$	51.68%	67.74%	70.09%	72.03%	73.05%	73.83%	74.61%	75.20%
$\mathbf{w}^{4\mathrm{y}}$	50.44%	66.10%	68.40%	70.29%	71.28%	72.05%	72.81%	73.39%
$\mathbf{w}^{\mathrm{inc}}$	50.38%	66.03%	68.33%	70.21%	71.21%	71.97%	72.73%	73.31%

TABLE 4

PREDICTED LOSS RATIOS $\widehat{C}_{q,i}^{u,\mathbf{w}}/v_q$.

$\mathbf{w}$	i = 1	2	3	4	5	6	7	8
$\mathbf{w}^{\mathrm{ALR}}$	50.12%	65.69%	67.97%	69.85%	70.84%	71.60%	72.35%	72.93%
$\mathbf{w}^{\mathrm{RF}}$	51.68%	67.74%	70.09%	72.03%	73.05%	73.83%	74.61%	75.20%
$\mathbf{w}^{4\mathrm{y}}$	50.44%	65.93%	68.53%	70.17%	71.17%	71.93%	72.69%	73.27%
$\mathbf{w}^{\mathrm{inc}}$	50.38%	65.79%	68.18%	69.98%	70.98%	71.72%	72.47%	73.05%

of the prediction $\widehat{C}_{q,n}^{*,\mathbf{w}}$ for the random variable $C_{q,n}$, given $\mathcal{D}^{\mathbf{v}}$.

As usual we split the mean squared error $\mathrm{mse}(\widehat{C}_{q,n}^{*,\mathbf{w}})$ into two components, the *process variance* $\mathrm{pvar}(\widehat{C}_{q,n}^{*,\mathbf{w}})$ and the *squared parameter estimation error* $\mathrm{spee}(\widehat{C}_{q,n}^{*,\mathbf{w}})$

$$\begin{aligned}\mathrm{mse}\left(\widehat{C}_{q,n}^{*,\mathbf{w}}\right) &= \mathrm{Var}(C_{q,n}\,|\,\mathcal{D}^{\mathbf{v}}) + [\widehat{C}_{q,n}^{*,\mathbf{w}} - \mathrm{E}(C_{q,n}\,|\,\mathcal{D}^{\mathbf{v}})]^2 \\ &= \underbrace{\mathrm{Var}(C_{q,n})}_{\text{process variance}} + \underbrace{[\widehat{C}_{q,n}^{*,\mathbf{w}} - \mathrm{E}(C_{q,n})]^2}_{\text{squared parameter estimation error}} \\ &=: \mathrm{pvar}(\widehat{C}_{q,n}^{*,\mathbf{w}}) + \mathrm{spee}(\widehat{C}_{q,n}^{*,\mathbf{w}}).\end{aligned}$$

Here, we have used the fact that $C_{q,n}$ and $\mathcal{D}^{\mathbf{v}}$ are independent. The calculation of the process variance is straight forward (see Appendix B).

Estimator 5.1 (Process variance). *We estimate* $\mathrm{pvar}(\widehat{C}_{q,n}^{*,\mathbf{w}})$ *by*

$$\widehat{\mathrm{pvar}}(\widehat{C}_{q,n}^{*,\mathbf{w}}) = \widehat{v}_q\left(\widehat{\sigma}_0^{\mathbf{w}}\right)^2 \widehat{f}_1^2 \ldots \widehat{f}_{n-1}^2 + v_q \sum_{j=1}^{n-1} \widehat{f}_0^{\mathbf{w}} \widehat{f}_1 \ldots \widehat{f}_{j-1} \widehat{\sigma}_j^2 \widehat{f}_{j+1}^2 \ldots \widehat{f}_{n-1}^2.$$

The squared parameter estimation error cannot be calculated directly. Simply replacing $\mathrm{E}(C_{q,n})$ by the estimator $\widehat{C}_{q,n}^{*,\mathbf{w}}$ would yield zero. Therefore, we have to analyze bias and volatility of the estimator $\widehat{C}_{q,n}^{*,\mathbf{w}}$ for $\mathrm{E}(C_{q,n})$. A nice overview of the commonly used approaches is found in Wüthrich and Merz (2008), Section 3.2.3. One of the most popular techniques is the *Conditional Resampling Approach* that has been introduced by Buchwalder *et al.* (2006). It can be shown that the volatility of $\widehat{C}_{q,n}^{f,\mathbf{w}}$ (as measured by this approach) is less than or equal to the volatility of $\widehat{C}_{q,n}^{u,\mathbf{w}}$ and that both predictors have the same bias as estimators for $\mathrm{E}(C_{q,n})$. Therefore, the predictor $\widehat{C}_{q,n}^{f,\mathbf{w}}$ should be preferred to $\widehat{C}_{q,n}^{u,\mathbf{w}}$.

Let us assume now that we have estimators $\widehat{\varepsilon}_j$ for the standard deviations ε_j of e_j. Let $\widehat{\boldsymbol{\varepsilon}} := (\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_{n+1})^t$. We define

$$\varphi(x) := 66x^4 + 12x^3 + 3x^2 + x + 1,$$
$$\psi(x) := 29297x^6 + 3492x^5 + 430x^4 + 55x^3 + 7x^2 + x,$$

and

$$\Phi(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}) := \varphi\left(\frac{\sum_{j=1}^{n+1}\widehat{\varepsilon}_j^2\left(\sum_{i=1}^{n\wedge j} w_i\widehat{v}_i\right)^2}{\left(\sum_{i=1}^{n} w_i\widehat{v}_i\right)^2}\right),$$
$$\Psi(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}) := \psi\left(\frac{\sum_{j=1}^{n+1}\widehat{\varepsilon}_j^2\left(\sum_{i=1}^{n\wedge j} w_i\widehat{v}_i\right)^2}{\left(\sum_{i=1}^{n} w_i\widehat{v}_i\right)^2}\right),$$

(with $n \wedge j := \min(n, j)$). In Appendix C we show that $\Phi(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w})$ and $\Psi(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w})$ can be used as estimators for

$$\mathrm{E}\left(\frac{\sum_{i=1}^{n} w_i v_i}{\sum_{i=1}^{n} w_i\widehat{v}_i}\right) \quad \text{and} \quad \mathrm{Var}\left(\frac{\sum_{i=1}^{n} w_i v_i}{\sum_{i=1}^{n} w_i\widehat{v}_i}\right),$$

respectively. The following estimators for the squared parameter estimation errors of $\widehat{C}_{q,n}^{f,\mathbf{w}}$ and $\widehat{C}_{q,n}^{u,\mathbf{w}}$, respectively, are derived in Appendix B. We use an approach similar to Braun (2004) since the resulting formulas are slightly simpler than those obtained with the Conditional Resampling Approach.

Estimator 5.2 (Squared parameter estimation error of $\widehat{C}_{q,n}^{f,\mathbf{w}}$). *For* $\mathrm{spee}(\widehat{C}_{q,n}^{f,\mathbf{w}})$ *we have the estimator*

$$\widehat{\mathrm{spee}}\big(\widehat{C}_{q,n}^{f,\mathbf{w}}\big) = \left(\frac{v_q}{\sum_{i=1}^{n} w_i\widehat{v}_i}\right)^2 \left[\Phi(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w})^2\widehat{\mathrm{Var}}\left(\sum_{i=1}^{n} w_i C_{i,1}\widehat{f}_1\cdots\widehat{f}_{n-1}\right) + \Psi(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w})\left(\sum_{i=1}^{n} w_i C_{i,1}\widehat{f}_1\cdots\widehat{f}_{n-1}\right)^2\right] + \big(\widehat{C}_{q,n}^{f,\mathbf{w}}\big)^2\big[\Phi(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}) - 1\big]^2$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

with

$$\widehat{\mathrm{Var}}\left(\sum_{i=1}^{n} w_i C_{i,1}\widehat{f}_1\cdots\widehat{f}_{n-1}\right)=\sum_{i=1}^{n} w_i^2\widehat{v}_i\left(\widehat{\sigma}_0^{\mathbf{w}}\right)^2\widehat{f}_1^2\ldots\widehat{f}_{n-1}^2+\left(\sum_{i=1}^{n} w_i C_{i,1}\right)^2$$

$$\times\sum_{j=1}^{n-1}\widehat{f}_1^2\cdots\widehat{f}_{j-1}^2\frac{\widehat{\sigma}_j^2}{\sum_{\kappa=1}^{n-j}C_{\kappa,j}}\widehat{f}_{j+1}^2\cdots\widehat{f}_{n-1}^2.$$

Estimator 5.3 (Squared parameter estimation error of $\widehat{C}_{q,n}^{u,\mathbf{w}}$). *For* $\mathrm{spee}(\widehat{C}_{q,n}^{u,\mathbf{w}})$ *we have the estimator*

$$\widehat{\mathrm{spee}}\left(\widehat{C}_{q,n}^{u,\mathbf{w}}\right)=\left(\frac{v_q}{\sum_{i=1}^{n}w_i\widehat{v}_i}\right)^2\left[\Phi(\widehat{\boldsymbol{\varepsilon}},\mathbf{w})^2\widehat{\mathrm{Var}}\left(\sum_{i=1}^{n}w_i C_{i,1}\widehat{f}_1^{\mathbf{w}}\cdots\widehat{f}_{n-1}^{\mathbf{w}}\right)\right.$$

$$\left.+\Psi(\widehat{\boldsymbol{\varepsilon}},\mathbf{w})\left(\sum_{i=1}^{n}w_i C_{i,1}\widehat{f}_1^{\mathbf{w}}\cdots\widehat{f}_{n-1}^{\mathbf{w}}\right)^2\right]+\left(\widehat{C}_{q,n}^{u,\mathbf{w}}\right)^2[\Phi(\widehat{\boldsymbol{\varepsilon}},\mathbf{w})-1]^2$$

with

$$\widehat{\mathrm{Var}}\left(\sum_{i=1}^{n} w_i C_{i,1}\widehat{f}_1^{\mathbf{w}}\cdots\widehat{f}_{n-1}^{\mathbf{w}}\right)=\sum_{i=1}^{n} w_i^2\widehat{v}_i\left(\widehat{\sigma}_0^{\mathbf{w}}\right)^2(\widehat{f}_1^{\mathbf{w}})^2\ldots(\widehat{f}_{n-1}^{\mathbf{w}})^2$$

$$+\left(\sum_{i=1}^{n} w_i C_{i,1}\right)^2\sum_{j=1}^{n-1}(\widehat{f}_1^{\mathbf{w}})^2\cdots(\widehat{f}_{j-1}^{\mathbf{w}})^2\frac{\sum_{i=1}^{n-j}(\lambda_{i,j}^{\mathbf{w}})^2/C_{i,j}}{(\sum_{\kappa=1}^{n-j}\lambda_{i,j}^{\mathbf{w}})^2}\widehat{\sigma}_j^2(\widehat{f}_{j+1}^{\mathbf{w}})^2\cdots(\widehat{f}_{n-1}^{\mathbf{w}})^2,$$

where we have used the coefficients

$$\lambda_{i,j}^{\mathbf{w}}=\left(w_i+\frac{\sum_{k=n-j+1}^{n}w_k\widehat{C}_{k,j}}{\sum_{k=1}^{n-j}C_{k,j}}\right)C_{i,j},$$

from Remark 4.1.

Note that we have replaced f_j by $\widehat{f}_j$ in Estimator 5.2 and by $\widehat{f}_j^{\mathbf{w}}$ in Estimator 5.3 since we want to use the same estimators for f_j that are used in the predictors. Therefore, these formulas do *not* guarantee $\widehat{\mathrm{spee}}(\widehat{C}_{q,n}^{f,\mathbf{w}})\leq\widehat{\mathrm{spee}}(\widehat{C}_{q,n}^{u,\mathbf{w}})$. But, at least, we normally have

$$\widehat{\mathrm{spee}}(\widehat{C}_{q,n}^{f,\mathbf{w}})/\widehat{C}_{q,n}^{f,\mathbf{w}}\leq\widehat{\mathrm{spee}}(\widehat{C}_{q,n}^{u,\mathbf{w}})/\widehat{C}_{q,n}^{u,\mathbf{w}}.$$

Example 5.4. We consider the portfolio from Example 4.2. The average payment in the first development year of a German MTPL claim is approximately $\widehat{x}=2$ (in EUR 1,000). For the coefficient of variation c of individual claims we use $\widehat{c}=2.5$ which has been estimated from individual claims payment data of a German motor portfolio.

TABLE 5

RESULTS FOR VARIOUS VECTORS $\mathbf{w}$ WITH $\widehat{\varepsilon}_i = 0.5\%$, $\widehat{x} = 2$ AND $\widehat{c} = 2.5$.

$\mathbf{w}$	$\mathbf{w}^{\mathrm{ALR}}$	$\mathbf{w}^{\mathrm{RF}}$	$\mathbf{w}^{\mathrm{4y}}$	$\mathbf{w}^{\mathrm{inc}}$
$\widehat{C}_{q,n}^{f,\mathbf{w}}/\widehat{v}_q$	72.93%	75.20%	73.39%	73.31%
$\sqrt{\widehat{\mathrm{pvar}}(\widehat{C}_{q,n}^{f,\mathbf{w}})}/\widehat{C}_{q,n}^{f,\mathbf{w}}$	4.02%	3.95%	4.00%	4.01%
$\sqrt{\widehat{\mathrm{spee}}(\widehat{C}_{q,n}^{f,\mathbf{w}})}/\widehat{C}_{q,n}^{f,\mathbf{w}}$	1.89%	2.27%	1.89%	1.85%
$\widehat{\mathrm{s.e.}}(\widehat{C}_{q,n}^{f,\mathbf{w}})/\widehat{C}_{q,n}^{f,\mathbf{w}}$	4.44%	4.56%	4.43%	4.41%
$\widehat{C}_{q,n}^{u,\mathbf{w}}/\widehat{v}_i$	72.93%	75.20%	73.27%	73.05%
$\sqrt{\widehat{\mathrm{pvar}}(\widehat{C}_{q,n}^{u,\mathbf{w}})}/\widehat{C}_{q,n}^{u,\mathbf{w}}$	4.02%	3.95%	4.01%	4.02%
$\sqrt{\widehat{\mathrm{spee}}(\widehat{C}_{q,n}^{u,\mathbf{w}})}/\widehat{C}_{q,n}^{u,\mathbf{w}}$	1.89%	2.27%	2.20%	1.92%
$\widehat{\mathrm{s.e.}}(\widehat{C}_{q,n}^{u,\mathbf{w}})/\widehat{C}_{q,n}^{u,\mathbf{w}}$	4.44%	4.56%	4.57%	4.46%

For the first calculation we use $\widehat{\varepsilon}_1 = \cdots = \widehat{\varepsilon}_{n+1} = 0.5\%$ which is a rather low assumption for restatement uncertainty. Table 5 shows the predicted ultimate loss ratios $\widehat{C}_{q,n}^{*,\mathbf{w}}/\widehat{v}_q$ of the quotation year and the standard errors

$$\widehat{\mathrm{s.e.}}(\widehat{C}_{q,n}^{*,\mathbf{w}}) := \sqrt{\widehat{\mathrm{mse}}(\widehat{C}_{q,n}^{*,\mathbf{w}})} := \sqrt{\widehat{\mathrm{pvar}}(\widehat{C}_{q,n}^{*,\mathbf{w}}) + \widehat{\mathrm{spee}}(\widehat{C}_{q,n}^{*,\mathbf{w}})},$$

in percent of the predicted ultimate loss burden $\widehat{C}_{q,n}^{*,\mathbf{w}}$. In addition, we provide the square roots of process variance $\widehat{\mathrm{pvar}}(\widehat{C}_{q,n}^{*,\mathbf{w}})$ and squared parameter estimation error $\widehat{\mathrm{spee}}(\widehat{C}_{q,n}^{*,\mathbf{w}})$ (in percent of $\widehat{C}_{q,n}^{*,\mathbf{w}}$). The columns show the results for the selections $\mathbf{w}^{\mathrm{ALR}}$, $\mathbf{w}^{\mathrm{RF}}$, $\mathbf{w}^{\mathrm{4y}}$ and $\mathbf{w}^{\mathrm{inc}}$ of $\mathbf{w}$ that have been introduced in Example 4.2. For the ALR method and the RF method (first two columns) the predictions $\widehat{C}_{q,n}^{f,\mathbf{w}}$ and $\widehat{C}_{q,n}^{u,\mathbf{w}}$ and also the standard errors $\widehat{\mathrm{s.e.}}(\widehat{C}_{q,n}^{f,\mathbf{w}})$ and $\widehat{\mathrm{s.e.}}(\widehat{C}_{q,n}^{u,\mathbf{w}})$ are identical. Since restatement uncertainty is rather low, the mean squared error is dominated by the process variance and the estimated standard error for the ALR method is smaller than for the RF method. Moreover, we see that in columns 3 and 4 the standard errors (in percent of the predicted ultimate) for the predictions $\widehat{C}_{q,n}^{u,\mathbf{w}}$ are slightly larger than for the predictions $\widehat{C}_{q,n}^{f,\mathbf{w}}$. This is in line with our expectation.

Table 6 shows the results of the same calculation with $\widehat{\varepsilon}_1 = \cdots = \widehat{\varepsilon}_{n+1} = 4.0\%$ which is a rather high assumption for the restatement uncertainty. In this case, the mean squared error is dominated by the parameter estimation error. It is not surprising that we obtain the smallest standard error for the RF method in the second column.

6. ESTIMATION OF OPTIMAL WEIGHTS

In this section, we assume that the parameters for restatement uncertainty are known or can be estimated from an external source (e.g. market data or expert judgment). We derive a recursion that converges to approximately-optimal

TABLE 6

RESULTS FOR VARIOUS VECTORS $\mathbf{w}$ WITH $\widehat{\varepsilon}_i = 4.0\%$, $\widehat{x} = 2$ AND $\widehat{c} = 2.5$.

$\mathbf{w}$	$\mathbf{w}^{\mathrm{ALR}}$	$\mathbf{w}^{\mathrm{RF}}$	$\mathbf{w}^{4\mathrm{y}}$	$\mathbf{w}^{\mathrm{inc}}$
$\widehat{C}_{q,n}^{f,\mathbf{w}}/\widehat{v}_q$	72.93%	75.20%	73.39%	73.31%
$\sqrt{\widehat{\mathrm{pvar}}(\widehat{C}_{q,n}^{f,\mathbf{w}})}/\widehat{C}_{q,n}^{f,\mathbf{w}}$	4.02%	3.95%	4.00%	4.01%
$\sqrt{\widehat{\mathrm{spee}}(\widehat{C}_{q,n}^{f,\mathbf{w}})}/\widehat{C}_{q,n}^{f,\mathbf{w}}$	8.52%	6.12%	7.12%	7.51%
$\widehat{\mathrm{s.e.}}(\widehat{C}_{q,n}^{f,\mathbf{w}})/\widehat{C}_{q,n}^{f,\mathbf{w}}$	9.42%	7.29%	8.16%	8.51%
$\widehat{C}_{q,n}^{u,\mathbf{w}}/\widehat{v}_i$	72.93%	75.20%	73.27%	73.05%
$\sqrt{\widehat{\mathrm{pvar}}(\widehat{C}_{q,n}^{u,\mathbf{w}})}/\widehat{C}_{q,n}^{u,\mathbf{w}}$	4.02%	3.95%	4.01%	4.02%
$\sqrt{\widehat{\mathrm{spee}}(\widehat{C}_{q,n}^{u,\mathbf{w}})}/\widehat{C}_{q,n}^{u,\mathbf{w}}$	8.52%	6.12%	7.21%	7.53%
$\widehat{\mathrm{s.e.}}(\widehat{C}_{q,n}^{u,\mathbf{w}})/\widehat{C}_{q,n}^{u,\mathbf{w}}$	9.42%	7.29%	8.25%	8.53%

weights. In consideration of Remark 4.1, we concentrate on the calculation of an optimal vector $\mathbf{w}$ for $\widehat{C}_{q,n}^{f,\mathbf{w}}$, i.e. optimal weights for $\widehat{f}_0^{\mathbf{w}}$. Let

$$T_i := \frac{C_{i,1}}{v_i}(1 - e_i - \cdots - e_{n+1}).$$

Then the T_i are unbiased estimators for f_0 with

$$\begin{aligned}
\mathrm{Var}(T_i) &= \frac{\mathrm{E}(C_{i,1}^2)}{v_i^2}\,\mathrm{Var}(1 - e_i - \cdots - e_{n+1}) + \frac{\mathrm{Var}(C_{i,1})}{v_i^2}\,\mathrm{E}(1 - e_i - \cdots - e_{n+1})^2 \\
&= \frac{\mathrm{E}(C_{i,1})^2 + \mathrm{Var}(C_{i,1})}{v_i^2}(\varepsilon_i^2 + \cdots + \varepsilon_{n+1}^2) + \frac{\mathrm{Var}(C_{i,1})}{v_i^2} \\
&= f_0^2(\varepsilon_i^2 + \cdots + \varepsilon_{n+1}^2) + \frac{\sigma_0^2}{v_i}(1 + \varepsilon_i^2 + \cdots + \varepsilon_{n+1}^2) \\
&=: s_{i,i}(\boldsymbol{\varepsilon}),
\end{aligned}$$

and for $i \neq j$

$$\begin{aligned}
\mathrm{Cov}(T_i, T_j) &= \frac{\mathrm{E}(C_{i,1}C_{j,1})}{v_i v_j}\,\mathrm{Cov}(1 - e_i - \cdots - e_{n+1}, 1 - e_j - \cdots - e_{n+1}) \\
&\quad + \frac{\mathrm{Cov}(C_{i,1}, C_{i,1})}{v_i v_j}\,\mathrm{E}(1 - e_i - \cdots - e_{n+1})\,\mathrm{E}(1 - e_j - \cdots - e_{n+1}) \\
&= \frac{\mathrm{E}(C_{i,1})\,\mathrm{E}(C_{j,1})}{v_i v_j}(\varepsilon_{\max(i,j)}^2 + \cdots + \varepsilon_{n+1}^2) + \frac{\mathrm{Cov}(C_{i,1}, C_{j,1})}{v_i v_j} \\
&= f_0^2(\varepsilon_{\max(i,j)}^2 + \cdots + \varepsilon_{n+1}^2) \\
&=: s_{i,j}(\boldsymbol{\varepsilon}),
\end{aligned}$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

i.e. the random vector $\mathbf{T} := (T_1, \dots, T_n)^t$ has the covariance matrix

$$\mathbf{S}(\boldsymbol{\varepsilon}) := \begin{pmatrix} s_{1,1}(\boldsymbol{\varepsilon}) & \cdots & s_{1,n}(\boldsymbol{\varepsilon}) \\ \vdots & \ddots & \vdots \\ s_{n,1}(\boldsymbol{\varepsilon}) & \cdots & s_{n,n}(\boldsymbol{\varepsilon}) \end{pmatrix}.$$

Lemma D.1 in Appendix D implies that $\mathbf{S}(\boldsymbol{\varepsilon})$ is positive definite and that the components of $\mathbf{S}(\boldsymbol{\varepsilon})^{-1}\mathbb{1}_n$ are non-negative. Therefore, optimal coefficients $\mathbf{g} = (g_1, \dots, g_n)^t$ for a convex combination $g_1 T_1 + \cdots + g_n T_n$ of the estimators $T_1, \dots, T_n$ are given by

$$\mathbf{g} := \frac{\mathbf{S}(\boldsymbol{\varepsilon})^{-1}\mathbb{1}_n}{\|\mathbf{S}(\boldsymbol{\varepsilon})^{-1}\mathbb{1}_n\|_1},$$

where $\|\cdot\|_1$ denotes the 1-norm (or Manhattan norm), i.e. the sum of the absolute values of the components (see Appendix E).

The estimators T_i are not observable. We only have the estimators $C_{i,1}/\widehat{v}_i$ for f_0. Since we have the linear approximation with respect to $e_i, \dots, e_{n+1}$ at $e_i = \cdots = e_{n+1} = 0$

$$\begin{aligned} \frac{C_{i,1}}{\widehat{v}_i} &= \frac{C_{i,1}}{v_i(1 + e_i + \cdots + e_{n+1})} \\ &\approx \frac{C_{i,1}}{v_i}(1 - e_i - \cdots - e_{n+1}) = T_i, \end{aligned}$$

the weights $\mathbf{g}$ are approximately-optimal for the estimators $C_{i,1}/\widehat{v}_i$. Since we have

$$\widehat{f_0^{\mathbf{w}}} = \frac{\sum_{i=1}^n w_i C_{i,1}}{\sum_{i=1}^n w_i \widehat{v}_i} = \sum_{i=1}^n \frac{w_i \widehat{v}_i}{\sum_{k=1}^n w_k \widehat{v}_k} \frac{C_{i,1}}{\widehat{v}_i},$$

the vector

$$\mathbf{w} := \frac{\widehat{\mathbf{V}}^{-1}\mathbf{g}}{\|\widehat{\mathbf{V}}^{-1}\mathbf{g}\|_1} = \frac{\widehat{\mathbf{V}}^{-1}\mathbf{S}(\boldsymbol{\varepsilon})^{-1}\mathbb{1}_n}{\|\widehat{\mathbf{V}}^{-1}\mathbf{S}(\boldsymbol{\varepsilon})^{-1}\mathbb{1}_n\|_1},$$

with $\widehat{\mathbf{V}} := \operatorname{diag}(\widehat{v}_1, \dots, \widehat{v}_n)$ is approximately-optimal for the calculation of $\widehat{f_0^{\mathbf{w}}}$.

Apart from $\boldsymbol{\varepsilon}$, the matrix $\mathbf{S}(\boldsymbol{\varepsilon})$ depends on the parameters f_0 and σ_0^2, which are not observable. We use the estimator

$$\widehat{\mathbf{S}}(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}) := \begin{pmatrix} \widehat{s}_{1,1}(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}) & \cdots & \widehat{s}_{1,n}(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}) \\ \vdots & \ddots & \vdots \\ \widehat{s}_{n,1}(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}) & \cdots & \widehat{s}_{n,n}(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}) \end{pmatrix},$$

with

$$\widehat{s}_{i,j}(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}) := (\widehat{f_0^{\mathbf{w}}})^2 (\widehat{\varepsilon}^2_{\max(i,j)} + \cdots + \widehat{\varepsilon}^2_{n+1}),$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

TABLE 7

RECURSION TO CALCULATE APPROXIMATELY-OPTIMAL WEIGHTS WITH THE ASSUMPTION $\widehat{\varepsilon}_i = 0.5\%$.

k	0	1	2	3
$w_1^{(k)}$	10.0%	5.3%	5.2%	5.2%
$w_2^{(k)}$	10.0%	5.8%	5.7%	5.7%
$w_3^{(k)}$	10.0%	6.8%	6.8%	6.8%
$w_4^{(k)}$	10.0%	8.6%	8.5%	8.5%
$w_5^{(k)}$	10.0%	11.2%	11.1%	11.1%
$w_6^{(k)}$	10.0%	14.9%	14.9%	14.9%
$w_7^{(k)}$	10.0%	20.1%	20.2%	20.2%
$w_8^{(k)}$	10.0%	27.4%	27.5%	27.5%
$\widehat{C}_{q,n}^{f,\mathbf{w}^{(k)}}/\widehat{v}_q$	72.93%	73.62%	73.63%	73.63%
$\widehat{\mathrm{s.e.}}(\widehat{C}_{q,n}^{f,\mathbf{w}^{(k)}})/\widehat{C}_{q,n}^{f,\mathbf{w}^{(k)}}$	4.44%	4.40%	4.40%	4.40%
$\widehat{C}_{q,n}^{u,\mathbf{w}^{(k)}}/\widehat{v}_q$	72.93%	73.57%	73.57%	73.57%
$\widehat{\mathrm{s.e.}}(\widehat{C}_{q,n}^{u,\mathbf{w}^{(k)}})/\widehat{C}_{q,n}^{u,\mathbf{w}^{(k)}}$	4.44%	4.43%	4.43%	4.43%

for $i \neq j$ and

$$\widehat{s}_{i,i}(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}) := (\widehat{f}_0^{\mathbf{w}})^2(\widehat{\varepsilon}_i^2 + \cdots + \widehat{\varepsilon}_{n+1}^2) + \frac{(\widehat{\sigma}_0^{\mathbf{w}})^2}{\widehat{v}_i}(1 + \widehat{\varepsilon}_i^2 + \cdots + \widehat{\varepsilon}_{n+1}^2).$$

The matrix $\widehat{\mathbf{S}}(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w})$ is positive definite and the components of $\widehat{\mathbf{S}}(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w})^{-1}\mathbb{1}_n$ are non-negative (see Appendix D). Since $\widehat{\mathbf{S}}(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w})$ depends on the vector $\mathbf{w}$ we need a recursion to estimate optimal weights for $\widehat{f}_0^{\mathbf{w}}$.

Recursion 6.1 (Estimation of optimal weights for given $\widehat{\boldsymbol{\varepsilon}}$). *Start with a vector $\mathbf{w}^{(0)}$ and use the iteration*

$$\mathbf{w}^{(k+1)} := \frac{\widehat{\mathbf{V}}^{-1}\widehat{\mathbf{S}}(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}^{(k)})^{-1}\mathbb{1}_n}{\|\widehat{\mathbf{V}}^{-1}\widehat{\mathbf{S}}(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}^{(k)})^{-1}\mathbb{1}_n\|_1}.$$

In Appendix F, we show that this recursion converges to a fixed point if the loss ratios $C_{i,1}/\widehat{v}_i$ do not fluctuate too much. In practice it is difficult to find an example where the recursion does not converge. Often two or three steps are sufficient to obtain a good approximation for the fixed-point, i.e. typically $\mathbf{w}^{(2)}$ or $\mathbf{w}^{(3)}$ can be used.

Example 6.2. We pick up the Examples 4.2 and 5.4. In Tables 7 and 8 we see the results of Recursion 6.1 with external estimates $\widehat{\varepsilon}_1 = \cdots = \widehat{\varepsilon}_{n+1} = 0.5\%$ and $\widehat{\varepsilon}_1 = \cdots = \widehat{\varepsilon}_{n+1} = 4.0\%$, respectively. We start with $\mathbf{w}^{(0)} := 10\% \cdot \mathbb{1}_n$. In both cases the recursion converges very quickly, we have $\mathbf{w}^{(3)} \approx \mathbf{w}^{(2)}$. If $\widehat{\varepsilon}_i = 0.5\%$ then we obtain substantial weights for the more mature accident years. If $\widehat{\varepsilon}_i = 4.0\%$ then $\mathbf{w}^{(3)}$ is close to $\mathbf{w}^{\mathrm{RF}}$, i.e. the “optimal” method is close to the RF method.

TABLE 8

RECURSION TO CALCULATE APPROXIMATELY-OPTIMAL WEIGHTS WITH THE ASSUMPTION $\widehat{\varepsilon}_i = 4.0\%$.

k	0	1	2	3
$w_1^{(k)}$	10.0%	0.0%	0.0%	0.0%
$w_2^{(k)}$	10.0%	0.0%	0.0%	0.0%
$w_3^{(k)}1$	10.0%	0.0%	0.0%	0.0%
$w_4^{(k)}$	10.0%	0.0%	0.0%	0.0%
$w_5^{(k)}$	10.0%	0.2%	0.1%	0.1%
$w_6^{(k)}$	10.0%	1.2%	1.2%	1.2%
$w_7^{(k)}$	10.0%	10.5%	10.3%	10.3%
$w_8^{(k)}$	10.0%	88.1%	88.4%	88.4%
$\widehat{C}_{q,n}^{f,\mathbf{w}^{(k)}}/\widehat{v}_q$	72.93%	75.31%	75.31%	75.31%
$\widehat{\text{s. e.}}(\widehat{C}_{q,n}^{f,\mathbf{w}^{(k)}})/\widehat{C}_{q,n}^{f,\mathbf{w}^{(k)}}$	9.42%	7.27%	7.27%	7.27%
$\widehat{C}_{q,n}^{u,\mathbf{w}^{(k)}}/\widehat{v}_q$	72.93%	75.26%	75.26%	75.26%
$\widehat{\text{s. e.}}(\widehat{C}_{q,n}^{u,\mathbf{w}^{(k)}})/\widehat{C}_{q,n}^{u,\mathbf{w}^{(k)}}$	9.42%	7.28%	7.28%	7.28%

7. ESTIMATION OF THE PARAMETERS FOR RESTATEMENT UNCERTAINTY

In this section, we use a maximum likelihood approach to estimate the parameters for restatement uncertainty. In the last section, we have used the linear approximation with respect to $e_i, \ldots, e_{n+1}$

$$\frac{C_{i,1}}{\widehat{v}_i} = \frac{C_{i,1}}{v_i(1 + e_i + \cdots + e_{n+1})} \approx \frac{C_{i,1}}{v_i}(1 - e_i - \cdots - e_{n+1}) = T_i.$$

We now interpret $C_{i,1}/\widehat{v}_i$ as a function of the variables $C_{i,1}, e_i, \ldots, e_{n+1}$ and use the linear approximation at $(f_0 v_i, 0, \ldots, 0)^t$, i.e.

$$\frac{C_{i,1}}{\widehat{v}_i} \approx \frac{C_{i,1}}{v_i} - f_0(e_i + \cdots + e_{n+1}) =: T_i'.$$

Then $T_1', \ldots, T_n'$ are unbiased estimators for f_0. We have

$$\operatorname{Var}(T_i') = f_0^2(\varepsilon_i^2 + \cdots + \varepsilon_{n+1}^2) + \frac{\sigma_0^2}{v_i} =: s_{i,i}'(\boldsymbol{\varepsilon}),$$

and for $i \neq j$

$$\operatorname{Cov}(T_i', T_j') = f_0^2(\varepsilon_{\max(i,j)}^2 + \cdots + \varepsilon_{n+1}^2) =: s_{i,j}'(\boldsymbol{\varepsilon}),$$

i.e. the covariance matrix of the vector $\mathbf{T}' := (T'_1, \ldots, T'_n)^t$ is given by

$$\mathbf{S}'(\boldsymbol{\varepsilon}) := \begin{pmatrix} s'_{1,1}(\boldsymbol{\varepsilon}) & \cdots & s'_{1,n}(\boldsymbol{\varepsilon}) \\ \vdots & \ddots & \vdots \\ s'_{n,1}(\boldsymbol{\varepsilon}) & \cdots & s'_{n,n}(\boldsymbol{\varepsilon}) \end{pmatrix}.$$

For motor quota shares the expected number of claims per accident year is typically very large. Taking the asymptotic normality of the compound Poisson process into account we can therefore assume that the distribution of $C_{i,1}$ is close to the normal distribution $N(f_0 v_i, \sigma_0^2 v_i)$ (CLT for compound Poisson distributions, see Kaas *et al.* (2001)). We have

$$\mathbf{T}' = \left(\begin{array}{ccc|cccc} v_1^{-1} & & 0 & -f_0 & \cdots & -f_0 & -f_0 \\ & \ddots & & & \ddots & \vdots & \vdots \\ 0 & & v_n^{-1} & 0 & & -f_0 & -f_0 \end{array}\right) \begin{pmatrix} C_{1,1} \\ \vdots \\ C_{n,1} \\ e_1 \\ \vdots \\ e_{n+1} \end{pmatrix},$$

i.e. $\mathbf{T}'$ is a linear transformation of the independent (and nearly normally distributed) random variables $C_{1,1}, \ldots, C_{n,1}, e_1, \ldots, e_{n+1}$. Therefore, the distribution of $\mathbf{T}'$ can be approximated by the multivariate normal distribution $N(f_0 \mathbb{1}_n, \mathbf{S}'(\boldsymbol{\varepsilon}))$. We use

$$\mathbf{F} := (C_{1,1}/\widehat{v}_1, \ldots, C_{n,1}/\widehat{v}_n)^t \approx \mathbf{T}',$$

and approximate $\mathbf{S}'(\boldsymbol{\varepsilon})$ by

$$\widehat{\mathbf{S}}'(\boldsymbol{\varepsilon}, \mathbf{w}) := \begin{pmatrix} \widehat{s}'_{1,1}(\boldsymbol{\varepsilon}, \mathbf{w}) & \cdots & \widehat{s}'_{1,n}(\boldsymbol{\varepsilon}, \mathbf{w}) \\ \vdots & \ddots & \vdots \\ \widehat{s}'_{n,1}(\boldsymbol{\varepsilon}, \mathbf{w}) & \cdots & \widehat{s}'_{n,n}(\boldsymbol{\varepsilon}, \mathbf{w}) \end{pmatrix},$$

with

$$\widehat{s}'_{i,j}(\boldsymbol{\varepsilon}, \mathbf{w}) := (\widehat{f}_0^{\mathbf{w}})^2 \big(\varepsilon^2_{\max(i,j)} + \cdots + \varepsilon^2_{n+1}\big),$$

for $i \neq j$ and

$$\widehat{s}'_{i,i}(\boldsymbol{\varepsilon}, \mathbf{w}) := (\widehat{f}_0^{\mathbf{w}})^2 \big(\varepsilon_i^2 + \cdots + \varepsilon^2_{n+1}\big) + \frac{(\widehat{\sigma}_0^{\mathbf{w}})^2}{\widehat{v}_i}.$$

Note that $\widehat{\mathbf{S}}'(\boldsymbol{\varepsilon}, \mathbf{w})$ is positive definite and thus invertible (cf. Appendix D). We obtain the approximation

$$\widehat{L}_{\mathbf{F}}^{\mathbf{w}}(\boldsymbol{\varepsilon}) := \frac{1}{\sqrt{(2\pi)^n \det\big(\widehat{\mathbf{S}}'(\boldsymbol{\varepsilon}, \mathbf{w})\big)}} \exp\left(-\frac{1}{2}(\mathbf{F} - \widehat{f}_0^{\mathbf{w}} \mathbb{1}_n)^t \widehat{\mathbf{S}}'(\boldsymbol{\varepsilon}, \mathbf{w})^{-1} (\mathbf{F} - \widehat{f}_0^{\mathbf{w}} \mathbb{1}_n)\right),$$

for the likelihood function $L_{\mathbf{F}}(\boldsymbol{\varepsilon})$. Note that the approximation depends on the choice of the weights $\mathbf{w}$.

Obviously, it does not make sense to estimate $n+1$ parameters $\varepsilon_1, \dots, \varepsilon_{n+1}$. We therefore assume that $\varepsilon_i := c_i \varepsilon$ with $\varepsilon > 0$ and *known* factors $c_i > 0$, i.e. $\boldsymbol{\varepsilon} = \mathbf{c}\varepsilon$ with $\mathbf{c} = (c_1, \dots, c_{n+1})^t$. The constants c_i provide a certain flexibility if the uncertainty of the index change is not the same for all accident years. In the case $\mathbf{c} = \mathbb{1}_{n+1}$ we simply have $\varepsilon_1 = \dots = \varepsilon_{n+1} = \varepsilon$.

Recursion 7.1 (Estimation of restatement uncertainty and optimal weights). *Choose an initial estimator $\widehat{\varepsilon}^{(0)}$ for ε and an initial vector $\mathbf{w}^{(0)}$. Given $\mathbf{w}^{(k)}$ we determine $\widehat{\varepsilon}^{(k+1)}$ such that*

$$\widehat{L}_{\mathbf{F}}^{\mathbf{w}^{(k)}}(\mathbf{c}\widehat{\varepsilon}^{(k+1)}) = \max\{\widehat{L}_{\mathbf{F}}^{\mathbf{w}^{(k)}}(\mathbf{c}\varepsilon) \mid \varepsilon \in [0, \infty)\}.$$

Then we use the matrix $\widehat{\mathbf{S}}(\mathbf{c}\widehat{\varepsilon}^{(k+1)}, \mathbf{w}^{(k)})$ from Section 6 to estimate optimal weights $\mathbf{w}^{(k+1)}$ with the new estimator $\widehat{\varepsilon}^{(k+1)}$ for ε

$$\mathbf{w}^{(k+1)} := \frac{\widehat{\mathbf{V}}^{-1}\widehat{\mathbf{S}}(\mathbf{c}\widehat{\varepsilon}^{(k+1)}, \mathbf{w}^{(k)})^{-1}\mathbb{1}_n}{\|\widehat{\mathbf{V}}^{-1}\widehat{\mathbf{S}}(\mathbf{c}\widehat{\varepsilon}^{(k+1)}, \mathbf{w}^{(k)})^{-1}\mathbb{1}_n\|_1}.$$

Remark 7.2. *The likelihood function $\lambda^{(k)}\colon [0, \infty) \to [0, \infty)$, $\varepsilon \mapsto \widehat{L}_{\mathbf{F}}^{\mathbf{w}^{(k)}}(\mathbf{c}\varepsilon)$ has a global maximum since $\lambda^{(k)}(0) > 0$ and $\lambda^{(k)}(\varepsilon) \to 0$ for $\varepsilon \to \infty$. We have calculated a number of cases and $\lambda^{(k)}$ was always either unimodal or strictly decreasing. However, there might be situations where $\lambda^{(k)}$ has several local maxima or even several global maxima. In such cases, we recommend not to use the method, since the results might be misleading.*

We omit a systematic analysis regarding convergence of Recursion 7.1 since it is much more technical and difficult than for Recursion 6.1. We have calculated a lot of cases and did not have any issues with convergence of this recursion. Typically $\widehat{\varepsilon}^{(k)}$ and $\mathbf{w}^{(k)}$ are stable after $k = 3$ iterations.

Remark 7.3. *In calculations with $n = 10$ and $\mathbf{c} = \mathbb{1}_{n+1}$ we have observed that the recursion provides quite reliable estimates for ε if the coefficients of variation $\mathrm{CV}(C_{i,1})$ are small compared to ε. The results are still good if $\mathrm{CV}(C_{i,1})$ and ε are similar in size. Therefore, the algorithm works quite well for segments with low stochastic volatility of the claims burden (like large German motor quota shares). Nevertheless, we recommend to check whether the estimated parameter for restatement uncertainty is plausible.*

For segments with high stochastic volatility, i.e. segments where the coefficients of variation $\mathrm{CV}(C_{i,1})$ are large compared to ε, the results of the algorithm are less reliable. The algorithm then often yields $\widehat{\varepsilon}^{(k)} = 0$ and $\mathbf{w}^{(k)} = \mathbb{1}_n/\|\mathbb{1}_n\|_1$ for $k \geq 1$ (which corresponds to the ALR method). In such segments, however, the influence of restatement uncertainty is less material than in segments with low stochastic volatility. Here, we suggest to use a reasonable external estimator $\widehat{\varepsilon}$ for ε and to apply Recursion 6.1 from Section 6.

TABLE 9

RECURSION TO ESTIMATE $\boldsymbol{\varepsilon} = \mathbb{1}_{n+1} \cdot \varepsilon$ AND CORRESPONDING OPTIMAL WEIGHTS.

k	0	1	2	3
$\varepsilon^{(k)}$	4.00%	2.36%	2.05%	2.05%
$w_1^{(k)}$	10.0%	0.0%	0.0%	0.0%
$w_2^{(k)}$	10.0%	0.0%	0.0%	0.0%
$w_3^{(k)}$	10.0%	0.1%	0.1%	0.1%
$w_4^{(k)}$	10.0%	0.3%	0.5%	0.5%
$w_5^{(k)}$	10.0%	1.2%	1.7%	1.7%
$w_6^{(k)}$	10.0%	4.6%	5.9%	5.9%
$w_7^{(k)}$	10.0%	18.7%	20.5%	20.5%
$w_8^{(k)}$	10.0%	75.1%	71.2%	71.2%
$\widehat{C}_{q,n}^{f,\mathbf{w}^{(k)}} / \widehat{v}_q$	72.93%	75.23%	75.17%	75.17%
$\widehat{\text{s.e.}}(\widehat{C}_{q,n}^{f,\mathbf{w}^{(k)}}) / \widehat{C}_{q,n}^{f,\mathbf{w}^{(k)}}$	9.42%	5.56%	5.29%	5.29%
$\widehat{C}_{q,n}^{u,\mathbf{w}^{(k)}} / \widehat{v}_q$	72.93%	75.21%	75.16%	75.16%
$\widehat{\text{s.e.}}(\widehat{C}_{q,n}^{u,\mathbf{w}^{(k)}}) / \widehat{C}_{q,n}^{u,\mathbf{w}^{(k)}}$	9.42%	5.59%	5.33%	5.33%

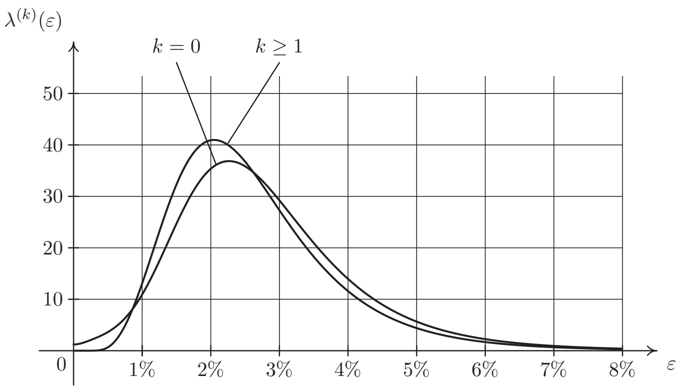

FIGURE 7: Graphs of the likelihood functions $\lambda^{(k)}$ for $k = 0$ and $k \geq 1$.

Example 7.4. We finalize the preceding examples. Table 9 contains the results of Recursion 7.1. We have assumed that $\varepsilon_1 = \cdots = \varepsilon_{n+1}$, i.e. $\boldsymbol{\varepsilon} = \mathbb{1}_{n+1}\varepsilon$. We start with $\widehat{\varepsilon}^{(0)} = 4.0\%$ and $\mathbf{w}^{(0)} := 10\% \cdot \mathbb{1}_n$. Again, the recursion converges very quickly, we have $\widehat{\varepsilon}^{(3)} \approx \widehat{\varepsilon}^{(2)} \approx 2.05\%$ and $\mathbf{w}^{(3)} \approx \mathbf{w}^{(2)}$. In Figure 7 we have plotted the likelihood functions $\lambda^{(k)}$ (as defined in Remark 7.2) for $k = 0$ and $k \geq 1$ (for $k \geq 1$ the $\lambda^{(k)}$ are nearly identical).

We conclude that for our example the "optimal" method mainly relies on the loss ratios of the last two accident years in the observation period. This is due to the fact that the average individual claims size $\widehat{x}$ and the coefficient of variation

$\widehat{c}$ are relatively small. This is typical for large MTPL quota shares in Germany where the loss ratios are to a large extent driven by the premium cycle.

ACKNOWLEDGEMENTS

The author would like to thank the referees for careful reading and useful comments that helped to improve the presentation of this paper.

REFERENCES

BRAUN, CH. (2004) The prediction error of the chain ladder method applied to correlated run-off triangles. *Astin Bulletin*, **34**(2), 399–423.

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.

FRISHMAN, F. (1971) *On the Arithmetic Means and Variances of Products and Ratios of Random Variables*. Springfield NA: National Technical Information Service.

HAAS, A. (2013) Rentabilität von Kraftfahrt-Quoten im Niedrigzins-Umfeld. *Versicherungswirtschaft*, **2003**(22), 32–35.

JONES, B. D. (2002) An introduction to premium trend. *CAS Study Note*.

KAAS, R., GOOVAERTS, M., DHAENE, J. and DENUIT, M. (2001) *Modern Actuarial Risk Theory.* Boston: Kluwer Academic Publishers.

KLUGMAN, S. A., PANJER, H. A. and WILLMOT, G. E. (2004) *Loss Models: From Data to Decisions.* 2nd Edition. New York: Wiley.

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

MACK, TH. (2002) *Schadenversicherungsmathematik.* Karlsruhe: VVW.

RIEGEL, U. (2014) A bifurcation approach for attritional and large losses in chain ladder calculations. *Astin Bulletin*, **44**(1), 127–172.

WÜTHRICH, M. V. and MERZ, M. (2008) *Stochastic Claims Reserving Methods in Insurance.* Chichester: John Wiley & Sons.

ULRICH RIEGEL
Swiss Re Europe S.A., Niederlassung Deutschland, 81911 München, Germany
E-mail: Ulrich_Riegel@swissre.com

APPENDIX A

ADDITIONAL DETAILS REGARDING THE APPROXIMATION FROM SECTION 3

In Section 3, we use the approximation

$$\breve{v}_i = v_i(1+e_i)\dots(1+e_{n+1}) \approx v_i(1+e_i+\dots+e_{n+1}) = \widehat{v}_i.$$

The following lemma shows that $(\breve{v}_1, \ldots, \breve{v}_{n+1})^t$ and $(\widehat{v}_1, \ldots, \widehat{v}_{n+1})^t$ have very similar covariance matrices.

Lemma A.1. *Let* $\varepsilon_1, \ldots, \varepsilon_{n+1} > 0$ *and* $\varepsilon := \max\{\varepsilon_1, \ldots, \varepsilon_{n+1}\}$. *Then we have* $\mathrm{Cov}(\widehat{v}_i, \widehat{v}_j) > 0$ *and*

$$1 \leq \frac{\mathrm{Cov}(\breve{v}_i, \breve{v}_j)}{\mathrm{Cov}(\widehat{v}_i, \widehat{v}_j)} \leq \frac{(1+\varepsilon^2)^{n+1} - 1}{(n+1)\varepsilon^2},$$

for all $i, j \in \{1, \ldots, n+1\}$.

For instance, Lemma A.1 yields for $n = 10$ and $\varepsilon = 5\%$

$$1 \leq \frac{\mathrm{Cov}(\breve{v}_i, \breve{v}_j)}{\mathrm{Cov}(\widehat{v}_i, \widehat{v}_j)} \leq 1.012.$$

Proof of Lemma A.1. For $i = 1, \ldots, n+1$ let $d_i = e_{n+2-i}$ and $\delta_i = \varepsilon_{n+2-i}$. We define

$$D_k^M := (1+d_1)\cdots(1+d_k) \quad \text{and} \quad D_k^A := 1 + d_1 + \cdots + d_k.$$

Then we have

$$\breve{v}_i = v_i \cdot D^M_{n+2-i} \quad \text{and} \quad \widehat{v}_i = v_i \cdot D^A_{n+2-i},$$

for $i = 1, \ldots, n+1$. Since the d_i are independent with $\mathrm{E}(d_i) = 0$ we obtain for $i \geq j$

$$\mathrm{Cov}(\breve{v}_i, \breve{v}_j) = v_i v_j \,\mathrm{Cov}(D^M_{n+2-i}, D^M_{n+2-j}) = v_i v_j \,\mathrm{Var}(D^M_{n+2-i}) \quad \text{and}$$
$$\mathrm{Cov}(\widehat{v}_i, \widehat{v}_j) = v_i v_j \,\mathrm{Cov}(D^A_{n+2-i}, D^A_{n+2-j}) = v_i v_j \,\mathrm{Var}(D^A_{n+2-i}),$$

i.e.

$$\frac{\mathrm{Cov}(\breve{v}_i, \breve{v}_j)}{\mathrm{Cov}(\widehat{v}_i, \widehat{v}_j)} = \frac{\mathrm{Var}(D^M_{n+2-i})}{\mathrm{Var}(D^A_{n+2-i})}.$$

Let $k, l \in \mathbb{N}$, $1 \leq i_1 < \cdots < i_k$ and $1 \leq j_1 < \cdots < j_l$. Taking $E(d_i) = 0$ into account, we can use Lemma D.2 from Riegel (2014) and induction to show that

$$\mathrm{Cov}(d_{i_1}\cdots d_{i_k},\, d_{j_1}\cdots d_{j_l}) = \begin{cases} \delta_{i_1}^2 \cdots \delta_{i_k}^2 & \text{if } k = l \text{ and } i_\nu = j_\nu \text{ for all } \nu, \\ 0 & \text{else.} \end{cases}$$

Using this fact, it is easy to show that

$$\mathrm{Var}(D_k^M) = (1+\delta_1^2)\cdots(1+\delta_k^2) - 1 \quad \text{and}$$
$$\mathrm{Var}(D_k^A) = \delta_1^2 + \cdots + \delta_k^2.$$

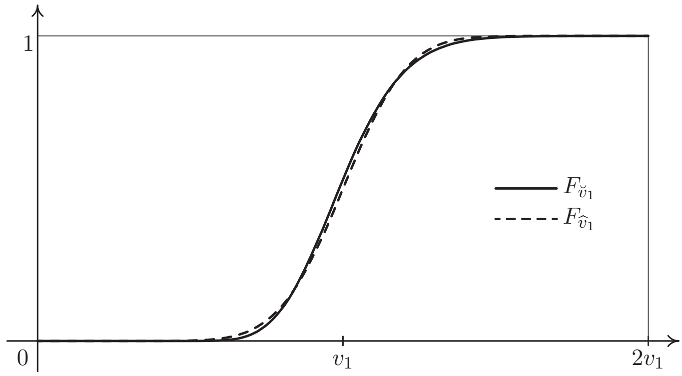

FIGURE 8: CDFs of $\breve{v}_1$ and $\widehat{v}_1$ for $n = 10$ and $\varepsilon_1 = \cdots = \varepsilon_{n+1} = 5\%$.

Applying the chain rule, we obtain for $i \in \{1, \ldots, k\}$

$$\frac{\partial}{\partial \delta_i} \frac{\mathrm{Var}(D_k^M)}{\mathrm{Var}(D_k^A)} = \frac{2\delta_i \prod_{j \le k, j \ne i} (1 + \delta_j^2)(\sum_{\nu=1}^{k} \delta_\nu^2) - 2\delta_i \left[\prod_{j=1}^{k} (1 + \delta_j^2) - 1\right]}{(\sum_{j=1}^{n} \delta_j^2)^2} > 0.$$

Consequently, $\delta_i = \varepsilon_{n+2-i} \le \varepsilon$ for $i = 1, \ldots, k$ implies

$$\frac{\mathrm{Var}(D_k^M)}{\mathrm{Var}(D_k^A)} \le \frac{(1 + \varepsilon^2)^k - 1}{k\varepsilon^2}.$$

With

$$\frac{(1+\varepsilon^2)^k - 1}{k\varepsilon^2} = \frac{\sum_{\nu=0}^{k} \binom{k}{\nu} \varepsilon^{2\nu} - 1}{k\varepsilon^2} = 1 + \sum_{\nu=2}^{k} \frac{(k-1)!}{(k-\nu)!\nu!} \varepsilon^{2(\nu-1)}$$

$$> 1 + \sum_{\nu=2}^{k-1} \frac{(k-2)!}{(k-\nu-1)!\nu!} \varepsilon^{2(\nu-1)} = \frac{(1+\varepsilon^2)^{(k-1)} - 1}{(k-1)\varepsilon^2},$$

we see that

$$\frac{\mathrm{Var}(D_k^M)}{\mathrm{Var}(D_k^A)} \le \frac{(1+\varepsilon^2)^{n+1} - 1}{(n+1)\varepsilon^2},$$

for $k = 1, \ldots, n + 1$. □

Figure 8 shows the distribution functions $F_{\breve{v}_1}$ and $F_{\widehat{v}_1}$ for $n = 10$ and $\varepsilon_1 = \cdots = \varepsilon_{n+1} = 5\%$. The deviation between the two distribution functions is certainly acceptable for our purpose. Let δ_i, D_i^M and D_i^A be defined as in the proof of Lemma A.1. Similar arguments as in the proof show that D_i^A and $D_i^M - D_i^A$ are uncorrelated and that

$$\frac{\mathrm{Var}(D_k^M - D_k^A)}{\mathrm{Var}(D_k^A)}$$

is a strictly increasing function with respect to k and $\delta_1, \ldots, \delta_k$. For $n \leq 10$ and $\varepsilon_1, \ldots, \varepsilon_{n+1} < 5\%$ we therefore expect that for each i the approximation of $F_{\check{v}_i}$ by $F_{\widehat{v}_i}$ is better that the approximation of $F_{\check{v}_1}$ by $F_{\widehat{v}_1}$ in Figure 8. This qualitative expectation is strongly supported by plots of the distribution functions for smaller values of n and ε_i where the approximations look much better.

Appendix B

Deduction of the estimators for the mean squared error

In this appendix, we deduct the estimators for process variance and squared parameter estimation error from Section 5. We have

$$\operatorname{Var}(C_{q,1}) = \sigma_0^2 v_q,$$

and

$$\begin{aligned}\operatorname{Var}(C_{q,j+1}) &= \operatorname{Var}(\mathrm{E}(C_{q,j+1} \mid \mathcal{B}_j)) + \mathrm{E}(\operatorname{Var}(C_{q,j+1} \mid \mathcal{B}_j)) \\ &= \operatorname{Var}(f_j C_{q,j}) + \mathrm{E}(\sigma_j^2 C_{q,j}) \\ &= f_j^2 \operatorname{Var}(C_{q,j}) + \sigma_j^2 v_q f_0 \ldots f_{j-1}.\end{aligned}$$

By induction we obtain

$$\operatorname{pvar}(\widehat{C}_{q,n}^{*,\mathbf{w}}) = \operatorname{Var}(C_{q,n}) = v_q \sum_{j=0}^{n-1} f_0 \ldots f_{j-1} \sigma_j^2 f_{j+1}^2 \ldots f_{n-1}^2.$$

Replacing the unknown parameters f_j, σ_j^2 by their estimators $\widehat{f}_j, \widehat{\sigma}_j^2$ (for $j \geq 1$) and $\widehat{f}_0^{\mathbf{w}}$, $\left(\widehat{\sigma}_0^{\mathbf{w}}\right)^2$ (for $j = 0$) we obtain Estimator 5.1. For the estimation of the squared parameter estimation error we use the approximation

$$\begin{aligned}\operatorname{spee}(\widehat{C}_{q,n}^{*,\mathbf{w}}) &\approx \mathrm{E}(\operatorname{spee}(\widehat{C}_{q,n}^{*,\mathbf{w}})) \\ &= \mathrm{E}([\widehat{C}_{q,n}^{*,\mathbf{w}} - \mathrm{E}(C_{q,n})]^2) \\ &= \mathrm{E}([\widehat{C}_{q,n}^{*,\mathbf{w}} - \mathrm{E}(\widehat{C}_{q,n}^{*,\mathbf{w}}) + \mathrm{E}(\widehat{C}_{q,n}^{*,\mathbf{w}}) - \mathrm{E}(C_{q,n})]^2) \\ &= \operatorname{Var}(\widehat{C}_{q,n}^{*,\mathbf{w}}) + [\mathrm{E}(\widehat{C}_{q,n}^{*,\mathbf{w}}) - \mathrm{E}(C_{q,n})]^2 \\ &= \operatorname{Var}(\widehat{C}_{q,n}^{*,\mathbf{w}}) + v_q^2[\mathrm{E}(\widehat{f}_0^{\mathbf{w}}) f_1 \ldots f_{n-1} - f_0 f_1 \ldots f_{n-1}]^2 \\ &= \operatorname{Var}(\widehat{C}_{q,n}^{*,\mathbf{w}}) + v_q^2 \left[\mathrm{E}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i} \frac{\sum_{i=1}^n w_i \widehat{C}_{i,1}}{\sum_{i=1}^n w_i v_i}\right) - f_0\right]^2 f_1^2 \ldots f_{n-1}^2\end{aligned}$$

$$= \operatorname{Var}(\widehat{C}^{*,\mathbf{w}}_{q,n}) + v_q^2 \left[\mathrm{E}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i} \right) f_0 - f_0 \right]^2 f_1^2 \ldots f_{n-1}^2$$

$$= \operatorname{Var}(\widehat{C}^{*,\mathbf{w}}_{q,n}) + \mathrm{E}(C_{q,n})^2 \left[\mathrm{E}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i} \right) - 1 \right]^2 .$$

For the calculation of $\operatorname{Var}(\widehat{C}^{*,\mathbf{w}}_{q,n})$ we have to differentiate between $\widehat{C}^{f,\mathbf{w}}_{q,j}$ and $\widehat{C}^{u,\mathbf{w}}_{q,j}$. With

$$\begin{aligned}
\operatorname{Var}(\widehat{C}^{f,\mathbf{w}}_{q,n}) &= \operatorname{Var}\left(\frac{v_q}{\sum_{i=1}^n w_i \widehat{v}_i} \sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{n-1} \right) \\
&= \left(\frac{v_q}{\sum_{i=1}^n w_i v_i} \right)^2 \operatorname{Var}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i} \sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{n-1} \right) \\
&= \left(\frac{v_q}{\sum_{i=1}^n w_i v_i} \right)^2 \left[\mathrm{E}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i} \right)^2 \operatorname{Var}\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{n-1} \right) \right. \\
&\quad \left. + \operatorname{Var}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i} \right) \mathrm{E}\left(\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{n-1} \right)^2 \right) \right],
\end{aligned}$$

we obtain

$$\begin{aligned}
\operatorname{spee}(\widehat{C}^{f,\mathbf{w}}_{q,n}) \approx{}& \left(\frac{v_q}{\sum_{i=1}^n w_i v_i} \right)^2 \left[\mathrm{E}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i} \right)^2 \operatorname{Var}\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{n-1} \right) \right. \\
&\left. + \operatorname{Var}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i} \right) \mathrm{E}\left(\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{n-1} \right)^2 \right) \right] \\
&+ \mathrm{E}(C_{q,n})^2 \left[\mathrm{E}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i} \right) - 1 \right]^2 .
\end{aligned}$$

We replace

$$\mathrm{E}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i} \right) \quad \text{and} \quad \operatorname{Var}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i} \right)$$

by the estimators $\Phi(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w})$ and $\Psi(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w})$. For details cf. Appendix C. Moreover, we replace

$$\frac{v_q}{\sum_{i=1}^n w_i v_i} \quad \text{and} \quad \mathrm{E}\left(\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{n-1} \right)^2 \right)$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

by the estimators

$$\frac{v_q}{\sum_{i=1}^n w_i \widehat{v}_i} \quad \text{and} \quad \left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{n-1}\right)^2,$$

and use the estimator $\widehat{C}_{q,n}^{f,\mathbf{w}}$ for $\mathrm{E}(C_{q,n})$. Apart from the formula for the estimator

$$\widehat{\mathrm{Var}}\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{n-1}\right) \quad \text{of} \quad \mathrm{Var}\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{n-1}\right),$$

we then obtain Estimator 5.2. With

$$\mathrm{Var}\left(\sum_{i=1}^n w_i C_{i,1}\right) = \sum_{i=1}^n w_i^2 \,\mathrm{Var}(C_{i,1}) = \sum_{i=1}^n w_i^2 \sigma_0^2 v_i,$$

and

$$\begin{aligned}
&\mathrm{Var}\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_j\right) \\
&= \mathrm{Var}\left(\mathrm{E}\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_j \,\middle|\, \mathcal{D}_j\right)\right) + \mathrm{E}\left(\mathrm{Var}\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_j \,\middle|\, \mathcal{D}_j\right)\right) \\
&= \mathrm{Var}\left(f_j \sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{j-1}\right) + \mathrm{E}\left(\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{j-1}\right)^2 \mathrm{Var}(\widehat{f}_j \mid \mathcal{D}_j)\right) \\
&= f_j^2 \,\mathrm{Var}\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{j-1}\right) + \mathrm{E}\left(\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{j-1}\right)^2 \frac{\sigma_j^2}{\sum_{\kappa=1}^{n-j} C_{\kappa,j}}\right) \\
&= f_j^2 \,\mathrm{Var}\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{j-1}\right) + \mathrm{E}\left(\frac{\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{j-1}\right)^2}{\sum_{\kappa=1}^{n-j} C_{\kappa,j}}\right) \sigma_j^2
\end{aligned}$$

we obtain by induction

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{n-1}\right) &= \sum_{i=1}^n w_i^2 v_i \sigma_0^2 f_1^2 \cdots f_{n-1}^2 \\
&+ \sum_{j=1}^{n-1} \mathrm{E}\left(\frac{\left(\sum_{i=1}^n w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{j-1}\right)^2}{\sum_{\kappa=1}^{n-j} C_{\kappa,j}}\right) \sigma_j^2 f_{j+1}^2 \cdots f_{n-1}^2.
\end{aligned}$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

We estimate

$$\mathrm{E}\left(\frac{\left(\sum_{i=1}^{n} w_i C_{i,1}\widehat{f}_1\cdots\widehat{f}_{j-1}\right)^2}{\sum_{\kappa=1}^{n-j} C_{\kappa,j}}\right) \quad \text{by} \quad \frac{\left(\sum_{i=1}^{n} w_i C_{i,1}\widehat{f}_1\cdots\widehat{f}_{j-1}\right)^2}{\sum_{\kappa=1}^{n-j} C_{\kappa,j}},$$

replace f_j and σ_j^2 by their estimators $\widehat{f}_j, \widehat{\sigma}_j^2$ (for $j \geq 1$) and $\widehat{f}_0^{\mathbf{w}}, \left(\widehat{\sigma}_0^{\mathbf{w}}\right)^2$ (for $j = 0$) and obtain

$$\begin{aligned}
\widehat{\mathrm{Var}}&\left(\sum_{i=1}^{n} w_i C_{i,1}\widehat{f}_1\cdots\widehat{f}_{n-1}\right)\\
&= \sum_{i=1}^{n} w_i^2 \widehat{v}_i \left(\widehat{\sigma}_0^{\mathbf{w}}\right)^2 \widehat{f}_1^2\ldots\widehat{f}_{n-1}^2\\
&\quad + \sum_{j=1}^{n-1}\left(\sum_{i=1}^{n} w_i C_{i,1}\widehat{f}_1\cdots\widehat{f}_{j-1}\right)^2 \frac{\widehat{\sigma}_j^2}{\sum_{\kappa=1}^{n-j} C_{\kappa,j}}\widehat{f}_{j+1}^2\cdots\widehat{f}_{n-1}^2.
\end{aligned}$$

This concludes the deduction of Estimator 5.2. For Estimator 5.3 we only derive the formula for

$$\widehat{\mathrm{Var}}\left(\sum_{i=1}^{n} w_i C_{i,1}\widehat{f}_1^{\mathbf{w}}\cdots\widehat{f}_{n-1}^{\mathbf{w}}\right).$$

The rest is completely analogous to Estimator 5.2. Recall from Remark 4.1 that

$$\widehat{f}_j^{\mathbf{w}} = \sum_{i=1}^{n-j}\frac{\lambda_{i,j}^{\mathbf{w}}}{\sum_{k=1}^{n-j}\lambda_{k,j}^{\mathbf{w}}}\cdot\frac{C_{i,j+1}}{C_{i,j}} \quad \text{with} \quad \lambda_{i,j}^{\mathbf{w}} = \left(w_i + \frac{\sum_{k=n-j+1}^{n} w_k \widehat{C}_{k,j}}{\sum_{k=1}^{n-j} C_{k,j}}\right) C_{i,j}.$$

Using

$$\mathrm{Var}(\widehat{f}_j^{\mathbf{w}} \mid \mathcal{D}_j) = \sum_{i=1}^{n-j}\frac{(\lambda_{i,j}^{\mathbf{w}})^2}{(\sum_{k=1}^{n-j}\lambda_{i,j}^{\mathbf{w}})^2}\mathrm{Var}\left(\frac{C_{i,j+1}}{C_{i,j}}\,\middle|\,\mathcal{D}_j\right) = \frac{\sum_{i=1}^{n-j}(\lambda_{i,j}^{\mathbf{w}})^2/C_{i,j}}{(\sum_{k=1}^{n-j}\lambda_{i,j}^{\mathbf{w}})^2}\sigma_j^2$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

we obtain

$$
\begin{aligned}
&\operatorname{Var}\left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_j^{\mathbf{w}}\right) \\
&\quad = \operatorname{Var}\left(\mathrm{E}\left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_j^{\mathbf{w}} \,\middle|\, \mathcal{D}_j\right)\right) \\
&\qquad + \mathrm{E}\left(\operatorname{Var}\left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_j^{\mathbf{w}} \,\middle|\, \mathcal{D}_j\right)\right) \\
&\quad = \operatorname{Var}\left(f_j \sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_{j-1}^{\mathbf{w}}\right) \\
&\qquad + \mathrm{E}\left(\left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_{j-1}^{\mathbf{w}}\right)^2 \operatorname{Var}(\widehat{f}_j^{\mathbf{w}} \mid \mathcal{D}_j)\right) \\
&\quad = f_j^2 \operatorname{Var}\left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_{j-1}^{\mathbf{w}}\right) \\
&\qquad + \mathrm{E}\left(\left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_{j-1}^{\mathbf{w}}\right)^2 \frac{\sum_{i=1}^{n-j} (\lambda_{i,j}^{\mathbf{w}})^2 / C_{i,j}}{(\sum_{\kappa=1}^{n-j} \lambda_{i,j}^{\mathbf{w}})^2} \sigma_j^2\right) \\
&\quad = f_j^2 \operatorname{Var}\left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_{j-1}^{\mathbf{w}}\right) \\
&\qquad + \mathrm{E}\left(\left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_{j-1}^{\mathbf{w}}\right)^2 \frac{\sum_{i=1}^{n-j} (\lambda_{i,j}^{\mathbf{w}})^2 / C_{i,j}}{(\sum_{\kappa=1}^{n-j} \lambda_{i,j}^{\mathbf{w}})^2}\right) \sigma_j^2,
\end{aligned}
$$

and by induction

$$
\begin{aligned}
\operatorname{Var}\left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1 \cdots \widehat{f}_{n-1}\right) &= \sum_{i=1}^{n} w_i^2 v_i \sigma_0^2 f_1^2 \cdots f_{n-1}^2 \\
+ \sum_{j=1}^{n-1} \mathrm{E}\left(\left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_{j-1}^{\mathbf{w}}\right)^2 \frac{\sum_{i=1}^{n-j} (\lambda_{i,j}^{\mathbf{w}})^2 / C_{i,j}}{(\sum_{\kappa=1}^{n-j} \lambda_{i,j}^{\mathbf{w}})^2}\right) &\sigma_j^2 f_{j+1}^2 \cdots f_{n-1}^2.
\end{aligned}
$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

We estimate

$$\mathrm{E}\left(\left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_{j-1}^{\mathbf{w}}\right)^2 \frac{\sum_{i=1}^{n-j} (\lambda_{i,j}^{\mathbf{w}})^2 / C_{i,j}}{(\sum_{\kappa=1}^{n-j} \lambda_{i,j}^{\mathbf{w}})^2}\right)$$

by

$$\left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_{j-1}^{\mathbf{w}}\right)^2 \frac{\sum_{i=1}^{n-j} (\lambda_{i,j}^{\mathbf{w}})^2 / C_{i,j}}{(\sum_{\kappa=1}^{n-j} \lambda_{i,j}^{\mathbf{w}})^2},$$

replace f_j and σ_j^2 by the estimators $\widehat{f}_j^{\mathbf{w}}$ and $\widehat{\sigma}_j^2$ and obtain

$$\begin{aligned}\widehat{\mathrm{Var}}\left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_{n-1}^{\mathbf{w}}\right) &= \sum_{i=1}^{n} w_i^2 \widehat{v}_i (\widehat{\sigma}_0^{\mathbf{w}})^2 (\widehat{f}_1^{\mathbf{w}})^2 \ldots (\widehat{f}_{n-1}^{\mathbf{w}})^2 \\ &+ \sum_{j=1}^{n-1} \left(\sum_{i=1}^{n} w_i C_{i,1} \widehat{f}_1^{\mathbf{w}} \cdots \widehat{f}_{j-1}^{\mathbf{w}}\right)^2 \frac{\sum_{i=1}^{n-j} (\lambda_{i,j}^{\mathbf{w}})^2 / C_{i,j}}{(\sum_{\kappa=1}^{n-j} \lambda_{i,j}^{\mathbf{w}})^2} \widehat{\sigma}_j^2 (\widehat{f}_{j+1}^{\mathbf{w}})^2 \cdots (\widehat{f}_{n-1}^{\mathbf{w}})^2.\end{aligned}$$

Appendix C

Moments of the reciprocal truncated normal distribution

For the calculation of the mean squared error we need an estimator for mean and variance of the random variable

$$\frac{\sum_{i=1}^{n} w_i v_i}{\sum_{i=1}^{n} w_i \widehat{v}_i}.$$

In our simple time series model for restatement uncertainty we have

$$\sum_{i=1}^{n} w_i \widehat{v}_i = \sum_{i=1}^{n} w_i v_i + \sum_{i=1}^{n} w_i v_i \sum_{j=i}^{n+1} e_j = \sum_{i=1}^{n} w_i v_i + \sum_{j=1}^{n+1} e_j \sum_{i=1}^{n \wedge j} w_i v_i.$$

Since the error terms e_i are normally distributed and independent $\sum_{i=1}^{n} w_i \widehat{v}_i$ is normally distributed with mean $\sum_{i=1}^{n} w_i v_i$ and

$$\mathrm{Var}\left(\sum_{i=1}^{n} w_i \widehat{v}_i\right) = \mathrm{Var}\left(\sum_{j=1}^{n+1} e_j \sum_{i=1}^{n \wedge j} w_i v_i\right) = \sum_{j=1}^{n+1} \varepsilon_j^2 \left(\sum_{i=1}^{n \wedge j} w_i v_i\right)^2,$$

i.e. the reciprocal random variable

$$\frac{\sum_{i=1}^{n} w_i \widehat{v}_i}{\sum_{i=1}^{n} w_i v_i},$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

of $(\sum_{i=1}^{n} w_i v_i)/(\sum_{i=1}^{n} w_i \widehat{v}_i)$ is normally distributed with mean 1. As mentioned by Frishman (1971), in the real world, few if any, random variables are truly normally distributed since they typically cannot range from $-\infty$ to $+\infty$. In practice, the normal distribution is always only an approximation and can at best be a good approximation around the mean of the data. Therefore, it is typically not critical to replace a normal distribution by a truncated normal distribution. In our case, an adequate truncation is definitely no issue since a pricing actuary would notice extremely large deviations from the mean (e.g. if $\sum_{i=1}^{n} w_i \widehat{v}_i$ is close to zero or even negative). In this case, the method would not be used.

Therefore, we are interested in mean and variance of the reciprocal truncated normal distribution. Let X_σ be normally distributed with mean $\mu = 1$ and standard deviation $\sigma > 0$, i.e. $X_\sigma \sim N(1, \sigma^2)$. Then the reciprocal variable $1/X_\sigma$ is not integrable due to the behavior at $\{X_\sigma = 0\}$. Let Φ_{1,σ^2} denote the CDF of X_σ and let $k > 0$. Let $\Psi_{1,\sigma^2}^{k\sigma}$ be the CDF of the distribution which is truncated from below at $1 - k\sigma$ and from above at $1 + k\sigma$, i.e.

$$\Psi_{1,\sigma^2}^{k\sigma}(x) := \frac{\Phi_{1,\sigma^2}(x) - \Phi_{1,\sigma^2}(1 - k\sigma)}{\Phi_{1,\sigma^2}(1 + k\sigma) - \Phi_{1,\sigma^2}(1 - k\sigma)} \quad \text{for } 1 - k\sigma \leq x \leq 1 + k\sigma.$$

Let $Y_\sigma^{k\sigma}$ be a random variable with CDF $\Psi_{1,\sigma^2}^{k\sigma}$. The random variable

$$Z_\sigma^{k\sigma} := \tau(Y_\sigma^{k\sigma} - 1) + 1,$$

with

$$\tau := \frac{\sigma}{\sqrt{\mathrm{Var}(Y_\sigma^{k\sigma})}},$$

has the same mean and standard deviation as X_σ. If k is not too small (e.g. $k \geq 3$) then $\tau \approx 1$ and the distributions of $Z_\sigma^{k\sigma}$ and X_σ are very similar around the mean. $Z_\sigma^{k\sigma}$ only ranges from $1 - k\tau\sigma$ to $1 + k\tau\sigma$. Let

$$Z_k := \frac{Z_\sigma^{k\sigma} - 1}{\sigma}.$$

If $k\tau\sigma < 1$ then mean and variance of $1/Z_\sigma^{k\sigma}$ can be expanded into power series

$$\mathrm{E}(1/Z_\sigma^{k\sigma}) = 1 + \sum_{i=2}^{\infty} \sigma^{2i}\, \mathrm{E}(Z_k^{2i}),$$

$$\mathrm{Var}(1/Z_\sigma^{k\sigma}) = \sigma^2 + \sum_{i=2}^{\infty} \sigma^{2i}\, \mathrm{Var}(Z_k^i) + 2\sum_{i=1}^{\infty} \sum_{j=i+1}^{\infty} \sigma^{i+j}(-1)^{i+j}\, \mathrm{Cov}(Z_k^i, Z_k^j),$$

(see Frishman (1971)). We use $k = 3$. Then for $0 \leq \sigma \leq 0.3$ the first terms of the power series provide good approximations

$$\mathrm{E}(1/Z_\sigma^{3\sigma}) \approx 66\sigma^8 + 12\sigma^6 + 3\sigma^4 + \sigma^2 + 1 = \varphi(\sigma^2)$$
$$\mathrm{Var}(1/Z_\sigma^{3\sigma}) \approx 29297\sigma^{12} + 3492\sigma^{10} + 430\sigma^8 + 55\sigma^6 + 7\sigma^4 + \sigma^2 = \psi(\sigma^2),$$

with

$$\varphi(x) := 66x^4 + 12x^3 + 3x^2 + x + 1 \quad \text{and}$$
$$\psi(x) := 29297x^6 + 3492x^5 + 430x^4 + 55x^3 + 7x^2 + x.$$

Let $\overline{Z}_\sigma^u$ denote the random variable obtained from X_σ by truncation from below at $u > 0$. For $0 < \sigma < 0.3$ and $0.05 \leq u \leq 1 - 3\sigma$ numerical calculations show that

$$\mathrm{E}(1/\overline{Z}_\sigma^u) \approx \mathrm{E}(1/Z_\sigma^{3\sigma}) \quad \text{and} \quad \mathrm{Var}(1/\overline{Z}_\sigma^u) \approx \mathrm{Var}(1/Z_\sigma^{3\sigma}),$$

i.e. the result does not depend materially on the choice of the truncation.

Applying these results to our situation, we obtain

$$\mathrm{E}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i}\right) \approx \varphi\left(\mathrm{Var}\left(\frac{\sum_{i=1}^n w_i \widehat{v}_i}{\sum_{i=1}^n w_i v_i}\right)\right) = \varphi\left(\frac{\sum_{j=1}^{n+1} \varepsilon_j^2 \left(\sum_{i=1}^{n\wedge j} w_i v_i\right)^2}{\left(\sum_{i=1}^n w_i v_i\right)^2}\right)$$

and

$$\mathrm{Var}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i}\right) \approx \psi\left(\mathrm{Var}\left(\frac{\sum_{i=1}^n w_i \widehat{v}_i}{\sum_{i=1}^n w_i v_i}\right)\right) = \psi\left(\frac{\sum_{j=1}^{n+1} \varepsilon_j^2 \left(\sum_{i=1}^{n\wedge j} w_i v_i\right)^2}{\left(\sum_{i=1}^n w_i v_i\right)^2}\right).$$

If we replace the parameters ε_i and v_i by their estimators $\widehat{\varepsilon}_i$ and $\widehat{v}_i$ we obtain the estimators

$$\mathrm{E}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i}\right) \approx \Phi(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}) \quad \text{and} \quad \mathrm{Var}\left(\frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i \widehat{v}_i}\right) \approx \Psi(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}).$$

Appendix D

A technical lemma from linear algebra

In this appendix, we provide a technical lemma from linear algebra that guarantees the asserted properties of the variance matrices in Sections 6 and 7.

Lemma D.1. *For $k = 1, \ldots, n$ let*

$$\mathbf{A}_k := \begin{pmatrix} a_1 & a_2 & \cdots & a_{k-1} & a_k \\ a_2 & a_2 & \cdots & a_{k-1} & a_k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{k-1} & a_{k-1} & \cdots & a_{k-1} & a_k \\ a_k & a_k & \cdots & a_k & a_k \end{pmatrix} \quad \textit{and} \quad \mathbf{D}_k := \mathrm{diag}(d_1, \ldots, d_k),$$

with $a_1 > a_2 > \cdots > a_n > 0$ *and* $d_1, \ldots, d_n \geq 0$. *Then* $\mathbf{A}_n + \mathbf{D}_n$ *is positive definite. Moreover, the components of* $\mathbf{w}^{(n)} := (\mathbf{A}_n + \mathbf{D}_n)^{-1}\mathbb{1}_n$ *are non-negative.*

Proof. We use induction to show that $\det(\mathbf{A}_k) > 0$ for $k = 1, \ldots, n$. We have $\det(\mathbf{A}_1) = a_1 > 0$. Assume that $\det(\mathbf{A}_{k-1}) > 0$. We consider $\mathbf{A}_k$ and subtract $\frac{a_k}{a_{k-1}}$ times row $k-1$ from row k. Then we obtain

$$\det(\mathbf{A}_k) = \det \left(\begin{array}{ccc|c} & \mathbf{A}_{k-1} & & \begin{matrix} * \\ \vdots \\ * \end{matrix} \\ \hline 0 & \cdots & 0 & a_k - \frac{a_k}{a_{k-1}} a_k \end{array} \right) = \det(\mathbf{A}_{k-1}) \left(a_k - \frac{a_k}{a_{k-1}} a_k \right) > 0.$$

Since the leading principal minors $\det(\mathbf{A}_k)$ of $\mathbf{A}_n$ are all positive, we conclude that $\mathbf{A}_n$ is positive definite.

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be an orthonormal eigenbasis of $\mathbf{A}_n$ and $\lambda_1, \ldots, \lambda_n > 0$ the corresponding eigenvalues. Since $d_i \geq 0$, we obtain for an arbitrary non-zero vector $\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$

$$\mathbf{v}^t(\mathbf{A}_n + \mathbf{D}_n)\mathbf{v} = \underbrace{\lambda_1 x_1^2 + \cdots + \lambda_n x_n^2}_{>0} + \underbrace{\mathbf{v}^t\mathbf{D}_n\mathbf{v}}_{\geq 0} > 0,$$

i.e. $\mathbf{A}_n + \mathbf{D}_n$ is positive definite.

We use induction to show that the components of $\mathbf{w}^{(k)} := (\mathbf{A}_k + \mathbf{D}_k)^{-1}\mathbb{1}_k$ are non-negative for $k = 1, \ldots, n$. For $k = 1$ this is obvious. Assume that the components of $\mathbf{w}^{(k-1)}$ are non-negative. From the first $k-1$ equations of the system $(\mathbf{A}_k + \mathbf{D}_k)\mathbf{w}^{(k)} = \mathbb{1}_k$ we see that

$$\mathbf{w}^{(k)} = \begin{pmatrix} (1 - a_k\lambda)\mathbf{w}^{(k-1)} \\ \\ \lambda \end{pmatrix},$$

with a unique $\lambda \in \mathbb{R}$. Since the components of $\mathbf{w}^{(k-1)}$ are non-negative it remains to show that $0 \leq \lambda \leq 1/a_k$. We define

$$\chi(\lambda) := \langle a_k\mathbb{1}_{k-1}\,,\ (1 - a_k\lambda)\mathbf{w}^{(k-1)} \rangle + (a_k + d_k)\lambda,$$

where $\langle\, \cdot\,,\ \cdot\, \rangle$ denotes the standard scalar product of $\mathbb{R}^{k-1}$. Then the last equation of the system $(\mathbf{A}_k + \mathbf{D}_k)\mathbf{w}^{(k)} = \mathbb{1}_k$ can be written as

$$\chi(\lambda) = 1.$$

We have

$$\chi(0) = \langle a_k\mathbb{1}_{k-1}\,,\ \mathbf{w}^{(k-1)} \rangle < 1 \quad \text{and} \quad \chi(1/a_k) = \frac{a_k + d_k}{a_k} \geq 1.$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

Since $\lambda \mapsto \chi(\lambda)$ is continuous, the intermediate value theorem guarantees the existence of a $\lambda \in (0, 1/a_k]$ with $\chi(\lambda) = 1$. $\square$

Remark D.2. *The assertion of Lemma D.1 remains true if we assume* $a_1 \geq \cdots \geq a_n > 0$ *and* $d_1, \ldots, d_n > 0$ *instead of* $a_1 > \cdots > a_n > 0$ *and* $d_1, \ldots, d_n \geq 0$.

APPENDIX E

OPTIMAL WEIGHTS FOR CONVEX COMBINATIONS OF CORRELATED ESTIMATORS

For independent and unbiased estimators T_i of a parameter t it is well known that the weights of T_i in a convex combination should be chosen in proportion to $1/\operatorname{Var}(T_i)$ (see e.g. Mack (2002)). In this appendix we provide the corresponding result for correlated estimators since it is probably less well known.

Lemma E.1. *Let* $T_1, \ldots, T_n$ *be unbiased estimators for* t*, let* $\mathbf{T} := (T_1, \ldots, T_n)^t$ *and assume that the covariance matrix*

$$\mathbf{S} := \operatorname{Cov}(\mathbf{T}),$$

is positive definite. Then a linear combination $g_1 T_1 + \cdots + g_n T_n$ *with* $g_1 + \cdots + g_n = 1$ *has minimal variance (amongst all such linear combinations) if and only if* $\mathbf{g} = (g_1, \ldots, g_n)^t$ *satisfies*

$$\mathbf{g} = \frac{\mathbf{S}^{-1}\mathbb{1}_n}{\langle \mathbf{S}^{-1}\mathbb{1}_n, \mathbb{1}_n \rangle},$$

where $\langle \cdot, \cdot \rangle$ *denotes the standard scalar product in* $\mathbb{R}^n$.

Proof. Let $f(g_1, \ldots, g_n) := \operatorname{Var}(\sum_{i=1}^n g_i T_i)$. We calculate the minimum of f under the boundary condition $\varphi(g_1, \ldots, g_n) := \sum_{i=1}^n g_i = 1$. We have

$$\operatorname{grad}_f(\mathbf{g}) = 2\mathbf{S}\mathbf{g} \quad \text{and} \quad \mathbf{H}_f(\mathbf{g}) = 2\mathbf{S},$$

(where $\mathbf{H}_f$ denotes the Hessian matrix of f). Moreover, we have

$$\operatorname{grad}_\varphi(\mathbf{g}) = \mathbb{1}_n.$$

The Lagrange condition for a local extremum is

$$\mathbf{S}\mathbf{g} = \lambda \mathbb{1}_n,$$

i.e.

$$\mathbf{g} = \lambda \mathbf{S}^{-1}\mathbb{1}_n$$

where λ is chosen such that $\sum_{i=1}^n g_i = 1$, i.e.

$$\lambda = \frac{1}{\langle \mathbf{S}^{-1}\mathbb{1}_n, \mathbb{1}_n \rangle}.$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

Note that $\langle \mathbf{S}^{-1}\mathbb{1}_n, \mathbb{1}_n\rangle \neq 0$ since $\mathbf{S}$ is positive definite. As $\mathbf{H}_f = 2\mathbf{S}$ is positive definite, we see that f has a local minimum at $\mathbf{g}$. It is a unique global minimum since it is the only local minimum and since $f(\mathbf{g}) \to +\infty$ for $\|\mathbf{g}\|_2 \to +\infty$. $\square$

Remark E.2. *If the components of* $\mathbf{S}^{-1}\mathbb{1}_n$ *are non-negative then we have*

$$\frac{\mathbf{S}^{-1}\mathbb{1}_n}{\langle \mathbf{S}^{-1}\mathbb{1}_n, \mathbb{1}_n\rangle} = \frac{\mathbf{S}^{-1}\mathbb{1}_n}{\|\mathbf{S}^{-1}\mathbb{1}_n\|_1},$$

where $\|\cdot\|_1$ *denotes the 1-norm in* $\mathbb{R}^n$.

Appendix F

Convergence of the recursion from Section 6

In this appendix, we first show that Recursion 6.1 converges to a fixed point if the loss ratios $C_{i,1}/\widehat{v}_i$ do not fluctuate too much. Afterwards we use a numerical calculation to illustrate this result for the case $n = 10$, $\widehat{v}_1 = \cdots = \widehat{v}_q$ and $\widehat{\boldsymbol{\varepsilon}} = \widehat{\varepsilon}\mathbb{1}_{n+1}$.

Lemma F.1. *Assume that restated premiums* $\widehat{v}_1, \ldots, \widehat{v}_q$ *and estimators* $\widehat{c}$, $\widehat{x}$ *and* $\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_{n+1}$ *are given. For arbitrary* $a > 0$ *there exists* $b > a$ *with the following property: If* $C_{i,1}/\widehat{v}_i \in [a, b]$ *for* $i = 1, \ldots, n$ *then Recursion 6.1 converges to a fixed point.*

Proof. Let

$$\mathbb{R}^n_+ := \{(w_1, \ldots, w_n) \mid w_i \geq 0, \sum_{i=1}^n w_i > 0\} \quad \text{and}$$

$$\Delta_{n-1} := \{(w_1, \ldots, w_n) \mid w_i \geq 0, \sum_{i=1}^n w_i = 1\}.$$

We define

$$F\colon \mathbb{R}^n_+ \to \mathbb{R},\ \mathbf{w} \mapsto \widehat{f_0^{\mathbf{w}}} = \frac{\sum_{i=1}^n w_i C_{i,1}}{\sum_{i=1}^n w_i \widehat{v}_i},$$

and

$$f := F|\Delta_{n-1}.$$

Moreover, we define $\mathbf{S}^*(t) := (s^*_{i,j}(t))_{i,j=1,\ldots,n}$ with

$$s^*_{i,i}(t) := t^2(\widehat{\varepsilon}_i^2 + \cdots + \widehat{\varepsilon}_{n+1}^2),$$

and

$$s^*_{i,j}(t) := t^2(\widehat{\varepsilon}_i^2 + \cdots + \widehat{\varepsilon}_{n+1}^2) + \frac{t(1+\widehat{c}^2)\widehat{x}}{\widehat{v}_i} \cdot (1 + \widehat{\varepsilon}_i^2 + \cdots + \widehat{\varepsilon}_{n+1}^2),$$

for $i \neq j$. Eventually, we consider the curve

$$\alpha\colon (0,\infty) \to \Delta_{n-1},\ t \mapsto \frac{\widehat{\mathbf{V}}^{-1}\mathbf{S}^*(t)^{-1}\mathbb{1}_n}{\|\widehat{\mathbf{V}}^{-1}\mathbf{S}^*(t)^{-1}\mathbb{1}_n\|_1}.$$

Since

$$\widehat{\mathbf{S}}(\widehat{\boldsymbol{\varepsilon}}, \mathbf{w}) = \mathbf{S}^*(\widehat{f_0^{\mathbf{w}}}),$$

we then have

$$\mathbf{w}^{(k+1)} = (\alpha \circ F)(\mathbf{w}^{(k)}).$$

If we can show that the mapping $\alpha \circ f\colon \Delta_{n-1} \to \Delta_{n-1}$ is contracting, then we can apply Brewer's fixed point theorem to see that Recursion 6.1 converges to a fixed point.

The continuous function $\|\dot{\alpha}\|_2$ assumes a global maximum C on the compact interval $[a, 2a]$. Let $m := \max\{\widehat{v}_i/\widehat{v}_j \mid i, j = 1, \ldots, n\}$ and choose $b \in (a, 2a]$ such that

$$b - a \leq \frac{1}{Cm\sqrt{n-1}}.$$

We have

$$\operatorname{grad}_F(\mathbf{w}) = \frac{1}{\sum_{j=1}^n w_j \widehat{v}_j} \begin{pmatrix} \widehat{v}_1 \left(\frac{C_{1,1}}{\widehat{v}_1} - \widehat{f_0^{\mathbf{w}}}\right) \\ \vdots \\ \widehat{v}_n \left(\frac{C_{n,1}}{\widehat{v}_n} - \widehat{f_0^{\mathbf{w}}}\right) \end{pmatrix}.$$

For $\mathbf{w} \in \Delta_{n-1}$ the gradient $\operatorname{grad}_f(\mathbf{w})$ is the orthogonal projection of $\operatorname{grad}_F(\mathbf{w})$ to the tangent space $T_{\mathbf{w}}\Delta_{n-1}$. We therefore have

$$\|\operatorname{grad}_f(\mathbf{w})\|_2 \leq \|\operatorname{grad}_F(\mathbf{w})\|_2 = \sqrt{\frac{1}{(\sum_{j=1}^n w_j \widehat{v}_j)^2} \sum_{i=1}^n \widehat{v}_i^2 \left(\frac{C_{i,1}}{\widehat{v}_i} - \widehat{f_0^{\mathbf{w}}}\right)^2}.$$

Let $a_0 := \min\{C_{i,1}/\widehat{v}_i \mid i = 1, \ldots, n\}$, $b_0 := \max\{C_{i,1}/\widehat{v}_i \mid i = 1, \ldots, n\}$. The function

$$\chi : [a_0, b_0] \to [0, \infty),\ x \mapsto \sum_{i=1}^n \left(\frac{C_{i,1}}{\widehat{v}_i} - t\right)^2$$

is convex and therefore $\chi(t) \leq \max\{\chi(a_0), \chi(b_0)\}$ for all $t \in [a_0, b_0]$. We obtain

$$\sum_{i=1}^n \left(\frac{C_{i,1}}{\widehat{v}_i} - \widehat{f_0^{\mathbf{w}}}\right)^2 \leq (n-1)(b_0 - a_0)^2.$$

Since $\widehat{v}_i^2/(\sum_{i=1}^n w_i \widehat{v}_i)^2 \leq m^2$ for $\mathbf{w} \in \Delta_{n-1}$ we conclude that

$$\|\operatorname{grad}_f(\mathbf{w})\|_2 \leq m\sqrt{n-1}(b_0 - a_0).$$

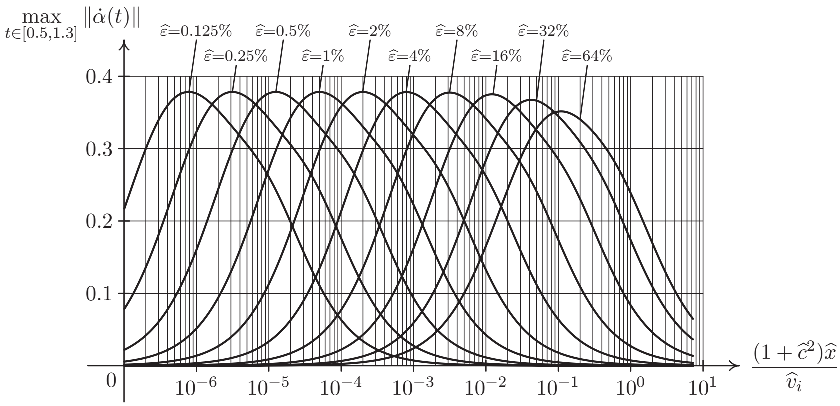

FIGURE 9: Plot of $(1+\widehat{c}^2)\widehat{x}/\widehat{v}_i \mapsto \max_{t\in[0.5,1.3]}\{\|\dot\alpha(t)\|_2\}$ for $n=10$, $\widehat{v}_1=\cdots=\widehat{v}_q$ and various $\widehat{\boldsymbol\varepsilon}=\widehat{\varepsilon}\mathbb{1}_{n+1}$.

If $C_{i,1}/\widehat{v}_i \in [a,b]$ for all i, i.e. $[a_0,b_0]\subset[a,b]$, then we have

$$\|\dot\alpha(f(\mathbf{w}))\|_2 \cdot \|\operatorname{grad}_f(\mathbf{w})\|_2 \le Cm\sqrt{n-1}(b-a)\le 1,$$

for all $\mathbf{w}\in\Delta_{n-1}$ which implies that $\alpha\circ f$ is contracting. $\square$

Lemma F.1 does not provide an indication on how large b can be chosen for given a. In order to obtain a better feeling we consider an example. Let $n=10$, $\widehat{v}_1=\cdots=\widehat{v}_n$ and $\widehat{\boldsymbol\varepsilon}=\widehat{\varepsilon}\mathbb{1}_{n+1}$. As lower bound for the loss ratios $C_{i,1}/\widehat{v}_i$ we choose $a=50\%$. We will use a numerical calculation to see that the recursion always converges if $C_{i,1}/\widehat{v}_i\in[50\%,130\%]$ for all i. Hence, we can choose $b=130\%$ in Lemma F.1. Note that this condition for $C_{i,1}/\widehat{v}_i$ is sufficient but not necessary. In practice, it is not easy to find an example where the recursion does not converge.

We use the notation from the proof of Lemma F.1. If we consider the definition of $\mathbf{S}^*(t)$ we see that the curve $\alpha\colon(0,\infty)\to\Delta_{n-1}$ depends on the parameters $(1+\widehat{c}^2)\widehat{x}/\widehat{v}_i$ and $\widehat\varepsilon$ (we omit these parameters in the notation). In Figure 9 we have plotted the graphs of

$$\frac{(1+\widehat{c}^2)\widehat{x}}{\widehat{v}_i}\mapsto \max_{t\in[0.5,1.3]}\|\dot\alpha(t)\|_2,$$

for a number of values for $\widehat\varepsilon$ on logarithmic paper. For small values of $\widehat\varepsilon$ we have

$$\frac{1+\widehat\varepsilon_i+\cdots+\widehat\varepsilon_{n+1}}{\widehat\varepsilon_i+\cdots+\widehat\varepsilon_{n+1}}\approx\frac{1}{\widehat\varepsilon_i+\cdots+\widehat\varepsilon_{n+1}},$$

which implies that a change of $\widehat\varepsilon$ results approximately in a shift of the graph (using logarithmic scale). From Figure 9 it seems to be quite obvious that

$$\max_{t\in[0.5,1.3]}\|\dot\alpha(t)\|_2\le 0.4=:C,$$

https://doi.org/10.1017/asb.2015.2 Published online by Cambridge University Press

for arbitrary $\widehat{x}$, $\widehat{c}$, $\widehat{v}_1 = \cdots = \widehat{v}_n$ and $\widehat{\varepsilon}$ (although this is not a strict proof of course). Since $\widehat{v}_1 = \cdots = \widehat{v}_n$ we have $m = 1$. Like in the proof of Lemma F.1 we obtain with $n = 10$

$$\|\dot{\alpha}(f(\mathbf{w})\|_2 \cdot \| \operatorname{grad}_f(\mathbf{w})\|_2 \leq Cm\sqrt{n-1}(130\% - 50\%) = 0.4 \cdot 1 \cdot 3 \cdot 0.8 = 0.96 < 1,$$

which implies that Recursion 6.1 converges to a fixed point.