On a combination of multiplicative and additive stochastic loss reserving methods

Philipp Arbenz, Robert Salzmann

Abstract We introduce the hybrid chain ladder (HCL) method, a distribution-free stochastic loss reserving method that allows for a weighted combination of two approaches. The first approach is data driven resembling the Chain-Ladder (CL) method. The second approach uses expert estimates of ultimate losses in a similar way as the Bornhuetter-Ferguson (BF) method. The HCL method provides a class of models that allows one to study mixtures of the two approaches. Since the CL method is susceptible to outliers whereas the BF method is very robust, mixing the two approaches becomes of practical relevance when the actuary has concerns about the quality of the data or knowledge of particular events that cause unusual effects. We give predictors for the ultimate claims and estimators for the prediction error and the uncertainty in the claims development result. An implementation of the method in an Excel spreadsheet is available at www.RiskLab.ch/hclmethod.

Keywords. Stochastic claims reserving, distribution-free method, Chain-Ladder, Bornhuetter-Ferguson, mean square error of prediction, claims development result.

1. INTRODUCTION

The Chain-Ladder (CL) and Bornhuetter-Ferguson (BF) methods are standard methods in claims reserving. When only the CL and BF methods are viable choices in reserving practice, choosing only one of them may not be adequate for a given data set. For instance, claims reserving triangles in which CL is appropriate for some accident years, whereas BF is suitable for the others, are frequently encountered in practice.

There are many different approaches to embed CL and BF into stochastic frameworks, which allow one to deduce statistically consistent claims predictors and assess the uncertainty of the claims development. However, stochastic model versions of the CL and BF methods generally impose incompatible assumptions on the claims development. For instance, claim increments are assumed to be independent in the BF model given in Mack (2008b), whereas they are correlated in the CL model given in Mack (1993). Therefore, switching between CL and BF on an accident year-wise basis may lead to inconsistencies.

In this paper we introduce the hybrid chain ladder (HCL) method, a class of distribution-free stochastic loss reserving models that allows for weighted combinations of two methodologies. The first methodology is multiplicative in structure, resembling the CL model as of Mack (1993), and relies purely on the claims data. The second methodology has an additive structure, similar to the BF model in Mack (2008b), and uses a prior estimate of ultimate losses. The weights between the two methodologies can be set differently for each accident and development year.

We have implemented HCL in a ready-to-use Excel spreadsheet which can be downloaded at

`http://www.RiskLab.ch/hclmethod`

This spreadsheet allows testing of all features of the method and provides several case studies including the one discussed in this paper.

Organisation of the paper. Section 2 motivates our modelling approach. Section 3 introduces the model and derives the structure of the mean and the variance of the claims process. Section 4 defines estimators of the unknown parameters and predictors for future claims. Section 5 and Section 6 give estimators of the uncertainty in ultimate claims predictions and the claims development result, respectively. We extend the model in Section 7 to incorporate uncertainty in the choice of prior ultimates. Section 8 discusses a possible choice of the remaining model parameters. A case study is presented in Section 9. We outline limitations and possible extensions in Section 10 and conclude in Section 11.

2. MOTIVATION

We first introduce some notation. Let $C_{i,j}$ denote the cumulative claims of accident year $i = 1,\ldots,I$ and development year $j = 0,\ldots,J$ with $J \le I - 1$. We assume that the ultimate claim for accident year i is given by $C_{i,J}$, i.e., there is no further development after development year J. Moreover, denote by

$$\mathrm{D}_I = \left\{ C_{i,j} : 1 \le i \le I, 0 \le j \le J, i + j \le I \right\},$$

the trapezoid of observations up to the I-th accounting year, illustrated in Figure 1.

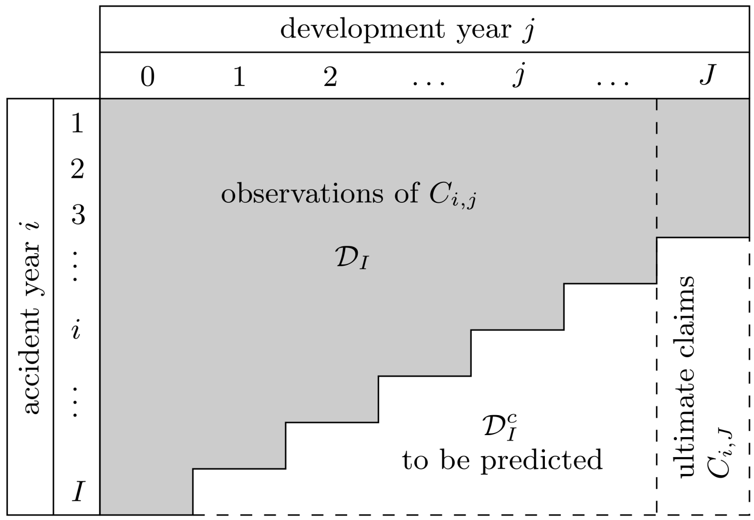

Figure 1: The trapezoid D_I of cumulative claims known in accounting year I. The lower right triangle is yet unknown and has to be predicted.

Stochastic reserving methods aim to predict the yet unknown lower right triangle D_I^c, i.e., $C_{i,j} \notin \mathrm{D}_I$, along with a measure of uncertainty. In particular we are interested in the prediction of the ultimate claim $C_{i,J}$ for each accident year $i \in \{I-J+1,\ldots,I\}$. For the distribution-free model versions of the CL and BF method considered here, this is done as follows. In brief, CL predicts $C_{i,J}$ by

$$\hat{C}_{i,J}^{CL} = C_{i,I-i}\hat{f}_{I-i+1}\cdots\hat{f}_J,$$

where the $\hat{f}_j$ are estimators of the age-to-age factors f_j. These factors indicate the average relative increase of the cumulative claims in one accident year from one development year to the next development year.

On the other hand, BF predicts $C_{i,J}$ by

$$\hat{C}_{i,J}^{BF} = C_{i,I-i} + \mu_i(\hat{\gamma}_{I-i+1} + \ldots + \hat{\gamma}_J),$$

where the μ_i is a prior estimate of $\mathrm{E}[C_{i,J}]$ and the $\hat{\gamma}_j$ are estimates of γ_j, which denote the fraction of claims expected in development year j.

The CL predictor has a multiplicative structure, whereas the BF predictor is additive. Moreover the two approaches represent two extreme positions of data reliance and expert opinion. These differences impose restrictions on the applicability of the two approaches. In practice, one often encounters claims reserving triangles in which CL is appropriate for some accident years, whereas BF is suitable for the others. As stated in Neuhaus (1992), this can be caused by data sparsity, e.g., triangles which have missing entries or zero cumulatives for recent accident years. For instance small or negative values of $C_{i,j}$ for early development years cause problems when applying CL as e.g., discussed in Busse et al. (2010). Another reason is illustrated in the following example. Suppose we are given a short-tailed insurance runoff triangle D_I. Assume the triangle behaves nicely to apply CL, except that in accident year k, the diagonal value $C_{k,I-k}$ is twice as large as $f_{I-k}C_{k,I-k-1}$. The latter value represents the expectation of $C_{k,I-k}$ in the CL model, given the information in the previous accounting year. In the following, we illustrate different possible underlying reasons for such an observation $C_{k,I-k}$, which imply different appropriate reserves.

- The increase is caused by a legal change leading to higher claims payments in all remaining development years of accident year k. In this case, CL is appropriate, as $\hat{C}_{k,J}^{CL}$ scales proportionally to $C_{k,I-k}$.
- The increase $C_{k,I-k} - f_{I-k}C_{k,I-k-1}$ is caused by a single event ("outlier"), which is not systematic and is not expected to happen again. In this case, the predicted ultimate should be increased by this difference, which is realised by BF.
- The increase is due to an exceptional early commutation, which causes a claim to be paid earlier than usual. In this case, the predicted ultimate from the previous accounting year should not be changed at all.

The reserving actuary often knows the underlying reasons for seemingly unusual behaviour in a triangle. However, such information cannot be extracted from D_I nor from prior estimates of the ultimate claims. Therefore, actuarial judgement is necessary and appropriate when choosing the reserving method.

Due to these reasons, it is common practice for reserving actuaries to switch accident year-wise between CL and BF, i.e., decide for each accident year separately whether the reserves are determined according to $\hat{C}_{i,J}^{CL}$ or $\hat{C}_{i,J}^{BF}$. Such approaches are easily applied, as both CL and BF can be quickly implemented in a spreadsheet. Commercial reserving software such as EMB ResQ™ and Milliman ReservePro® allow the user to set reserves as a weighted average between different reserving methodologies, with weights different for each accident year.

The crucial assumptions behind CL as given in Mack (1993) are that $\mathrm{E}[C_{i,j} \mid C_{i,j-1}] = f_j C_{i,j-1}$ and $var(C_{i,j} \mid C_{i,j-1}) = \sigma_j^2 C_{i,j-1}$, where the σ_j^2 are variance parameters. BF as given in Mack (2008b) assumes increments to be independent, such that $\mathrm{E}[C_{i,j} \mid C_{i,j-1}] = C_{i,j-1} + \mu_i \gamma_j$ and $var(C_{i,j} - C_{i,j-1}) = \sigma_j^2 \mu_i$. As a consequence, we can calculate the conditional covariance of consecutive increments:

$$cov\left(C_{i,j+2} - C_{i,j+1}, C_{i,j+1} - C_{i,j} \middle| C_{i,j}\right) = \begin{cases} \left(f_{j+2} - 1\right)\sigma_{j+1}^2 C_{i,j}, & \text{for CL,} \\ 0, & \text{for BF.} \end{cases}$$

This shows that the assumptions behind distribution-free model versions of the CL and BF method considered here are incompatible. However, these assumptions are applied to the whole triangle when estimating parameters. Thus, when switching accident year-wise between CL and BF, one cannot avoid applying two conflicting sets of assumptions on the triangle D_I.

Due to the different nature of the two methodologies, there is no straightforward stochastic representation of an accident year-wise weighting of CL and BF within a distribution-free model.

Reserving models that allow for a combination of CL-style and BF-style reserves have already been studied in the literature. A first credibility approach is given by the Benktander-Hovinen method, see Benktander (1976). In Neuhaus (1992) and Mack (2000) this credibility mixture of CL and BF is further studied but they do not specify a concrete parametric model and do not deduce a parameter estimation error. Alai (2010) provides a solution within a generalised linear model (GLM) framework. Bayesian approaches are proposed in Verrall (2004) and Section 4.3.2 of Wüthrich and Merz (2008), respectively.

The HCL method as proposed here provides a class of distribution-free models that allows for a weighting between a multiplicative and an additive behaviour. The following three aspects highlight the main differences between HCL and the models referenced above.

1) In the existing literature the weight between CL and BF is generally interpreted as a credibility weight, which is chosen to minimise the prediction uncertainty. This approach neglects special situations encountered in practice as discussed above. On the contrary, the HCL method assumes these weights to be known. In practice the reserving actuary determines through the choice of the weights

whether a multiplicative structure as in CL or an additive structure as in BF is more appropriate to predict the ultimate claim.

2) The HCL method provides a stochastic model that allows parameter estimation, claims prediction, the calculation of the conditional mean square error of prediction (MSEP), and the calculation of uncertainty in the claims development result (CDR). Furthermore, an easy-to-use implementation is available.

3) HCL is based on a distribution-free framework. This enables us to avoid assumptions needed in distribution based methods, such as positivity of incremental claims (in contrast to over-dispersed Poisson and GLM) or distributional assumption (in contrast to Bayesian methods). Distribution-free models are of limited use to answer questions about the tail behaviour of random quantities. However they are widely used and well understood in the insurance industry. They provide a first glance about the scale of the considered business and the corresponding uncertainty. As described in point 2) the HCL model provides all quantities that are usually considered when studying distribution-free models. This allows a direct comparison with other distribution-free models.

3. THE HCL MODEL

In this section we introduce the distribution-free Markov chain model defining the HCL model. Let $(\gamma_0, \gamma_1, \ldots, \gamma_J)$ denote the incremental claims pattern and define the cumulative pattern as $\beta_j = \sum_{k=0}^{j} \gamma_k$, $j = 0, \ldots, J$. We can then interpret the assumptions on the conditional expectation of an increment as

$$\mathrm{E}[C_{i,j} \mid C_{i,j-1}] = \beta_j / \beta_{j-1} C_{i,j-1} = (1 + \gamma_j / \beta_{j-1}) C_{i,j-1},$$

for CL and as

$$\mathrm{E}[C_{i,j} \mid C_{i,j-1}] = C_{i,j-1} + \gamma_j \mu_i,$$

for BF. The main idea of HCL is to take a weighted average of the above expressions

$$\begin{aligned} \mathrm{E}[C_{i,j} \mid C_{i,j-1}] &= \alpha_{i,j}\left(1+\frac{\gamma_j}{\beta_{j-1}}\right)C_{i,j-1} + (1-\alpha_{i,j})\left(C_{i,j-1}+\gamma_j\mu_i\right) \\ &= C_{i,j-1} + \gamma_j\left(\alpha_{i,j}\frac{C_{i,j-1}}{\beta_{j-1}} + (1-\alpha_{i,j})\mu_i\right), \end{aligned}$$

where $\alpha_{i,j}$ denotes the weight for accident year i and development year j.

Model 3.1 (HCL Model). *There exist parameters*

- γ_j *and* $\sigma_j^2 > 0$ *for* $j = 0,\ldots,J$,
- $\beta_j > 0$ *for* $j = 0,\ldots,J-1$,
- $\alpha_{i,j} \in [0,1]$ *for* $i = 1,\ldots,I$ *and* $j = 1,\ldots,J$,
- $\mu_i > 0$ *for* $i = 1,\ldots,I$,

such that $(C_{1,j})_{j=0,\ldots,J},\ldots,(C_{I,j})_{j=0,\ldots,J}$ *are independent Markov processes with*

- $\mathrm{E}[C_{i,0}] = \gamma_0\mu_i$ *and* $\mathrm{E}[C_{i,j} \mid C_{i,j-1}] = C_{i,j-1} + \gamma_j\left(\alpha_{i,j}\frac{C_{i,j-1}}{\beta_{j-1}} + (1-\alpha_{i,j})\mu_i\right)$ *for* $j = 1,\ldots,J$ *and*
- $var(C_{i,0}) = \sigma_0^2\mu_i$ *and* $var(C_{i,j} \mid C_{i,j-1}) = \sigma_j^2\mu_i$ *for* $j = 1,\ldots,J$. ■

Note that although the parameters γ_j and β_j are closely related, we assume them to be distinct and independent. The reason for this assumption is to allow a derivation of claims predictors and MSEP analogous to the deductions in Mack (1993) for CL. The precise relationship and the way to link the estimates of the γ_j and β_j are explained in Section 8.1.

The expectation $\mathrm{E}[C_{i,j} - C_{i,j-1} \mid C_{i,j-1}]$ of the incremental claim $C_{i,j} - C_{i,j-1}$ conditional on the previous year $C_{i,j-1}$ is equal to the product of γ_j and $m_{i,j}$, where

- γ_j is the fraction of the ultimate claim expected in development year j,
- $m_{i,j}$ is a volume measure of accident year i, defined by $m_{i,0} = \mu_i$ and

$$m_{i,j} = \alpha_{i,j}\frac{C_{i,j-1}}{\beta_{j-1}} + \left(1-\alpha_{i,j}\right)\mu_i \quad \text{for} \quad j > 0.$$

The $m_{i,j}$ represent a weighted average of

- μ_i, which is a prior estimate of the expected ultimate claim $\mathrm{E}[C_{i,J}]$ that is independent of the claims process, and
- $C_{i,j-1}/\beta_{j-1}$, which scales linearly in the previous observation $C_{i,j-1}$. The proportionality factor β_{j-1} is an estimate of the ratio of cumulative claims up to year $j-1$ to ultimate claims.

We have that, for $\alpha_{i,j}$ close to 0, $\mathrm{E}[C_{i,j} \mid C_{i,j-1}]$ inherits similar features as in the BF model of Mack (2008b), whereas for $\alpha_{i,j}$ close to 1, $\mathrm{E}[C_{i,j} \mid C_{i,j-1}]$ is similar to the distribution-free version of the CL model presented in Mack (1993). Indeed, we show below that in the extreme cases $\alpha_{i,j}=1$ or $\alpha_{i,j}=0$ for all $j=I-i+1,\ldots,J$ we have

$$\mathrm{E}[C_{i,J} \mid \mathrm{D}_I] = \mathrm{E}[C_{i,J} \mid C_{i,I-i}] = \begin{cases} C_{i,I-i} \displaystyle\prod_{I-i<j\le J} \frac{\beta_{j-1}+\gamma_j}{\beta_{j-1}}, & \text{if } \alpha_{i,j}=1, \\ C_{i,I-i} + \mu_i \displaystyle\sum_{I-i<j\le J} \gamma_j, & \text{if } \alpha_{i,j}=0. \end{cases}$$

In particular, for $\alpha_{i,j}=1$ we get a multiplicative structure as in the CL model whereas for $\alpha_{i,j}=0$ we get an additive structure in correspondence to BF-type models.

In Section 8.2 we give a discussion about the choice of the $\alpha_{i,j}$ which corresponds to a prior choice of the model within the class of HCL models. Compared to the situation where only CL and BF are admissible models, in the HCL model, the choice of the $\alpha_{i,j}$ replaces the selection process deciding between CL and BF.

The following theorems deduce the structure of mean and variance of the $C_{i,j}$. We identify empty sums with 0 and empty products with 1. To simplify notation, let

$$\xi_{i,j} = 1 + \alpha_{i,j} \frac{\gamma_j}{\beta_{j-1}} \quad \text{for} \quad i=1,\ldots,I,\ j=1,\ldots,J.$$

The $\xi_{i,j}$ take a similar role as the age-to-age factors f_j in the CL model in representing the influence of the realisation of $C_{i,j-1}$ on the expectation of $C_{i,j}$.

Theorem 3.2. *For* $0 \le k \le j \le J$ *we have*

$$\mathrm{E}[C_{i,j} \mid C_{i,k}] = C_{i,k} \prod_{k<m\le j} \xi_{i,m} + \mu_i \sum_{k<n\le j} \left((1-\alpha_{i,n})\gamma_n \prod_{n<m\le j} \xi_{i,m} \right). \tag{3.1}$$

Proof. The case $k = j$ is trivial. We assume by induction that (3.1) holds for some fixed j and k. Then

$$\begin{aligned}
\mathrm{E}[C_{i,j} \mid C_{i,k-1}] &= \mathrm{E}\left[\mathrm{E}[C_{i,j} \mid C_{i,k}] \middle| C_{i,k-1}\right] \\
&= \mathrm{E}\left[C_{i,k} \prod_{k<m\leq j} \xi_{i,m} + \mu_i \sum_{k<n\leq j} \left((1-\alpha_{i,n})\gamma_n \prod_{n<m\leq j} \xi_{i,m}\right) \middle| C_{i,k-1}\right] \\
&= \left(\gamma_k (1-\alpha_{i,k})\mu_i + C_{i,k-1}\xi_{i,k}\right) \prod_{k<m\leq j} \xi_{i,m} + \mu_i \sum_{k<n\leq j} \left((1-\alpha_{i,n})\gamma_n \prod_{n<m\leq j} \xi_{i,m}\right) \\
&= C_{i,k-1} \prod_{k-1<m\leq j} \xi_{i,m} + \mu_i \sum_{k-1<n\leq j} \left((1-\alpha_{i,n})\gamma_n \prod_{n<m\leq j} \xi_{i,m}\right).
\end{aligned}$$

■

Other possibilities to express the conditional expectation $\mathrm{E}[C_{i,j} \mid C_{i,k}]$ are

$$\mathrm{E}[C_{i,j} \mid C_{i,k}] = C_{i,k} + \sum_{k<n\leq j} \gamma_n \left((1-\alpha_{i,n})\mu_i + \alpha_{i,n} \frac{\mathrm{E}[C_{i,n-1} \mid C_{i,k}]}{\beta_{n-1}}\right),$$

$$\mathrm{E}[C_{i,j} \mid C_{i,k}] = C_{i,k} + \sum_{k<n\leq j} \gamma_n \left(\alpha_{i,n} \frac{C_{i,k}}{\beta_{n-1}} + (1-\alpha_{i,n})\mu_i\right) \prod_{n<m\leq j} \xi_{i,m}.$$

Theorem 3.3. *For* $0 \leq k \leq j \leq J$ *we have*

$$var(C_{i,j} \mid C_{i,k}) = \mu_i \sum_{k<n\leq j} \sigma_n^2 \prod_{n<m\leq j} \xi_{i,m}^2. \tag{3.2}$$

Proof. The case $k = j$ is trivial. We assume by induction that (3.2) holds for some j and k. By the law of total variance we have

$$\begin{aligned}
var(C_{i,j} \mid C_{i,k-1}) &= \mathrm{E}\left[var(C_{i,j} \mid C_{i,k}) \middle| C_{i,k-1}\right] + var\left(\mathrm{E}[C_{i,j} \mid C_{i,k}] \middle| C_{i,k-1}\right) \\
&= \mathrm{E}\left[\mu_i \sum_{k<n\leq j} \sigma_n^2 \prod_{n<m\leq j} \xi_{i,m}^2 \middle| C_{i,k-1}\right] + var\left(C_{i,k} \prod_{k<m\leq j} \xi_{i,m} \middle| C_{i,k-1}\right) \\
&= \mu_i \sum_{k<n\leq j} \sigma_n^2 \prod_{n<m\leq j} \xi_{i,m}^2 + \mu_i \sigma_k^2 \left(\prod_{k<m\leq j} \xi_{i,j}\right)^2 \\
&= \mu_i \sum_{k-1<n\leq j} \sigma_n^2 \prod_{n<m\leq k} \xi_{i,m}^2.
\end{aligned}$$

■

We see that mean and variance are similar in structure to the CL model, in the sense that they can be expressed in terms of some volume terms inflated by factors $\xi_{i,j}$ (resembling age-to-age factors). Similar results are also obtained in the model given in Schnieper (1991), see also Section 10.2 in Wüthrich and Merz (2008).

Note that for fixed i, the variances $var(C_{i,j} \mid C_{i,j-1})$ all have the same proportionality factor μ_i. Hence the σ_j^2 can be directly compared, unlike in the CL model, where the proportionality factors are different.

4. PARAMETER ESTIMATION AND CLAIMS PREDICTION

In the following section we define estimators for γ_j and σ_j^2 based on the information given in D_I. These will imply a predictor of $\mathrm{C}_{i,J}$. For the deduction of estimators of the model parameters and other quantities, we will first assume that the parameters β_j, μ_i and $\alpha_{i,j}$ are known constants and hence $m_{i,j}$ can be calculated, conditionally given $C_{i,j-1}$. This assumption on the μ_i will be weakened in Section 7 and the estimation of the β_j and an approach to setting the $\alpha_{i,j}$ will be illustrated in Section 8.

First, we define $\Gamma_{i,0} = C_{i,0} / \mu_i$ and

$$\Gamma_{i,j} = \frac{C_{i,j} - C_{i,j-1}}{m_{i,j}} \quad \text{for} \quad j>0. \tag{4.1}$$

Moreover, we define the weights $\omega_{i,j}$ where $\omega_{i,0} = \mu_i$ and

$$\omega_{i,j} = \frac{\sigma_j^2}{\mathrm{var}(\Gamma_{i,j} \mid C_{i,j-1})} = \frac{m_{i,j}^2}{\mu_i} \quad \text{for} \quad j>0.$$

In case there are i and j such that $C_{i,j-1} < 0$, we assume that $\alpha_{i,j}$ is small enough (e.g. $\alpha_{i,j} = 0$) to ensure $m_{i,j} > 0$.

In the following, conditioning on $C_{i,-1}$ denotes conditioning on the empty set, i.e., $\mathrm{E}[\Gamma_{i,0} \mid C_{i,-1}] = \mathrm{E}[\Gamma_{i,0}]$. Note that $\mathrm{E}[\Gamma_{i,j} \mid C_{i,j-1}] = \gamma_j$ and that the $\omega_{i,j}$ are independent of

σ_j^2. Using the $\Gamma_{i,j}$, we define the following linear estimators of γ_j and σ_j^2 based on D_I. For $j = 0,\ldots,J$, let

$$\hat{\gamma}_j^I = \frac{1}{\Omega_j^I}\sum_{i=1}^{I-j}\omega_{i,j}\Gamma_{i,j}, \tag{4.2}$$

$$\hat{\sigma}_j^2 = \frac{1}{I-j-1}\left(\sum_{i=1}^{I-j}\omega_{i,j}(\Gamma_{i,j}-\hat{\gamma}_j^I)^2\right),$$

where $\Omega_j^I = \sum_{i=1}^{I-j}\omega_{i,j}$. The superscript I indicates the fact that $\hat{\gamma}_j^I$ is based on the data in D_I. Later, we will define an estimator $\hat{\gamma}_j^{I+1}$ based on the data available in accounting year $I+1$. If $J = I-1$, there is only one observation in the last development year and the denominator in $\hat{\sigma}_J^2$ becomes zero. In that case, we propose to estimate σ_J^2 with the extrapolation used by Mack (1993), i.e., $\hat{\sigma}_J^2 = \min\{\hat{\sigma}_{J-2}^2, \hat{\sigma}_{J-1}^2, \hat{\sigma}_{J-1}^4 / \hat{\sigma}_{J-2}^2\}$.

The following theorem shows several nice properties of the estimators $\hat{\gamma}_j^I$ and $\hat{\sigma}_j^2$, which are analogous to the characteristics of the estimators $\hat{f}_j$ of the age-to-age factors f_j in the CL model.

Theorem 4.1. *Let* $\mathrm{B}_k = \{C_{i,j} : C_{i,j} \in \mathrm{D}_I, j \le k\}$ *and note that* $\mathrm{B}_{-1} = \varnothing$. *We have that*

1. $\mathrm{E}[\hat{\gamma}_j^I] = \gamma_j$.
2. $\mathrm{cov}(\hat{\gamma}_j^I, \hat{\gamma}_k^I) = 0$ *for* $j \ne k$.
3. *Among all linear estimators,* $\hat{\gamma}_j^I$ *has minimal variance and* $var(\hat{\gamma}_j^I \mid B_{j-1}) = \sigma_j^2 / \Omega_j^I$.
4. $\mathrm{E}[\hat{\sigma}_j^2] = \sigma_j^2$.

Proof. (1.) For $j = 0$, the claim immediately follows from the definition of Model 3.1. For $j \ge 1$, unbiasedness of $\hat{\gamma}_j^I$ directly follows from $\mathrm{E}[\Gamma_{i,j}] = \mathrm{E}[\mathrm{E}[\Gamma_{i,j} \mid C_{i,j-1}]] = \gamma_j$ and $1/\Omega_j^I \sum_{i=1}^{I-j}\omega_{i,j} = 1$.

(2.) Let $\mathrm{B}_k = \{C_{i,j} : C_{i,j} \in \mathrm{D}_I, j \le k\}$ and note that $\mathrm{B}_{-1} = \varnothing$. Suppose $j < k$ without loss of generality. As $\hat{\gamma}_k^I$ is B_k-measurable, we get

$$cov(\hat{\gamma}_j^I, \hat{\gamma}_k^I) = \mathrm{E}[\hat{\gamma}_j^I \hat{\gamma}_k^I] - \mathrm{E}[\hat{\gamma}_j^I]\mathrm{E}[\hat{\gamma}_k^I] = \mathrm{E}\left[\mathrm{E}[\hat{\gamma}_j^I \hat{\gamma}_k^I \mid \mathrm{B}_{k-1}]\right] - \gamma_j \gamma_k$$
$$= \mathrm{E}\left[\hat{\gamma}_j^I \mathrm{E}[\hat{\gamma}_k^I \mid \mathrm{B}_{k-1}]\right] - \gamma_j \gamma_k = \mathrm{E}[\hat{\gamma}_j^I]\gamma_k - \gamma_j \gamma_k = 0.$$

(3.) This is a direct consequence of Lemma 3.4 in Wüthrich and Merz (2008).

(4.) Note that $\mathrm{B}_{-1} = \varnothing$. For $\mathrm{E}[\hat{\sigma}_j^2 \mid \mathrm{B}_{j-1}]$, we obtain

$$\mathrm{E}\left[\hat{\sigma}_j^2 \middle| \mathrm{B}_{j-1}\right] = \frac{1}{I-j-1} \sum_{i=1}^{I-j} \omega_{i,j} \mathrm{E}\left[\left(\Gamma_{i,j} - \hat{\gamma}_j^I\right)^2 \middle| \mathrm{B}_{j-1}\right].$$

For the conditional expectations in the sum above, we have

$$\mathrm{E}\left[\left(\Gamma_{i,j} - \hat{\gamma}_j^I\right)^2 \middle| \mathrm{B}_{j-1}\right] = var\left(\Gamma_{i,j} \middle| \mathrm{B}_{j-1}\right) - 2cov\left(\Gamma_{i,j}, \hat{\gamma}_j^I \middle| \mathrm{B}_{j-1}\right) + var\left(\hat{\gamma}_j^I \middle| \mathrm{B}_{j-1}\right) = \frac{\sigma_j^2}{\omega_{i,j}} - \frac{\sigma_j^2}{\Omega_j^I}.$$

Summing up yields $\mathrm{E}[\hat{\sigma}_j^2 \mid \mathrm{B}_{j-1}] = \sigma_j^2$. From the tower property for conditional expectations, we get $\mathrm{E}[\hat{\sigma}_j^2] = \mathrm{E}[\mathrm{E}[\hat{\sigma}_j^2 \mid \mathrm{B}_{j-1}]] = \sigma_j^2$. ■

In order to predict the ultimate claim $C_{i,J}$, we replace all the γ_j in $\mathrm{E}[C_{i,J} \mid \mathrm{D}_I] = \mathrm{E}[C_{i,J} \mid C_{i,I-i}]$ (see (3.1)) with the corresponding $\hat{\gamma}_j^I$:

$$\hat{C}_{i,J}^I = C_{i,I-i} \prod_{I-i<m\le J} \hat{\xi}_{i,m}^I + \mu_i \sum_{I-i<n\le J} \left((1-\alpha_{i,n})\hat{\gamma}_n^I \prod_{n<m\le J} \hat{\xi}_{i,m}^I \right) \tag{4.3}$$

where $\hat{\xi}_{i,j}^I = 1 + \alpha_{i,j}\hat{\gamma}_j^I / \beta_{j-1}$.

In the extreme case where $\alpha_{i,j} = 0$, for $j = I-i+1,\ldots,J$, we have

$$\hat{C}_{i,J}^I = C_{i,I-i} + \mu_i \sum_{I-i<j\le J} \hat{\gamma}_j^I. \tag{4.4}$$

This resembles the expression in the pure BF case. If furthermore $\alpha_{i,j} = 0$ for all i and j, the estimates $\hat{\gamma}_j^I$ are equal to the raw (unsmoothed) estimates deduced in Mack (2008b), hence also the predictions $\hat{C}_{i,J}^I$ coincide.

If conversely $\alpha_{i,j} = 1$ for $j = I - i + 1, \ldots, J$, we have

$$\hat{C}_{i,J}^{I} = C_{i,I-i} \prod_{I-i<j\le J} \frac{\beta_{j-1} + \hat{\gamma}_j^I}{\beta_{j-1}}. \tag{4.5}$$

The estimates $\hat{\gamma}_j^I$ are not equal to the CL estimates $\hat{\gamma}_j^{CL}$ of γ_j, even if $\alpha_{i,j} = 1$ for all i and j. However, numerical examples show that for triangles with few outliers, they are very close and hence also the predictions $\hat{C}_{i,J}^I$ are close to the CL predictions $\hat{C}_{i,J}^{CL}$.

To simplify notation we rewrite (3.1) and (4.3) as

$$\mathrm{E}[C_{i,J} \mid \mathrm{D}_I] = \sum_{I-i\le n\le J} \kappa_{i,n}^I \prod_{n<m\le J} \xi_{i,m} \quad \text{and} \quad \hat{C}_{i,J}^I = \sum_{I-i\le n\le J} \hat{\kappa}_{i,n}^I \prod_{n<m\le J} \hat{\xi}_{i,m}^I,$$

where

$$\kappa_{i,n}^I = \begin{cases} C_{i,I-i}, & \text{for } n = I - i, \\ \mu_i (1 - \alpha_{i,n}) \gamma_n, & \text{for } n > I - i, \end{cases}$$
$$\hat{\kappa}_{i,n}^I = \begin{cases} C_{i,I-i}, & \text{for } n = I - i, \\ \mu_i (1 - \alpha_{i,n}) \hat{\gamma}_n^I, & \text{for } n > I - i. \end{cases}$$

Of course, the properties of $\hat{\gamma}_j^I$ shown in Theorem 4.1 directly translate to analogous properties of $\hat{\xi}_{i,j}^I$ and $\hat{\kappa}_{i,n}^I$, which are linear transformations of $\hat{\gamma}_j^I$.

5. ASSESSING THE PREDICTION ERROR

In this section, we provide a measure of uncertainty of the predicted aggregate ultimate claims $\sum_{i=1}^{I} \hat{C}_{i,J}^I$. This is important, as we do not only want to determine the reserves but also measure their precision in predicting the outstanding claims. A popular measure of uncertainty among actuaries is the so called conditional mean square error of prediction (MSEP), defined as

$$msep\left(\sum_{i=1}^{I} \hat{C}_{i,J}^I \right) = \mathrm{E}\left[\left(\sum_{i=1}^{I} \left(\hat{C}_{i,J}^I - C_{i,J} \right) \right)^2 \middle| \mathrm{D}_I \right],$$

which can be decomposed into process variance and parameter estimation error (PEE)

$$msep\left(\sum_{i=1}^{I}\hat{C}_{i,J}^{I}\right)=var\left(\sum_{i=1}^{I}C_{i,J}\middle|\mathrm{D}_I\right)+\left(\sum_{i=1}^{I}\left(\hat{C}_{i,J}^{I}-\mathrm{E}[C_{i,J}\mid\mathrm{D}_I]\right)\right)^2.$$

Note that the above expression corresponds to the MSEP of the aggregate reserves as, conditionally on D_I, the reserves are a deterministic shift of $\sum_{i=1}^{I}\hat{C}_{i,J}^{I}$. An analogous decomposition holds for the MSEP of one accident year, i.e., $msep(\hat{C}_{i,J}^{I})$. For $i \le I-J$, we have $C_{i,J} \in \mathrm{D}_I$. Hence, we can omit the $i \le I-J$ terms in the sum above.

The process variance can be decoupled into

$$var\left(\sum_{i=1}^{I}C_{i,J}\mid\mathrm{D}_I\right)=\sum_{i=1}^{I}var(C_{i,J}\mid\mathrm{D}_I)$$

as accident years are assumed to be independent. For estimation, we replace the unknown quantities in (3.2) by their respective D_I estimates and get

$$\widehat{var}(C_{i,J}\mid D_I)=\mu_i\sum_{I-i<n\le J}\hat{\sigma}_n^2\prod_{n<m\le J}(\hat{\xi}_{i,m}^{I})^2. \tag{5.1}$$

In order to estimate the PEE, we use a similar approach as used in Theorem 3 in Mack (1993). That is, we linearise the PEE and decompose it into a sum of expectations of which most drop out due to pairwise independence under a suitable change of conditioning, see Appendix A. To that end, for $i \ge I-J+1$ and $n = I-i,\ldots,J$, we define $\Psi_{i,n}=\kappa_{i,n}^{I}\prod_{n<m\le J}\xi_{i,m}$, its D_I estimator $\hat{\Psi}_{i,n}^{I}=\hat{\kappa}_{i,n}^{I}\prod_{n<m\le J}\hat{\xi}_{i,m}^{I}$ and

$$\psi_{i,n,k}=\begin{cases}(\hat{\kappa}_{i,n}^{I}-\kappa_{i,n}^{I})\prod_{n<m\le J}\xi_{i,m}, & \text{for } k=n,\\ \hat{\kappa}_{i,n}^{I}\left(\prod_{n<m<k}\hat{\xi}_{i,m}^{I}\right)\left(\hat{\xi}_{i,k}^{I}-\xi_{i,k}\right)\left(\prod_{k<m\le J}\xi_{i,m}\right), & \text{for } n<k\le J.\end{cases}$$

Note that using the above defined variables, we can decompose $\mathrm{E}[C_{i,J}\mid\mathrm{D}_I]$ and $\hat{C}_{i,J}^{I}$, as $\mathrm{E}[C_{i,J}\mid\mathrm{D}_I]=\sum_{n=I-i}^{J}\Psi_{i,n}$, $\hat{C}_{i,J}^{I}=\sum_{n=I-i}^{J}\hat{\Psi}_{i,n}^{I}$ and $\hat{\Psi}_{i,n}^{I}-\Psi_{i,n}=\sum_{k=n}^{J}\psi_{i,n,k}$. We can now rewrite the aggregate PEE as

$$
\begin{aligned}
pee_\Sigma &= \left(\sum_{i=1}^{I} \left(\hat{C}_{i,J}^I - \mathrm{E}[C_{i,J} \mid \mathrm{D}_I] \right) \right)^2 \\
&= \sum_{\substack{I-J+1 \le i_1 \le I \\ I-i_1 \le n_1 \le J}} \sum_{\substack{I-J+1 \le i_2 \le I \\ I-i_2 \le n_2 \le J}} \left(\hat{\Psi}_{i_1,n_1}^I - \Psi_{i_1,n_1} \right) \left(\hat{\Psi}_{i_2,n_2}^I - \Psi_{i_2,n_2} \right) \\
&= \sum_{\substack{I-J+1 \le i_1 \le I \\ I-i_1 \le n_1 \le J \\ n_1 \le k_1 \le J}} \sum_{\substack{I-J+1 \le i_2 \le I \\ I-i_2 \le n_2 \le J \\ n_2 \le k_2 \le J}} \psi_{i_1,n_1,k_1} \psi_{i_2,n_2,k_2}.
\end{aligned}
$$

For the one-year PEE, which is defined by $pee_i = (\hat{C}_{i,J}^I - \mathrm{E}[C_{i,J} \mid \mathrm{D}_I])^2$ we find an analogous formula.

In Appendix A we show that by a suitable change of conditioning and replacing all unknown quantities with their estimates we get the estimators

$$
\widehat{\psi_{i_1,n_1,k_1} \psi_{i_2,n_2,k_2}} = \begin{cases} 0, & \text{if } k_1 \ne k_2, \\ \hat{\Psi}_{i_1,n_1}^I \hat{\Psi}_{i_2,n_2}^I b_{i_1,n_1,k} b_{i_2,n_2,k} \dfrac{\hat{\sigma}_k^2}{\Omega_k^I} & \text{if } k_1 = k_2 = k, \end{cases}
$$

where

$$
b_{i,n,k} = \begin{cases} \dfrac{\alpha_{i,k}}{\beta_{k-1} \hat{\xi}_{i,k}^I}, & \text{if } k > n, \\ \dfrac{1}{\hat{\gamma}_k^I} & \text{if } k = n, n > I - i, \\ 0, & \text{if } k = n = I - i. \end{cases}
$$

This leads to the following estimates

$$
\begin{aligned}
\widehat{pee}_i &= \sum_{\substack{I-i \le n_1 \le J \\ n_1 \le k_1 \le J}} \sum_{\substack{I-i \le n_2 \le J \\ n_2 \le k_2 \le J}} \widehat{\psi_{i,n_1,k_1} \psi_{i,n_2,k_2}}, \\
\widehat{pee}_\Sigma &= \sum_{\substack{I-J+1 \le i_1 \le I \\ I-i_1 \le n_1 \le J \\ n_1 \le k_1 \le J}} \sum_{\substack{I-J+1 \le i_2 \le I \\ I-i_2 \le n_2 \le J \\ n_2 \le k_2 \le J}} \widehat{\psi_{i,n_1,k_1} \psi_{i,n_2,k_2}},
\end{aligned}
$$

for the one-year and aggregate PEE.

We use the direct and well known approach of Mack (1993) in order to get an estimation of the PEE. Note that there are alternative procedures to estimate the PEE such as the conditional approach given in Buchwalder et al. (2006), which lead to similar estimates.

The MSEP is finally estimated by $\widehat{msep}(\hat{C}_{i,J}^{I}) = \widehat{var}(C_{i,J} \mid D_I) + \widehat{pee}_i$ and

$$\widehat{msep}\left(\sum_{i=1}^{I} \hat{C}_{i,J}^{I}\right) = \sum_{i=1}^{I} \widehat{var}(C_{i,J} \mid D_I) + \widehat{pee}_{\Sigma}.$$

6. CLAIMS DEVELOPMENT RESULT

After each year, i.e., at the end of an accounting year, we receive additional information on the claims runoff. This means that the available data D_I is augmented with an additional diagonal to $D_{I+1} = \{C_{i,j} : 1 \le i \le I, 0 \le j \le J, i + j \le I + 1\}$. According to this new information, we can update our prediction of the ultimate claim $C_{i,J}$. New solvency regimes (e.g., Solvency II and the Swiss Solvency Test) require protection for each year (one-year perspective) against the risk of possible changes in subsequent ultimate claim predictions. In this section, we provide estimates of the uncertainty in those changes.

Similar to (4.3) we define $\hat{C}_{i,J}^{I+1}$ for $i \ge I - J + 1$ which is an estimate of $\mathrm{E}\left[C_{i,J} \middle| D_{I+1}\right]$ at time $I + 1$, based on D_{I+1}:

$$\hat{C}_{i,J}^{I+1} = C_{i,I-i+1} \prod_{I-i+1<m\le J} \hat{\xi}_{i,m}^{I+1} + \mu_i \sum_{I-i+1<n\le J} \left((1-\alpha_{i,n}) \hat{\gamma}_n^{I+1} \prod_{n<m\le J} \hat{\xi}_{i,m}^{I+1} \right). \tag{6.1}$$

The $\hat{\gamma}_j^{I+1}$ are estimates of γ_j based on the information in D_{I+1}, as $\hat{\gamma}_0^{I+1} = \hat{\gamma}_0^{I}$ and for $j \ge 1$

$$\hat{\gamma}_j^{I+1} = \frac{1}{\Omega_j^{I+1}} \sum_{i=1}^{I-j+1} \omega_{i,j} \Gamma_{i,j} = \frac{\Omega_j^{I}}{\Omega_j^{I+1}} \hat{\gamma}_j^{I} + \frac{\omega_{I-j+1,j}}{\Omega_j^{I+1}} \Gamma_{I-j+1,j},$$

where $\Omega_j^{I+1} = \sum_{i=1}^{I-j+1} \omega_{i,j} = \Omega_j^{I} + \omega_{I-j+1,j}$ and $\hat{\xi}_{i,j}^{I+1} = 1 + \alpha_{i,j} \hat{\gamma}_j^{I+1} / \beta_{j-1}$. Note that Ω_j^{I+1} is D_I-measurable.

The CDR viewed from time I denotes the difference in estimated ultimate claims for subsequent accounting years

$$CDR_{\Sigma} = \sum_{i=1}^{I} \left(\hat{C}_{i,J}^{I} - \hat{C}_{i,J}^{I+1} \right),$$

and for a single accident year i we have $CDR_i = \hat{C}_{i,J}^{I} - \hat{C}_{i,J}^{I+1}$. Note that for given D_I and $i \le I - J$, we have $CDR_i = 0$, whereas for $i \ge I - J + 1$, CDR_i is stochastic viewed from time I.

Note that other publications such as Wüthrich et al. (2009) define the CDR as $\sum_{i=1}^{I} (\mathrm{E}[C_{i,J} \mid \mathrm{D}_I] - \mathrm{E}[C_{i,J} \mid \mathrm{D}_{I+1}])$, while our definition is an estimate thereof. For a detailed interpretation of the CDR in a Bayesian setting we refer to Bühlmann et al. (2009).

As we consider both $\hat{C}_{i,J}^{I}$ and $\hat{C}_{i,J}^{I+1}$ to be conditional best-estimates, we estimate $\widehat{CDR_i} = 0$. We want to assess the uncertainty in the CDR, measured by its second moment, $CDRU_\Sigma = \mathrm{E}[(CDR_\Sigma)^2 \mid \mathrm{D}_I]$ as well as the accident year-wise uncertainty $CDRU_i = \mathrm{E}[(CDR_i)^2 \mid \mathrm{D}_I]$. To do so, we define

$$\hat{\kappa}_{i,n}^{I+1} = \begin{cases} C_{i,I-i+1}, & \text{for } n = I - i + 1, \\ \mu_i (1 - \alpha_{i,n}) \hat{\gamma}_n^{I+1}, & \text{for } n > I - i + 1, \end{cases}$$
$$\hat{\Psi}_{i,n}^{I+1} = \hat{\kappa}_{i,n}^{I+1} \prod_{n < m \le J} \hat{\xi}_{i,m}^{I+1}.$$

With the above formulas and (6.1) we get $\hat{C}_{i,J}^{I+1} = \sum_{I-i+1 \le n \le J} \hat{\Psi}_{i,n}^{I+1}$. For the CDR we then have

$$CDR_i = \sum_{I-i \le n \le J} \hat{\Psi}_{i,n}^{I} - \sum_{I-i+1 \le n \le J} \hat{\Psi}_{i,n}^{I+1} = \hat{\Psi}_{i,I-i}^{I} + \sum_{I-i+1 \le n \le J} \left(\hat{\Psi}_{i,n}^{I} - \hat{\Psi}_{i,n}^{I+1} \right).$$

We now rewrite $CDR_i = \sum_{I-i+1 \le n \le J} \Theta_{i,n}$ and $\Theta_{i,n} = \sum_{k=n}^{J} \theta_{i,n,k}$, where

$$\Theta_{i,n} = \begin{cases} \hat{\Psi}_{i,n-1}^{I} + \hat{\Psi}_{i,n}^{I} - \hat{\Psi}_{i,n}^{I+1}, & \text{for } n = I - i + 1, \\ \hat{\Psi}_{i,n}^{I} - \hat{\Psi}_{i,n}^{I+1}, & \text{for } n > I - i + 1, \end{cases}$$

$$\theta_{i,n,k} = \begin{cases} e_{i,n,k} \prod_{n < m \le J} \hat{\xi}_{i,m}^{I}, & \text{for } k = n, \\ \hat{\kappa}_{i,n}^{I+1} \prod_{n < m < k} \hat{\xi}_{i,m}^{I+1} e_{i,n,k} \prod_{k < m \le J} \hat{\xi}_{i,m}^{I}, & \text{for } k > n, \end{cases}$$

with

$$e_{i,n,k} = \begin{cases} \alpha_{i,k} / \beta_{k-1}(\hat{\gamma}_k^I - \hat{\gamma}_k^{I+1}), & \text{for } k > n, \\ \mu_i(1-\alpha_{i,n})\hat{\gamma}_n^I - \mu_i(1-\alpha_{i,n})\hat{\gamma}_n^{I+1}, & \text{for } k = n, n > I-i+1, \\ C_{i,I-i}\hat{\xi}_{i,n-1}^I + \mu_i(1-\alpha_{i,n})\hat{\gamma}_n^I - C_{i,I-i+1}, & \text{for } k = n = I-i+1. \end{cases}$$

The $CDRU_\Sigma$ expressed in terms of $\theta_{i,n,k}$ is given by

$$\begin{aligned} CDRU_\Sigma &= \mathrm{E}\left[\left(\sum_{I-J+1\le i\le I}\ \sum_{I-i+1\le n\le J}\ \sum_{n\le k\le J} \theta_{i,n,k}\right)^2 \middle| \mathrm{D}_I\right] \\ &= \sum_{\substack{I-J+1\le i_1\le I \\ I-i_1+1\le n_1\le J \\ n_1\le k_1\le J}}\ \sum_{\substack{I-J+1\le i_2\le I \\ I-i_2+1\le n_2\le J \\ n_2\le k_2\le J}} \mathrm{E}\left[\theta_{i_1,n_1,k_1}\theta_{i_2,n_2,k_2} \middle| \mathrm{D}_I\right], \end{aligned}$$

where the $i \le I-J$ terms have been omitted since $CDR_i = 0$ in that case.

In Appendix B, we prove by the help of approximations and a suitable change of conditioning, that we can estimate $\mathrm{E}[\theta_{i_1,n_1,k_1}\theta_{i_2,n_2,k_2} \mid \mathrm{D}_I]$ with

$$\hat{\mathrm{E}}\left[\theta_{i_1,n_1,k_1}\theta_{i_2,n_2,k_2} \mid \mathrm{D}_I\right] = \begin{cases} 0, & \text{for } k_1 \ne k_2, \\ \widetilde{\Psi}_{i_1,n_1}^{I+1} g_{i_1,n_1,k_1} \widetilde{\Psi}_{i_2,n_2}^{I+1} g_{i_2,n_2,k_2} \mu_{I-k_1+1}\hat{\sigma}_{k_1}^2, & \text{for } k_1 = k_2, \end{cases}$$

where

$$\widetilde{\Psi}_{i,n}^{I+1} = \begin{cases} \hat{\Psi}_{i,n}^{I}, & \text{for } n > I-i+1, \\ \hat{\Psi}_{i,n-1}^{I} + \hat{\Psi}_{i,n}^{I}, & \text{for } n = I-i+1, \end{cases}$$

$$g_{i,n,k} = \begin{cases} \dfrac{1}{\hat{\xi}_{i,k}^{I}} \dfrac{\alpha_{i,k}}{\beta_{k-1}} \dfrac{\omega_{I-k+1,k}}{\Omega_k^{I+1}} \dfrac{1}{m_{I-k+1,k}}, & \text{for } k > n, \\ \dfrac{1}{\hat{\gamma}_n^{I}} \dfrac{\omega_{I-k+1,k}}{\Omega_k^{I+1}} \dfrac{1}{m_{I-k+1,k}}, & \text{for } k = n, n > I-i+1, \\ \dfrac{1}{C_{i,I-i}\hat{\xi}_{i,I-i+1}^{I} + \mu_i(1-\alpha_{i,I-i+1})\hat{\gamma}_{I-i+1}^{I}}, & \text{for } k = n = I-i+1. \end{cases}$$

The terms with $k = n = I-i+1$ cover the variability of the next diagonal $\{C_{i,I-I+1} : i = I-J+1,\ldots,I\}$, i.e., they can be seen as the process variance components of the CDR. The other terms cover the variability of the γ_j estimates.

With those approximations we define estimates of $CDRU_\Sigma$ and $CDRU_i$ by

$$\widehat{CDRU}_\Sigma = \sum_{\substack{I-J+1\le i_1\le I\\ I-i_1+1\le n_1\le J\\ n_1\le k_1\le J}} \sum_{\substack{I-J+1\le i_2\le I\\ I-i_2+1\le n_2\le J\\ n_2\le k_2\le J}} \widehat{E}\left[\theta_{i_1,n_1,k_1}\theta_{i_2,n_2,k_2} \mid D_I\right],$$

$$\widehat{CDRU}_i = \sum_{\substack{I-i+1\le n_1\le J\\ n_1\le k_1\le J}} \sum_{\substack{I-i+1\le n_2\le J\\ n_2\le k_2\le J}} \widehat{E}\left[\theta_{i,n_1,k_1}\theta_{i,n_2,k_2} \mid D_I\right].$$

When $\alpha_{i,j} = 0$ for all i and j, the HCL γ_j estimates as defined in (4.2) are equivalent to the raw estimates $\hat{\gamma}_j$ as defined in Mack (2008b) and Saluz et al. (2011). Hence, our formula can be used to estimate the $CDRU$ of the pure BF method. To our awareness, the only other work deducing such an estimate is Saluz (2010).

7. ESTIMATING THE μ_i AND INCORPORATING THEIR UNCERTAINTY

The HCL method requires prior estimates μ_i of the expected ultimate claims $\mathrm{E}[C_{i,J}]$. In most cases, the product of (planned) premium volume and expected loss ratio is considered as a viable estimate. Since the prior estimates μ_i for the HCL method are of the same nature as the prior estimates for the BF method, the same policy guidelines as in the BF method should apply. Namely, the μ_i should be independent of D_I.

The source of information that determines the μ_i often comprises inaccuracy. Therefore, the PEE should take this additional uncertainty into account. In the BF model, Mack (2008b) embeds the μ_i in the distribution-free framework, by assuming some variance $var(\mu_i)$. A similar approach is used by Saluz et al. (2011) in a distributional framework. In order to estimate the PEE, Mack (2008b) assumes $\hat{\gamma}_j$ and μ_i to be independent. As the μ_i appear in the quotient of $\Gamma_{i,j}$, we avoid this assumption and present a simple scenario-based approach instead, which naturally captures the dependence between the $\hat{\gamma}_j$ and μ_i. Of course, the scenarios of μ_i can be fitted to any assumption on mean and variance.

We consider $(\mu_1,\ldots,\mu_I)$ to be a realisation of a random vector $\boldsymbol{\mu}$ taking values in a finite set of scenarios $\{\mu^1,\ldots,\mu^S\}$ and reformulate Model 3.1 as being conditional on $(\mu_1,\ldots,\mu_I)=\mu^s$ for some $s\in\{1,\ldots,S\}$. The equations deduced in the previous sections continue to hold under the following model (but must be thought of as being conditional on the μ_i).

Note that we could also assume $\boldsymbol{\mu}$ to have a continuous distribution. Then the calculation would require simulation.

Model 7.1 (HCL Model with uncertain μ_i). *There exist parameters*

- γ_j *and* $\sigma_j^2>0$ *for* $j=0,\ldots,J$,
- $\beta_j>0$, *for* $j=0,\ldots,J-1$,
- $\alpha_{i,j}\in[0,1]$ *for* $i=1,\ldots,I$ *and* $j=1,\ldots,J$,
- *and a discrete random vector* $\boldsymbol{\mu}\in(0,\infty)^I$ *satisfying* $\mathrm{P}[\boldsymbol{\mu}=\mu^s]=p^s>0$, $s\in\{1,2,\ldots,S\}$ *for* $\mu^s\in(0,\infty)^I$ *and* $\sum_{s=1}^{S}p_s=1$.

Conditionally on $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_I)$, *the* $(C_{1,j})_{j=0,\ldots,J},\ldots,(C_{I,j})_{j=0,\ldots,J}$ *are independent Markov processes* *with*

- $\mathrm{E}[C_{i,0} \mid \boldsymbol{\mu}] = \gamma_0 \mu_i$ *and* $\mathrm{E}[C_{i,j} \mid C_{i,j-1}, \boldsymbol{\mu}] = C_{i,j-1} + \gamma_j \left(\alpha_{i,j} \dfrac{C_{i,j-1}}{\beta_{j-1}} + (1-\alpha_{i,j})\mu_i \right)$ *for* $j = 1,\ldots,J$;
- $var(C_{i,0} \mid \boldsymbol{\mu}) = \sigma_0^2 \mu_i$ *and* $var(C_{i,j} \mid C_{i,j-1}, \boldsymbol{\mu}) = \sigma_j^2 \mu_i$ *for* $j = 1,\ldots,J$. ■

We can now run the estimation procedure for each possible realisation of the random vector $\boldsymbol{\mu}$. For fixed μ_i, we are in the framework of Model 3.1, which allows us to calculate estimates of ultimate claims etc. For the case $\boldsymbol{\mu} = \mu^s$, these estimates are denoted with a superscript s, such as $\hat{C}_{i,J}^{I,s}$, $\widehat{var}^{s}(C_{i,J} \mid D_I)$, $\widehat{pee}_{\Sigma}^{s}$ and $\widehat{CDRU}_i^{s}$.

We calculate unconditional estimates by averaging out over the different possible outcomes of $\boldsymbol{\mu}$, using weights given by the probabilities p_s.

- Because $\mathrm{E}[C_{i,J}] = \mathrm{E}[\mathrm{E}[C_{i,J} \mid \boldsymbol{\mu}]]$, we estimate

$$\hat{C}_{i,J}^{I} = \sum_{s=1}^{S} \hat{C}_{i,J}^{I,\mu_s} p_s.$$

- By the law of total variance,

$$var(C_{i,J} \mid \mathrm{D}_I) = \mathrm{E}[var(C_{i,J} \mid \mathrm{D}_I, \boldsymbol{\mu})] + var(\mathrm{E}[C_{i,J} \mid \mathrm{D}_I, \boldsymbol{\mu}]),$$

thus

$$\widehat{var}(C_{i,J} \mid D_I) = \sum_{s=1}^{S} \widehat{var}^{s}(C_{i,J} \mid D_I) p_s + \sum_{s=1}^{S} \left(\hat{C}_{i,J}^{I,s} - \hat{C}_{i,J}^{I} \right)^2 p_s.$$

- For the PEE we have that $\hat{C}_{i,J}^{I,s}$ is close to $\mathrm{E}[C_{i,J} \mid \mathrm{D}_I, \mu_s]$, which allows to approximate $\widehat{pee}_i = \sum_{s=1}^{S} \widehat{pee}_i^{s} p_s$ and $\widehat{pee}_{\Sigma} = \sum_{s=1}^{S} \widehat{pee}_{\Sigma}^{s} p_s$, respectively.
- As $\mathrm{E}[CDR_i \mid \mu] \approx 0$, the uncertainty in the CDR can be estimated by $\widehat{CDRU}_i = \sum_{s=1}^{S} \widehat{CDRU}_i^{s} p_s$ and $\widehat{CDRU}_{\Sigma} = \sum_{s=1}^{S} \widehat{CDRU}_{\Sigma}^{s} p_s$, respectively.

Note that the distribution of the μ_i can be fitted to any assumption on mean and variance. To determine an adequate amount of uncertainty in the μ_i, we suggest the following possible sources of information:

- Experts can determine scenarios μ^s based on socio-economic scenarios. Using expert judgement unavoidably involves psychological effects, due to the expert's subjectivity. These effects should be taken care of, see for instance Meyer and Booker (2001).

- Literature suggests to use a coefficient of variation,

$$CoV(\mu_i) = \sqrt{var(\mu_i)} \,/\, \mathrm{E}[\mu_i],$$

  between 5% and 10%, see Section 6.6.2 in Wüthrich and Merz (2008). This is also based on regulatory guidelines, see FINMA (2006).

- Saluz et al. (2011) suggest a data-based approach in order to determine the uncertainty in the μ_i. The two case studies considered in Saluz et al. (2011) induce a $CoV(\mu_i)$ between 5% and 10%.

- Regulators may enforce socio-economic scenarios, see FINMA (2006).

- Industry data, such as provided by the *Schedule P* database in the US, can be used to determine a prudent amount of uncertainty in the μ_i, see NAIC (2010).

Note that we assume that the scenarios implied by $\boldsymbol{\mu}$ do not change over time.

8. COMMENTS FOR PRACTITIONERS

In Model 3.1 the parameters β_j and $\alpha_{i,j}$ are assumed to be fixed and given constants. In this section, we discuss practical approaches to setting the β_j and $\alpha_{i,j}$.

8.1 Setting the γ_j's and β_j's

It might be surprising that in Model 3.1, γ_j and β_j are introduced as distinct parameters although their role is very similar in describing the runoff pattern of the claims. This is done to allow a deduction of MSEP and CDRU estimates similar to the approach given in Mack (1993). In the following, we consider the relation between these two parameters. Although the β_j are regarded in Model 3.1 as exogenous estimates, we propose an approach for estimation based on D_I.

For given prior estimates μ_i and a known runoff pattern γ_j, we would expect a realistic stochastic claims reserving method to satisfy

$$\mathrm{E}\left[C_{i,j}\right] = \left(\gamma_0 + \ldots + \gamma_j\right)\mu_i, \quad for \quad all \quad 0 \le j \le J, \tag{8.1}$$

$$\mathrm{E}\left[C_{i,J}\right] = \mu_i. \tag{8.2}$$

This means that the γ_j and μ_i provide best-estimates for the runoff process.

The following lemma yields a relation between the β_j and γ_j such that (8.1) and (8.2) hold.

Lemma 8.1. *Suppose* $\gamma_0 + \ldots + \gamma_J = 1$ *and* $\gamma_0 + \ldots + \gamma_j = \beta_j$ *for all* $j < J$. *Then (8.1) and (8.2) hold.*

Proof. Fix some i. For $j = 0$, $\mathrm{E}[C_{i,0}] = \gamma_0 \mu_i$ holds by definition. We proceed by induction on j.

$$\begin{aligned}
\mathrm{E}\left[C_{i,j}\right] &= \mathrm{E}\left[\mathrm{E}\left[C_{i,j} \middle| C_{i,j-1}\right]\right] = \mathrm{E}\left[C_{i,j-1}\xi_{i,j} + \gamma_j(1-\alpha_{i,j})\mu_i\right] \\
&= (\gamma_0 + \ldots + \gamma_{j-1})\mu_i\left(1 + \alpha_{i,j}\frac{\gamma_j}{\beta_{j-1}}\right) + \gamma_j(1-\alpha_{i,j})\mu_i \\
&= (\gamma_0 + \ldots + \gamma_{j-1})\mu_i + \mu_i\alpha_{i,j}\gamma_j\left(\frac{(\gamma_0 + \ldots + \gamma_{j-1})}{\beta_{j-1}} - 1\right).
\end{aligned}$$

The second summand above is equal to zero if $\gamma_0 + \ldots + \gamma_{j-1} = \beta_{j-1}$. ∎

Due to Lemma 8.1, we propose the following adaptions to the HCL method to be used in practice.

- For initial estimates of the $\hat{\gamma}_j^I$, use (4.2). Then, smooth and/or rescale the $\hat{\gamma}_j^I$ such that $\sum_{j=0}^{J} \hat{\gamma}_j^I = 1$ holds. Smoothing techniques are described for instance in Mack (2008b) and Section 11.2 in Wüthrich and Merz (2008). This adaption may also comprise the inclusion of tail factors, see for instance Mack (1999).

- Fix the β_j such that $\beta_j = \sum_{k=0}^{j} \hat{\gamma}_k^I$ holds. This can be achieved by running the whole estimation procedure several times and iteratively plugging $\sum_{k=0}^{j} \hat{\gamma}_k^I$ into β_j. In our numerical examples, this took no more than 5 iterations to converge to a precision of 10^{-10}.

This approach attaches additional data dependence and uncertainty to the β_j. However, we assume that these additional sources of errors are negligible compared to the error terms covered in Section 5, 6 and 7. One rationale to do so is that the estimated β_j is most stable for late development years. For small j, when the relative error of the β_j is potentially high, the $\alpha_{i,j}$ are in general rather small when chosen as proposed in Section 8.2, which diminishes the final impact. Indeed, numerical tests confirm this statement.

8.2 Setting the $\boldsymbol{\alpha}_{i,j}$

From a practical perspective, the choice of the $\alpha_{i,j}$ reflects the actuary's assessment of the behaviour of the claims process. At the same time, the $\alpha_{i,j}$ control the reliance of the HCL method on the data D_I. We can distinguish two ways of how the $\alpha_{i,j}$ influence the final result:

- For $i + j \le I$, the $\alpha_{i,j}$ control the data reliance in the *parameter estimation*. Each $\alpha_{i,j}$ gives a measure of the predictive power of $C_{i,j-1} / \beta_{j-1}$ as an estimator of the ultimate claim $C_{i,J}$.

- For $i + j > I$, the $\alpha_{i,j}$ control the predictive power of the diagonal value $C_{i,I-i}$ of the ultimate claim $C_{i,J}$. In other words, how sensitive the *prediction* $\hat{C}_{i,J}^I$ is to deviations of $C_{i,I-i}$.

Recognising the two distinct manners of influence, the following presents a simple proposal to set the $\alpha_{i,j}$.

- For $i + j \le I$, in general the runoff becomes more stable for higher development years, i.e., $\alpha_{i,j}$ increases to 1 for increasing j. A simple and practical approach that satisfies this requirement, is to set

$$\alpha_{i,j} = \beta_{j-1} \quad \text{for all} \quad i + j \leq I.$$

We highlight that the $\alpha_{i,j}$ as given above are development year-wise constant. Note that this approach is tail-adaptive, in the sense that for triangles with a longer runoff, the weight of the μ_i in the estimation of the $\hat{\gamma}_j^I$ for initial years will be higher than in triangles with a shorter tail.

- On the other hand, for $i + j > I$, a possible choice would be to set the $\alpha_{i,j}$ accident year-wise constant as

$$\alpha_{i,j} = \tilde{\alpha}_i \quad \text{for all} \quad i + j > I,$$

where $\tilde{\alpha}_i \in [0,1]$ for $I - J < i \leq I$ have to be determined by the actuary. In that way, the data-reliance of the reserves estimates can be set separately for each accident year. We highlight that the HCL method is closely related to the Benktander-Hovinen method, see Benktander (1976), when setting $\tilde{\alpha}_i = \beta_{I-i}$.

This approach is illustrated in Figure 2. Note that it reduces the number of α-parameters that have to be determined by the actuary from $I \times J$ to J. The actuary can choose the $\tilde{\alpha}_i$ according to judgment, whether a multiplicative or an additive structure is more appropriate for a particular accident year. Possible reasons to come to such a choice are given in the example in Section 2. However, there are many other possible approaches to choosing the $\alpha_{i,j}$.

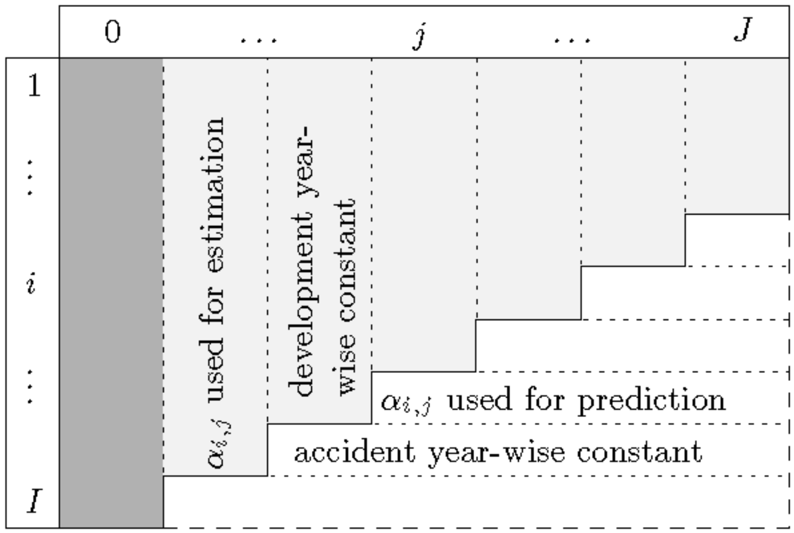

Figure 2: A possible practical approach to set the $\alpha_{i,j}$: For $i+j \leq I$ development year-wise constant and increasing, for $i+j > I$ accident year-wise constant according to the actuary's chosen $\tilde{\alpha}_i$.

One drawback affecting the robustness of the CL method is that the $C_{i,j}$ appear in the denominator of the parameter estimators. These estimators can strongly deviate if some $C_{i,j}$ are small, zero, or even negative for early development years in long-tailed claims triangles, see Busse et al. (2010). In the HCL model, the $C_{i,j-1}$ also appear in the denominator of the $\Gamma_{i,j}$, but for early development years their weight (namely $\alpha_{i,j}$) is generally low for long-tailed triangles, if the $\alpha_{i,j}$ are set as outlined above. Therefore, the problem with unstable parameter estimators does not occur.

Considering the case where only CL and BF models are admissible choices, a "super-expert" chooses one and declares it as the true model. In the HCL model, the prior selection of the $\alpha_{i,j}$ corresponds to the model choice. That is in the deduction of the estimates of *msep* and *CDRU*, the $\alpha_{i,j}$ are assumed to be constant and known. If model risk is to be taken into account, the sensitivity of the results with respect to changes in the $\alpha_{i,j}$ should be checked and the additional risk should be considered. By no means should the choice of the $\alpha_{i,j}$ be motivated by a minimizing argument for the reserves or a measure of uncertainty, as it is done in credibility approaches, see Neuhaus (1992).

For sufficiently small x_j, we have $\prod(1+x_j)-1 \approx \sum x_j$. Applying these approximations to (4.3), we get for the reserves of a single accident year i that

$$\hat{C}_{i,J}^{I} - C_{i,I-i} \approx \tilde{\alpha}_i \left(C_{i,I-i} \prod_{I-i<j\leq J} \frac{\beta_{j-1} + \hat{\gamma}_j^I}{\beta_{j-1}} - C_{i,I-i} \right) + \left(1-\tilde{\alpha}_i\right)\left(\mu_i \sum_{I-i<j\leq J} \hat{\gamma}_j^I \right). \tag{8.3}$$

Recalling (4.4) and (4.5), we see that the reserves $\hat{C}_{i,J}^{I} - C_{i,I-i}$ are approximately the weighted average between the completely data-reliant approach and the completely expert-reliant approach.

Equation (8.3) also shows that the HCL claims estimate is very similar to the Benktander-Hovinen estimate, except for a different interpretation of the weights, see Mack (2000) or Section 4.1.1 in Wüthrich and Merz (2008).

9. CASE STUDY

In this section, we present a case study illustrating the use of the HCL method applied to the triangle given in Table 1. We analyse the same loss triangle and μ_i as in Mack (2006) and Mack (2008b). However, our results cannot be directly compared to the results therein, as Mack (2008b) manually increases the γ_j for the last development years and also adds a tail factor. The triangle represents claims from a general liability excess line of business, which is very long-tailed with more than 80% of the claims payments being expected in the development years $j > 2$. In general CL-type methods do not work robustly for data coming from business subject to strong volatility and long-tailed development. Therefore, we assume that the classical CL method is not an adequate model for the triangle considered in this case study. The results of the CL method applied to this dataset is given in the Excel sheet, which is available at `www.RiskLab.ch/hclmethod`.

$i \mid j$	0	1	2	3	4	5	6	7	8	9	10	11	12	μ_i	$\tilde{\alpha}_i$
1	234	4'877	11'126	14'656	21'195	23'932	26'478	28'293	28'628	28'738	28'756	28'782	28'781	32'299.9	
2	1'994	6'930	11'755	17'935	25'594	27'545	32'655	33'266	34'042	34'451	34'499	35'826		40'279.1	1
3	-75	3'133	10'986	18'113	23'473	27'349	30'775	32'215	33'498	33'565	35'181			40'634.6	1
4	236	2'438	6'563	11'566	15'755	24'819	27'021	29'085	32'329	33'508				39'604.3	1
5	976	5'695	15'092	28'345	34'451	39'426	42'475	47'194	49'909					58'440.6	1
6	-730	2'623	15'527	26'169	42'660	51'546	58'774	67'286						81'346.9	1
7	539	5'777	20'678	45'543	65'817	83'586	116'520							163'258.7	1
8	725	15'625	50'301	93'896	146'517	173'997								268'150.6	0
9	312	6'754	50'350	139'052	177'864									331'893.1	0
10	2'988	12'909	33'266	67'851										193'519.8	0
11	260	7'441	29'643											169'559.7	0
12	994	4'043												157'381.6	0
13	2'411													156'150.7	0
$\hat{\gamma}_j \approx$	0.7%	4.8%	13.9%	20.8%	16.6%	11.8%	13.9%	7.6%	4.6%	1.4%	1.7%	2.2%	0%		

Table 1: A triangle of cumulative paid losses $C_{i,j}$ and prior ultimates μ_i from a general liability excess line of business. This triangle is the same as used in Mack (2006) and Mack (2008b). On the right, we give the $\tilde{\alpha}_i$ used in model HCL 1 and HCL 2. As an approximation for the $\hat{\gamma}_j$ of all models, we give the $\hat{\gamma}_j$ of HCL 1.

We compare three different setups for Model 3.1 (HCL 1, HCL 3 and HCL 4) and one setup for Model 7.1 (HCL 2):

- *HCL 1:* Model 3.1 with $\alpha_{i,j}$ set according to the proposals in Section 8.2. The chosen $\tilde{\alpha}_i$ are given in Table 1: $\tilde{\alpha}_i = 1$ for $i = 1,\ldots,7$ and $\tilde{\alpha}_i = 0$ for $i = 8,\ldots,13$.
- *HCL 2:* Model 7.1 with three scenarios for $\boldsymbol{\mu}$ (i.e., $S = 3$) and $\alpha_{i,j}$ as in HCL 1. The first scenario μ^1 is equal to the μ_i given in Table 1. Second and third scenarios are $\mu_i^2 = 1.1 \cdot \mu_i^1$ and $\mu_i^3 = 0.9 \cdot \mu_i^1$, respectively. The scenario probabilities are $p_1 = 0.6$ and $p_2 = p_3 = 0.2$. These scenarios correspond to $\mathrm{E}[\boldsymbol{\mu}] = \mu^1$ and $\mathrm{CoV}(\mu_i) = 6.3\%$.
- *HCL 3:* Model 3.1 with $\alpha_{i,j} = 0$ for all i and j. This model can be seen as representing BF.
- *HCL 4:* Model 3.1 with $\alpha_{i,j} = 1$ for all i and j. This model is similar to CL.

- *Mack BF:* The BF model as given in Mack (2008b), with the assumptions that $CoV(\mu_i) = 10\%$, $\hat{\rho}^U_{i,j} = 1/(1+|i-j|)$, $\hat{\rho}^z_{i,j}$ as given in Mack (2008a). Furthermore, the $\hat{\gamma}_j$ are rescaled to sum up to 1 but no further smoothing techniques or tail factors are used.

	HCL 1 ($S=1, \tilde{\alpha}_i$)			HCL 2 ($S=2, \tilde{\alpha}_i$)			HCL 3 ($S=3, \alpha_{I,j}=0$)			HCL 4 ($S=4, \alpha_{I,j}=1$)			Mack BF	
i	reserves	$\sqrt{msep}$	$\sqrt{CDRU}$	reserves	$\sqrt{msep}$	$\sqrt{CDRU}$	reserves	$\sqrt{msep}$	$\sqrt{CDRU}$	reserves	$\sqrt{msep}$	$\sqrt{CDRU}$	reserves	$\sqrt{msep}$
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	-1	1'294	864	-1	1'297	866	-1	1'273	849	-2	1'392	930	-1	1'273
3	799	1'708	890	799	1'711	891	842	1'684	875	956	1'822	934	842	1'684
4	1'385	1'984	922	1'384	1'987	922	1'476	1'947	886	1'660	2'097	947	1'476	1'947
5	2'820	2'770	652	2'819	2'776	652	2'930	2'686	618	3'388	2'935	683	2'930	2'686
6	7'440	4'178	1'786	7'436	4'194	1'790	7'661	3'934	1'593	8'990	4'503	1'970	7'661	3'934
7	24'806	8'291	3'647	24'792	8'356	3'661	27'282	7'890	3'146	30'297	9'271	4'275	27'282	7'890
8	84'355	18'646	10'138	84'414	20'052	10'167	81'821	16'390	8'955	98'794	24'308	14'815	81'821	13'638
9	143'623	23'893	7'368	143'686	26'654	7'419	140'449	20'905	6'484	171'007	34'793	15'524	140'449	16'489
10	115'799	17'650	7'086	115'823	19'746	7'165	114'154	15'844	6'855	131'612	32'404	20'859	114'154	13'329
11	136'677	18'598	8'704	136'685	20'915	8'800	135'915	17'081	8'484	166'073	55'113	43'260	135'915	14'693
12	148'719	18'173	3'819	148'720	20'673	3'911	148'522	16'873	4'163	84'930	89'384	73'585	148'522	14'680
13	155'088	18'540	3'905	155'089	21'106	3'916	155'060	17'299	3'970	270'331	173'332	130'123	155'060	15'100
Σ	821'509	89'253	18'226	821'644	106'548	18'365	816'112	79'146	17'011	968'036	236'197	158'553	816'112	61'922

Table 2: Reserves estimates, square root of MSEP and square root of the uncertainty in the CDR for triangle given in Table 1. We use the models HCL 1 (one scenario, $\alpha_{i,j}$ through $\tilde{\alpha}_i$), HCL 2 (three scenarios, $\alpha_{i,j}$ through $\tilde{\alpha}_i$), HCL 3 (one scenario, $\alpha_{i,j} = 0$), HCL 4 (one scenario, $\alpha_{i,j} = 1$) and Mack BF. The rows 1 to 13 denote accident year-wise results and the row designated by $\sum$ shows the results on an aggregate basis.

For these four different HCL setups and Mack BF we summarise in Table 2 the reserves, the uncertainty in ultimate claims in terms of $\sqrt{msep}$ and the uncertainty of the CDR in terms of $\sqrt{CDRU}$. For Mack BF, no results on the *CDR* are available. Moreover we list the $\hat{\gamma}^I_j$ of HCL 1, to which the $\hat{\gamma}^I_j$ of the other models are very close. Additional details, such as all $\hat{\gamma}^I_j$, $\hat{\sigma}^2_j$ and $\widehat{var}$, are given in the Excel sheet.

For the reserves, we see that HCL 1, 2, and 3 are very close. HCL 4 (and also CL, presented only in the Excel sheet) have significantly more unstable reserves estimates. The largest differences are in accident year $i = 12$ and $i = 13$, caused by the fact that $C_{12,1} \ll (\hat{\gamma}_0 + \hat{\gamma}_1)\mu_{12} \approx 8656$ and $C_{13,0} \gg \hat{\gamma}_0 \mu_{13} \approx 1093$. The $\hat{\gamma}_j$ for HCL 4 are larger for late development years and smaller for earlier ones. This leads to the systematically larger reserves of HCL 4. The reserves of Mack BF are equal to HCL 3.

For the uncertainties, we have that the μ_i generally have more explanatory power for the claims process than the $C_{i,j}$, as it is used in HCL 4 and CL. This leads to much smaller $\hat{\sigma}_j^2$ for HCL 1, 2, and 3 than for HCL 4. Furthermore, HCL 4 is more sensitive to deviations from the expectation in the next accounting year. This explains why MSEP and CDRU for HCL 4 are significantly higher than in HCL 1, 2, and 3.

The MSEP of Mack BF is smaller than the MSEP of HCL 3 because Mack BF gives much lower estimates for the parameter estimation error in recent accident years. The estimated process variances are very close.

We see that the additional uncertainty in the μ_i in HCL 2 compared to HCL 1 increases the MSEP and CDRU but only marginally affects the reserves.

10. LIMITATIONS AND POSSIBLE EXTENSIONS

This section illustrates several possibilities to further extend and generalise our model we though refrain from working out the details. Most extensions to the classical CL model can also be translated to our model. Additional insight could be gained by reformulating our assumptions on mean and variance in a GLM or Bayesian framework, as it has been done for the CL model.

The distinction between the γ_j and β_j parameters in the model setup was made to allow a deduction of MSEP and CDRU estimates similar to Mack (1993). By using distributional assumptions (as e.g., in Saluz et al. (2011)), one could remove this distinction and furthermore avoid the necessity to smooth and rescale the γ_j's.

Accounting year payments and their uncertainty can be estimated with the approach given in Section 5.5 in Wüthrich et al. (2010). This allows a market-consistent valuation.

The HCL model can be formulated as a special case in the class of linear stochastic reserving methods (LSRM), introduced in Dahms (2010). Using the notation therein, the model can be extended to use more sources of information for predicting ultimate claims.

Model 3.1 can be interpreted within a time series model where $C_{i,j} = E[C_{i,j} \mid C_{i,j-1}] + \sqrt{\sigma_j^2 \mu_i} \varepsilon_{i,j}$ with i.i.d. residuals $\varepsilon_{i,j}$. Under certain assumptions, bootstrapping can then be used to estimate the density of the reserve distribution as well as risk measures. Time and data dependencies of the β_j and the $\alpha_{i,j}$ can also be taken into account. Dependencies with other triangles can be modelled with the approaches given in Merz and Wüthrich (2009) and Kirschner et al. (2008).

A different volume dependence of the process variance can be incorporated, for instance through assuming $var(C_{i,j} \mid C_{i,j-1}) = \sigma_j^2 \mu_i^{\phi}$ for some fixed ϕ.

Generalising from the CL model, we would expect the process variance to be dependent on $m_{i,j}$, i.e., $var(C_{i,j} \mid C_{i,j-1}) = \sigma_j^2 m_{i,j}$. However, this approach leads to complicated estimation formulas as $m_{i,j}$ is not $\mathcal{D}_I$-measurable for $i + j > I$, in contrast to μ_i. The effects of this issue can be seen in the model proposed in Schnieper (1991), see also Lemma 10.13 in Wüthrich and Merz (2008). Furthermore, the μ_i in $var(C_{i,j} \mid C_{i,j-1})$ merely represent a volume measure.

11. CONCLUSION

This paper discussed the HCL method which provides a class of distribution-free reserving models that combines two different claims reserving methodologies: a data-reliant method which is multiplicative in structure, resembling CL, and an expert-reliant method which is additive in structure, resembling BF. Contrary to the existing literature, the HCL method does not consider the combination of the two methodologies as a credibility mixture. Instead, the weighting between the two reserving methodologies is seen as a prior model selection reflecting the actuary's assessment of the development of the claims.

The approach to simultaneously use CL and BF to determine reserves is often used, but stochastic model versions of CL and BF generally make different underlying assumptions. The HCL method embeds this approach within a distribution-free framework, which provides estimators for parameters, reserves, uncertainty in ultimate claims and the uncertainty of the CDR. Moreover, we provide an estimator for the uncertainty of the CDR in the BF method.

As our method falls in the class of CL-type reserving methods, most modifications and extensions that are available for the CL method can also be applied to our method.

Acknowledgements

As former SCOR Fellow, Philipp Arbenz thanks SCOR for financial support. The authors thank Eric Dal Moro, Alois Gisler, Annina Saluz, Frédéric Schwach, Richard Verrall and Mario V. Wüthrich for their helpful comments.

Biographies of the Authors

Philipp Arbenz (philipp.arbenz@gmail.com) is currently an Actuarial Analyst at SCOR Switzerland. He completed his PhD in Mathematics at Risklab ETH Zürich between 2009-2012 under the supervision of Prof. Paul Embrechts. He has been awarded the SCOR Actuarial Fellowship for Switzerland in 2009 for the integrality of his PhD and the Walter-Saxer-Versicherungs-Hochschulpreis for his PhD Thesis in 2013. His thesis topic and his main research interests lie in multivariate risk modeling in non-life insurance.

Robert Salzmann (robert.salzmann@math.ethz.ch) is currently working as a Quantitative Finance Specialist in the financial industry. He completed his PhD in Mathematics at Risklab ETH Zürich between 2009-2012 under the supervision of Prof. Paul Embrechts and Prof. Mario V. Wüthrich. He has been awarded the Walter-Saxer-Versicherungs-Hochschulpreis for his Master Thesis in 2009 and the Polya-Prize for his Master diploma at ETH Zürich.

BIBLIOGRAPHY

Alai, D. (2010). Prediction uncertainty in a weighted method approach to claims reserving. Working paper.

Benktander, G. (1976). An approach to credibility in calculating IBNR for casualty excess reinsurance. The Actuarial Review, 312:7.

Buchwalder, M., Bühlmann, H., Merz, M., and Wüthrich, M. V. (2006). The mean square error of prediction in the chain ladder reserving model (Mack and Murphy revisited). Astin Bulletin, 36(2):521-571.

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

Busse, M., Müller, U., and Dacorogna, M. (2010). Robust estimation of reserve risk. Astin Bulletin, 40(2):453-489.

Dahms, R. (2012). Linear stochastic reserving models. Astin Bulletin, 42(1):1-34.

FINMA (2006). Technical document on the Swiss Solvency Test. Version 2, October 2006, available under www.finma.ch

Kirschner, G., Kerley, C., and Isaacs, B. (2008). Two approaches to calculating correlated reserve indications across multiple lines of business. Variance, 2(1):15-38.

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserves estimates. Astin Bulletin, 23(2):213-225.

Mack, T. (1994). Measuring the variability of chain ladder reserve estimates. CAS Forum, Spring 1994:101-182.

Mack, T. (1999). The standard error of chain ladder reserve estimates: recursive calculation and inclusion of a tail factor. Astin Bulletin, 29(2):361-366.

Mack, T. (2000). Credible claims reserves: the Benktander method. Astin Bulletin, 30(2):333-347.

Mack, T. (2006). Parameter estimation for Bornhuetter/Ferguson. CAS Forum, Fall 2006:141-157.

Mack, T. (2008a). Correction note to "The Prediction Error of Bornhuetter/Ferguson". Astin Bulletin, 38(2):669.

Mack, T. (2008b). The prediction error of Bornhuetter /Ferguson. Astin Bulletin, 38(1):87-103.

Merz, M. and Wüthrich, M. V. (2009). Prediction error of the multivariate additive loss reserving method for dependent lines of business. Variance, 3(1):131-151.

Meyer, M. and Booker, J. (2001). Eliciting and Analyzing Expert Judgement: A Practical Guide. Society for Industrial and Applied Mathematics, Philadelphia.

NAIC (2010). National Association of Insurance Commissioners: Schedule P. http://www.naic.org/store_idp_sched_p.htm

Neuhaus, W. (1992). Another pragmatic loss reserving method or Bornhuetter-Ferguson revisited. Scandinavian Actuarial Journal, 2:151-162.

Saluz, A. (2010). Stochastic Models and Mean Square Error of the Chain Ladder and Bornhuetter/Ferguson Reserving Method. Master's thesis, ETH Zurich.

Saluz, A., Gisler, A., and Wüthrich, M. V. (2011). Development pattern and prediction error for the stochastic Bornhuetter-Ferguson claims reserving model. Astin Bulletin, 41(2):279-317.

Schnieper, R. (1991). Separating true IBNR and IBNER claims. Astin Bulletin, 21(1):111-127.

Verrall, R. (2004). A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving. North American Actuarial Journal, 8(3):67-89.

Wüthrich, M. V., Bühlmann, H., and Furrer, H. (2010). Market-Consistent Actuarial Valuation. Springer.

Wüthrich, M. V. and Merz, M. (2008). Stochastic Claims Reserving Methods in Insurance. Wiley.

Wüthrich, M. V., Merz, M. and Lysenko, N. (2009). Uncertainty of the claims development result in the chain ladder method. Scandinavian Actuarial Journal, 109(1):63-84.

A PROOF OF SECTION 5

Theorem A.1. *For the* $\psi_{i,n,k}$*, the following holds:*

1. $\mathrm{E}[\psi_{i,n,k} \mid \mathrm{B}_{k-1}] = 0$,
2. *For* $k_1 < k_2$*, we have* $\mathrm{E}[\psi_{i_1,n_1,k_1}\psi_{i_2,n_2,k_2} \mid \mathrm{B}_{k_2-1}] = 0$,
3. *For* $k_1 = k_2 = k$*, we have*

$$\mathrm{E}[\psi_{i_1,n_1,k}\psi_{i_2,n_2,k} \mid \mathrm{B}_{k-1}] = \hat{\Psi}^I_{i_1,n_1}\hat{\Psi}^I_{i_2,n_2} b_{i_1,n_1,k} b_{i_2,n_2,k} \frac{\sigma_k^2}{\Omega_k^I}\left(\prod_{k<m\le J}\frac{\xi_{i_1,m}}{\hat{\xi}^I_{i_1,m}}\frac{\xi_{i_2,m}}{\hat{\xi}^I_{i_2,m}}\right),$$

where the $b_{i,n,k}$ *are defined as in Section 5.*

Proof. (1) Note that for all k, $\psi_{i,n,k}$, $\hat{\kappa}^I_{i,k}$ and $\hat{\xi}^I_{i,k}$ are B_k-measurable. Suppose that $n < k$, the other cases can be treated analogously. Then we have

$$\begin{aligned}\mathrm{E}\left[\psi_{i,n,k} \mid \mathrm{B}_{k-1}\right] &= \mathrm{E}\left[\hat{\kappa}^I_{i,n}\left(\prod_{n<m<k}\hat{\xi}^I_{i,m}\right)\left(\hat{\xi}^I_{i,k} - \xi_{i,k}\right)\left(\prod_{k<m\le J}\xi_{i,m}\right) \mid \mathrm{B}_{k-1}\right]\\ &= \hat{\kappa}^I_{i,n}\left(\prod_{n<m<k}\hat{\xi}^I_{i,m}\right)\mathrm{E}[\hat{\xi}^I_{i,k} - \xi_{i,k} \mid \mathrm{B}_{k-1}]\left(\prod_{k<m\le J}\xi_{i,m}\right) = 0.\end{aligned}$$

(2) $\mathrm{E}[\psi_{i_1,n_1,k_1}\psi_{i_2,n_2,k_2} \mid \mathrm{B}_{k_2-1}] = \psi_{i_1,n_1,k_1}\mathrm{E}[\psi_{i_2,n_2,k_2} \mid \mathrm{B}_{k_2-1}] = 0.$

(3) We only prove the equality for the case $k > n_1,\ k > n_2$, the other cases can be treated completely analogously:

$$\begin{aligned}\mathrm{E}\left[\psi_{i_1,n_1,k}\psi_{i_2,n_2,k} \mid \mathrm{B}_{k-1}\right] &= \mathrm{E}\Bigg[\hat{\kappa}^I_{i_1,n_1}\left(\prod_{n_1<m_1<k}\hat{\xi}^I_{i_1,m_1}\right)\left(\hat{\xi}^I_{i_1,k} - \xi_{i_1,k}\right)\left(\prod_{k<m_1\le J}\xi_{i_1,m_1}\right)\\ &\quad\times\hat{\kappa}^I_{i_2,n_2}\left(\prod_{n_2<m_2<k}\hat{\xi}^I_{i_2,m_2}\right)\left(\hat{\xi}^I_{i_2,k} - \xi_{i_2,k}\right)\left(\prod_{k<m_2\le J}\xi_{i_2,m_2}\right) \mid \mathrm{B}_{k-1}\Bigg]\\ &= \hat{\Psi}^I_{i_1,n_1}\hat{\Psi}^I_{i_2,n_2}\frac{\mathrm{E}[(\hat{\xi}^I_{i_1,k} - \xi_{i_1,k})(\hat{\xi}^I_{i_2,k} - \xi_{i_2,k}) \mid \mathrm{B}_{k-1}]}{\hat{\xi}^I_{i_1,k}\hat{\xi}^I_{i_2,k}}\left(\prod_{k<m\le J}\frac{\xi_{i_1,m}}{\hat{\xi}^I_{i_1,m}}\frac{\xi_{i_2,m}}{\hat{\xi}^I_{i_2,m}}\right).\end{aligned}$$

The last equation follows from multiplying and dividing by $\prod_{k \le m \le J} \hat{\xi}^I_{i_1,m} \hat{\xi}^I_{i_2,m}$. For instance, the multiplication with $\prod_{k \le m \le J} \hat{\xi}^I_{i_1,m}$ provides the missing factors for $\hat{\Psi}^I_{i_1,n_1}$, which is why the expectation above simplifies. For the remaining parts inside the expectation, we get

$$\mathrm{E}[(\hat{\xi}^I_{i_1,k} - \xi_{i_1,k})(\hat{\xi}^I_{i_2,k} - \xi_{i_2,k}) \mid \mathrm{B}_{k-1}] = \frac{\alpha_{i_1,k}}{\beta_{k-1}} \frac{\alpha_{i_2,k}}{\beta_{k-1}} var(\hat{\gamma}^I_k \mid \mathrm{B}_{k-1}) = \frac{\alpha_{i_1,k}}{\beta_{k-1}} \frac{\alpha_{i_2,k}}{\beta_{k-1}} \frac{\sigma^2_k}{\Omega^I_k}.$$

■

B PROOF OF SECTION 6

In order to derive an approximation for $\mathrm{E}[\theta_{i_1,n_1,k_1}\theta_{i_2,n_2,k_2} \mid \mathrm{D}_I]$, we first apply a change of conditioning. That is, we will condition on sets like G_k, where G_k is D_I joined with the new diagonal of D_{I+1}, which is cut off at $j = k$, i.e.,

$$\mathrm{G}_k = \mathrm{D}_I \cup \{C_{I-j+1,j} : 0 \le j \le k\}.$$

The set G_k is illustrated in Figure 3.

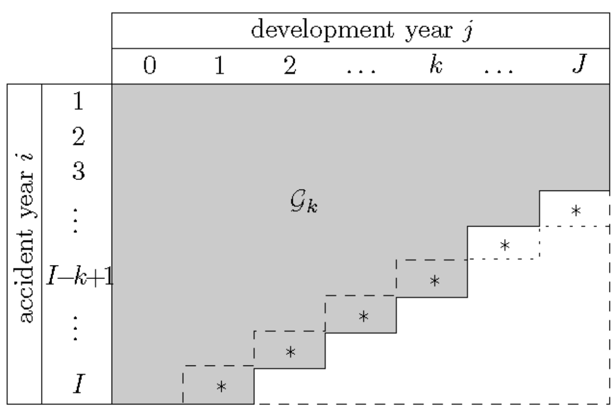

Figure 3. G_k which is D_I joined with the new diagonal of D_{I+1}, which is cut off at $j = k$.

As a first approximation, we set

$$\mathrm{E}\left[\theta_{i_1,n_1,k_1}\theta_{i_2,n_2,k_2} \middle| \mathrm{D}_I\right] \approx \mathrm{E}\left[\theta_{i_1,n_1,k_1}\theta_{i_2,n_2,k_2} \middle| \mathrm{G}_{\max\{k_1,k_2\}-1}\right].$$

Note that $\theta_{i,n,k}$ is G_k-measurable where $\theta_{i,n,k}$ is given in Section 6. Hence,

$$\mathrm{E}\left[\theta_{i_1,n_1,k_1}\theta_{i_2,n_2,k_2}\middle|\mathcal{G}_{\max\{k_1,k_2\}-1}\right]=\hat{\Psi}^{I+1}_{i_1,n_1}h_{i_1,n_1,k_1}\hat{\Psi}^{I+1}_{i_2,n_2}h_{i_2,n_2,k_2}\mathrm{E}\left[e_{i_1,n_1,k_1}e_{i_2,n_2,k_2}\middle|\mathcal{G}_{\max\{k_1,k_2\}-1}\right],$$

Where

$$h_{i,n,k}=\begin{cases}\dfrac{1}{\hat{\xi}^{I+1}_{i,k}}\prod_{k<m\le J}\left(\hat{\xi}^{I}_{i,m}/\hat{\xi}^{I+1}_{i,m}\right), & \text{for } k>n,\\ \dfrac{1}{\hat{\kappa}^{I+1}_{i,n}}\prod_{n<m\le J}\left(\hat{\xi}^{I}_{i,m}/\hat{\xi}^{I+1}_{i,m}\right), & \text{for } k=n.\end{cases}$$

Note that the terms taken out in the above expectation contain terms which are not $\mathcal{G}_{\max\{k_1,k_2\}-1}$ measurable. A closer inspection reveals that all these terms cancel out in (B.1) because they appear as factors in the $\hat{\Psi}^{I+1}_{i,n}$ and as reciprocal in the $h_{i,n,k}$.

In a next step, we deduce an approximation of $\mathrm{E}[e_{i_1,n_1,k_1}e_{i_2,n_2,k_2}\,|\,\mathcal{G}_{\max\{k_1,k_2\}-1}]$. The difference of $\hat{\gamma}^I_j$ and the unknown γ_j is not measurable and is best estimated by 0. For the consideration of the distortion $\hat{\gamma}^I_j-\hat{\gamma}^{I+1}_j$, we approximate $e_{i,n,k}$ with $\tilde{e}_{i,n,k}$, by replacing all $\hat{\gamma}^I_j$ in $e_{i,n,k}$ by γ_j. To do this we introduce

$$\tilde{\gamma}_j=\frac{\Omega^I_j}{\Omega^{I+1}_j}\gamma_j+\frac{\omega_{I-j+1,j}}{\Omega^{I+1}_j}\Gamma_{I-j+1,j}\approx\hat{\gamma}^{I+1}_j,$$

and hence,

$$\tilde{e}_{i,n,k}=\begin{cases}\alpha_{i,k}/\beta_{k-1}(\gamma_k-\tilde{\gamma}_k), & \text{for } k>n,\\ \mu_i(1-\alpha_{i,n})(\gamma_n-\tilde{\gamma}_n), & \text{for } k=n, n>I-i+1,\\ C_{i,I-i}\xi_{i,n-1}+\mu_i(1-\alpha_{i,n})\gamma_n-C_{i,I-i+1}, & \text{for } k=n=I-i+1.\end{cases}$$

Using the above definitions for $e_{i,n,k}$, we approximate

$$E\left[e_{i_1,n_1,k_1}e_{i_2,n_2,k_2}\middle|G_{\max\{k_1,k_2\}-1}\right]\approx E\left[\tilde{e}_{i_1,n_1,k_1}\ \tilde{e}_{i_2,n_2,k_2}\middle|G_{\max\{k_1,k_2\}-1}\right].$$

The following theorem summarizes several properties of $\tilde{e}_{i,n,k}$, that are similar to Theorem A.1. We will use these properties to estimate the conditional expectation $E[\tilde{e}_{i_1,n_1,k_1} \; \tilde{e}_{i_2,n_2,k_2} \mid G_{\max\{k_1,k_2\}-1}]$.

Theorem B.1. *For the* $\tilde{e}_{i,n,k}$ *, the following results holds*

1. $E[\tilde{e}_{i,n,k} \mid G_{k-1}] = 0$.
2. *For* $k_1 < k_2$*, we have* $E[\tilde{e}_{i_1,n_1,k_1} \tilde{e}_{i_2,n_2,k_2} \mid G_{k_2-1}] = 0$.
3. *For* $k_1 = k_2 = k$*, we have*

$$E[\tilde{e}_{i_1,n_1,k_1} \tilde{e}_{i_2,n_2,k_2} \mid G_{k_2-1}] = d_{i_1,n_1,k_1} d_{i_2,n_2,k_2} \mu_{I-k+1} \sigma_k^2,$$

where

$$d_{i,n,k} = \begin{cases} \dfrac{\alpha_{i,k}}{\beta_{k-1}} \dfrac{\omega_{I-k+1,k}}{\Omega_k^{I+1}} \dfrac{1}{m_{I-k+1,k}}, & \text{for } k > n, \\ \mu_i (1-\alpha_{i,n}) \dfrac{\omega_{I-k+1,k}}{\Omega_k^{I+1}} \dfrac{1}{m_{I-k+1,k}}, & \text{for } k = n, n > I-i+1, \\ 1, & \text{for } k = n = I-i+1. \end{cases}$$

Proof. (1) This follows directly from $\mathrm{E}[\Gamma_{I-k+1,k} \mid \mathrm{G}_{k-1}] = \gamma_k$.

(2) $\tilde{e}_{i,n,k}$ is G_k -measurable. Hence we have the equality

$$E[\tilde{e}_{i_1,n_1,k_1} \tilde{e}_{i_2,n_2,k_2} \mid G_{k_2-1}] = \tilde{e}_{i_1,n_1,k_1} E[\tilde{e}_{i_2,n_2,k_2} \mid G_{k_2-1}] = 0$$

(3) All stochasticity contained in $\tilde{e}_{i,n,k}$ with respect to G_{k-1} is contained in $C_{I-k+1,k}$. Furthermore, $\tilde{e}_{i,n,k}$ depends linearly on $C_{I-k+1,k}$. Hence, it remains to show that the proportionality factor is equal to $-d_{i,n,k}$, i.e., $\tilde{e}_{i,n,k} \mid G_k = -d_{i,n,k} C_{I-k+1,k} + (const)$ which can be done by simple calculations. ■

In the last step to get an approximation of $\mathrm{E}[\theta_{i_1,n_1,k_1} \theta_{i_2,n_2,k_2} \mid \mathrm{D}_I]$ we replace all unknown quantities by their estimates based on D_I. For the two different cases $n = I-i+1$ and $n > I-i+1$, we have

$$\begin{aligned}
\hat{\Psi}^{I+1}_{i,n} &= C_{i,I-i+1} \prod_{n<m\le J} \hat{\xi}^{I+1}_{i,m} \approx (C_{i,I-i}\hat{\xi}_{i,I-i+1} + \mu_i(1-\alpha_{i,I-i+1})\hat{\gamma}_{I-i+1} \prod_{n<m\le J} \hat{\xi}^{I}_{i,m} \\
&= \tilde{\Psi}^{I+1}_{i,n} = \hat{\Psi}^{I}_{i,n-1} + \hat{\Psi}^{I}_{i,n}, \quad for \quad n = I-i+1,
\end{aligned}$$

$$\begin{aligned}
\hat{\Psi}^{I+1}_{i,n} &= \mu_i(1-\alpha_{i,n})\hat{\gamma}^{I+1}_{n} \prod_{n<m\le J} \hat{\xi}^{I+1}_{i,m} \approx \mu_i(1-\alpha_{i,n})\hat{\gamma}^{I}_{n} \prod_{n<m\le J} \hat{\xi}^{I}_{i,m} \\
&= \tilde{\Psi}^{I+1}_{i,n} = \hat{\Psi}^{I}_{i,n-1}, \quad for \quad n > I-i+1,
\end{aligned}$$

An analogous approximation can be used to deduce the approximation

$$h_{i,n,k} d_{i,n,k} \approx g_{i,n,k}.$$

We illustrate the case $k > n$, the other cases being analogous. For $k > n$, we have

$$\begin{aligned}
h_{i,n,k} d_{i,n,k} &= \frac{1}{\hat{\xi}^{I+1}_{i,k}} \prod_{k<m\le J} \left(\hat{\xi}^{I}_{i,m} / \hat{\xi}^{I+1}_{i,m}\right) \frac{\alpha_{i,k}}{\beta_{k-1}} \frac{\omega_{I-k+1,k}}{\Omega^{I+1}_{k}} \frac{1}{m_{I-k+1,k}}, \\
&\approx \frac{1}{\hat{\xi}^{I}_{i,k}} \frac{\alpha_{i,k}}{\beta_{k-1}} \frac{\omega_{I-k+1,k}}{\Omega^{I+1}_{k}} \frac{1}{m_{I-k+1,k}} = g_{i,n,k}.
\end{aligned}$$

The approach used here is similar to the one used in Appendix A, but is different to the approach used in Bühlmann et al. (2009).