.

Claim reservation in non-life insurance

by

Jostein Paulsen

University of Copenhagen, Denmark

August 24, 2016

Chapter 1
Claim reservation

One of the major tasks for actuaries is to predict future claims on premiums received. Accidents are not always reported immediately, sometimes due to procrastination of the claimant, but more often because the damage causing the claim is not observed immediately, and it can take years before it is observed, and then more years until it is settled. Such long term effects are often occurring in lines of business that incur human life, like workers compensation. Workers may be exposed to environmental hazards, and it can take years before the health effects are realized. After that it can take a long time with litigation until the claims are finally settled. A line of business with a much shorter runoff is vehicle insurance for cars, but even here there may be unobserved damages to the cars that are not discovered immediately.

When the claim, or accident, is reported, it can still take a long time until it is settled, like in workers compensation mentioned above. Usually the company will set aside a reserve for any reported claim, and this may or may not be sufficient to cover the actual claimsize.

The chapter starts with a general introduction to the problem and a few useful formulas. Then in Chapter 1.2 the chain-ladder method, probably the most popular of them all, is introduced, both with the standard nonparametric approach as well as some parametric approaches. In the nonparametric subsection the famous Mack formula for the mean squared error is proved.

The other classical method, the Bornhuetter-Ferguson method, is presented in Chapter 1.3. Again there is both a nonparametric and a parametric approach. In the latter case a paramtetric model that yields the same reserves as the chain-ladder method is presented.

There is a hybrid method of the two main methods already covered, the Benktander-Hovinen method, and this is briefly discussed in Chapter 1.4.

All these chapters are mainly focussed on expected outstanding claims, but in Chapter 1.5 we discuss how to calculate risk margins or solvency capital requirements (SCR), with main emphasis on bootstrap methods.

1.1 Introduction

There are several kinds of unpaid claims, the four most common definitions are:

IBNR: Incurrred But Not Reported.
RBNS: Reported But Not Settled.
IBNS: Incurred But Not Settled.
CBNI: Covered But Not Incurred.

Here IBNR+RBNS=IBNS. Furthermore, IBNR+CBNI are total outstanding claims that are unknown at present. The RBNS will be the sum of the payments plus reserves set aside for known, but unsettled, claims, and this will vary year by year for the same claim as reserves are matched with real payments.

The aim is to predict total claims, or equivalently, total outstanding claims. These are made up by IBNR+CBNI plus the changes in RBNS. To do so, historical data are used. Assume that J runoff years after the accident year, all claims are fully settled. Assume there are data for accident years $0,1,\ldots,I$, where we at present are at the end of year I. Therefore, for accident year i, there are data for runoff years $0,\ldots,j^*(i)$, where $j^*(i) = (I-i) \wedge J$. But runoff years $j^*(i)+1,\ldots,J$ are in the future, so their contribution to the total payoff must be estimated. In order to have at least one year in full runoff, it is assumed that $I \geq J$.

An alternative to predicting total outstanding claims on premiums received is to predict separately IBNR, IBNS and CBNI. This will not change any of the analysis of this chapter.

Table 1.1 shows the problem at hand with $I = J+1$. This is called a runoff triangle, as the observed part makes a (almost) triangle.

Table 1.1 Runoff triangle

Accident years i	Development years j: $0\ 1\ 2\ \cdots\quad j\quad \cdots\ J$
0	
1	
2	Observations of $C_{i,j}$ or $X_{i,j}$
$\vdots$	
i	
$\vdots$	$C_{i,j}$ or $X_{i,j}$ unknown
$I-1$	
I	

For accident year i and runoff year j, total claims registered is C_{ij}. Typically, but not always, C_{ij} will increase with j. The reason for a potential decrease is that reserved claims often turn out to be smaller than expected, or even zero. Such a decrease will most likelely take place at the end of the runoff period. If the triangle is for paid claims, or for number of claims reported, an eventual decrease is much less likely.

In a runoff situation it is possible that underwriting stopped k years ago, so that all accident years have been observed $k+1$ years. In that case we may just lump the first $k+1$ development years together and consider this as a "new year" in the table.

Another possible situation is when $I = J$, but claims are not fully settled until runoff year J^* say, where $J^* > J$. In this case the actuary will have to resort to some kind of a priori specification of the runoff for the last $J^* - J$ years.

The year to year changes are given by

$$X_{ij} = C_{ij} - C_{i,j-1}, \quad i = 0, \dots, I, \;\; j = 0, \dots, J,$$

with $C_{i,-1} = 0$ so that $X_{i0} = C_{i0}$.

If the portfolios of the various accident years are very different, to obtain homogeneity it may be a good idea to work with the claim ratio, i.e. claims divided by total premiums, instead of claims directly. This may also take care of some problems with inflation, but if the runoff J is very long it may be necessary to adjust for inflation. If the premium level fluctuates a lot, an alternative to dividing by premiums for each accident year is to divide by some kind of volume measure.

It is common to exclude large claims in the triangles to avoid large fluctuations. This does not mean that such claims should not be considered in a reserving context, just that they have to be treated separately and probably on a more ad hoc basis.

Assume we are at year I, and want to predict future claims, or what amounts to the same, total claims including unreported claims. The information available is

$$\mathscr{D} = \sigma\{C_{ij},\, i+j \le I,\, j \le J\} = \sigma\{X_{ij},\, i+j \le I,\, j \le J\}.$$

The reader should not be intimidated by the notation here, it is just for notational convenience. In fact

$$E[Y|\mathscr{D}] = E[Y|C_{ij},\, i+j \le I,\, j \le J] = E[Y|X_{ij},\, i+j \le I,\, j \le J].$$

Also define the total information up to runoff year k,

$$\mathscr{E}_k = \sigma\{C_{ij},\, i \le I,\, j \le k\} = \sigma\{X_{ij},\, i \le I,\, j \le k\}.$$

Here $\mathscr{E}_0 \subset \mathscr{D}$, but for $k > 0$, $\mathscr{D}$ and $\mathscr{E}_k$ are different. With $\mathscr{D} \vee \mathscr{E}_k$ we mean the combined information in $\mathscr{D}$ and $\mathscr{E}_k$. In particular $\mathscr{D} \vee \mathscr{E}_0 = \mathscr{D}$. The rules for conditional expectation and variance are readily transferred to information sets. So with $\mathscr{G} \subset \mathscr{D} \vee \mathscr{E}_k$, for any integrable random variable X,

$$\begin{aligned} E[X|\mathscr{G}] &= E[E[X|\mathscr{D} \vee \mathscr{E}_k]|\mathscr{G}], \\ \mathrm{Var}[X|\mathscr{G}] &= E[\mathrm{Var}[X|\mathscr{D} \vee \mathscr{E}_k]|\mathscr{G}] + \mathrm{Var}[E[X|\mathscr{D} \vee \mathscr{E}_k]|\mathscr{G}]. \end{aligned} \tag{1.1}$$

With $\mathscr{D} \wedge \mathscr{E}_k$ we mean the information in both $\mathscr{D}$ and $\mathscr{E}_k$, so that

$$\mathscr{D} \wedge \mathscr{E}_k = \sigma\{C_{ij},\, i+j \le I,\, j \le k\} = \sigma\{X_{ij},\, i+j \le I,\, j \le k\}.$$

In particcular $\mathscr{D} \wedge \mathscr{E}_J = \mathscr{D}$.

In all further developments we use the convention for $j < k$,

$$\prod_{i=k}^{j} a_i = 1.$$

With $j^*(i) = (I-i) \wedge J$ it is easy to see that for all sequences $\{a_{ij}\}$,

$$\sum_{i=0}^{I} \sum_{j=0}^{j^*(i)} a_{ij} = \sum_{j=0}^{J} \sum_{i=0}^{I-j} a_{ij}. \tag{1.2}$$

$$\sum_{i=0}^{I-1} \sum_{j=0}^{j^*(i)-1} a_{ij} = \sum_{j=0}^{J-1} \sum_{i=0}^{I-j-1} a_{ij}. \tag{1.3}$$

The total number of observations is

$$n = \sum_{j=0}^{J}\sum_{i=0}^{I-j} 1 = \sum_{j=0}^{J}(I-j+1) = (J+1)\left(I+1-\frac{1}{2}J\right). \tag{1.4}$$

Total claims over the whole period is denoted by

$$C_T = \sum_{i=0}^{I} C_{iJ} = \sum_{i=0}^{I}\sum_{j=0}^{J} X_{ij},$$

and total outstanding claims at time I is

$$C_O = \sum_{i=0}^{I}(C_{iJ} - C_{i,j^*(i)}) = \sum_{i=I-J+1}^{I}\ \sum_{j=I-i+1}^{J} X_{ij}.$$

Thus,

$$\mathrm{Var}[C_T|\mathscr{D}] = \mathrm{Var}[C_O|\mathscr{D}].$$

If accident years are independent,

$$\mathrm{Var}[C_T|\mathscr{D}] = \sum_{i=0}^{I}\mathrm{Var}[C_{iJ}|\mathscr{D}].$$

If in addition increments are independent,

$$\mathrm{Var}[C_T|\mathscr{D}] = \sum_{i=I-J+1}^{I}\ \sum_{j=I-i+1}^{J}\mathrm{Var}[X_{ij}].$$

Another quantity that has attracted some interest is the mean squared error of an estimator $\hat{C}_T$ of C_T. Since it is an estimator, $\hat{C}_T$ will be a function of elements of $\mathscr{D}$, and so

$$\mathrm{MSE}[\hat{C}_T|\mathscr{D}] = E[(C_T - \hat{C}_T)^2|\mathscr{D}] = \mathrm{Var}[C_T|\mathscr{D}] + (E[C_T|\mathscr{D}] - \hat{C}_T)^2. \tag{1.5}$$

The most difficult part to estimate here is the second term, and a lot of effort has gone into this estimation as will be commented on below.

1.2 The chain-ladder method

Throughout this section the following two assumptions are made.

CL1: Different accident years are independent, i.e. C_{ij} and C_{kl} are independent whenever $i \neq k$.
CL2: There are positive factors $f_0, \ldots, f_{J-1}$ so that

$$E[C_{i,j+1}|\mathscr{E}_j] = f_j C_{ij}, \quad j = 0, \ldots, J-1.$$

According to CL2, if claims for a given accident year up to runoff year j have been large, the claims for the next runoff year can be expected to be large as well when $f_j > 1$.

Total claims for year i is then estimated as

$$C_{iJ}^{\mathrm{CL}} = E[C_{iJ}|\mathscr{D}],$$

so that $C_{iJ}^{\mathrm{CL}} = C_{iJ}$ when $i \le I-J$. For $i > I-J$ we get by (1.1),

$$E[C_{iJ}|\mathscr{D}] = E[E[C_{iJ}|\mathscr{D} \vee \mathscr{E}_{J-1}]|\mathscr{D}] = f_{J-1}E[C_{i,J-1}|\mathscr{D}].$$

Continuing gives the estimator

$$C_{iJ}^{\mathrm{CL}} = C_{i,I-i} \prod_{j=I-i}^{J-1} f_j. \tag{1.6}$$

Total expected claims are then

$$C_T^{\mathrm{CL}} = \sum_{i=0}^{I} C_{iJ}^{\mathrm{CL}}.$$

To say something about variance, one more assumption is needed.

CL3: There are positive factors $\sigma_0^2, \ldots, \sigma_{J-1}^2$ so that

$$\mathrm{Var}[C_{i,j+1}|\mathscr{E}_j] = \sigma_j^2 C_{ij}, \quad j = 0, \ldots, J-1.$$

In (1.13) below an unbiased estimator of σ_j^2 is given. However, as this is based on only a few observations, and more so as j increases, it can be quite unstable. As C_{ij} normally increases with j, while f_j decreases towards 1 and the uncertainty in $C_{i,j+1}$ given $\mathscr{E}_j$ decreases, a simplifying parametric form would be

$$\sigma_j^2 = (f_j - 1)\tau^2, \quad j = 0, \ldots, J-1. \tag{1.7}$$

For notational simplicity, define

$$g_{ik} = \mathrm{Var}[C_{ik}|\mathscr{D}].$$

Clearly $g_{iJ} = 0$ for $i \le I-J$. Using the rule of conditional variance (1.1), for $k > I-i$,

$$g_{ik} = E[\mathrm{Var}[C_{ik}|\mathscr{D} \vee \mathscr{E}_{k-1}]|\mathscr{D}] + \mathrm{Var}[E[C_{ik}|\mathscr{D} \vee \mathscr{E}_{k-1}]|\mathscr{D}] = E[\sigma_{k-1}^2 C_{i,k-1}|\mathscr{D}] + \mathrm{Var}[f_{k-1}C_{i,k-1}|\mathscr{D}].$$

Therefore, with

$$C_{ik}^{\mathrm{CL}} \stackrel{\mathrm{def}}{=} E[C_{ik}|\mathscr{D}] = C_{i,I-i} \prod_{j=I-i}^{k-1} f_j = C_{iJ}^{\mathrm{CL}} \prod_{j=k}^{J-1} f_j^{-1}, \tag{1.8}$$

we get the nice recursion to calculate g_{iJ},

$$g_{ik} = \sigma_{k-1}^2 C_{i,k-1}^{\mathrm{CL}} + f_{k-1}^2 g_{i,k-1}, \quad k = I-i+1, \ldots, J,$$

with initial value $g_{i,I-i} = 0$. An alternative is to solve the recursion directly, again using that $g_{i,I-i} = 0$,

$$
\begin{aligned}
g_{iJ} &= \sigma_{J-1}^2 C_{i,J-1}^{\mathrm{CL}} + f_{J-1}^2 g_{i,J-1} \\
&= \sigma_{J-1}^2 C_{i,J-1}^{\mathrm{CL}} + f_{J-1}^2 (\sigma_{J-2}^2 C_{i,J-2}^{\mathrm{CL}} + f_{J-2}^2 g_{i,J-2}) \\
&= \sigma_{J-1}^2 C_{i,J-1}^{\mathrm{CL}} + f_{J-1}^2 \sigma_{J-2}^2 C_{i,J-2}^{\mathrm{CL}} + f_{J-2}^2 f_{J-1}^2 g_{i,J-2} \\
&= \cdots \\
&= \sum_{k=I-i}^{J-1} \left(\prod_{j=k+1}^{J-1} f_j^2 \right) \sigma_k^2 C_{ik}^{\mathrm{CL}} \\
&= \sum_{k=I-i}^{J-1} \left(\prod_{j=k}^{J-1} f_j^2 \right) \frac{\sigma_k^2}{f_k^2} C_{ik}^{\mathrm{CL}} \\
&= \sum_{k=I-i}^{J-1} \left(\frac{C_{iJ}^{\mathrm{CL}}}{C_{ik}^{\mathrm{CL}}} \right)^2 \frac{\sigma_k^2}{f_k^2} C_{ik}^{\mathrm{CL}} \\
&= (C_{iJ}^{\mathrm{CL}})^2 \sum_{k=I-i}^{J-1} \frac{\sigma_k^2}{f_k^2 C_{ik}^{\mathrm{CL}}}.
\end{aligned}
\tag{1.9}
$$

Here (1.8) was used in the second last equality.

Due to the assumption CL1, total conditional variance for outstanding claims is then

$$
\mathrm{Var}[C_O|\mathscr{D}] = \mathrm{Var}\left[\sum_{i=0}^{I} C_{iJ} \middle| \mathscr{D} \right] = \sum_{i=0}^{I} g_{iJ}.
$$

Remark 1.1. From

$$
f_j C_{ij} = E[C_{i,j+1}|\mathscr{E}_j] = C_{ij} + E[X_{i,j+1}|\mathscr{E}_j]
$$

the assumption CL2 is equivalent to

CL2': There are positive factors $f_0, \ldots, f_{J-1}$ so that

$$
E[X_{i,j+1}|\mathscr{E}_j] = (f_j - 1) C_{ij}, \quad j = 0, \ldots, J-1.
$$

Similarly, CL3 is equivalent to

CL3': There are positive factors $\sigma_0^2, \ldots, \sigma_{J-1}^2$ so that

$$
\mathrm{Var}[X_{i,j+1}|\mathscr{E}_j] = \sigma_j^2 C_{ij}, \quad j = 0, \ldots, J-1.
$$

Remark 1.2. A special case that satisfies CL1-CL3 is

$$
C_{i,j+1} = C_{ij} f_j + \sigma_j \sqrt{C_{ij}}\, e_{i,j+1}, \tag{1.10}
$$

where the e_{ij} are assumed i.i.d. with expectation zero. This model has a consistency issue as it is possible that C_{ij} is negative, in which case $C_{i,j+1}$ cannot be defined. When the probability for this to happen is so small that it is negligible in practice, we will see later that the model can give meaningful results.

1.2.1 A nonparametric approach to chain-ladder estimation

In practice the f_j and σ_j^2 are unknown and need to be estimated. Starting with f_j, define the individual development factors

$$F_{ij} = \frac{C_{i,j+1}}{C_{ij}}.$$

Then

$$E[F_{ij}|\mathscr{E}_j] = \frac{E[C_{i,j+1}|\mathscr{E}_j]}{C_{ij}} = f_j.$$

In particular $E[F_{ij}] = E[E[F_{ij}|\mathscr{E}_j]] = f_j$. A reasonable estimator is therefore

$$\hat{f}_j = \sum_{i=0}^{I-j-1} a_i F_{ij} \quad \text{subject to} \quad \sum_{i=0}^{I-j-1} a_i = 1 \quad \text{and} \quad \mathrm{Var}[\hat{f}_j|\mathscr{E}_j] \ \text{ is minimized.}$$

The condition $\sum_{i=0}^{I-j-1} a_i = 1$ gives that $E[\hat{f}_j|\mathscr{E}_j] = f_j$ and so

$$\mathrm{Var}[\hat{f}_j] = E[\mathrm{Var}[\hat{f}_j|\mathscr{E}_j]] + \mathrm{Var}[E[\hat{f}_j|\mathscr{E}_j]] = E[\mathrm{Var}[\hat{f}_j|\mathscr{E}_j]].$$

Therefore, if $\hat{f}_j$ satisfies $\sum_{i=0}^{I-j-1} a_i = 1$ and minimizes $\mathrm{Var}[\hat{f}_j|\mathscr{E}_j]$, then it also minimizes the unconditional $\mathrm{Var}[\hat{f}_j]$. Also, since accident years are independent and only those F_{ij} with $i \le I-j-1$ are used to calculate $\hat{f}_j$, we get

$$E[\hat{f}_j|\mathscr{D} \wedge \mathscr{E}_j] = E[\hat{f}_j|\mathscr{E}_j] = f_j \quad \text{and} \quad \mathrm{Var}[\hat{f}_j|\mathscr{D} \wedge \mathscr{E}_j] = \mathrm{Var}[\hat{f}_j|\mathscr{E}_j].$$

We have that

$$\mathrm{Var}[F_{ij}|\mathscr{E}_j] = \frac{1}{C_{ij}^2}\mathrm{Var}[C_{i,j+1}|\mathscr{E}_j] = \frac{1}{C_{ij}}\sigma_j^2.$$

To find the a_j, we use that $a_0 = 1 - \sum_{i=1}^{I-j-1} a_i$ so with $\mathbf{a} = (a_1, \ldots, a_{I-j-1})$,

$$h(\mathbf{a}) \stackrel{\text{def}}{=} \mathrm{Var}[\hat{f}_j|\mathscr{E}_j] = \sum_{i=0}^{I-j-1} \frac{\sigma_j^2}{C_{ij}} a_i^2 = \sum_{i=1}^{I-j-1} \frac{\sigma_j^2}{C_{ij}} a_i^2 + \frac{\sigma_j^2}{C_{0j}} \left(1 - \sum_{i=1}^{I-j-1} a_i\right)^2.$$

Then for $k \ge 1$,

$$\frac{\partial}{\partial a_k} h(\mathbf{a}) = 2\frac{\sigma_j^2}{C_{kj}} a_k - 2\frac{\sigma_j^2}{C_{0j}} a_0 = 0 \ \Rightarrow\ a_k = C_{kj}\frac{1}{C_{0j}} a_0.$$

As this clearly also holds for $k = 0$, we can sum up to get

$$1 = \sum_{k=0}^{I-j-1} a_k = C_{\bullet j}\frac{1}{C_{0j}} a_0 \quad \Rightarrow \quad a_0 = \frac{C_{0j}}{C_{\bullet j}},$$

where here and below,

$$C_{\bullet j} = \sum_{i=0}^{I-j-1} C_{ij}.$$

Therefore, plugging a_0 into the equation for a_k,

$$a_k = \frac{C_{kj}}{C_{\bullet j}},$$

and so the minimum variance conditional unbiased estimator is

$$\hat{f}_j = \sum_{k=0}^{I-j-1} \frac{C_{ij}}{C_{\bullet j}} \frac{C_{i,j+1}}{C_{ij}} = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{ij}}. \tag{1.11}$$

For clarity we will sometimes write $\hat{f}_j^{\text{CL}}$. From (1.8) a natural estimator of C_{iJ}^{CL} is

$$\hat{C}_{iJ}^{\text{CL}} = C_{i,I-i} \prod_{j=I-i}^{J-1} \hat{f}_j.$$

From this it is also easy to see that

$$\text{Var}[\hat{f}_j|\mathscr{E}_j] = \frac{\sigma_j^2}{C_{\bullet j}}. \tag{1.12}$$

For $j < k$,

$$E[\hat{f}_j \hat{f}_k|\mathscr{E}_j] = E[E[\hat{f}_j \hat{f}_k|\mathscr{E}_k]|\mathscr{E}_j] = f_k E[\hat{f}_j|\mathscr{E}_j] = f_j f_k.$$

Note that it is the same if we condition on $\mathscr{D} \wedge \mathscr{E}_j$ and $\mathscr{D} \wedge \mathscr{E}_k$ instead, and therefore,

$$E\left[\prod_{j=I-i}^{J-1} \hat{f}_j \middle| \mathscr{D} \wedge \mathscr{E}_{I-i}\right] = \prod_{j=I-i}^{J-1} f_j,$$

and consequently

$$E[\hat{C}_{iJ}^{\text{CL}}|\mathscr{D} \wedge \mathscr{E}_{I-i}] = C_{iJ}^{\text{CL}}.$$

As an estimator for σ_j^2 we set

$$\hat{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} C_{ij} \left(F_{ij} - \hat{f}_j\right)^2. \tag{1.13}$$

We will show that $E[\hat{\sigma}_j^2|\mathscr{E}_j] = \sigma_j^2$ (which is equivalent to $E[\hat{\sigma}_j^2|\mathscr{D} \wedge \mathscr{E}_j] = \sigma_j^2$), so that in particular $E[\hat{\sigma}_j^2] = \sigma_j^2$. To this end note that first that

$$\text{Cov}(F_{ij}, \hat{f}_j) = \frac{1}{C_{ij} C_{\bullet j}} \text{Cov}\left(C_{i,j+1}, \sum_{k=0}^{I-j-1} C_{k,j+1}\right) = \frac{\sigma_j^2}{C_{\bullet j}} = \text{Var}[\hat{f}_j|\mathscr{E}_j].$$

Therefore

$$\begin{aligned} E\left[\left(F_{ij} - \hat{f}_j\right)^2 |\mathscr{E}_j\right] &= E\left[(F_{ij} - f_j)^2|\mathscr{E}_j\right] + E[(\hat{f}_j - f_j)^2|\mathscr{E}_j] - 2E\left[(F_{ij} - f_j)(\hat{f}_j - f_j)|\mathscr{E}_j\right] \\ &= \text{Var}[F_{ij}|\mathscr{E}_j] + \text{Var}[\hat{f}_j|\mathscr{E}_j] - 2\text{Cov}(F_{ij}, \hat{f}_j|\mathscr{E}_j) \\ &= \left(\frac{1}{C_{ij}} - \frac{1}{C_{\bullet j}}\right)\sigma_j^2. \end{aligned}$$

Adding up gives

$$E[\hat{\sigma}_j^2|\mathscr{E}_j] = \frac{1}{I-j-1}\sum_{i=0}^{I-j-1} C_{ij}\left(\frac{1}{C_{ij}} - \frac{1}{C_{\bullet j}}\right)\sigma_j^2 = \sigma_j^2,$$

as was to be proved. Using (1.9) we can define for $i > I-J$,

$$\widehat{\mathrm{Var}}[C_{iJ}|\mathscr{D}] = (\hat{C}_{iJ}^{\mathrm{CL}})^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2 \hat{C}_{ij}^{\mathrm{CL}}}, \tag{1.14}$$

where $\hat{C}_{ik}^{\mathrm{CL}}$ is defined in (1.8), but with $\hat{f}_j$ instead of f_j.

Setting $\widehat{\mathrm{Var}}[C_{iJ}|\mathscr{D}] = 0$ for $i \le I-J$, the estimated variance of outstanding claims is then

$$\widehat{\mathrm{Var}}[C_O] = \widehat{\mathrm{Var}}\left[\sum_{i=I-J+1}^{I} C_{iJ}\middle|\mathscr{D}\right] = \sum_{i=I-J+1}^{I} \widehat{\mathrm{Var}}[C_{iJ}|\mathscr{D}]. \tag{1.15}$$

If $I \le J$ we cannot define $\hat{\sigma}_{J-1}^2$, but as σ_{J-1}^2 is likely to be small, any reasonable small estimate will do, for example $\hat{\sigma}_{J-1}^2 = \frac{1}{2}\hat{\sigma}_{J-2}^2$. One possibility is to set

$$\hat{\sigma}_{J-1}^2 = \min\left\{\frac{\hat{\sigma}_{J-2}^4}{\hat{\sigma}_{J-3}^2}, \min\{\hat{\sigma}_{J-3}^2, \hat{\sigma}_{J-2}^2\}\right\}. \tag{1.16}$$

To expand on this, if $I-J$ is small, some subjective interference in the estimation of f_j and σ_j^2 when j is close to J may be a good idea.

As of the second part of (1.5), this is easily seen to equal

$$(\hat{C}_T - E[C_T|\mathscr{D}])^2 = (E[C_T|\mathscr{D}] - \hat{C}_T)^2 = \left(\sum_{i=I-J+1}^{I} C_{i,I-i}\left(\prod_{j=I-i}^{J-1} f_j - \prod_{j=I-i}^{J-1} \hat{f}_j\right)\right)^2.$$

Note that we cannot replace $\hat{f}_j$ with f_j here, since then the whole expression becomes zero. The first analytical approach to this problem was given by Mack, and since his formula is in widespread use, we will deduce it here. To this end let

$$R_i = \prod_{j=I-i}^{J-1} f_j - \prod_{j=I-i}^{J-1} \hat{f}_j = \sum_{j=I-i}^{J-1} S_{ij},$$

where

$$S_{ij} = \left(\prod_{k=I-i}^{j-1} \hat{f}_k\right)(f_j - \hat{f}_j)\left(\prod_{k=j+1}^{J-1} f_k\right),$$

where as usual $\prod_{k=m}^{m-1} a_k = 1$. This gives

$$(\hat{C}_T - E[C_T|\mathscr{D}])^2 = \left(\sum_{i=I-J+1}^{I} C_{i,I-i}R_i\right)^2 = \sum_{i=I-J+1}^{I} C_{i,I-i}^2 R_i^2 + 2\sum_{i=I-J+1}^{I-1}\sum_{l=i+1}^{I} C_{i,I-i}C_{l,I-l}R_iR_l.$$

So far so well, but we are still left with the unknown f_j, and setting these equal to $\hat{f}_j$ gives zero as before. Therefore, some kind of conditional expectations are necessary. Mack's suggestion is to condition in each expression with the highest $\mathscr{E}_j$ before it gets deterministic (or equivalently the highest $\mathscr{E}_j \wedge \mathscr{D}$). To explain, consider

$$R_i^2 = \sum_{j=I-i}^{J-1} S_{ij}^2 + 2 \sum_{j=I-i}^{J-2} \sum_{k=j+1}^{J-1} S_{ij} S_{ik}.$$

In S_{ij}^2, if we condition with $\mathscr{E}_{j+1}$ it is deterministic, but if we condition with $\mathscr{E}_j$ we get by (1.12),

$$E[S_{ij}^2 | \mathscr{E}_j] = \left(\prod_{k=I-i}^{j-1} \hat{f}_k^2 \right) \frac{\sigma_j^2}{C_{\bullet j}} \left(\prod_{k=j+1}^{J-1} f_k^2 \right).$$

Also, for $j < k$,

$$E[S_{ij} S_{ik} | \mathscr{E}_k] = S_{ij} E[S_{ik} | \mathscr{E}_k] = 0.$$

Therefore, replacing f_j by $\hat{f}_j$, Mack estimates $C_{i,I-i}^2 R_i^2$ by

$$C_{i,I-i}^2 \left(\prod_{k=I-i}^{J-1} \hat{f}_k^2 \right) \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2 C_{\bullet j}} = \left(\hat{C}_{iJ}^{\mathrm{CL}} \right)^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2 C_{\bullet j}}. \tag{1.17}$$

As above, for $i < l$,

$$\begin{aligned} R_i R_l &= \sum_{j=I-i}^{J-1} \sum_{k=I-l}^{J-1} S_{ij} S_{lk} \\ &= \sum_{j=I-i}^{J-1} \sum_{k=I-l}^{J-1} \left(\prod_{m=I-i}^{j-1} \hat{f}_m \right) (f_j - \hat{f}_j) \left(\prod_{m=j+1}^{J-1} f_m \right) \\ &\quad \times \left(\prod_{m=I-l}^{k-1} \hat{f}_k \right) (f_k - \hat{f}_k) \left(\prod_{m=k+1}^{J-1} f_m \right). \end{aligned}$$

After conditioning, terms of the form $(f_j - \hat{f}_j)(f_k - \hat{f}_k)$ only count when $j = k$. Therefore, it is sufficient to consider

$$\sum_{j=I-i}^{J-1} \left(\prod_{m=I-i}^{j-1} \hat{f}_m \right) \left(\prod_{m=j+1}^{J-1} f_m \right) \left(\prod_{m=I-l}^{j-1} \hat{f}_m \right) \left(\prod_{m=j+1}^{J-1} f_m \right) (f_j - \hat{f}_j)^2.$$

Taking expectations as before, and then replacing f_j by $\hat{f}_j$ gives the contribution from $C_{i,I-i} C_{l,I-l} R_i R_l$,

$$C_{i,I-i} C_{l,I-l} \sum_{j=I-i}^{J-1} \left(\prod_{m=I-i}^{J-1} \hat{f}_m \right) \left(\prod_{m=I-l}^{J-1} \hat{f}_m \right) \frac{\sigma_j^2}{\hat{f}_j^2 C_{\bullet j}} = \hat{C}_{iJ}^{\mathrm{CL}} \hat{C}_{lJ}^{\mathrm{CL}} \sum_{j=I-l}^{J-1} \frac{\sigma_j^2}{\hat{f}_j^2 C_{\bullet j}}. \tag{1.18}$$

Now we can combine (1.14) and (1.17) to obtain for $i > I - J$,

$$\widehat{\mathrm{MSE}}[C_{iJ} | \mathscr{D}] = \left(\hat{C}_{iJ}^{\mathrm{CL}} \right)^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \left(\frac{1}{\hat{C}_{ij}^{\mathrm{CL}}} + \frac{1}{C_{\bullet j}} \right). \tag{1.19}$$

Combining this with (1.15) and (1.18) gives

$$\widehat{\mathrm{MSE}}[C_O] = \sum_{i=I-J+1}^{I} \widehat{\mathrm{MSE}}[C_{iJ}|\mathscr{D}] + 2\sum_{i=I-J+1}^{I-1} \hat{C}_{iJ}^{\mathrm{CL}} \left(\sum_{l=i+1}^{I} \hat{C}_{lJ}^{\mathrm{CL}} \right) \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \frac{1}{C_{\bullet j}}. \tag{1.20}$$

The contribution from the bias term $(E[C_T|\mathscr{D}] - \hat{C}_T)^2$ to the MSE is

$$\sum_{i=I-J+1}^{I} \left(\hat{C}_{iJ}^{\mathrm{CL}}\right)^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \frac{1}{C_{\bullet j}} + 2\sum_{i=I-J+1}^{I-1} \hat{C}_{iJ}^{\mathrm{CL}} \left(\sum_{l=i+1}^{I} \hat{C}_{lJ}^{\mathrm{CL}} \right) \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \frac{1}{C_{\bullet j}}. \tag{1.21}$$

This is the controversial part of formulas (1.19) and (1.20) as we used different conditionings for each term of the formula. This has of course led to discussions and alternatives.

Remark 1.3. An alternative to the assumption CL3 is obtained by the multiplicative model

$$C_{i,j+1} = C_{ij} F_{ij}, \quad i = 0, \ldots, I, \ j = 0, \ldots, J-1, \tag{1.22}$$

where the F_{ij} are independent with $E[F_{ij}] = f_j$ and $\mathrm{Var}[F_{ij}] = \sigma_j^2$. This gives

$$E[C_{i,j+1}|\mathscr{E}_j] = C_{ij} f_j \quad \text{and} \quad \mathrm{Var}[C_{i,j+1}|\mathscr{E}_j] = \sigma_j^2 C_{ij}^2.$$

With $\tilde{f}_j = \sum_{i=0}^{I-j-1} a_i F_{ij}$ you are asked in Problem 1.5 to prove that the minimizer of $\mathrm{Var}[\tilde{f}_j|\mathscr{E}_j]$ subject to $E[\tilde{f}_j|\mathscr{E}_j] = f_j$ is given as

$$\hat{f}_j = \frac{1}{I-j} \sum_{i=0}^{I-j-1} F_{ij}. \quad j = 0, \ldots, J-1. \tag{1.23}$$

As this is just an average, an unbiased estimator for σ_j^2 is

$$\hat{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} (F_{ij} - \hat{f}_j)^2,$$

which can be compared to (1.13).

1.2.2 Some parametric approaches to chain-ladder estimation

A natural question is how realistic are the assumptions CL2 and CL3, will there be any natural models that satisfy these assumptions? We will give a few examples.

Example 1.1. Assume that given $\mathscr{E}_j$,

$$C_{i,j+1} \sim \mathscr{N}(C_{ij} f_j, C_{ij} \sigma_j^2), \quad i = 0, \ldots, I, \ j = 0, \ldots, J-1. \tag{1.24}$$

This model clearly satisfies CL2 and CL3, and in addition we assume CL1. There is a consistency issue since it is possible that C_{ij} is negative, in which case $C_{i,j+1}$ cannot be defined. We will assume that the probability for this to happen is so small that it is negligible.

We will estimate the parameters using maximimum likelihood. To this end, set $\mathbf{C}_i = (C_{i1}, \ldots, C_{i,j^*(i)})$. As this model has the Markov property, we get

$$
\begin{aligned}
f_{\mathbf{C}_i|C_{i0}}(\mathbf{c}_i|c_{i0}) &= \prod_{j=0}^{j^*(i)-1} f_{C_{i,j+1}|C_{ij}}(c_{i,j+1}|c_{ij}) \\
&= \prod_{j=0}^{j^*(i)-1} (2\pi)^{-\frac{1}{2}}(c_{ij}\sigma_j^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2c_{ij}\sigma_j^2}(c_{i,j+1}-c_{ij}f_j)^2\right\}.
\end{aligned}
$$

Taking the logarithm and summing over i, using (1.3) gives with $\mathbf{f} = (f_0, \ldots, f_{J-1})$, $\boldsymbol{\sigma}^2 = (\sigma_0^2, \ldots, \sigma_{J-1}^2)$ and $\mathbf{c} = (\mathbf{c}_0, \ldots, \mathbf{c}_{I-1})$,

$$
\begin{aligned}
l(\mathbf{f}, \boldsymbol{\sigma}^2; \mathbf{c}) = &-\frac{n_{J-1}}{2}\log 2\pi - \frac{1}{2}\sum_{j=0}^{J-1}\sum_{i=0}^{I-j-1}\log c_{ij} - \frac{1}{2}\sum_{j=0}^{J-1}(I-j)\log\sigma_j^2 \\
&-\frac{1}{2}\sum_{j=0}^{J-1}\frac{1}{\sigma_j^2}\sum_{i=0}^{I-j-1}\frac{(c_{i,j+1}-c_{ij}f_j)^2}{c_{ij}},
\end{aligned}
\tag{1.25}
$$

where

$$
n_k = \sum_{j=0}^{k}(I-j) = \left(I - \frac{1}{2}k\right)(k+1).
$$

Taking the partial derivative w.r.t. f_j and setting this equal to zero shows that the MLE of f_j is the chain-ladder estimator $\hat{f}_j$ given by (1.11). To find the MLE of σ_j^2, note first that

$$
\sum_{i=0}^{I-j-1}\frac{(C_{i,j+1}-C_{ij}\hat{f}_j)^2}{C_{ij}} = \sum_{i=0}^{I-j-1} C_{ij}\left(\frac{C_{i,j+1}}{C_{ij}} - \hat{f}_j\right)^2 = (I-j-1)\hat{\sigma}_j^2,
$$

where $\hat{\sigma}_j^2$ is given in (1.13). Therefore

$$
l(\hat{\mathbf{f}}, \boldsymbol{\sigma}^2; \mathbf{c}) = -\frac{n_{J-1}}{2}\log 2\pi - \frac{1}{2}\sum_{j=0}^{J-1}\sum_{i=0}^{I-j-1}\log c_{ij} - \frac{1}{2}\sum_{j=0}^{J-1}(I-j)\log\sigma_j^2 - \frac{1}{2}\sum_{j=0}^{J-1}\frac{(I-j-1)\hat{\sigma}_j^2}{\sigma_j^2}.
$$

Taking the derivative w.r.t. σ_j^2 and setting this equal to zero gives the MLE

$$
\tilde{\sigma}_j^2 = \frac{I-j-1}{I-j}\hat{\sigma}_j^2,
$$

the usual relation between the MLE and the unbiased estimator in the normal model. Let us now assume that σ_j^2 is of the form (1.7), i.e. $\sigma_j^2 = (f_j - 1)\tau^2$. A full MLE with $\mathbf{f}$ and τ^2 must be done numerically, but an alternative is to insert the $\hat{\mathbf{f}}$ already found into the log likelihood. This gives, merging all terms that do not contain τ^2 into the constant a,

$$
l(\hat{\mathbf{f}}, \tau^2; \mathbf{c}) = a - \frac{1}{2}n_{J-1}\log\tau^2 - \frac{1}{2\tau^2}\sum_{j=0}^{J-1}\frac{(I-j-1)\hat{\sigma}_j^2}{\hat{f}_j - 1}.
$$

Taking the derivative w.r.t. τ^2 gives the estimator

$$
\hat{\tau}^2 = \frac{1}{n_{J-1}}\sum_{j=0}^{J-1}\frac{I-j-1}{\hat{f}_j - 1}\hat{\sigma}_j^2.
$$

As $\hat{f}_j$ can be quite unreliable, and close to 1, for j close to J, an alternative is to use for $k < J$,

$$\hat{\tau}_k^2 = \frac{1}{n_k} \sum_{j=0}^{k} \frac{I-j-1}{\hat{f}_j - 1} \hat{\sigma}_j^2. \tag{1.26}$$

An alternative to (1.10) that contains less parameters is

$$C_{i,j+1} = C_{ij} f_j + \sigma \sqrt{C_{ij}} \frac{1}{p_{ij}} e_{i,j+1}, \tag{1.27}$$

where the p_{ij} are known volume factors, and the $e_{i,j+1}$ are i.i.d. $\mathcal{N}(0,1)$. This model still suffers from inconsistency since C_{ij} may become negative, but again we ignore this issue. The log likelihood is now, lumping all terms that do not contain $\mathbf{f}$ or σ into the constant a,

$$l(\mathbf{f}, \sigma; \mathbf{c}) = a - \frac{1}{2} n_{J-1} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{j=0}^{J-1} \sum_{i=0}^{I-j-1} p_{ij} \frac{(c_{i,j+1} - c_{ij} f_j)^2}{c_{ij}}.$$

Taking the derivative w.r.t. f_j gives

$$\hat{f}_j = \frac{\sum_{i=0}^{I-j-1} p_{ij} C_{i,j+1}}{\sum_{i=0}^{I-j-1} p_{ij} C_{ij}}.$$

If $p_{ij} = p_j$, independent of i, this is the usual chain-ladder estimator. Plugging this estimator into the log likelihood gives the MLE of σ^2,

$$\tilde{\sigma}^2 = \frac{1}{n_{J-1}} \sum_{j=0}^{J-1} \sum_{i=0}^{I-j-1} p_{ij} \frac{(C_{i,j+1} - C_{ij} \hat{f}_j)^2}{C_{ij}}.$$

It can be natural to instead use the adjusted

$$\hat{\sigma}^2 = \frac{n_{J-1}}{n_{J-1} - J} \tilde{\sigma}^2.$$

Example 1.2. Here is a somewhat more advanced example. Assume that the increments can be written as

$$X_{ij} = \sum_{k=1}^{N_{ij}} Y_{ijk},$$

where the Y_{ijk} are all independent with distribution function $F_j(y; \boldsymbol{\theta}_j)$. Thus for fixed j, the Y_{ijk} are i.i.d.. Furthermore, conditionally on $\mathcal{E}_{j-1}$, we assume that $N_{ij} \sim \mathrm{Po}(\lambda_j C_{i,j-1})$. As in the last example $C_{i,-1}$ must be set a priori. It can for example be the premium level for accident year i. Note that these assumptions imply CL1.

According to this model, the year to year dependency in the runoff is due to the claim frequency only.

Let $\mu_j = E[Y_{ijk}]$ and $\kappa_j = E[Y_{ijk}^2]$. Then

$$E[X_{i,j+1} | \mathcal{E}_j] = \mu_{j+1} \lambda_{j+1} C_{ij} \quad \text{and} \quad \mathrm{Var}[X_{i,j+1} | \mathcal{E}_j] = \kappa_{j+1} \lambda_{j+1} C_{ij}.$$

Therefore, by Remark 1.1,

$$f_j = 1 + \mu_{j+1}\lambda_{j+1} \quad \text{and} \quad \sigma_j^2 = \kappa_{j+1}\lambda_{j+1}.$$

Combining these two expressions gives

$$\sigma_j^2 = (f_j - 1)\frac{\kappa_{j+1}}{\mu_{j+1}},$$

and so if $\frac{\kappa_{j+1}}{\mu_{j+1}} \equiv \tau^2$, for example if the Y_{ijk} are all i.i.d., we have the variance form (1.7).

We can of course use the nonparametric estimators (1.11) and (1.13), but an alternative is to use maximum likelihood. To show how this is done let $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_J)$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_J)$. Also let $\mathbf{Y}_{N_{ij}} = (Y_{ij1}, \ldots, Y_{ijN_{ij}})$. Then the total information vector for accident year i is

$$\mathbf{W}_i = (N_{i0}, \mathbf{Y}_{N_{i0}}, N_{i1}, \mathbf{Y}_{N_{i1}}, \ldots, N_{i,j^*(i)}, \mathbf{Y}_{N_{i,j^*(i)}}).$$

Also, when we write $\mathbf{W}_i(\mathbf{Z}-)$ we shall mean all data in $\mathbf{W}_i$ up to but not including $\mathbf{Z}$. For example

$$\mathbf{W}_i(\mathbf{Y}_{N_{i1}}-) = (N_{i0}, \mathbf{Y}_{N_{i0}}, N_{i1}).$$

Assume that the Y_{ijk} have density f_j (w.r.t. some measure μ). Then

$$\begin{aligned} f_{\mathbf{W}_i}(\mathbf{w}_i) &= f_{\mathbf{Y}_{N_{i,j^*(i)}}|\mathbf{W}_i(\mathbf{Y}_{N_{i,j^*(i)}}-)}(\mathbf{y}_{n_{i,j^*(i)}}|\mathbf{w}_i(\mathbf{y}_{n_{i,j^*(i)}}-)) \\ &\quad \times P(N_{i,j^*(i)} = n_{i,j^*(i)}|\mathbf{W}_i(N_{i,j^*(i)}-) = \mathbf{w}_i(n_{i,j^*(i)}-)) \\ &\quad \times \cdots \times f_{\mathbf{Y}_{N_{i0}}|N_{i0}}(\mathbf{y}_{n_{i0}}|n_{i0})P(N_{i0} = n_{i0}) \\ &= \left(\prod_{j=0}^{j^*(i)} \frac{(c_{i,j-1}\lambda_j)^{n_{ij}}}{n_{ij}!} e^{-c_{i,j-1}\lambda_j}\right)\left(\prod_{j=0}^{j^*(i)}\prod_{k=1}^{n_{ij}} f_j(y_{ijk};\boldsymbol{\theta}_j)\right). \end{aligned}$$

Now take the product over all accident years and then the logarithm. Using (1.2) gives the log likelihood

$$l(\boldsymbol{\theta}, \boldsymbol{\lambda}; \mathbf{w}) = \sum_{j=0}^{J} l_j^{\theta}(\boldsymbol{\theta}_j; \mathbf{w}) + \sum_{j=0}^{J} l_j^{\lambda}(\lambda_j; \mathbf{w}). \tag{1.28}$$

Here

$$l_j^{\theta}(\boldsymbol{\theta}_j; \mathbf{w}) = \sum_{i=0}^{I-j}\sum_{k=1}^{n_{ij}} \log f_j(y_{ijk}; \boldsymbol{\theta}_j) \tag{1.29}$$

and

$$l_j^{\lambda}(\lambda_j; \mathbf{w}) = \sum_{i=0}^{I-j}(n_{ij}\log c_{i,j-1} + n_{ij}\log\lambda_j - c_{i,j-1}\lambda_j - \log n_{ij}!). \tag{1.30}$$

From (1.28) it is clear that all the MLE's of λ_j and $\boldsymbol{\theta}_j$ can be found by separate maximization. The $\boldsymbol{\theta}_j$ are found by maximizing (1.29), while the λ_j by maximizing (1.30). However, the latter can easily be solved to give

$$\hat{\lambda}_j = \frac{\sum_{i=0}^{I-j} N_{ij}}{\sum_{i=0}^{I-j} C_{i,j-1}}.$$

Assume that the $Y_{ijk} \sim \mathcal{N}(\mu_j, \tau_j^2)$. Then

$$\hat{\mu}_j = \frac{1}{\sum_{i=0}^{I-j} N_{ij}} \sum_{i=0}^{I-j} \sum_{k=1}^{N_{ij}} Y_{ijk} = \frac{\sum_{i=0}^{I-j} X_{ij}}{\sum_{i=0}^{I-j} N_{ij}}.$$

Therefore,

$$1 + \hat{\mu}_j \hat{\lambda}_j = 1 + \frac{\sum_{i=0}^{I-j} X_{ij}}{\sum_{i=0}^{I-j} C_{i,j-1}} = \frac{\sum_{i=0}^{I-j} C_{ij}}{\sum_{i=0}^{I-j} C_{i,j-1}} = \hat{f}_{j-1}^{\mathrm{CL}},$$

i.e. the ordinary chain-ladder estimator (1.11). However, the maximum likelihood estimator of the variance σ_j^2 will be different from that of (1.13). In fact

$$\hat{\kappa}_j = \frac{1}{\sum_{i=0}^{I-j} N_{ij}} \sum_{i=0}^{I-j} \sum_{k=1}^{N_{ij}} Y_{ijk}^2,$$

so

$$\hat{\sigma}_{j-1}^2 = \hat{\kappa}_j \hat{\lambda}_j = \frac{\sum_{i=0}^{I-j} \sum_{k=1}^{N_{ij}} Y_{ijk}^2}{\sum_{i=0}^{I-j} C_{i,j-1}}.$$

A drawback with this model is that in order to calculate $\hat{\sigma}_j^2$ it is necessary to know the individual claims.

1.3 The Bornhuetter-Ferguson method

This is another classical method, and we will cover it in some depth. Again the presentation is split into a nonparametric approach and a parametric approach.

1.3.1 Introduction and the nonparametric Bornhuetter-Ferguson model

The following two assumptions are made throughout this section.

BF1: Different accident years are independent, i.e. C_{ij} and C_{kl} are independent whenever $i \neq k$.
BF2: There are positive factors $\mu_0, \dots, \mu_I$ and $\beta_0, \dots, \beta_J$ with $\beta_J = 1$ so that for all i, j,

$$\begin{aligned} E[C_{i0}] &= \mu_i \beta_0, \\ E[C_{ik} | \mathscr{E}_j] &= C_{ij} + \mu_i(\beta_k - \beta_j), \quad k = j+1, \dots, J. \end{aligned}$$

With $j = 0$ and k in the second equation in BF2 replaced by j, we get

$$E[C_{ij}] = \mu_i \beta_0 + \mu_i(\beta_j - \beta_0) = \mu_i \beta_j, \tag{1.31}$$

and in particuar $E[C_{iJ}] = \mu_i \beta_J = \mu_i$. Thus μ_i is the expected total claims for accident year i. The update is

$$C_{iJ}^{\mathrm{BF}} = E[C_{iJ} | \mathscr{D}] = C_{i,I-i} + \mu_i(1 - \beta_{I-i}). \tag{1.32}$$

Let

$$\gamma_j = \beta_j - \beta_{j-1}, \quad j = 0, \dots, J,$$

with $\beta_{-1} = 0$. Then $\sum_{j=0}^{J} \gamma_j = 1$, and an equivalent definition of BF2 is

BF2': There are positive factors $\mu_0, \ldots, \mu_I$ and factors $\gamma_0, \ldots, \gamma_J$ with $\sum_{j=0}^{J} \gamma_j = 1$ so that for all i, j,

$$\begin{aligned} E[X_{i0}] &= \mu_i \gamma_0, \\ E[X_{ik}|\mathscr{E}_j] &= \mu_i \gamma_k, \quad k = j+1, \ldots, J. \end{aligned}$$

To make a comparison with the chain-ladder model, in the latter set

$$E[C_{iJ}] = E[C_{ij}] \prod_{k=j}^{J-1} f_k \;\Rightarrow\; E[C_{ij}] = E[C_{iJ}] \prod_{k=j}^{J-1} f_k^{-1}.$$

If we let

$$\beta_j = \prod_{k=j}^{J-1} f_k^{-1}, \quad j = 0, \ldots, J-1, \tag{1.33}$$

and set $\mu_i = E[C_{iJ}]$, the two methods give the same values for the unconditional $E[C_{ij}]$. However, for the conditional values, with β_j and μ_i as just given, we get by (1.6),

$$\begin{aligned} C_{iJ}^{\mathrm{CL}} &= C_{i,I-i} \prod_{j=I-i}^{J-1} f_j \\ &= C_{i,I-i} + C_{i,I-i} \left(\prod_{j=I-i}^{J-1} f_j \right) \left(1 - \prod_{j=I-i}^{J-1} f_j^{-1} \right) \\ &= C_{i,I-i} + C_{iJ}^{\mathrm{CL}} (1 - \beta_{I-i}). \end{aligned} \tag{1.34}$$

The differences between (1.34) and (1.32) is that (1.32) sticks to the original μ_i, while (1.34) continuously updates the right hand side with C_{iJ}^{CL}.

The relation (1.33) can be used to estimate the β_j, by letting

$$\hat{\beta}_j = \prod_{k=j}^{J-1} \hat{f}_k^{-1}, \quad j = 0, \ldots, J-1,$$

where the $\hat{f}_j$ are the chain-ladder estimates. As of the μ_i, in a pure Bornhuetter-Ferguson approach they are determined a priori, i.e. they are set according to a "best guess" principle. One approach is to set $\mu_i = p_i q_i$ where p_i is the known total premium for year i, while q_i is an estimate of the unknown claims ratio.

1.3.2 Parametric Bornhuetter-Ferguson models

There are many alternative approaches to estimation of both the μ_i and the β_j using the data. Here are a few examples.

Example 1.3 (Normally distributed increments). Assume that the increments X_{ij} are independent with distribution

$$X_{ij} \sim \mathscr{N}\left(\mu_i \gamma_j, \frac{\sigma^2}{p_{ij}} \right), \quad i = 0, \ldots, I, \;\; j = 0, \ldots, J,$$

where $\sum_{j=0}^{J}\gamma_j = 1$. The p_{ij} are known weights and it is reasonable to let $p_{ij} = p_j$, i.e. only depending on the runoff year. This model can also be written as

$$X_{ij} = \mu_i\gamma_j + \frac{\sigma}{\sqrt{p_{ij}}}e_{ij}, \tag{1.35}$$

where the e_{ij} are i.i.d. $\mathscr{N}(0,1)$.

Set $\boldsymbol{\mu} = (\mu_0,\ldots,\mu_I)$ and $\boldsymbol{\gamma} = (\gamma_0,\ldots,\gamma_J)$. We will here deviate from the pure Bornhuetter-Ferguson principle and treat all parameters as unknown. The MLE are then given as the minimizer of

$$Q(\boldsymbol{\mu},\boldsymbol{\gamma}) \stackrel{\text{def}}{=} \sum_{i=0}^{I}\sum_{j=0}^{j^*(i)} p_{ij}(X_{ij}-\mu_i\gamma_j)^2 = \sum_{j=0}^{J}\sum_{i=0}^{I-j} p_{ij}(X_{ij}-\mu_i\gamma_j)^2,$$

subject to the constraint $\sum_{j=0}^{J}\gamma_j = 1$. Since for any positive a, $\mu_i\gamma_j = (a\mu_i)(a^{-1}\gamma_j)$, without the constraint $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ are not uniquely determined. A way to get around this and include the constraint is to let $\boldsymbol{\mu}^0 = (\mu_0^0,\ldots,\mu_I^0)$ and $\boldsymbol{\gamma}^0 = (\gamma_0^0,\ldots,\gamma_J^0)$ with $\gamma_0^0 = 1$. Then if we can minimize $Q(\boldsymbol{\mu}^0,\boldsymbol{\gamma}^0)$ to yield $\hat{\boldsymbol{\mu}}^0$ and $\hat{\boldsymbol{\gamma}}^0$ with $\hat{\gamma}_0^0 = 1$, the $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\gamma}}$ that minimize $Q(\boldsymbol{\mu},\boldsymbol{\gamma})$ subject to the constraint are given by

$$\hat{\mu}_i = \hat{\mu}_i^0\left(\sum_{j=0}^{J}\hat{\gamma}_j^0\right), \quad i=0,\ldots,I \quad \text{and} \quad \hat{\gamma}_j = \frac{\hat{\gamma}_j^0}{\sum_{j=0}^{J}\hat{\gamma}_j^0}, \quad j=0,\ldots,J. \tag{1.36}$$

Taking the derivatives of $Q(\boldsymbol{\mu}^0,\boldsymbol{\gamma}^0)$ gives

$$\begin{aligned}
-\frac{1}{2}\frac{\partial}{\partial\mu_i^0}Q(\boldsymbol{\mu}^0,\boldsymbol{\gamma}^0) &= \sum_{j=0}^{j^*(i)} p_{ij}\gamma_j^0(X_{ij}-\mu_i^0\gamma_j^0), \\
-\frac{1}{2}\frac{\partial}{\partial\gamma_j^0}Q(\boldsymbol{\mu}^0,\boldsymbol{\gamma}^0) &= \sum_{i=0}^{I-j} p_{ij}\mu_i^0(X_{ij}-\mu_i^0\gamma_j^0).
\end{aligned}$$

Setting these equal to zero yields the updating equations

$$\mu_i^0 = \frac{\sum_{j=0}^{j^*(i)} p_{ij}\gamma_j^0 X_{ij}}{\sum_{j=0}^{j^*(i)} p_{ij}(\gamma_j^0)^2}, \quad i=0,\ldots,I \tag{1.37}$$

and

$$\gamma_j^0 = \frac{\sum_{i=0}^{I-j} p_{ij}\mu_i^0 X_{ij}}{\sum_{i=0}^{I-j} p_{ij}(\mu_i^0)^2}, \quad j=1,\ldots,J. \tag{1.38}$$

To solve this system, start with a $\boldsymbol{\gamma}^0$ so that $\gamma_0^0 = 1$, use (1.37) to update $\boldsymbol{\mu}^0$, then use (1.38) to update $\boldsymbol{\gamma}^0$ and continue until convergence. This will usually work well.

If a pure Bornhuetter-Ferguson approach is preferred with known μ_i, the γ_j are easily estimated using (1.38) (no iterations are necessary).

By assumption,

$$\mathrm{Var}[C_T|\mathscr{D}] = \sigma^2\sum_{i=I-J+1}^{I}\sum_{j=J-i+1}^{J}\frac{1}{p_{ij}}.$$

Since for the true parameters $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$,

$$E\left[\sum_{i=0}^{I}\sum_{j=0}^{j^*(i)} p_{ij}(X_{ij}-\mu_i\gamma_j)^2\right] = n\sigma^2,$$

where $n = \sum_{i=0}^{I}(1+j^*(i))$ is the total number of data, and there are $I+J+1$ estimated parameters (the constraint reduces the number of estimated parameters by one), an estimator for σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n-I-J-1}\sum_{i=0}^{I}\sum_{j=0}^{j^*(i)} p_{ij}(X_{ij}-\hat{\mu}_i\hat{\gamma}_j)^2. \tag{1.39}$$

Example 1.4 (A gamma model). Assume the increments X_{ij} are independent and gamma distributed with

$$X_{ij} \sim G_G(\mu_i\gamma_j,\phi), \quad i=0,\ldots,I,\ j=0,\ldots,J.$$

This means that

$$E[X_{ij}] = \mu_i\gamma_j \quad \text{and} \quad \mathrm{Var}[X_{ij}] = \phi(\mu_i\gamma_j)^2.$$

Thus the standard deviation is proportional to the expectation, and so we can let ϕ be a constant, not depending on j. If we write $\alpha_i = \log\mu_i$ and $\beta_j = \log\gamma_j$, then

$$\log(E[X_{ij}]) = \alpha_i + \beta_j,$$

so this can be considered as a gamma distributed GLM with logarithmic link function. With $\mathbf{X} = (X_{00},\ldots,X_{0,J},\ldots,X_{I,0})$, the density is

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=0}^{I}\prod_{j=0}^{j^*(i)} \frac{1}{(\phi\mu_i\gamma_j)^{\phi^{-1}}\Gamma(\phi^{-1})} x_{ij}^{\phi^{-1}-1} e^{-\frac{x_{ij}}{\phi\mu_i\gamma_j}}.$$

Taking the logarithm gives the log likelihood

$$\begin{aligned} l(\boldsymbol{\mu},\boldsymbol{\gamma},\phi;\mathbf{x}) = {} & n\phi^{-1}\log\phi^{-1} - n\log\Gamma(\phi^{-1}) - \phi^{-1}\sum_{i=0}^{I}(j^*(i)+1)\log\mu_i + \phi^{-1}\sum_{j=0}^{J}(I-j+1)\log\gamma_j \\ & +(\phi^{-1}-1)\sum_{i=0}^{I}\sum_{j=0}^{j^*(i)}\log x_{ij} - \phi^{-1}\sum_{i=0}^{I}\frac{1}{\mu_i}\sum_{j=0}^{j^*(i)}\frac{x_{ij}}{\gamma_j}, \end{aligned}$$

where n is the total number of observations, see (1.4). Taking the partial derivatives gives

$$\begin{aligned} \frac{\partial}{\partial\mu_i} l(\boldsymbol{\mu},\boldsymbol{\gamma},\phi;\mathbf{x}) &= -\phi^{-1}\frac{j^*(i)+1}{\mu_i} + \phi^{-1}\frac{1}{\mu_i^2}\sum_{j=0}^{j^*(i)}\frac{x_{ij}}{\gamma_j}, \\ \frac{\partial}{\partial\gamma_j} l(\boldsymbol{\mu},\boldsymbol{\gamma},\phi;\mathbf{x}) &= -\phi^{-1}\frac{I-j+1}{\gamma_j} + \phi^{-1}\frac{1}{\gamma_j^2}\sum_{i=0}^{I-j}\frac{x_{ij}}{\mu_i}. \end{aligned}$$

As in the last example this system of overparametrized, but as there we can set $\gamma_0 = 1$. Then setting the partial derivatives above equal to zero gives the equations

$$
\begin{aligned}
\mu_i &= \frac{1}{j^*(i)+1} \sum_{j=0}^{j^*(i)} \frac{X_{ij}}{\gamma_j}, \quad i = 0, \dots, I, \\
\gamma_j &= \frac{1}{I-j+1} \sum_{i=0}^{I-j} \frac{X_{ij}}{\mu_i}, \quad j = 1, \dots, J.
\end{aligned} \tag{1.40}
$$

Iterating these equations should give the MLE. Like in (1.36) these estimates can be scaled so that the $\hat{\gamma}_j$ sum to one. Inserting the $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\gamma}}$ into the log likelihood, this can be maximized w.r.t. ϕ (or rather w.r.t. ϕ^{-1}). Alternatively, as is typical in GLM

$$
\hat{\phi} = \frac{1}{n-I-J-1} \sum_{i=0}^{I} \sum_{j=0}^{j^*(i)} \frac{(X_{ij} - \hat{\mu}_i \hat{\gamma}_j)^2}{(\hat{\mu}_i \hat{\gamma}_j)^2} = \frac{1}{n-I-J-1} \sum_{i=0}^{I} \sum_{j=0}^{j^*(i)} \left(\frac{X_{ij}}{\hat{\mu}_i \hat{\gamma}_j} - 1 \right)^2 .
$$

Here is another model that is in the spirit of Bornhuetter-Ferguson.

Example 1.5 (A lognormal model). Assume the increments X_{ij} are independent and lognormally distributed with

$$
Y_{ij} \stackrel{\text{def}}{=} \log X_{ij} \sim \mathcal{N}(\alpha_i + \beta_j, \sigma^2), \quad i = 0, \dots, I,\ j = 0, \dots, J.
$$

Thus the Y_{ij} is a two-way analysis of variance model without interaction. The expectation is

$$
E[X_{ij}] = e^{\alpha_i + \beta_j + \frac{1}{2}\sigma^2} \stackrel{\text{def}}{=} \mu_i \gamma_j,
$$

where

$$
\mu_i = e^{\alpha_i + \frac{1}{2}\sigma^2} \sum_{k=0}^{J} e^{\beta_k} \quad \text{and} \quad \gamma_j = \frac{e^{\beta_j}}{\sum_{k=0}^{J} e^{\beta_k}}.
$$

As usual with the lognormal distribution

$$
\operatorname{Var}[X_{ij}] = (e^{\sigma^2} - 1) E[X_{ij}]^2.
$$

It is clear that the system is overspecified, so to make it well specified we set $\beta_0 = 0$. With $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_I)$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_J)$ and $\tau = \sigma^2$, the loglikelihood becomes

$$
l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau; \mathbf{y}) = c(\mathbf{y}) - \frac{n}{2} \log \tau - \frac{1}{\tau} Q(\boldsymbol{\alpha}, \boldsymbol{\beta}),
$$

where $c(\mathbf{y}) = -\frac{n}{2} \log 2\pi - \sum_{i=0}^{I} \sum_{j=0}^{j^*(i)} y_{ij}$ and

$$
Q(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=0}^{I} \sum_{j=0}^{j^*(i)} (y_{ij} - \alpha_i - \beta_j)^2 = \sum_{j=0}^{J} \sum_{i=0}^{I-j} (y_{ij} - \alpha_i - \beta_j)^2.
$$

Taking the partial derivatives gives

$$
\begin{aligned}
\frac{1}{2} \frac{\partial}{\partial \alpha_i} Q(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= -\sum_{j=0}^{j^*(i)} (y_{ij} - \alpha_i - \beta_j) = (j^*(i)+1)(\bar{y}_{i\bullet} - \alpha_i) + \sum_{j=0}^{j^*(i)} \beta_j, \quad i = 0, \dots, I, \\
\frac{1}{2} \frac{\partial}{\partial \beta_j} Q(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= -\sum_{i=0}^{I-j} (y_{ij} - \alpha_i - \beta_j) = (I-j+1)(\bar{y}_{\bullet j} - \beta_j) + \sum_{i=0}^{I-j} \alpha_i, \quad j = 1, \dots, J.
\end{aligned}
$$

Here

$$\bar{y}_{i\bullet} = \frac{1}{j^*(i)+1}\sum_{j=0}^{j^*(i)} y_{ij} \quad \text{and} \quad \bar{y}_{\bullet j} = \frac{1}{I-j+1}\sum_{i=0}^{I-j} y_{ij}.$$

Setting these equal to zero yields the iterative equations

$$\alpha_i = \bar{y}_{i\bullet} - \frac{1}{j^*(i)+1}\sum_{j=0}^{j^*(i)} \beta_j, \quad i=0,\dots,I,$$

$$\beta_j = \bar{y}_{\bullet j} - \frac{1}{I-j+1}\sum_{i=0}^{I-j} \alpha_i, \quad j=1,\dots,J.$$

An alternative to this iterative solution is to directly minimize $Q(\boldsymbol{\alpha},\boldsymbol{\beta})$ using a computer program. Once the MLE $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are found, the MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=0}^{I}\sum_{j=0}^{j^*(i)}(y_{ij}-\hat{\alpha}_i-\hat{\beta}_j)^2,$$

where $\hat{\beta}_0 = 0$. As usual it may be a good idea to correct for the number of estimated parameters, giving

$$\tilde{\sigma}^2 = \frac{1}{n-I-J-1}\sum_{i=0}^{I}\sum_{j=0}^{j^*(i)}(y_{ij}-\hat{\alpha}_i-\hat{\beta}_j)^2.$$

1.4 Benktander-Hovinen methods

If there is uncertainty whether a chain-ladder or a Bornhuetter-Ferguson model should be used, one possibility is to go for a linear combination of the two. This is called the Benktander-Hovinen method, i.e.

$$\hat{C}_{iJ}^{\mathrm{BH}} = c_i\hat{C}_{iJ}^{\mathrm{CL}} + (1-c_i)\hat{C}_{iJ}^{\mathrm{BF}}, \quad 0 \le c \le 1.$$

The choice $c_i = 1$ gives the chain-ladder method, while $c_i = 0$ gives the Bornhuetter-Ferguson method. As the level of information at time I about C_{iJ} decreases with increasing i, and chain-ladder updates the estimate continuously, it is reasonable to let c_i decrease with i. Benktander proposed to set $c_i = \beta_{j^*(i)}$, but as this must be estimated we get

$$\hat{C}_{iJ}^{\mathrm{BH}} = \hat{\beta}_{j^*(i)}\hat{C}_{iJ}^{\mathrm{CL}} + (1-\hat{\beta}_{j^*(i)})\hat{C}_{iJ}^{\mathrm{BF}}. \tag{1.41}$$

When i is close to I, $\hat{\beta}_{j^*(i)}$ is small, so that $\hat{C}_{iJ}^{\mathrm{BH}}$ is not very different from $\hat{C}_{iJ}^{\mathrm{BF}}$. But the biggest contributions to total outstanding claims are when i is close to I, and therefore this method gives total outstanding claims close to those obtained by the Bornhuetter-Ferguson method.

If the $\hat{\beta}_j$ are estimated from the chain-ladder method, i.e.

$$\hat{\beta}_j = \prod_{k=j}^{J-1}\hat{f}_k^{-1},$$

this becomes

$$\hat{C}_{iJ}^{\mathrm{BH}} = C_{i,j^*(i)} + (1-\hat{\beta}_{j^*(i)})\hat{C}_{iJ}^{\mathrm{BF}}.$$

1.5 Solvency margins

Until now most attention has been focussed on calculating expected outstanding claims. In a risk management context, we are also interested in other quantities, such as value at risk or expected shortfall, where estimation uncertainty can be included for the parametric models. We have seen various methods to calculate the mean squared errors, so an approximation to for example an α quantile is

$$\hat{C}_O + z_{1-\alpha}\sqrt{\widehat{\mathrm{MSE}}[C_O]}, \tag{1.42}$$

where $P(Z > z_{1-\alpha}) = 1-\alpha$ for Z standard normal.

We will now have a look at bootstrap methods to calculate useful quantities.

1.5.1 Parametric bootstrap

Instead of trying to give general descriptions, it is easier to illustrate by using examples. Here we will give one example for the chain-ladder method and one for the Bornhuetter-Ferguson method.

Example 1.6. As an example let us use the model of Example 1.1, i.e.

$$C_{i,j+1} = C_{ij} f_j + \sigma_j \sqrt{C_{ij}}\, e_{i,j+1}, \tag{1.43}$$

where the e_{ij} are i.i.d. $\mathcal{N}(0,1)$. Due to the multiplicative structure, the C_{iJ} are not Gaussian. Here is how we can do the bootstrap:

1. Find estimates for $\mathbf{f} = (f_0, \ldots, f_{J-1})$ and $\boldsymbol{\sigma}^2 = (\sigma_0^2, \ldots, \sigma_{J-1}^2)$. The natural estimators for $\mathbf{f}$ is the chain-ladder estimators, and for $\boldsymbol{\sigma}^2$ we can use (1.13) and (1.16).
2. Generate $e_{ij},\ j = j^*(i)+1, \ldots, J,\ i = I-J+1, \ldots, I$. Starting with $C_{i,I-i}$ we can use (1.43) iteratively to get the simulated C_{iJ}. If one genereated C_{ij} becomes negative, we can start over again for that particular i.
3. Set
$$C_{O,1} = \sum_{i=I-J+1}^{I} (C_{iJ} - C_{i,I-i}).$$
Then continue to generate $C_{O,2}, \ldots, C_{O,B}$ in the same way, where B can be quite large.
4. Use the generated $C_{O,k}$ to calculate the relevant quantities.

This procedure does not take into account the fact that the parameters are estimated. To do so, we use that

$$\begin{bmatrix} \hat{\mathbf{f}} \\ \hat{\boldsymbol{\sigma}}^2 \end{bmatrix} \approx \mathcal{N}_{2J}\left(\begin{bmatrix} \mathbf{f} \\ \boldsymbol{\sigma}^2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{f}} & \boldsymbol{\Sigma}_{\mathbf{f}\boldsymbol{\sigma}^2} \\ \boldsymbol{\Sigma}'_{\mathbf{f}\boldsymbol{\sigma}^2} & \boldsymbol{\Sigma}_{\boldsymbol{\sigma}^2} \end{bmatrix} \right).$$

Taking the Bayesian point of view and using the Hessian from the log likelihood to estimate the covariance matrix, this can be written as

$$\begin{bmatrix} \mathbf{f} \\ \boldsymbol{\sigma}^2 \end{bmatrix} \approx \mathcal{N}_{2J}\left(\begin{bmatrix} \hat{\mathbf{f}} \\ \hat{\boldsymbol{\sigma}}^2 \end{bmatrix}, \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{\mathbf{f}} & \hat{\boldsymbol{\Sigma}}_{\mathbf{f}\boldsymbol{\sigma}^2} \\ \hat{\boldsymbol{\Sigma}}'_{\mathbf{f}\boldsymbol{\sigma}^2} & \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\sigma}^2} \end{bmatrix} \right). \tag{1.44}$$

We can now add the following step after Step 1 above:

1a. For each simulation of $C_{O,k}$, generate $(\mathbf{f}, \boldsymbol{\sigma}^2)$ according to (1.44).

This will probably result in more negative C_{ij}, but unless the frequency of such C_{ij} is high, it is not a big problem since we are concerned with the large C_{ij}, not the small. If this should be a problem, an alternative is to use the same procedure with the model of Problem 1.8. If sufficient data are available, the model in Example 1.2 can also be used.

To get a better approximation in (1.44) it may be an idea to use the normal approximation for $\log \sigma_j^2$, i.e. simulate from

$$\begin{bmatrix} \mathbf{f} \\ \log \boldsymbol{\sigma}^2 \end{bmatrix} \approx \mathcal{N}_{2J}\left(\begin{bmatrix} \hat{\mathbf{f}} \\ \log \hat{\boldsymbol{\sigma}}^2 \end{bmatrix}, \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{\mathbf{f}} & \hat{\boldsymbol{\Sigma}}_{\mathbf{f},\log \boldsymbol{\sigma}^2} \\ \hat{\boldsymbol{\Sigma}}'_{\mathbf{f},\log \boldsymbol{\sigma}^2} & \hat{\boldsymbol{\Sigma}}_{\log \boldsymbol{\sigma}^2} \end{bmatrix} \right). \tag{1.45}$$

Indeed, using the log likelihood (1.25) with $\tau = \log \sigma^2$, some straightforward differentiation shows that for $j = 0, \ldots, J-1$,

$$\begin{aligned} \frac{\partial^2}{\partial f_j^2} l(\hat{\mathbf{f}}, \hat{\boldsymbol{\tau}}; \mathbf{c}) &= -\frac{c_{\bullet j}}{\hat{\sigma}_j^2}, \quad j = 0, \ldots, J-1, \\ \frac{\partial^2}{\partial \tau_j^2} l(\hat{\mathbf{f}}, \hat{\boldsymbol{\tau}}; \mathbf{c}) &= -\frac{1}{2}(I-j-1), \quad j = 0, \ldots, J-2. \end{aligned}$$

Here as usual $c_{\bullet j} = \sum_{i=0}^{I-j-1} c_{ij}$. All other second order partial derivatives are zero when the estimated $(\hat{\mathbf{f}}, \hat{\boldsymbol{\tau}})$ are inserted. This means that we can generate independent for $j = 0, \ldots, J-1$,

$$f_j \sim \mathcal{N}\left(\hat{f}_j, \frac{\hat{\sigma}_j^2}{c_{\bullet j}}\right) \quad \text{and} \quad \tau_j \sim \mathcal{N}\left(\hat{\tau}_j, \frac{2}{I-j-1}\right).$$

Actually, for $j = J-1$ the variance was set to $\sigma_{J-1}^2 = 2$.

Example 1.7. In this example we will study the Bornhuetter-Ferguson model (1.35), i.e. the increments

$$X_{ij} = \mu_i \gamma_j + \frac{\sigma}{\sqrt{p_{ij}}} e_{ij},$$

where $\sum_{j=0}^{J} \gamma_j = 1$, the e_{ij} are i.i.d. $\mathcal{N}(0,1)$ and the p_{ij} are known weights.

Assuming first that all parameters are known, the simulation is straightforward. To include parameter uncertainty, set $\boldsymbol{\gamma}_{-0} = (\gamma_1, \ldots, \gamma_J)$. Then let $\gamma_0 = 1$ be fixed and continue as follows:

1. Calculate $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\gamma}}_{-0}$ and $\hat{\sigma}^2$. Here we can use (1.37)-(1.39).
2. Compute the log likelihood $l(\boldsymbol{\mu}, \boldsymbol{\gamma}_{-0}, \tau; \mathbf{x})$ where $\tau = \log \sigma^2$. Calculate the Hessian, and plug in the MLE $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\gamma}}_{-0}, \hat{\tau})$. Call this $\hat{\boldsymbol{\Sigma}}$.
3. Generate the parameters from

$$\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\gamma}_{-0} \\ \tau \end{bmatrix} \sim \mathcal{N}_{I+J+2}\left(\begin{bmatrix} \hat{\boldsymbol{\mu}} \\ \hat{\boldsymbol{\gamma}}_{-0} \\ \hat{\tau} \end{bmatrix}, \hat{\boldsymbol{\Sigma}} \right).$$

These can be scaled so that the γ_j sums to one, but it is not necessary since they always appear in the form $\mu_i \gamma_j$.

4. The rest is as in the last example.

Let us now take a pure Bornhuetter-Ferguson approach and assume that $\boldsymbol{\mu}$ is known, so that only $\boldsymbol{\gamma}$ and σ^2 need to be simulated.

1. Assume that $\boldsymbol{\gamma}$ and σ^2 are independent.
2. As
$$(n-I-J-1)\frac{\hat{\sigma}^2}{\sigma^2} \approx \chi^2(n-I-J-1),$$
we can generate σ^2 as
$$\sigma^2 = \frac{\hat{\sigma}^2}{n-I-J-1}V,$$
where $V \sim \chi^2(n-I-J-1)$.
3. To simulate $\boldsymbol{\gamma}$ that must satisfy $\sum_{j=0}^{J}\gamma_j = 1$, we can use the Dirichlet distribution. According to that problem, the following recipe works:
   - Generate independent $Y_0, \ldots, Y_J$ where $Y_j \sim G(\hat{\gamma}_j, 1)$.
   - Set
   $$\gamma_j = \frac{Y_j}{\sum_{j=0}^{J} Y_j}, \quad j = 0, \ldots, J.$$

   Then $\boldsymbol{\gamma} \sim \mathrm{Dir}_{J+1}(\hat{\boldsymbol{\gamma}})$. (Or rather $(\gamma_0, \ldots, \gamma_{J-1}) \sim \mathrm{Dir}_{J+1}(\hat{\boldsymbol{\gamma}})$).
4. The rest is now as before.

1.5.2 Nonparametric bootstrap

There are several approaches to do nonparametric bootstrap, and we will not try to cover them all here. Instead of trying to write down a general, and probably confusing procedure, we will look at a couple of examples that should be sufficient to explain the basics of the nonparametric bootstrap method.

Example 1.8 (A chain ladder approach). Assume that the accumulated claims follow
$$C_{i,j+1} = f_j C_{ij} + \sigma_j \sqrt{C_{ij}} e_{i,j+1},$$
where the e_{ij} are i.i.d. from some unknown distribution F. To initiate the procedure we do as follows.

1. Estimate $\hat{f}_j$ and $\hat{\sigma}_j$ by using e.g. (1.11) as well as (1.13) and (1.16).
2. Calculate
$$\hat{e}_{i,j+1} = \frac{C_{i,j+1} - \hat{f}_j C_{ij}}{\hat{\sigma}_j \sqrt{C_{ij}}}, \quad j = 0, \ldots, j^*(i) - 1,\ i = 0, \ldots, I-1.$$

The proper bootstrap procedure goes as follows:

3. Sample randomly with replacement from the $\hat{e}_{ij}$ found in Step 2,
$$\hat{e}^*_{ij}, \quad j = 0, \ldots, j^*(i) - 1,\ i = 0, \ldots, I-1.$$

These can be scaled so that they have variance 1.

4. Let iteratively in j,

$$C^*_{i,j+1} = \hat{f}_j C^*_{ij} + \hat{\sigma}_j \sqrt{C^*_{ij}} \hat{e}^*_{i,j+1}, \quad j = 0, \ldots, j^*(i) - 1,\ i = 0, \ldots, I-1.$$

5. Use the C^*_{ij} obtained in Step 4 to calculate new $\hat{f}_j$ and $\hat{\sigma}_j$ with the same procedure as in Step 1.
6. Sample randomly with replacement from the $\hat{e}_{ij}$ found in Step 2,

$$\hat{e}^*_{ij}, \quad j = j^*(i), \ldots, J-1,\ i = I-J+1, \ldots, I.$$

7. Let iteratively in j

$$C^*_{i,j+1} = \hat{f}_j C^*_{ij} + \hat{\sigma}_j \sqrt{C^*_{ij}} \hat{e}^*_{i,j+1}, \quad j = j^*(i), \ldots, J-1,\ i = I-J+1, \ldots, I.$$

8. Calculate

$$C^*_O = \sum_{i=I-J+1}^{I} (C^*_{iJ} - C^*_{i,I-i}).$$

9. Repeat steps 3-7 to obtain $C_{O,1}, \ldots, C_{O,B}$. These can be used to calculate the required quantiles.

If instead of steps 6-8 we let

$$C^*_O = \sum_{i=I-J+1}^{I} C_{i,I-i} \left(\prod_{j=I-i}^{J-1} \hat{f}^*_j - 1 \right), \tag{1.46}$$

the result will be a bootstrap procedure that only accounts for estimation uncertainty but not development uncertainty.

It is also suggested in the literature to replace Step 4 by

4a. Let iteratively in j,

$$C^*_{i,j+1} = \hat{f}_j C_{ij} + \hat{\sigma}_j \sqrt{C_{ij}} \hat{e}^*_{i,j+1}, \quad j = 0, \ldots, j^*(i) - 1,\ i = 0, \ldots, I-1. \tag{1.47}$$

This results in less variation in the C^*_{ij}, and thus less variation in the predictions.

Example 1.9 (A Bornhuetter-Ferguson approach). This is basically the same as in the last example. The residuals are now estimated as

$$\hat{e}_{ij} = \frac{\sqrt{p_{ij}}}{\hat{\sigma}} (X_{ij} - \hat{\mu}_i \hat{\gamma}_j), \quad j = 0, \ldots, j^*(i),\ i = 0, \ldots, I.$$

These can be standardized to have variance 1.

1.6 Problems

1.1. Let X_j, $j = 0, \ldots, J$ be incremental losses, so that $C_j = \sum_{i=0}^{j} X_j$. Assume that

$$\mathbf{Z} \stackrel{\text{def}}{=} \left(\frac{X_0}{C_J}, \ldots, \frac{X_{J-1}}{C_J} \right) \sim \mathrm{Dir}_d(\boldsymbol{\alpha}),$$

and that $\mathbf{Z}$ and C_J are independent. Let f_C be the density of C_J, and let

$$\gamma_j = \frac{\alpha_j}{\sum_{i=0}^{J} \alpha_J} \quad \text{and} \quad \beta_j = \sum_{i=0}^{j} \gamma_j, \quad j = 0, \dots, J,$$

so that $\beta_J = 1$. Finally, set $\mu = E[C_J]$.

a) What is $E[C_j|C_J]$?
b) Show that

$$f_{C_j|C_J}(x|y) = \frac{1}{B(a,b)} x^{a-1} y^{1-a-b} (y-x)^{b-1}, \quad 0 \le x \le y,$$

where

$$a = \sum_{i=0}^{j} \alpha_i, \quad b = \sum_{i=j+1}^{J} \alpha_i \quad \text{and} \quad \alpha_\bullet = a + b = \sum_{i=0}^{J} \alpha_j.$$

c) What is the conditional density $f_{C_J|C_j}(y|x)$? Use this to write down an expression for $E[C_J|C_j]$.

In order to find the maximum likelihood estimators of the unknown parameters based on a runoff triangle, we need for all $j \le J$ the joint density of $\mathbf{X}_j = (X_0, \dots, X_j)$.

d) Find the conditional density $f_{\mathbf{X}_j|C_J}(\mathbf{x}|y)$ and use this to write down an expression for the unconditional density $f_{\mathbf{X}_j}(\mathbf{x})$.

1.2. In 1.2 assume the Y_{ijk} are independent with $Y_{ijk} \sim \text{LN}(\gamma_j, \tau_j^2)$. What are the MLE of f_j and σ_j^2 in this case?

1.3. An alternative to the estimates $\hat{f}_j$ in the chain-ladder method would be to minimize the mean squared error

$$E\left[\left(\sum_{i=0}^{I-j-1} b_i F_{ij} - f_j \right)^2 \middle| \mathscr{E}_j \right].$$

Find the minimizing $\tilde{\mathbf{b}} = (\tilde{b}_0, \dots, \tilde{b}_{I-j-1})$ and explain why the estimator

$$\tilde{f}_j = \sum_{i=0}^{I-j-1} \tilde{b}_i F_{ij}$$

is of limited use.

1.4. Here we again consider the chain-ladder method, and let

$$R_j = \sum_{i=0}^{I-j-1} C_{ij},$$

so that

$$\hat{f}_j = \frac{R_{j+1} + C_{I-j-1,j+1}}{R_j}.$$

Under the assumptions CL1 and CL2, we saw that $E[\hat{f}_j \hat{f}_{j+1} | \mathscr{E}_j] = f_j f_{j+1}$ which implies that $\text{Cov}(\hat{f}_j, \hat{f}_{j+1} | \mathscr{E}_j) = 0$, i.e. $\hat{f}_j$ and $\hat{f}_{j+1}$ are conditionally uncorrelated.

In addition to CL1 and CL2, also assume CL3.

a) Use a suitable conditioning to show that

$$\mathrm{Cov}(\hat{f}_j^2, \hat{f}_{j+1}^2 | \mathscr{E}_j) = \sigma_{j+1}^2 \mathrm{Cov}\left(\hat{f}_j^2, \frac{1}{R_{j+1}} \middle| \mathscr{E}_j\right).$$

b) Show that

$$\mathrm{Cov}(\hat{f}_j^2, \hat{f}_{j+1}^2 | \mathscr{E}_j) < 0.$$

In particular this means that $\hat{f}_j$ and $\hat{f}_{j+1}$ are not conditionally independent.

1.5. Prove (1.23).

1.6. In order to reduce the number of unknown parameters in the model $E[X_{ij}] = \mu_i \gamma_j$, assume that

$$\mu_i = \boldsymbol{\beta}' \mathbf{z}_i, \quad i = 0, \ldots, I,$$

where as usual $\mathbf{z}_i$ is a known covariate of dimension p while the p-vector $\boldsymbol{\beta}$ must be estimated. For example with $\mathbf{z}_i = (1, i)'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)'$, $\mu_i = \beta_0 + \beta_1 i$, so the last term may represent a growth factor. To have the product $(\boldsymbol{\beta}' \mathbf{z}_i)\gamma_j$ well defined we can as usual set $\gamma_0 = 1$ and then rescale afterwards.

Assume that

$$X_{ij} \sim \mathscr{N}\left((\boldsymbol{\beta}' \mathbf{z}_i)\gamma_j, \frac{1}{p_{ij}}\sigma^2\right), \quad j = 0, \ldots, j^*(i),\ i = 0, \ldots, I.$$

Find suitable updating equations for the MLE of $\boldsymbol{\beta}$ and $(\gamma_1, \ldots, \gamma_J)$. Also give a reasonable estimator of σ^2.

1.7. In this problem we will have a look at a model that replaces accident years with calendar years. So assume the increments X_{ij} are independent with

$$E[X_{ij}] = p_i \lambda_{i+j} \gamma_j, \quad i = 0, \ldots, I,\ j = 0, \ldots, J.$$

Here the p_i are known volumes, while the λ_k and γ_j are unknown. We see that λ_{i+j} is a calendar year effect, so the model can be useful in an inflationary world. This is called, or a version of it, the separation method. Let

$$R_i = \sum_{j=0}^{i \wedge J} \frac{1}{p_{i-j}} X_{i-j,j} \quad \text{and} \quad S_j = \sum_{i=0}^{I-j} \frac{1}{p_i} X_{ij}.$$

a) Show that

$$E[R_i] = \lambda_i \sum_{j=0}^{i \wedge J} \gamma_j \quad \text{and} \quad E[S_j] = \gamma_j \sum_{i=0}^{I-j} \lambda_{i+j} = \gamma_j \sum_{k=j}^{I} \lambda_k.$$

b) Explain why

$$\hat{\lambda}_i \stackrel{\text{def}}{=} R_i, \quad i = J, \ldots, I$$

is unbiased for λ_i.

c) Explain why

$$\hat{\gamma}_J = \frac{S_J}{\sum_{k=J}^{I} \hat{\lambda}_k}$$

and then recursively for $j = J-1, J-2, \ldots, 0,$

$$\hat{\lambda}_j = \frac{R_j}{1 - \sum_{k=j+1}^{J} \hat{\gamma}_k}$$

and

$$\hat{\gamma}_j = \frac{S_j}{\sum_{k=j}^{I} \hat{\lambda}_j}$$

are reasonable estimators.

A problem with this model is that we cannot fill out the entire square unless we have estimates for λ_j for $j = I+1, \ldots, I+J$. One easy approach is to assume there is an inflation factor δ so that

$$\lambda_{k+1} = (1+\delta)\lambda_k + e_{k+1}.$$

Assume the e_k are i.i.d. $\mathscr{N}(0, \sigma^2)$ and that we have observed $\lambda_0, \ldots, \lambda_I$. Then

$$f(\lambda_0, \ldots, \lambda_I) = \left(\prod_{i=1}^{I} f(\lambda_i | \lambda_{i-1}) \right) f(\lambda_0).$$

To make the likelihood simpler, we drop the last term so that,

$$L_0(\delta, \sigma^2; \boldsymbol{\lambda}) = \prod_{i=1}^{I} f(\lambda_i | \lambda_{i-1})$$

is the likelihood conditional on λ_0.

d) Show that the conditional MLE obtained by maximizing $L_0(\delta, \sigma^2)$ is given as

$$\hat{\delta} = \frac{\sum_{i=1}^{I} (\lambda_i - \lambda_{i-1})\lambda_{i-1}}{\sum_{i=1}^{I} \lambda_{i-1}^2}$$

and

$$\hat{\sigma}^2 = \frac{1}{I} \sum_{i=1}^{I} (\lambda_i - (1+\hat{\delta})\lambda_{i-1})^2.$$

As the λ_i are unknown, we can set

$$\hat{\hat{\delta}} = \frac{\sum_{i=1}^{I} (\hat{\lambda}_i - \hat{\lambda}_{i-1})\hat{\lambda}_{i-1}}{\sum_{i=1}^{I} \hat{\lambda}_{i-1}^2}$$

and

$$\hat{\hat{\sigma}}^2 = \frac{1}{I} \sum_{i=1}^{I} (\hat{\lambda}_i - (1+\hat{\hat{\delta}})\hat{\lambda}_{i-1})^2.$$

This gives the prediction

$$\hat{\hat{\lambda}}_{I+j} = (1+\hat{\hat{\delta}})^j \hat{\lambda}_I, \quad j = 1, \ldots, J.$$

1.8. Similar to Example 1.1, assume that given $\mathscr{E}_j$,

$$C_{i,j+1} \sim G_G\left(C_{ij} f_j, \frac{\phi_j}{C_{ij}} \right), \quad i = 0, \ldots, I, \; j = 0, \ldots, J-1.$$

Show that the MLE of f_j equals the chain-ladder estimator $\hat{f}_j^{\mathrm{CL}}$.

1.9. Explain how you can simulate in the model of Example 1.2.