Stochastic Claims Reserving Methods in Non-Life Insurance

Mario V. Wüthrich [1]
Department of Mathematics
ETH Zürich

Michael Merz [2]
Faculty of Economics
University Tübingen

Version 1.1

[1]ETH Zürich, CH-8092 Zürich, Switzerland.
[2]University Tübingen, D-72074 Tübingen, Germany.

Contents

Chapter 1

Introduction and Notation

1.1 Claims process

In this lecture we consider claims reserving for a branch of insurance called

Non-Life Insurance.

Sometimes, this branch is also called “General Insurance” (UK), or “Property and Casualty Insurance (US)”.
This branch usually contains all kind of insurance products except life insurance products. This separation is mainly for two reasons: 1) Life insurance products are rather different from non-life insurance contracts, e.g. the terms of a contract, the type of claims, etc. This implies that life and non-life products are modelled rather differently. 2) Moreover, in many countries, e.g. in Switzerland, there is a strict legal separation between life insurance and non-life insurance products. This means that a company for non-life insurance products is not allowed to sell life products, and on the other hand a life insurance company can besides life products only sell health and disability products. Every Swiss company which sells both life and non-life products has at least two legal entities.
The branch non-life insurance contains the following lines of business (LoB):

- Motor insurance (motor third party liability, motor hull)
- Property insurance (private and commercial property against fire, water, flooding, business interruption, etc.)
- Liability insurance (private and commercial liability including director and officers (D&O) liability insurance)
- Accident insurance (personal and collective accident including compulsory accident insurance and workmen’s compensation)

- Health insurance (private personal and collective health)
- Marine insurance (including transportation)
- Other insurance products, like aviation, travel insurance, legal protection, credit insurance, epidemic insurance, etc.

A non-life insurance policy is a contract among two parties, the insurer and the insured. It provides to the insurer a fixed amount of money (called premium), to the insured a financial coverage against the random occurrence of well-specified events (or at least a promise that he gets a well-defined amount in case such an event happens). The right of the insured to these amounts (in case the event happens) constitutes a **claim** by the insured on the insurer.
The amount which the insurer is obliged to pay in respect of a claim is known as **claim amount** or **loss amount**. The payments which make up this claim are known as

- claims payments,
- loss payments,
- paid claims, or
- paid losses.

The history of a typical claim may look as follows:

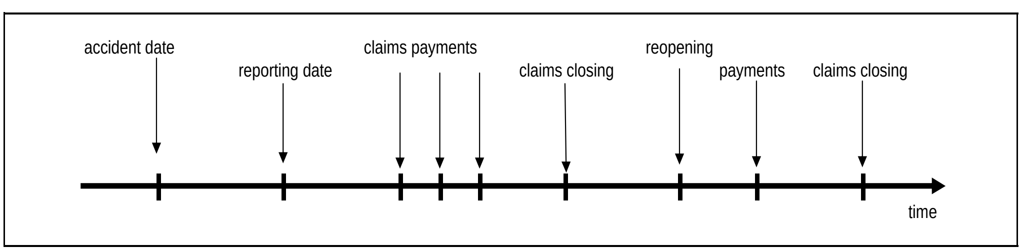


Figure 1.1: Typical time line of a non-life insurance claim

This means that usually the insurance company is not able to settle a claim immediately, this is mainly due to two reasons:

1. Usually, there is a reporting delay (time-lag between claims occurrence and claims reporting to the insurer). The reporting of a claim can take several years, especially in liability insurance (e.g. asbestos or environmental pollution claims), see also Example 1.1.

2. After the reporting it can take several years until a claim gets finally settled. In property insurance we usually have a rather fast settlement whereas in liability or bodily injury claims it often takes a lot of time until the total degree of a claim is clear and known (and can be settled).

3. It can also happen that a closed claim needs to be reopend due to new (unexpected) new developments or in case a relapse happens.

1.1.1 Accounting principle and accident year

There are different premium accounting principles: i) premium booked, ii) premium written, iii) premium earned. It depends on the kind of business written, which principle should be chosen. W.l.o.g. we concentrate in the present manuscript on the premium earned principle:
Usually an insurance company closes its books at least once a year. Let us assume that we close our book always on December 31. How should we show a one-year contract which was written on October 1 2006 with two premium installments paid on October 1 2006 and April 1 2007?
We assume that

- premium written 2006 = 100,
- premium booked 2006 = 50 (= premium received in 2006),
- pipeline premium 31.12.2006 = 50 (= premium which will be received in 2007), which gives premium booked 2007 = 50.

If we assume that the risk exposure is distributed uniformly over time (pro rata temporis), this implies that

- premium earned 2006 = 25 (= premium used for exposure in 2006),
- unearned premium reserve UPR 31.12.2006 = 75 (= premium which will be used for exposure in 2007), which gives premium earned 2007 = 75.

If the exposure is not pro rata temporis, then of course we have a different split of the premium earned into the different accounting years. In order to have a consistent financial statement it is now important that the accident date and the premium accounting principle are compatible (via the exposure pattern). Hence all claims which have accident year 2006 have to be matched to the premium earned 2006, i.e. the claims 2006 have to be paid by the premium earned 2006, whereas the claims with accident year later than 2006 have to be paid by the unearned premium reserve UPR 31.12.2006.

Hence on the one hand we have to build premium reserves for future exposures, but on the other hand we need also to build claims reserves for unsettled claims of past exposures. There are two different types of claims reserves for past exposures:

1. IBNyR reserves (incurred but not yet reported): We need to build claims reserves for claims which have occurred before 31.12.2006, but which have not been reported by the end of the year (i.e. the reporting delay laps into the next accounting years).

2. IBNeR reserves (incurred but not enough reported): We need to build claims reserves for claims which have been reported before 31.12.2006, but which have not been settled yet, i.e. we still expect payments in the future, which need to be financed by the already earned premium.

Example 1.1 (Reporting delay)

accident year	number of reported claims, non-cumulative according to reporting delay										
	reporting period										
	0	1	2	3	4	5	6	7	8	9	10
0	368	191	28	8	6	5	3	1	0	0	1
1	393	151	25	6	4	5	4	1	2	1	0
2	517	185	29	17	11	10	8	1	0	0	1
3	578	254	49	22	17	6	3	0	1	0	0
4	622	206	39	16	3	7	0	1	0	0	0
5	660	243	28	12	12	4	4	1	0	0	0
6	666	234	53	10	8	4	6	1	0	0	0
7	573	266	62	12	5	7	6	5	1	0	1
8	582	281	32	27	12	13	6	2	1	0	
9	545	220	43	18	12	9	5	2	0		
10	509	266	49	22	15	4	8	0			
11	589	210	29	17	12	4	9				
12	564	196	23	12	9	5					
13	607	203	29	9	7						
14	674	169	20	12							
15	619	190	41								
16	660	161									
17	660										

Table 1.1: claims development triangle for number of IBNyR cases (source [75])

1.1.2 Inflation

The following subsection on inflation follows Taylor [75].

Claims costs are often subject to inflation. Usually it is not the typical inflation, like salary or price inflation. Inflation is very specific to the LoB chosen. For example in the LoB accident inflation is driven by medical inflation, whereas for

the LoB motor hull inflation is driven by the technical complexity of car repairing techniques. The essential point is that claims inflation may continue beyond the occurrence date of the accident up to the point of its final payments/settlement.
If X_{t_i} denote the positive single claims payments at time t_i expressed in money value at time t_1, then the total claim amount is in money value at time t_1 given by

$$C_1 = \sum_{i=1}^{\infty} X_{t_i}. \tag{1.1}$$

If $\lambda(\cdot)$ denotes the index which measures the claims inflation, the actual claim amount (nominal) is

$$C = \sum_{i=1}^{\infty} \frac{\lambda(t_i)}{\lambda(t_1)} X_{t_i}. \tag{1.2}$$

Whenever λ is an increasing function we observe that C is bigger than C_1. Of course, in practice we only observe the unindexed payments $X_{t_i}\lambda(t_i)/\lambda(t_1)$ and in general it is difficult to estimate an index function such that we obtain indexed values X_{t_i}. Finding an index function $\lambda(\cdot)$ is equivalent in defining appropriate deflators φ, which is a well-known concept in market consistent actuarial valuation, see e.g. Wüthrich-Bühlmann-Furrer [91].
The basic idea between indexed values C_1 is that, if two sets of payments relate to identical circumstances except that there is a time translation in the payment, their indexed values will be the same, whereas the unindexed values are not the same: For $c > 0$ we assume that

$$\tilde{X}_{t_i+c} = X_{t_i}. \tag{1.3}$$

For λ increasing we have that

$$C_1 = \sum_i X_{t_i} = \sum_i \tilde{X}_{t_i+c} = \tilde{C}_1 \tag{1.4}$$

$$\tilde{C} = \sum_{i=1}^{\infty} \frac{\lambda(t_i+c)}{\lambda(t_1)} \tilde{X}_{t_i+c} = \sum_{i=1}^{\infty} \frac{\lambda(t_i+c)}{\lambda(t_1)} X_{t_i} > C, \tag{1.5}$$

whenever λ is an increasing function (we have assumed (1.3)). This means that the unindexed values differ by the factor $\lambda(t_i+c)/\lambda(t_i)$. However in practice this ratio turns often out to be even of a different form, namely

$$\left(1+\tilde{\psi}(t_i, t_i+c)\right) \cdot \frac{\lambda(t_i+c)}{\lambda(t_i)}, \tag{1.6}$$

meaning that over the time interval $[t_i, t_i+c]$ claim costs are inflated by an additional factor $\left(1+\tilde{\psi}(t_i, t_i+c)\right)$ above the "natural" inflation. This additional inflation is referred to as **superimposed inflation** and can be caused e.g. by changes in the jurisdiction and an increased claims awareness of the insured. We will not further discuss this in the sequel.

1.2 Structural framework to the claims reserving problem

In this section we present a mathematical framework for claims reserving. For this purpose we follow Arjas [5]. Observe that in this subsection all actions of a claim are ordered according to their notification at the insurance company. From a statistical point of view this makes perfect sense, however from an accounting point of view, one should order the claims rather to their occurence/accident date, this has been done e.g. in Norberg [58, 59]. Of course, there is a one-to-one relation between the two concepts.

We assume that we have N claims within a fixed time period with reporting dates $T_1, \ldots T_N$ (assume that they are ordered, $T_i \le T_{i+1}$ for all i). Fix the i-th claim. Then $T_i = T_{i,0}, T_{i,1}, \ldots, T_{i,j}, \ldots, T_{i,N_i}$ denotes the sequence of dates, where some action on claim i is observed, at time $T_{i,j}$ we have for example a payment, a new estimation of the claims adjuster or other new information on claim i. T_{i,N_i} denotes the final settlement of the claim. Assume that $T_{i,N_i+k} = \infty$ for $k \ge 1$.

We specify the events that take place at time $T_{i,j}$ by

$$X_{i,j} = \begin{cases} \text{payment at time } T_{i,j} \text{ for claim } i, \\ 0, \text{ if there is no payment at time } T_{i,j}, \end{cases} \tag{1.7}$$

$$I_{i,j} = \begin{cases} \text{new information available at } T_{i,j} \text{ for claim } i, \\ \emptyset, \text{ if there is no new information at time } T_{i,j}. \end{cases} \tag{1.8}$$

We set $X_{i,j} = 0$ and $I_{i,j} = \emptyset$ whenever $T_{i,j} = \infty$.

With this structure we can define various interesting processes, moreover our claims reserving problem splits into several subproblems. For every i we obtain a marked point processes.

- **Payment process of claim i.** $(T_{i,j}, X_{i,j})_{j\ge 0}$ defines the following cumulative payment process

  $$C_i(t) = \sum_{j: T_{i,j} \le t} X_{i,j}. \tag{1.9}$$

  Moreover $C_i(t) = 0$ for $t < T_i$. The total ultimate claim amount is given by

  $$C_i(\infty) = C_i(T_{i,N_i}) = \sum_{j \ge 0} X_{i,j}. \tag{1.10}$$

  The total claims reserves for claim i at time t for the future liabilities (outstanding claim at time t) are given by

  $$R_i(t) = C_i(\infty) - C_i(t) = \sum_{j: T_{i,j} > t} X_{i,j}. \tag{1.11}$$

- **Information process of claim** i is given by $(T_{i,j}, I_{i,j})_{j\geq 0}$.
- **Settlement process of claim** i is given by $(T_{i,j}, I_{i,j}, X_{i,j})_{j\geq 0}$.

We denote the aggregated processes of all claims i by

$$C(t) = \sum_{i=1}^{N} C_i(t), \tag{1.12}$$

$$R(t) = \sum_{i=1}^{N} R_i(t).$$

$C(t)$ denotes all payments up to time t for all N claims, and $R(t)$ denotes the outstanding claims payments (reserves) at time t for these N claims.
We consider now claims reserving as a prediction problem. Let

$$\mathcal{F}_t^N = \sigma\left\{(T_{i,j}, I_{i,j}, X_{i,j})_{i\geq 1, j\geq 0} : T_{i,j} \leq t\right\} \tag{1.13}$$

be the information available at time t. This σ-field is obtained from the information available at time t from the claims settlement processes. Often there is additional exogenous information $\mathcal{E}_t$ at time t (change of legal practice, high inflation, job market infromation, etc.). Therefore we define the information which the insurance company has at time t by

$$\mathcal{F}_t = \sigma\left(\mathcal{F}_t^N \otimes \mathcal{E}_t\right). \tag{1.14}$$

Problem. Estimate the conditional distributions

$$\mu_t(\cdot) = P\left[C(\infty) \in \cdot \,|\mathcal{F}_t\right], \tag{1.15}$$

with the first two moments

$$M_t = E\left[C(\infty)|\mathcal{F}_t\right], \tag{1.16}$$

$$V_t = \operatorname{Var}\left(C(\infty)|\mathcal{F}_t\right). \tag{1.17}$$

1.2.1 Fundamental properties of the reserving process

Because of

$$C(\infty) = C(t) + R(t), \tag{1.18}$$

we have that

$$M_t = C(t) + E\left[R(t)|\mathcal{F}_t\right] \stackrel{def.}{=} C(t) + m_t, \tag{1.19}$$

$$V_t = \operatorname{Var}\left(R(t)|\mathcal{F}_t\right). \tag{1.20}$$

Lemma 1.2 *M_t is an $\mathcal{F}_t$-martingale, i.e. for $t > s$ we have that*

$$E\left[M_t\middle|\,\mathcal{F}_s\right] = M_s, \qquad a.s. \tag{1.21}$$

Proof. The proof is clear (successive forecasts).

□

Lemma 1.3 *The variance process V_t is an $\mathcal{F}_t$-supermartingale, i.e. for $t > s$ we have that*

$$E\left[V_t\middle|\,\mathcal{F}_s\right] \le V_s, \qquad a.s. \tag{1.22}$$

Proof. Using Jensen's inequality for $t > s$ we have a.s. that

$$\begin{aligned} E\left[V_t\middle|\,\mathcal{F}_s\right] &= E\left[\mathrm{Var}\left(C(\infty)|\mathcal{F}_t\right)\middle|\,\mathcal{F}_s\right] \\ &= E\left[E\left[C^2(\infty)\middle|\,\mathcal{F}_t\right]\middle|\,\mathcal{F}_s\right] - E\left[E\left[C(\infty)\middle|\,\mathcal{F}_t\right]^2\middle|\,\mathcal{F}_s\right] \\ &\le E\left[C^2(\infty)\middle|\,\mathcal{F}_s\right] - E\left[E\left[C(\infty)\middle|\,\mathcal{F}_t\right]\middle|\,\mathcal{F}_s\right]^2 \\ &= \mathrm{Var}\left(C(\infty)|\mathcal{F}_s\right) = V_s. \end{aligned} \tag{1.23}$$

□

Consider $u > t$ and define the increment from t to u by

$$M(t,u) = M_u - M_t. \tag{1.24}$$

Then, a.s., we have that

$$\begin{aligned} E\left[M(t,u)M(u,\infty)\middle|\,\mathcal{F}_t\right] &= E\left[M(t,u)E\left[M(u,\infty)\middle|\,\mathcal{F}_u\right]\middle|\,\mathcal{F}_t\right] \\ &= E\left[M(t,u)\left(E\left[C(\infty)\middle|\,\mathcal{F}_u\right] - M_u\right)\middle|\,\mathcal{F}_t\right] = 0. \end{aligned} \tag{1.25}$$

This implies that $M(t,u)$ and $M(u,\infty)$ are uncorrelated, which is the well-known property that a we have **uncorrelated increments**.

First approach to the claims reserving problem. Use the martingale integral representation. This leads to the "innovation gains process", which determines M_t when updating $\mathcal{F}_t$.

- This theory is well-understood.
- One has little idea about the updating process.
- One has (statistically) not enough data.

Second approach to the claims reserving problem. For $t < u$ we have that $\mathcal{F}_t \subset \mathcal{F}_u$. Since M_t is an $\mathcal{F}_t$-martingale we have that

$$E\left[M(t,u)|\,\mathcal{F}_t\right] = 0 \qquad \text{a.s.} \tag{1.26}$$

We define the incremental payments within t and u by

$$X(t,u) = C(u) - C(t). \tag{1.27}$$

Hence we have that

$$\begin{aligned} M(t,u) &= E\left[C(\infty)|\,\mathcal{F}_u\right] - E\left[C(\infty)|\,\mathcal{F}_t\right] \\ &= C(u) + E\left[R(u)|\mathcal{F}_u\right] - \left(C(t) + E\left[R(t)|\mathcal{F}_t\right]\right) \\ &= X(t,u) + E\left[R(u)|\mathcal{F}_u\right] - E\left[C(u) - C(t) + R(u)|\,\mathcal{F}_t\right] \\ &= X(t,u) - E\left[X(t,u)|\mathcal{F}_t\right] + E\left[R(u)|\mathcal{F}_u\right] - E\left[R(u)|\mathcal{F}_t\right]. \end{aligned} \tag{1.28}$$

Henceforth we have the following two terms

1. prediction error for payments within $(t, t+1]$

$$X(t,t+1) - E\left[X(t,t+1)|\mathcal{F}_t\right]; \tag{1.29}$$

2. prediction error of reserves $R(t+1)$ when updating information

$$E\left[R(t+1)|\mathcal{F}_{t+1}\right] - E\left[R(t+1)|\mathcal{F}_t\right]. \tag{1.30}$$

1.2.2 Known and unknown claims

As in Subsection 1.1.1 we define IBNyR (incurred but not yet reported) claims and reported claims. The following process counts the number of reported claims,

$$N_t = \sum_{i \geq 1} 1_{\{T_i \leq t\}}. \tag{1.31}$$

Hence we can split the ultimate claim and the reserves at time t with respect to the fact whether we have a reported or an IBNyR claim by

$$R(t) = \sum_i R_i(t) \cdot 1_{\{T_i \leq t\}} + \sum_i R_i(t) \cdot 1_{\{T_i > t\}}, \tag{1.32}$$

where

$$\sum_i R_i(t) \cdot 1_{\{T_i \leq t\}} \qquad \text{reserves for at time } t \text{ reported claims,} \tag{1.33}$$

$$\sum_i R_i(t) \cdot 1_{\{T_i > t\}} \qquad \text{reserves for at time } t \text{ IBNyR claims.} \tag{1.34}$$

Hence we define

$$R_t^{rep} = E\left[\sum_i R_i(t)\cdot 1_{\{T_i\le t\}}\middle|\mathcal{F}_t\right] = E\left[\sum_{i=1}^{N_t} R_i(t)\middle|\mathcal{F}_t\right], \tag{1.35}$$

$$R_t^{IBNyR} = E\left[\sum_i R_i(t)\cdot 1_{\{T_i> t\}}\middle|\mathcal{F}_t\right] = E\left[\sum_{i=N_t+1}^{N} R_i(t)\middle|\mathcal{F}_t\right], \tag{1.36}$$

where N is total (random) number of claims. Hence we easily see that

$$R_t^{rep} = \sum_{i\le N_t} E\left[R_i(t)|\,\mathcal{F}_t\right], \tag{1.37}$$

$$R_t^{IBNyR} = E\left[\sum_{i=N_t+1}^{N} R_i(t)\middle|\mathcal{F}_t\right]. \tag{1.38}$$

R_t^{rep} denotes the at time t expected future payments for reported claims. This is often called "best estimate reserves at time t for reported claims". R_t^{IBNyR} are the at time t expected future payments for IBNyR claims (or "best estimate reserves for IBNyR claims").

Conclusions. (1.37)-(1.38) shows that the reserves for reported claims and the reserves for IBNyR claims are of rather different nature:

i) The reserves for reported claims should be determined individually, i.e. on a single claims basis. Often one has quite a lot of information on reported claims (e.g. case estimates), which asks for an estimate on single claims.

ii) The reserves for IBNyR claims can not be decoupled due to the fact that N is not known at time t (see (1.36)). Moreover we have no information on a single claims basis. This shows that IBNyR reserves should be determined on a collective basis.

Unfortunately most of the classical claims reserving methods do not distinguish reported claims from IBNyR claims, i.e. they estimate the claims reserves at the same time on both classes. In that context, we have to slightly disappoint the reader, because the most methods presented in this manuscript do also not make this distinction.

1.3 Outstanding loss liabilities, classical notation

In this subsection we introduce the classical claims reserving notation and terminology. In most cases outstanding loss liabilities are estimated in so-called claims development triangles which separates claims on two time axis.

In the sequel we always denote by

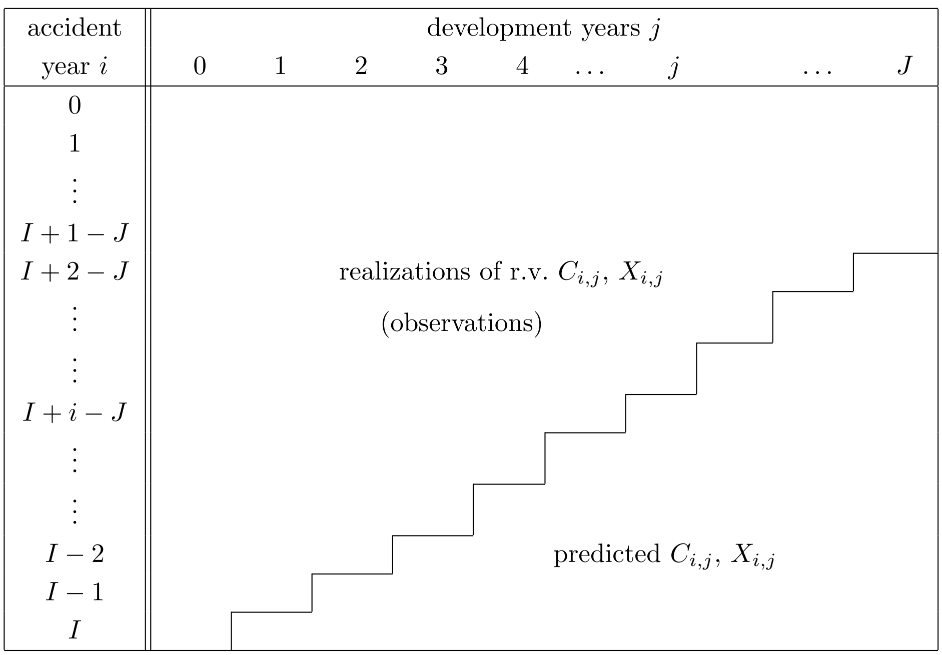
$$i = \text{accident year, year of occurrence,} \tag{1.39}$$

$$j = \text{development year, development period.} \tag{1.40}$$

For illustrative purposes we assume that: $X_{i,j}$ denotes all payments in development period j for claims with accident year i, i.e. this corresponds to the incremental claims payments for claims with accident year i done in accounting year $i+j$. Below, we see which other meaning $X_{i,j}$ can have.

In a claims development triangle accident years are usually on the vertical line whereas development periods are on the horizontal line (see also Table 1.1). Usually the loss development tables split into two parts the upper triangle/trapezoid where we have observations and the lower triangle where we want to estimate the outstanding payments. On the diagonals we always see the accounting years.

Hence the claims data have the following structure:



Data can be shown in cumulative form or in non-cumulative (incremental) form. Incremental data are always denoted by $X_{i,j}$ and cumulative data are given by

$$C_{i,j} = \sum_{k=0}^{j} X_{i,k}. \tag{1.41}$$

The incremental data $X_{i,j}$ may denote the incremental payments in cell (i,j), the number of reported claims with reporting delay j and accident year i or the change of reported claim amount in cell (i,j). For cumulative data $C_{i,j}$ we often use the terminology **cumulative payments**, **total number of reported claims** or

claims incurred (for cumulative reported claims). $C_{i,\infty}$ is often called **ultimate claim amount/load** of accident year i or total number of claims in year i.

$X_{i,j}$ incremental payments	$\Longleftrightarrow$	$C_{i,j}$ cumulative payments
$X_{i,j}$ number of reported claims with delay j	$\Longleftrightarrow$	$C_{i,j}$ total number of reported claims
$X_{i,j}$ change of reported claim amount	$\Longleftrightarrow$	$C_{i,j}$ claims incurred

Usually we have observations $\mathcal{D}_I = \{X_{i,j};\ i+j \le I\}$ in the upper trapezoid and $\mathcal{D}_I^c = \{X_{i,j};\ i+j > I\}$ needs to be estimated.
The payments in a single accounting year are

$$X_k = \sum_{i+j=k} X_{i,j}, \tag{1.42}$$

these are the payments in the $(k+1)$-st diagonal.
If $X_{i,j}$ denote incremental payments then the **outstanding loss liabilities** for accident year i at time j are given by

$$R_{i,j} = \sum_{k=j+1}^{\infty} X_{i,k} = C_{i,\infty} - C_{i,j}. \tag{1.43}$$

$R_{i,j}$ are also called **claims reserves**, this is essentially the amount we have to estimate (lower triangle) so that together with the past payments $C_{i,j}$ we obtain the whole claims load (ultimate claim) for accident year i.

1.4 General Remarks

If we consider loss reserving models, i.e. models which estimate the total claim amount there are always several possibilities to do so.

- Cumulative or incremental data
- Payments or claims incurred data
- Split small and large claims
- Indexed or unindexed data
- Number of claims and claims averages
- Etc.

Usually, different methods and differently aggregated data sets lead to very different results. Only an **experienced reserving actuary** is able to tell you which is an accurate/good estimate for future liabilities for a specific data set.
Often there are so many phenomenons in the data which need first to be understood before applying a method (we can not simply project the past to the future by applying one model).
With this in mind we describe different methods, but only practical experience will tell you which method should be applied in which situation. I.e. the focus of this manuscript lies on the mathematical description of stochastic models. We derive various properties of these models. The question of an **appropriate model choice** for a specific data set is **not treated here**. Indeed, this is probably one of the most difficult questions. Moreover, there is only very little literature on this topic, e.g. for the chain-ladder method certain aspects are considered in Barnett-Zehnwirth [7] and Venter [77].

Remark on claims figures.
When we speak about claims development triangles (paid or incurred data), these usually contain loss adjustment expenses, which can be allocated/attributed to single claims (and therefore are contained in the claims figures). Such expenses are called **allocated loss adjustment expenses** (ALAE). These are typically expenses for external lawyers, an external expertise, etc. Internal loss adjustment expenses (income of claims handling department, maintenance of claims handling system, management fees, etc.) are typically not contained in the claims figures and therefore have to be estimated separately. These costs are called **unallocated loss adjustment expenses** (ULAE). Below, in the appendix, we describe the New York-method (paid-to-paid method), which serves to estimate ULAE. The New York-method is a rather rough method which only works well in stationary situation. Therefore one could think of more sophisticated methods. Since usually, ULAE are rather small compared to the other claims payments, the New York-method is often sufficient in practical applications.

Chapter 2

Basic Methods

We start the general discussion on claims reserving with three standard methods:

1. Chain-ladder method
2. Bornhuetter-Ferguson method
3. Poisson model for claim counts

This short chapter has on the one hand illustrative purposes to give some ideas, how one can tackle the problem. It presents the easiest two methods (chain-ladder and Bornhuetter-Ferguson method). On the other hand one should realize that in practice these are the methods which are used most often (due to their simplicity). The chain-ladder method will be discussed in detail in Chapter 3, the Bornhuetter-Ferguson method will be discussed in detail in Chapter 4.

We assume that the last development period is given by J, i.e. $X_{i,j} = 0$ for $j > J$, and that the last observed accident year is given by I (of course we assume $(J \leq I)$).

2.1 Chain-ladder model (distribution free model)

The chain-ladder model is probably the most popular loss reserving technique. We give different derivations for the chain-ladder model. In this section we give a distribution-free derivation of the chain-ladder model (see Mack [49]). The conditional prediction error of the chain-ladder model will be treated in Chapter 3.
The classical actuarial literature often explains the chain-ladder method as a pure computational alogrithm to estimate claims reserves. It was only much later that actuaries started to think about stochastic models which generate the chain-ladder algorithm. The first who came up with a full stochastic model for the chain-ladder method was Mack [49]. In 1993, Mack [49] published one of the most famous

articles in claims reserving on the calculation of the standard error in the chain-ladder model.

Model Assumptions 2.1 (Chain-ladder model)

There exist development factors $f_0, \ldots, f_{J-1} > 0$ such that for all $0 \le i \le I$ and all $1 \le j \le J$ we have that

$$E\left[C_{i,j} \middle| C_{i,0}, \ldots, C_{i,j-1}\right] = E\left[C_{i,j} \middle| C_{i,j-1}\right] = f_{j-1} \cdot C_{i,j-1}, \tag{2.1}$$

and different accident years i are independent.

□

Remarks 2.2

- We assume independence of the accident years. We will see below that this assumption is done in almost all of the methods. It means that we have already eliminated accounting year effects in the data.
- In addition we could also do stronger assumptions for the sequences $C_{i,0}, C_{i,1}, \ldots$, namely that they form Markov chains. Moreover, observe that

  $$C_{i,j} \cdot \prod_{l=0}^{j-1} f_l^{-1} \tag{2.2}$$

  forms a martingale for $j \ge 0$.
- The factors f_j are called development factors, chain-ladder factors or age-to-age factors. It is the central object of interest in the chain-ladder method.

Lemma 2.3 *Let $\mathcal{D}_I = \{C_{i,j};\ i + j \le I, 0 \le j \le J\}$ be the set of observations (upper trapezoid). Under Model Assumptions 2.1 we have for all $I - J + 1 \le i \le I$ that*

$$E\left[C_{i,J} \middle| \mathcal{D}_I\right] = E\left[C_{i,J} \middle| C_{i,I-i}\right] = C_{i,I-i} \cdot f_{I-i} \cdots f_{J-1}. \tag{2.3}$$

Proof. This is an exercise using conditional expectations:

$$\begin{aligned} E\left[C_{i,J} \middle| C_{i,I-i}\right] &= E\left[C_{i,J} \middle| \mathcal{D}_I\right] \\ &= E\left[E\left[C_{i,J} \middle| C_{i,J-1}\right] \middle| \mathcal{D}_I\right] \\ &= f_{J-1} \cdot E\left[C_{i,J-1} \middle| \mathcal{D}_I\right]. \end{aligned} \tag{2.4}$$

If we iterate this procedure until we reach the diagonal $i + j = I$ we obtain the claim.

□

Lemma 2.3 gives an algorithm for estimating the expected ultimate claim given the observations $\mathcal{D}_I$. This algorithm is often called recursive algorithm. For known chain-ladder factors f_j we estimate the expected outstanding claims liabilities of accident year i based on $\mathcal{D}_I$ by

$$E\left[C_{i,J}\middle|\mathcal{D}_I\right] - C_{i,I-i} = C_{i,I-i} \cdot \left(f_{I-i} \cdots f_{J-1} - 1\right). \tag{2.5}$$

This corresponds to the "best estimate" reserves for accident year i at time I (based on the information $\mathcal{D}_I$). Unfortunately, in most practical applications the chain-ladder factors are not known and need to be estimated. We define

$$j^*(i) = \min\{J, I-i\} \qquad \text{and} \qquad i^*(j) = I - j, \tag{2.6}$$

these denote the last observations/indices on the diagonal. The age-to-age factors f_{j-1} are estimated as follows:

$$\widehat{f}_{j-1} = \frac{\sum_{k=0}^{i^*(j)} C_{k,j}}{\sum_{k=0}^{i^*(j)} C_{k,j-1}}. \tag{2.7}$$

Estimator 2.4 (Chain-ladder estimator) *The CL estimator for* $E\left[C_{i,j}\middle|\mathcal{D}_I\right]$ *is given by*

$$\widehat{C_{i,j}}^{CL} = \widehat{E}\left[C_{i,j}\middle|\mathcal{D}_I\right] = C_{i,I-i} \cdot \widehat{f}_{I-i} \cdots \widehat{f}_{j-1} \tag{2.8}$$

for $i + j > I$.

We define (see also Table 2.1)

$$\mathcal{B}_k = \{C_{i,j};\ i + j \le I, 0 \le j \le k\} \subseteq \mathcal{D}_I. \tag{2.9}$$

In fact, we have $\mathcal{B}_J = \mathcal{D}_I$, which is the set of all observations at time I.

Lemma 2.5 *Under Model Assumptions 2.1 we have that:*

a) $\widehat{f}_j$ *is, given* $\mathcal{B}_j$, *an unbiased estimator for* f_j, *i.e.* $E\left[\widehat{f}_j\middle|\mathcal{B}_j\right] = f_j$,

b) $\widehat{f}_j$ *is (unconditionally) unbiased for* f_j, *i.e.* $E\left[\widehat{f}_j\right] = f_j$,

c) $\widehat{f}_0, \ldots, \widehat{f}_{J-1}$ *are uncorrelated, i.e.* $E\left[\widehat{f}_0 \cdot \ldots \cdot \widehat{f}_{J-1}\right] = E\left[\widehat{f}_0\right] \cdot \ldots \cdot E\left[\widehat{f}_{J-1}\right]$,

d) $\widehat{C_{i,J}}^{CL}$ *is, given* $C_{i,I-i}$, *an unbiased estimator for* $E\left[C_{i,J}\middle|\mathcal{D}_I\right] = E\left[C_{i,J}\middle|C_{i,I-i}\right]$, *i.e.* $E\left[\widehat{C_{i,J}}^{CL}\middle|C_{I-i}\right] = E\left[C_{i,J}\middle|\mathcal{D}_I\right]$ *and*

accident year	number of reported claims, non-cumulative according to reporting delay — reporting period 0	1	2	3	4	5	6	7	8	9	10
0	368	191	28	8	6	5	3	1	0	0	1
1	393	151	25	6	4	5	4	1	2	1	0
2	517	185	29	17	11	10	8	1	0	0	1
3	578	254	49	22	17	6	3	0	1	0	0
4	622	206	39	16	3	7	0	1	0	0	0
5	660	243	28	12	12	4	4	1	0	0	0
6	666	234	53	10	8	4	6	1	0	0	0
7	573	266	62	12	5	7	6	5	1	0	1
8	582	281	32	27	12	13	6	2	1	0	
9	545	220	43	18	12	9	5	2	0		
10	509	266	49	22	15	4	8	0			
11	589	210	29	17	12	4	9				
12	564	196	23	12	9	5					
13	607	203	29	9	7						
14	674	169	20	12							
15	619	190	41								
16	660	161									
17	660										

Table 2.1: The set $\mathcal{B}_3$

e) $\widehat{C_{i,J}}^{CL}$ *is (unconditionally) unbiased for* $E[C_{i,J}]$*, i.e.* $E\left[\widehat{C_{i,J}}^{CL}\right] = E[C_{i,J}]$.

At the first sight, the uncorrelatedness of $\widehat{f}_j$ is surprising since neighboring estimators of the age-to-age factors depend on the same data.

Proof of Lemma 2.5. a) We have

$$E\left[\left.\widehat{f}_{j-1}\right|\mathcal{B}_{j-1}\right] = \frac{\sum_{k=0}^{i^*(j)} E\left[C_{k,j}\right|\mathcal{B}_{j-1}]}{\sum_{k=0}^{i^*(j)} C_{k,j-1}} = \frac{\sum_{k=0}^{i^*(j)} C_{k,j-1}\cdot f_{j-1}}{\sum_{k=0}^{i^*(j)} C_{k,j-1}} = f_{j-1}. \tag{2.10}$$

This immediately implies the conditional unbiasedness.

b) Follows immediately from a).

c) For the uncorrelatedness of the estimators we have for $j < k$

$$E\left[\widehat{f}_j\cdot\widehat{f}_k\right] = E\left[E\left[\left.\widehat{f}_j\cdot\widehat{f}_k\right|\mathcal{B}_k\right]\right] = E\left[\widehat{f}_j\cdot E\left[\left.\widehat{f}_k\right|\mathcal{B}_k\right]\right] = E\left[\widehat{f}_j\cdot f_k\right] = f_j\cdot f_k, \tag{2.11}$$

which implies the claim.

d) For the unbiasedness of the chain-ladder estimator we have

$$\begin{aligned} E\left[\left.\widehat{C_{i,J}}^{CL}\right|C_{i,I-i}\right] &= E\left[\left.C_{i,I-i}\cdot\widehat{f}_{I-i}\cdots\widehat{f}_{J-1}\right|C_{i,I-i}\right] \\ &= E\left[\left.C_{i,I-i}\cdot\widehat{f}_{I-i}\cdots\widehat{f}_{J-2}\cdot E\left[\left.\widehat{f}_{J-1}\right|\mathcal{B}_{J-1}\right]\right|C_{i,I-i}\right] \\ &= f_{J-1}\cdot E\left[\left.\widehat{C_{i,J-1}}^{CL}\right|C_{i,I-i}\right]. \end{aligned} \tag{2.12}$$

Iteration of this procedure leads to

$$E\left[\left.\widehat{C_{i,J}}^{CL}\right|C_{i,I-i}\right] = C_{i,I-i}\cdot f_{I-i}\cdots f_{J-1} = E\left[C_{i,J}\right|\mathcal{D}_I]. \tag{2.13}$$

e) Follows immediately from d).
This finishes the proof of this lemma.

□

Remarks 2.6

- Observe that we have proved in Lemma 2.5 that the estimators $\widehat{f}_j$ are uncorrelated. But pay attention to the fact that they are **not** independent. In fact, the squares of two successive estimators $\widehat{f}_j$ and $\widehat{f}_{j+1}$ are negatively correlated (see also Lemma 3.8 below). It is also this negative correlation which will lead to quite some discussions about estimation errors of our parameter estimates.

- Observe that Lemma 2.5 d) shows that we obtain unbiased estimators for the best estimate reserves $E\left[C_{i,J}\middle|\mathcal{D}_I\right]$.

Let us finish this section with an example.

Example 2.7 (Chain-ladder method)

	0	1	2	3	4	5	6	7	8	9
0	5'946'975	9'668'212	10'563'929	10'771'690	10'978'394	11'040'518	11'106'331	11'121'181	11'132'310	11'148'124
1	6'346'756	9'593'162	10'316'383	10'468'180	10'536'004	10'572'608	10'625'360	10'636'546	10'648'192	
2	6'269'090	9'245'313	10'092'366	10'355'134	10'507'837	10'573'282	10'626'827	10'635'751		
3	5'863'015	8'546'239	9'268'771	9'459'424	9'592'399	9'680'740	9'724'068			
4	5'778'885	8'524'114	9'178'009	9'451'404	9'681'692	9'786'916				
5	6'184'793	9'013'132	9'585'897	9'830'796	9'935'753					
6	5'600'184	8'493'391	9'056'505	9'282'022						
7	5'288'066	7'728'169	8'256'211							
8	5'290'793	7'648'729								
9	5'675'568									
$\widehat{f}_j$	1.4925	1.0778	1.0229	1.0148	1.0070	1.0051	1.0011	1.0010	1.0014	

Table 2.2: Observed historical cumulative payments $C_{i,j}$ and estimated chain-ladder factors $\widehat{f}_j$

	0	1	2	3	4	5	6	7	8	9	Reserves
0											
1										10'663'318	15'126
2									10'646'884	10'662'008	26'257
3								9'734'574	9'744'764	9'758'606	34'538
4							9'837'277	9'847'906	9'858'214	9'872'218	85'302
5						10'005'044	10'056'528	10'067'393	10'077'931	10'092'247	156'494
6					9'419'776	9'485'469	9'534'279	9'544'580	9'554'571	9'568'143	286'121
7				8'445'057	8'570'389	8'630'159	8'674'568	8'683'940	8'693'030	8'705'378	449'167
8			8'243'496	8'432'051	8'557'190	8'616'868	8'661'208	8'670'566	8'679'642	8'691'971	1'043'242
9		8'470'989	9'129'696	9'338'521	9'477'113	9'543'206	9'592'313	9'602'676	9'612'728	9'626'383	3'950'815
										Total	6'047'061

Table 2.3: Estimated cumulative chain-ladder payments $\widehat{C_{i,j}}^{CL}$ and estimated chain-ladder reserves $\widehat{C_{i,J}}^{CL} - C_{i,I-i}$

2.2 The Bornhuetter-Ferguson method

The Bornhuetter-Ferguson method is in general a very robust method, since it does not consider outliers in the observations. We will further comment on this in Chapter 4. The method goes back to Bornhuetter-Ferguson [10] which have published this method in 1972 in an article called "the actuary and IBNR".
The Bornhuetter-Ferguson method is usually understood as a pure algorithm to estimate reserves (this is also the way it was published in [10]). There are several possibilities to define an appropriate underlying stochastic model which motivates the BF method. Straightforward are for example the following assumptions:

Model Assumptions 2.8

- Different accident years i are independent.
- There exist parameters $\mu_0, \ldots, \mu_I > 0$ and a pattern $\beta_0, \ldots, \beta_J > 0$ with $\beta_J = 1$ such that for all $i \in \{0, \ldots, I\}$, $j \in \{0, \ldots, J-1\}$ and $k \in \{1, \ldots, J-j\}$

$$\begin{aligned} E[C_{i,0}] &= \mu_i \cdot \beta_0, \\ E[C_{i,j+k}|C_{i,0}, \ldots, C_{i,j}] &= C_{i,j} + \mu_i \cdot (\beta_{j+k} - \beta_j) . \end{aligned} \tag{2.14}$$

□

Then we have $E[C_{i,j}] = \mu_i \cdot \beta_j$ and $E[C_{i,J}] = \mu_i$. The sequence $(\beta_j)_j$ denotes the claims development pattern. If $C_{i,j}$ are cumulative payments, then β_j is the expected cumulative cashflow pattern (also called payout pattern). Such a pattern is often used, when one needs to build market-consistent/discounted reserves, where time values differ over time (see also Subsection 1.1.2 on inflation).
From this discussion we see that Model Assumptions 2.8 imply the following model assumptions.

Model Assumptions 2.9

- Different accident years i are independent.
- There exist parameters $\mu_0, \ldots, \mu_I > 0$ and a pattern $\beta_0, \ldots, \beta_J > 0$ with $\beta_J = 1$ such that for all $i \in \{0, \ldots, I\}$ and $j \in \{0, \ldots, J-1\}$

$$E[C_{i,j}] = \mu_i \cdot \beta_j. \tag{2.15}$$

□

Often the Bornhuetter-Ferguson method is explained with the help of Model Assumptions 2.9 (see e.g. Radtke-Schmidt [63], pages 37ff.). However, with Model Assumptions 2.9 we face some difficulties: Observe that

$$\begin{aligned} E\left[C_{i,J}\right|\mathcal{D}_I\right] &= E[C_{i,J}|C_{i,0},\ldots,C_{i,I-i}] \\ &= C_{i,I-i} + E\left[C_{i,J}-C_{i,I-i}\right|C_{i,0},\ldots,C_{i,I-i}\right]. \end{aligned} \tag{2.16}$$

If we have no additional assumptions, we do not exactly know, what we should do with this last term. If we would know that the incremental payment $C_{i,J} - C_{i,I-i}$ is independent from $C_{i,0}, \ldots, C_{i,I-i}$ then this would imply that

$$\begin{aligned} E\left[C_{i,J}\right|\mathcal{D}_I\right] &= E[C_{i,J}|C_{i,0},\ldots,C_{i,I-i}] \\ &= C_{i,I-i} + (1-\beta_{I-i})\cdot\mu_i, \end{aligned} \tag{2.17}$$

which also comes out of Model Assumptions 2.8.
In both model assumptions it remains to estimate the last term in (2.16)-(2.17). In the Bornhuetter-Ferguson method this is done as follows

Estimator 2.10 (Bornhuetter-Ferguson estimator) *The BF estimator is given by*

$$\widehat{C_{i,J}}^{BF} = \widehat{E}\left[C_{i,J}\right|\mathcal{D}_I\right] = C_{i,I-i} + \left(1-\widehat{\beta}_{I-i}\right)\cdot\widehat{\mu}_i \tag{2.18}$$

for $I-J+1 \le i \le I$, where $\widehat{\beta}_{I-i}$ is an estimate for β_{I-i} and $\widehat{\mu}_i$ is an a priori estimate for $E[C_{i,J}]$.

Comparison of Bornhuetter-Ferguson and chain-ladder estimator. From the Chain-ladder Assumptions 2.1 we have that

$$E[C_{i,j}] = E\left[E\left[C_{i,j}|C_{i,j-1}\right]\right] = f_{j-1}\cdot E[C_{i,j-1}] = E[C_{i,0}]\cdot\prod_{k=0}^{j-1} f_k, \tag{2.19}$$

$$E[C_{i,J}] = E[C_{i,0}]\cdot\prod_{k=0}^{J-1} f_k, \tag{2.20}$$

which implies

$$E[C_{i,j}] = \prod_{k=j}^{J-1} f_k^{-1}\cdot E[C_{i,J}]. \tag{2.21}$$

If we compare this to the Bornhuetter-Ferguson model (Model Assumptions 2.9) $E\left[C_{i,j}\right] = \beta_j\cdot\mu_i$ we find that

$$\prod_{k=j}^{J-1} f_k^{-1} \quad \text{"plays the role of"} \quad \beta_j, \tag{2.22}$$

since $\prod_{k=j}^{J-1} f_k^{-1}$ describes the proportion already paid from $\mu_i = E[C_{i,J}]$ after j development periods in the chain-ladder model. Therefore the two variables in (2.22) are often identified: this can be done with Model Assumptions 2.9, but not with Model Assumptions 2.8 (since Model Assumptions 2.8 are not implied by the chain-ladder assumptions nor vica versa). I.e. if one knows the chain-ladder factors f_j one constructs a development pattern β_k using the identity in (2.22) and vice-versa. Then the Bornhuetter-Ferguson estimator can be rewritten as follows

$$\widehat{C_{i,J}}^{BF} = C_{i,I-i} + \left(1 - \left(\widehat{\frac{1}{\prod_{j=I-i}^{J-1} f_j}}\right)\right) \cdot \widehat{\mu}_i. \tag{2.23}$$

On the other hand we have for the chain-ladder estimator that

$$\begin{aligned}
\widehat{C_{i,J}}^{CL} &= C_{i,I-i} \cdot \prod_{j=I-i}^{J-1} \widehat{f}_j \\
&= C_{i,I-i} + C_{i,I-i} \cdot \left(\prod_{j=I-i}^{J-1} \widehat{f}_j - 1\right) \\
&= C_{i,I-i} + \frac{\widehat{C_{i,J}}^{CL}}{\prod_{j=I-i}^{J-1} \widehat{f}_j} \cdot \left(\prod_{j=I-i}^{J-1} \widehat{f}_j - 1\right) \\
&= C_{i,I-i} + \left(1 - \frac{1}{\prod_{j=I-i}^{J-1} \widehat{f}_j}\right) \cdot \widehat{C_{i,J}}^{CL}. \qquad (2.24)
\end{aligned}$$

Hence the difference between the Bornhuetter-Ferguson method and the chain-ladder method is that for the Bornhuetter-Ferguson method we completely believe into our a priori estimate $\widehat{\mu}_i$, whereas in the chain-ladder method the a priori estimate is replaced by an estimate $\widehat{C_{i,J}}^{CL}$ which comes completely from the observations.

Parameter estimation.

- For μ_i we need an a priori estimate $\widehat{\mu}_i$. This is often a plan value from a strategic business plan. This value is estimated before one has any observations, i.e. it is a pure a priori estimate.

- For the still-to-come factor $(1-\beta_{I-i})$ one should also use an a priori estimate if one applies strictly the Bornhuetter-Ferguson method. This should be done independently from the observations. In most practical applications here one quits the path of the pure Bornhuetter-Ferguson method and one estimates the still-to-come factor from the data with the chain-ladder estimators: If $\widehat{f}_k$

denote the chain-ladder estimators (2.7) (see also (2.22)), then we set

$$\widehat{\beta}_j^{(CL)} = \widehat{\beta}_j = \widehat{\left(\frac{1}{\prod_{k=j}^{J-1} f_k} \right)} = \prod_{k=j}^{J-1} \frac{1}{\widehat{f}_k}. \tag{2.25}$$

In that case the Bornhuetter-Ferguson method and the chain-ladder method differ only in the choice of the estimator for the ultimate claim, i.e. a priori estimate vs. chain-ladder estimate (see (2.23) and (2.24)).

Example 2.11 (Bornhuetter-Ferguson method)

We revisit the example given in Table 2.2 (see Example 2.7).

	a priori estimate $\widehat{\mu}_i$	$\widehat{\beta}_{I-i}^{(CL)}$	estimator $\widehat{C_{i,J}}^{BF}$	estimator $\widehat{C_{i,J}}^{CL}$		BF reserves	CL reserves
0	11'653'101	100.0%	11'148'124	11'148'124			
1	11'367'306	99.9%	10'664'316	10'663'318		16'124	15'126
2	10'962'965	99.8%	10'662'749	10'662'008		26'998	26'257
3	10'616'762	99.6%	9'761'643	9'758'606		37'575	34'538
4	11'044'881	99.1%	9'882'350	9'872'218		95'434	85'302
5	11'480'700	98.4%	10'113'777	10'092'247		178'024	156'494
6	11'413'572	97.0%	9'623'328	9'568'143		341'305	286'121
7	11'126'527	94.8%	8'830'301	8'705'378		574'089	449'167
8	10'986'548	88.0%	8'967'375	8'691'971		1'318'646	1'043'242
9	11'618'437	59.0%	10'443'953	9'626'383		4'768'384	3'950'815
					Total	7'356'580	6'047'061

Table 2.4: Claims reserves from the Bornhuetter-Ferguson method and the chain-ladder method

We already see in this example, that using different methods can lead to substantial differences in the claims reserves.

2.3 Number of IBNyR claims, Poisson model

We close this chapter with the Poisson model, which is mainly used for claim counts. The remarkable thing in the Poisson model is, that it leads to the same reserves as the chain-ladder model (see Lemma 2.16). It was Mack [48], Appendix A, who has first proved that the chain-ladder reserves as maximum likelihood reserves for the Poisson model.

Model Assumptions 2.12 (Poisson model)

There exist parameters $\mu_0, \ldots, \mu_I > 0$ and $\gamma_0, \ldots, \gamma_J > 0$ such that the incremental quantities $X_{i,j}$ are independent Poisson distributed with

$$E[X_{i,j}] = \mu_i \cdot \gamma_j, \tag{2.26}$$

for all $i \le I$ and $j \le J$, and $\sum_{j=0}^{J} \gamma_j = 1$.

□

For the definition of the Poisson distribution we refer to the appendix, Section B.1.2.

The cumulative quantity in accident year i, $C_{i,J}$, is again Poisson-distributed with

$$E[C_{i,J}] = \mu_i. \tag{2.27}$$

Hence, μ_i is a parameter that stands for the expected number of claims in accident year i (exposure), whereas γ_j defines an expected reporting/cashflow pattern over the different development periods j. Moreover we have

$$\frac{E[X_{i,j}]}{E[X_{i,0}]} = \frac{\gamma_j}{\gamma_0}, \tag{2.28}$$

which is independent of i.

Lemma 2.13 *The Poisson model satisfies the Model Assumptions 2.8.*

Proof. The independence of different accident years follows from the independence of $X_{i,j}$. Moreover, we have that $E\left[C_{i,0}\right] = E\left[X_{i,0}\right] = \mu_i \cdot \beta_0$ with $\beta_0 = \gamma_0$ and

$$E\left[C_{i,j+k} \middle| C_{i,0}, \ldots, C_{i,j}\right] = C_{i,j} + \sum_{l=1}^{k} E\left[X_{i,j+l} \middle| C_{i,0}, \ldots, C_{i,j}\right] \tag{2.29}$$

$$= C_{i,j} + \mu_i \cdot \sum_{l=1}^{k} \gamma_{j+l} \;=\; C_{i,j} + \mu_i \cdot \left(\beta_{j+k} - \beta_j\right),$$

with $\beta_j = \sum_{l=0}^{j} \gamma_j$. This finishes the proof.

□

To estimate the parameters $(\mu_i)_i$ and $(\gamma_j)_j$ there are different methods, one possibility is to use the maximum likelihood estimators: The likelihood function for $\mathcal{D}_I = \{C_{i,j};\; i+j \le I, j \le J\}$, the σ-algebra generated by $\mathcal{D}_I$ is the same as the one generated by $\{X_{i,j};\; i+j \le I, j \le J\}$, is given by

$$L_{\mathcal{D}_I}(\mu_0, \ldots, \mu_I, \gamma_0, \ldots, \gamma_J) = \prod_{i+j \le I} \left(e^{-\mu_i \gamma_j} \cdot \frac{(\mu_i \gamma_j)^{X_{i,j}}}{X_{i,j}!} \right). \tag{2.30}$$

We maximize this log-Likelihood function by setting its $I+J+2$ partial derivatives w.r.t. the unknown parameters μ_j and γ_j equal to zero. Thus, we obtain on $\mathcal{D}_I$ that

$$\sum_{j=0}^{(I-i)\wedge J} \widehat{\mu}_i \cdot \widehat{\gamma}_j \;=\; \sum_{j=0}^{(I-i)\wedge J} X_{i,j} \;=\; C_{i,(I-i)\wedge J}, \tag{2.31}$$

$$\sum_{i=0}^{I-j} \widehat{\mu}_i \cdot \widehat{\gamma}_j \;=\; \sum_{i=0}^{I-j} X_{i,j}, \tag{2.32}$$

for all $i \in \{0, \dots, I\}$ and all $j \in \{0, \dots, J\}$ under the constraint that $\sum \widehat{\gamma}_j = 1$. This system has a unique solution and gives us the ML estimates for μ_i and γ_j.

Estimator 2.14 (Poisson ML estimator) *The ML estimator in the Poisson Model 2.12 is for $i + j > I$ given by*

$$\widehat{X_{i,j}}^{Poi} = \widehat{E}[X_{i,j}] = \widehat{\mu}_i \cdot \widehat{\gamma}_j, \tag{2.33}$$

$$\widehat{C_{i,J}}^{Poi} = \widehat{E}\left[C_{i,J} \middle| \mathcal{D}_I\right] = C_{i,I-i} + \sum_{j=I-i+1}^{J} \widehat{X_{i,j}}^{Poi}. \tag{2.34}$$

Observe that

$$\widehat{C_{i,J}}^{Poi} = C_{i,I-i} + \left(1 - \sum_{j=0}^{I-i} \widehat{\gamma}_j\right) \cdot \widehat{\mu}_i, \tag{2.35}$$

hence the Poisson ML estimator has the same form as the BF Estimator 2.10. However, here we use estimates for μ_i and γ_j that depend on the data.

Example 2.15 (Poisson ML estimator)

We revisit the example given in Table 2.2 (see Example 2.7).

	0	1	2	3	4	5	6	7	8	9
0	5'946'975	3'721'237	895'717	207'760	206'704	62'124	65'813	14'850	11'130	15'813
1	6'346'756	3'246'406	723'222	151'797	67'824	36'603	52'752	11'186	11'646	
2	6'269'090	2'976'223	847'053	262'768	152'703	65'444	53'545	8'924		
3	5'863'015	2'683'224	722'532	190'653	132'976	88'340	43'329			
4	5'778'885	2'745'229	653'894	273'395	230'288	105'224				
5	6'184'793	2'828'338	572'765	244'899	104'957					
6	5'600'184	2'893'207	563'114	225'517						
7	5'288'066	2'440'103	528'043							
8	5'290'793	2'357'936								
9	5'675'568									

Table 2.5: Observed historical incremental payments $X_{i,j}$

	0	1	2	3	4	5	6	7	8	9	$\widehat{\mu}_i$	estimated reserves
0											11'148'124	
1										15126	10'663'318	15'126
2									11133	15124	10'662'008	26'257
3								10506	10190	13842	9'758'606	34'538
4							50361	10628	10308	14004	9'872'218	85'302
5						69291	51484	10865	10538	14316	10'092'247	156'494
6					137754	65693	48810	10301	9991	13572	9'568'143	286'121
7				188846	125332	59769	44409	9372	9090	12348	8'705'378	449'167
8			594767	188555	125139	59677	44341	9358	9076	12329	8'691'972	1'043'242
9		2795422	658707	208825	138592	66093	49107	10364	10052	13655	9'626'383	3'950'815
$\widehat{\gamma}_j$	58.96%	29.04%	6.84%	2.17%	1.44%	0.69%	0.51%	0.11%	0.10%	0.14%		6'047'061

Table 2.6: Estimated $\widehat{\mu}_i$, $\widehat{\gamma}_j$, incremental payments $\widehat{X_{i,j}}^{Poi}$ and Poisson reserves

Remark. The expected reserve is the same as in the chain-ladder model on cumulative data (see Lemma 2.16 below).

2.3.1 Poisson derivation of the chain-ladder model

I this subsection we show that the Poisson model (Section 2.3) leads to the chain-ladder estimate for the reserves.

Lemma 2.16 *The Chain-ladder Estimate 2.4 and the ML Estimate 2.14 in the Poisson model 2.12 lead to the same reserve.*

In fact the Poisson ML model/estimate defined in Section 2.3 leads to a chain-ladder model (see formula (2.39)), moreover the ML estimators lead to estimators for the age-to-age factors which are the same as in the distribution-free chain-ladder model.

Proof. In the Poisson model 2.12 the estimate for $E\left[C_{i,j}|C_{i,j-1}\right]$ is given by

$$\widehat{\mu}_i \cdot \widehat{\gamma}_j + C_{i,j-1}. \tag{2.36}$$

If we iterate this procedure we obtain for $i > I - J$

$$\begin{aligned} \widehat{C_{i,J}}^{Poi} &= \widehat{E}\left[C_{i,J}|\mathcal{D}_I\right] = \widehat{\mu}_i \cdot \sum_{j=j^*(i)+1}^{J} \widehat{\gamma}_j + C_{i,I-i} \\ &= \widehat{\mu}_i \cdot \sum_{j=j^*(i)+1}^{J} \widehat{\gamma}_j + \sum_{j=0}^{j^*(i)} X_{i,j} = \widehat{\mu}_i \cdot \sum_{j=0}^{J} \widehat{\gamma}_j, \end{aligned} \tag{2.37}$$

where in the last step we have used (2.31). Using (2.31) once more we find that

$$\widehat{C_{i,J}}^{Poi} = \widehat{E}\left[C_{i,J}|\mathcal{D}_I\right] = C_{i,I-i} \cdot \frac{\sum_{j=0}^{J} \widehat{\gamma}_j}{\sum_{j=0}^{j^*(i)} \widehat{\gamma}_j}. \tag{2.38}$$

This last formula can be rewritten introducing additional factors

$$\begin{aligned} \widehat{C_{i,J}}^{Poi} &= C_{i,I-i} \cdot \frac{\sum_{j=0}^{J} \widehat{\gamma}_j}{\sum_{j=0}^{j^*(i)} \widehat{\gamma}_j} \\ &= C_{i,I-i} \cdot \frac{\sum_{j=0}^{j^*(i)+1} \widehat{\gamma}_j}{\sum_{j=0}^{j^*(i)} \widehat{\gamma}_j} \cdot \ldots \cdot \frac{\sum_{j=0}^{J} \widehat{\gamma}_j}{\sum_{j=0}^{J-1} \widehat{\gamma}_j}. \end{aligned} \tag{2.39}$$

If we use Lemma 2.17 below we see that on $\mathcal{D}_I$ we have that

$$\sum_{i=0}^{I-j} C_{i,j\wedge J} = \sum_{i=0}^{I-j} \widehat{\mu}_i \cdot \sum_{k=0}^{j\wedge J} \widehat{\gamma}_k. \tag{2.40}$$

Moreover using (2.32)

$$\sum_{i=0}^{I-j} C_{i,(j-1)\wedge J} = \sum_{i=0}^{I-j} \left(C_{i,j\wedge J} - X_{i,j} \cdot 1_{\{j \le J\}} \right) = \sum_{i=0}^{I-j} \widehat{\mu}_i \cdot \sum_{k=0}^{(j-1)\wedge J} \widehat{\gamma}_k. \tag{2.41}$$

But (2.40)-(2.41) immediately imply for $j \le J$ that

$$\frac{\sum_{k=0}^{j} \widehat{\gamma}_k}{\sum_{k=0}^{j-1} \widehat{\gamma}_k} = \frac{\sum_{i=0}^{I-j} C_{i,j}}{\sum_{i=0}^{I-j} C_{i,j-1}} = \widehat{f}_{j-1}. \tag{2.42}$$

Hence from (2.39) we obtain

$$\begin{aligned} \widehat{C_{i,J}}^{Poi} &= C_{i,I-i} \cdot \frac{\sum_{k=0}^{I-(j^*(i)+1)} C_{k,j^*(i)+1}}{\sum_{k=0}^{I-(j^*(i)+1)} C_{k,j^*(i)}} \cdot \ldots \cdot \frac{\sum_{k=0}^{I-J} C_{k,J}}{\sum_{k=0}^{I-J} C_{k,J-1}} \\ &= C_{i,I-i} \cdot \widehat{f}_{I-i} \cdots \widehat{f}_{J-1} = \widehat{C_{i,J}}^{CL}, \end{aligned} \tag{2.43}$$

which is the chain-ladder estimate (2.8). This finishes the proof of Lemma 2.16.

□

Lemma 2.17 *Under Model Assumptions 2.12 we have on $\mathcal{D}_I$ that*

$$\sum_{i=0}^{I-j} C_{i,j\wedge J} = \sum_{i=0}^{I-j} \widehat{\mu}_i \cdot \sum_{k=0}^{j\wedge J} \widehat{\gamma}_k. \tag{2.44}$$

Proof. We proof this by induction. Using (2.31) for $i = 0$ we have that

$$C_{0,I\wedge J} = \sum_{j=0}^{I\wedge J} X_{0,j} = \widehat{\mu}_0 \cdot \sum_{j=0}^{I\wedge J} \widehat{\gamma}_j, \tag{2.45}$$

which is the starting point of our induction ($j = I$). Induction step $j \to j-1$ (using (2.31)-(2.32)): In the last step we use the induction assuption, then

$$\begin{aligned} \sum_{i=0}^{I-(j-1)} C_{i,(j-1)\wedge J} &= \sum_{i=0}^{I-(j-1)} \left(C_{i,j\wedge J} + C_{i,(j-1)\wedge J} - C_{i,j\wedge J} \right) \qquad (2.46) \\ &= \sum_{i=0}^{I-j} C_{i,j\wedge J} - \sum_{i=0}^{I-(j-1)} X_{i,j} \cdot 1_{\{j\le J\}} + C_{I-j+1,j\wedge J} \\ &= \sum_{i=0}^{I-j} C_{i,j\wedge J} - \sum_{i=0}^{I-j} X_{i,j} \cdot 1_{\{j\le J\}} - X_{I-j+1,j} \cdot 1_{\{j\le J\}} + \sum_{k=0}^{j\wedge J} X_{I-j+1,k} \\ &= \sum_{i=0}^{I-j} C_{i,j\wedge J} - \sum_{i=0}^{I-j} X_{i,j} \cdot 1_{\{j\le J\}} + \sum_{k=0}^{(j-1)\wedge J} X_{I-j+1,k} \\ &= \sum_{i=0}^{I-j} \widehat{\mu}_i \cdot \sum_{k=0}^{j\wedge J} \widehat{\gamma}_k - \widehat{\gamma}_j \cdot 1_{\{j\le J\}} \cdot \sum_{i=0}^{I-j} \widehat{\mu}_i + \widehat{\mu}_{I-j+1} \cdot \sum_{k=0}^{(j-1)\wedge J} \widehat{\gamma}_k. \end{aligned}$$

Hence we find that

$$\begin{aligned}\sum_{i=0}^{I-(j-1)} C_{i,(j-1)\wedge J} &= \sum_{i=0}^{I-j} \widehat{\mu}_i \cdot \sum_{k=0}^{(j-1)\wedge J} \widehat{\gamma}_k + \widehat{\mu}_{I-j+1} \cdot \sum_{k=0}^{(j-1)\wedge J} \widehat{\gamma}_k \\ &= \sum_{i=0}^{I-(j-1)} \widehat{\mu}_i \cdot \sum_{k=0}^{(j-1)\wedge J} \widehat{\gamma}_k, \end{aligned} \tag{2.47}$$

which proves the claim (2.44).

□

Corollary 2.18 *Under Model Assumptions 2.12 we have for all $j \in \{0, \ldots, J\}$ that (see also* (2.25)*)*

$$\sum_{k=0}^{j} \widehat{\gamma}_k = \widehat{\beta}_j^{(CL)} = \prod_{k=j}^{J-1} \frac{1}{\widehat{f}_k}. \tag{2.48}$$

Proof. From (2.38) and (2.43) we obtain for all $i \geq I - J$

$$C_{i,I-i} \cdot \frac{\sum_{j=0}^{J} \widehat{\gamma}_j}{\sum_{j=0}^{I-i} \widehat{\gamma}_j} = \widehat{C_{i,J}}^{Poi} = \widehat{C_{i,J}}^{CL} = C_{i,I-i} \cdot \widehat{f}_{I-i} \cdot \ldots \cdot \widehat{f}_{J-1}. \tag{2.49}$$

Since $\sum \widehat{\gamma}_j = 1$ is normalized we obtain that

$$1 = \sum_{j=0}^{I-i} \widehat{\gamma}_j \cdot \prod_{j=I-i}^{J-1} \widehat{f}_j = \sum_{j=0}^{I-i} \widehat{\gamma}_j \cdot \left(\widehat{\beta}_{I-i}^{(CL)}\right)^{-1}, \tag{2.50}$$

which proves the claim.

□

Remarks 2.19

- Corollary 2.18 says that the chain-ladder method and the Poisson model method lead to the same cash-flow pattern $\widehat{\beta}_j^{(CL)}$ (and hence to the same Bornhuetter-Ferguson reserve if we use this cash-flow pattern for the estimate of β_j). Henceforth, if we use the cash-flow pattern $\widehat{\beta}_j^{(CL)}$ for the BF method, the BF method and the Poisson model only differ in the choice of the expected ultimate claim μ_i, since with (2.35) we obtain that

$$\widehat{C_{i,J}}^{Poi} = C_{i,I-i} + \left(1 - \widehat{\beta}_{I-i}^{(CL)}\right) \cdot \widehat{\mu}_i, \tag{2.51}$$

  where $\widehat{\mu}_i$ is the ML estimate given in (2.31)-(2.32).

- Observe that we have to solve a system of linear equations (2.31)-(2.32) to obtain the ML estimates $\widehat{\mu}_i$ and $\widehat{\gamma}_j$. This solution can easily be obtained with the help of the chain-ladder factors $\widehat{f}_j$ (see Corollary 2.18), namely

$$\widehat{\gamma}_l = \widehat{\beta}_l^{(CL)} - \widehat{\beta}_{l-1}^{(CL)} = \prod_{k=l}^{J-1} \frac{1}{\widehat{f}_k} \cdot \left(1 - 1/\widehat{f}_{l-1}\right), \tag{2.52}$$

and

$$\widehat{\mu}_i = \sum_{j=0}^{(I-i)\wedge J} X_{i,j} \Big/ \sum_{j=0}^{(I-i)\wedge J} \widehat{\gamma}_j. \tag{2.53}$$

Below we will see other ML methods and GLM models where the solution of the equations is more complicated, and where one applies algorithmic methods to find numerical solutions.

Chapter 3

Chain-ladder models

3.1 Mean square error of prediction

In the previous section we have only given an estimate for the mean/expected ultimate claim, of course we would also like to know, how good this estimate is. To measure the quality of the estimate we consider second moments.

Assume that we have a random variable X and a set of observations $\mathcal{D}$. Assume that $\widehat{X}$ is a $\mathcal{D}$-measurable estimator for $E[X|\mathcal{D}]$.

Definition 3.1 (Conditional Mean Square Error of Prediction) *The conditional mean square error of prediction of the estimator $\widehat{X}$ is defined by*

$$msep_{X|\mathcal{D}}(\widehat{X}) = E\left[\left(\widehat{X} - X\right)^2 \middle| \mathcal{D}\right]. \tag{3.1}$$

For a $\mathcal{D}$-measurable estimator $\widehat{X}$ we have

$$\mathrm{msep}_{X|\mathcal{D}}(\widehat{X}) = \mathrm{Var}\left(X|\,\mathcal{D}\right) + \left(\widehat{X} - E\left[X|\mathcal{D}\right]\right)^2. \tag{3.2}$$

The first term on the right-hand side of (3.2) is the so-called process variance (stochastic error), i.e. the variance which is within the stochastic model (pure randomness which can not be eliminated). The second term on the right-hand side of (3.2) is the parameter/estimation error. It reflects the uncertainty in the estimation of the parameters and of the expectation, respectively. In general, this estimation error becomes smaller the more observations we have. But pay attention: In many practical situations it does not completely disappear, since we try to predict future values with the help of past information, already a slight change in the model over time causes lots of problems (this is also discussed below in Section 3.3).

For the estimation error we would like to explicitly calculate the last term in (3.2). However, this can only be done if $E\left[X|\mathcal{D}\right]$ is known, but of course this term is in general not known (we estimate it with the help of $\widehat{X}$). Therefore, the derivation of an estimate for the parameter error is more sophisticated: One is interested into the quality of the estimate $\widehat{X}$, therefore one studies the possible fluctuations of $\widehat{X}$ around $E[X|\mathcal{D}]$.

- **Case 1.** We assume that X is independent of $\mathcal{D}$. This is e.g. the case if we have i.i.d. experiments where we want to estimate its average outcome. In that case we have that

$$E[X|\mathcal{D}] = E[X] \quad \text{and} \quad \mathrm{Var}(X|\mathcal{D}) = \mathrm{Var}(X). \tag{3.3}$$

If we consider the unconditional mean square error of prediction for the estimator $\widehat{X}$ we obtain

$$\mathrm{msep}_X(\widehat{X}) = E\left[\mathrm{msep}_{X|\mathcal{D}}(\widehat{X})\right] = \mathrm{Var}(X) + E\left[\left(\widehat{X} - E[X]\right)^2\right], \tag{3.4}$$

and if $\widehat{X}$ is an unbiased estimator for $E[X]$, i.e. $E\big[\widehat{X}\big] = E[X]$, we have

$$\mathrm{msep}_X(\widehat{X}) = E\left[\mathrm{msep}_{X|\mathcal{D}}(\widehat{X})\right] = \mathrm{Var}(X) + \mathrm{Var}\big(\widehat{X}\big). \tag{3.5}$$

Hence the parameter error is estimated by the variance of $\widehat{X}$.

Example. Assume X and $X_1, \ldots, X_n$ are i.i.d. with mean μ and variance $\sigma^2 < \infty$. Then we have for the estimator $\widehat{X} = \sum_{i=1}^n X_i/n$ that

$$\mathrm{msep}_{X|\mathcal{D}}(\widehat{X}) = \sigma^2 + \left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right)^2. \tag{3.6}$$

By the strong law of large numbers we know that the last term disappears a.s. for $n \to \infty$. In order to determine this term for finite n, one would like to explicitly calculate the distance between $\sum_{i=1}^n X_i/n$ and μ. However, since in general μ is not known, we can only give an estimate for that distance. If we calculate the unconditional mean square error of prediction we obtain

$$\mathrm{msep}_X(\widehat{X}) = \sigma^2 + \sigma^2/n. \tag{3.7}$$

Henceforth, we can say that the deviation of $\sum_{i=1}^n X_i/n$ around μ is in the average of order $\sigma/\sqrt{n}$. But unfortunately this doesn't tell anything about the estimation error for a specific realisation of $\sum_{i=1}^n X_i/n$. We will further discuss this below.

- **Case 2.** X is not independent of the observations $\mathcal{D}$. We have several time series examples below, where we do not have independence between different observations, e.g. in the distribution free version of the chain-ladder method.

In all these cases the situation is even more complicated. Observe that if we calculate the unconditional mean square error of prediction we obtain

$$\begin{aligned} \mathrm{msep}_X(\widehat{X}) &= E\left[\mathrm{msep}_{X|\mathcal{D}}(\widehat{X})\right] \qquad (3.8)\\ &= E\left[\mathrm{Var}\left(X|\,\mathcal{D}\right)\right] + E\left[\left(\widehat{X} - E\left[X|\,\mathcal{D}\right]\right)^2\right] \\ &= \mathrm{Var}\left(X\right) - \mathrm{Var}\left(E\left[X|\,\mathcal{D}\right]\right) + E\left[\left(\widehat{X} - E\left[X|\,\mathcal{D}\right]\right)^2\right] \\ &= \mathrm{Var}\left(X\right) + E\left[\left(\widehat{X} - E\left[X\right]\right)^2\right] \\ &\quad -2 \cdot E\left[\left(\widehat{X} - E\left[X\right]\right) \cdot \left(E\left[X|\,\mathcal{D}\right] - E\left[X\right]\right)\right]. \end{aligned}$$

If the estimator $\widehat{X}$ is unbiased for $E[X]$ we obtain

$$\mathrm{msep}_X(\widehat{X}) = \mathrm{Var}\left(X\right) + \mathrm{Var}\left(\widehat{X}\right) - 2 \cdot \mathrm{Cov}\left(\widehat{X}, E\left[X|\,\mathcal{D}\right]\right). \qquad (3.9)$$

This again tells something on the average estimation error but it doesn't tell anything on the quality of the estimate $\widehat{X}$ for a specific realization.

3.2 Chain-ladder method

We have already described the chain-ladder method in Subsection 2.1. The chain-ladder method can be applied to cumulative payments, to claims incurred, etc. It is the method which is most commonly applied because it is very simple, and often using appropriate estimates for the chain-ladder factors, one obtains reliable claims reserves.

The main deficiencies of the chain-ladder method are

- The homogeneity property need to be satisfied, e.g. we should not have any trends in the development factors (otherwise we have to transform our observations).

- For estimating old development factors (f_j for large j) there is only very little data available, which is maybe (in practice) no longer representative for younger accident years. E.g. assume that we have a claims development with $J = 20$ (years), and that $I = 2006$. Hence we estimate with today's information (accident years < 2006) what will happen with accident year 2006 in 20 years.

- For young accident years, very much weight is given to the observations, i.e. if we have an outlier on the diagonal, this outlier is projected right to the ultimate claim, which is not always appropriate. Therefore for younger accident years sometimes the Bornhuetter-Ferguson method is preferred (see also discussion in Subsection 4.2.4).

- In long-tailed branches/LoB the difference between the chain-ladder method on cumulative payments and claims incurred is often very large. This is mainly due to the fact that the homogeneity property is not fulfilled. Indeed, if we have new phenomenons in the data, usually claims incurred methods overestimates such effects, whereas estimates on paid data underestimate the effects since we only observe the new behavior over time. This is mainly due to the effect that the claims adjusters usually overestimate new phenomenons (which is reflected in the claims incurred figures), whereas in claims paid figures one observes new phenomenons only over time (when a claim is settled via payments).

- There is an extensive list of references on how the chain-ladder method should be applied in practice and where future research projects could be settled. We do not further discuss this here but only give two references [46] and [40] which refer to such questions. Moreover, we would mention that there is also literature on the appropriatness of the chain-ladder method for specific data sets, see e.g. Barnett-Zehnwirth [7] and Venter [77].

3.2.1 The Mack model

We define the chain-ladder model once more, but this time we extend the definition to the second moments, so that we are also able to give an estimate for the conditional mean square error of prediction for the chain-ladder estimator.
In the actuarial literature, the chain-ladder method is often understood as a purely computational algorithm and leaves the question open which probabilistic model would lead to that algorithm. It is Mack's merit [49] that he has given first an answer to that question (a first decisive step towards the formulas was done by Schnieper [69]).

Model Assumptions 3.2 (Chain-ladder, Mack [49])

- Different accident years i are independent.

- $(C_{i,j})_j$ is a Markov chain with: There exist factors $f_0, \ldots, f_{J-1} > 0$ and variance parameters $\sigma_0^2, \ldots, \sigma_{J-1}^2 > 0$ such that for all $0 \leq i \leq I$ and $1 \leq j \leq$

J we have that

$$E\left[C_{i,j}\middle|\, C_{i,j-1}\right] = f_{j-1} \cdot C_{i,j-1}, \tag{3.10}$$

$$\mathrm{Var}\left(C_{i,j}\middle|\, C_{i,j-1}\right) = \sigma^2_{j-1} \cdot C_{i,j-1}. \tag{3.11}$$

□

Remark. In Mack [49] there are slightly weaker assumptions, namely the Markov chain assumption is replaced by weaker assumptions on the first two moments of $(C_{i,j})_j$.

We recall the results from Section 2.1 (see Lemma 2.5):

- Choose the following estimators for the parameters f_j and σ_j^2:

$$\widehat{f}_j = \frac{\sum_{i=0}^{i^*(j+1)} C_{i,j+1}}{\sum_{i=0}^{i^*(j+1)} C_{i,j}} = \sum_{i=0}^{i^*(j+1)} \frac{C_{i,j}}{\sum_{k=0}^{i^*(j+1)} C_{k,j}} \cdot \frac{C_{i,j+1}}{C_{i,j}},$$

$$\widehat{\sigma}_j^2 = \frac{1}{i^*(j+1)} \cdot \sum_{i=0}^{i^*(j+1)} C_{i,j} \cdot \left(\frac{C_{i,j+1}}{C_{i,j}} - \widehat{f}_j\right)^2. \tag{3.12}$$

- $\widehat{f}_j$ is unconditionally and conditionally, given $\mathcal{B}_j$, unbiased for f_j.
- $\widehat{f}_0, \ldots, \widehat{f}_{J-1}$ are uncorrelated.

If we define the individual development factors by

$$F_{i,j+1} = \frac{C_{i,j+1}}{C_{i,j}}, \tag{3.13}$$

then the age-to-age factor estimates $\widehat{f}_j$ are weighted averages of $F_{i,j+1}$, namely

$$\widehat{f}_j = \sum_{i=0}^{i^*(j+1)} \frac{C_{i,j}}{\sum_{k=0}^{i^*(j+1)} C_{k,j}} \cdot F_{i,j+1}. \tag{3.14}$$

Lemma 3.3 *Under Assumptions 3.2 the estimator $\widehat{f}_j$ is the $\mathcal{B}_{j+1}$-measurable unbiased estimator for f_j, which has minimal conditional variance among all linear combinations of the unbiased estimators $\left(F_{i,j+1}\right)_{0\le i\le i^*(j+1)}$ for f_j, conditioned on $\mathcal{B}_j$, i.e.*

$$Var\left(\widehat{f}_j|\mathcal{B}_j\right) = \min_{\alpha_i\in\mathbb{R}} Var\left(\sum_{i=0}^{i^*(j+1)} \alpha_i \cdot F_{i,j+1}\middle|\, \mathcal{B}_j\right). \tag{3.15}$$

The conditional variance of $\widehat{f}_j$ is given by

$$Var\big(\widehat{f}_j|\mathcal{B}_j\big) = \sigma_j^2 / \sum_{i=0}^{i^*(j+1)} C_{i,j}. \tag{3.16}$$

We need the following lemma to proof the statement:

Lemma 3.4 *Assume that $P_1, \ldots, P_H$ are stochastically independent unbiased estimators for μ with variances $\sigma_1^2, \ldots, \sigma_H^2$. Then the minimum variance unbiased linear combination of the P_h is given by*

$$P = \frac{\sum_{h=1}^{H} (P_h/\sigma_h^2)}{\sum_{h=1}^{H} (1/\sigma_h^2)}, \tag{3.17}$$

with

$$Var(P) = \left(\sum_{h=1}^{H} (1/\sigma_h^2)\right)^{-1}. \tag{3.18}$$

Proof. See Proposition 12.1 in Taylor [75] (the proof is based on the method of Lagrange).

□

Proof of Lemma 3.3. Consider the individual development factors

$$F_{i,j+1} = \frac{C_{i,j+1}}{C_{i,j}}. \tag{3.19}$$

Conditioned on $\mathcal{B}_j$, $F_{i,j+1}$ are unbiased and independent estimators for f_j with

$$\mathrm{Var}(F_{i,j+1}|\mathcal{B}_j) = \mathrm{Var}(F_{i,j+1}|C_{i,j}) = \sigma_j^2/C_{i,j}. \tag{3.20}$$

With Lemma 3.4 the claim immediately follows with

$$\mathrm{Var}(\widehat{f}_j|\mathcal{B}_j) = \sigma_j^2 / \sum_{i=0}^{i^*(j+1)} C_{i,j}. \tag{3.21}$$

□

Lemma 3.5 *Under Assumptions 3.2 we have:*

a) $\widehat{\sigma}_j^2$ is, given $\mathcal{B}_j$, an unbiased estimator for σ_j^2, i.e. $E\left[\widehat{\sigma}_j^2 \middle| \mathcal{B}_j\right] = \sigma_j^2$,

b) $\widehat{\sigma}_j^2$ is (unconditionally) unbiased for σ_j^2, i.e. $E\left[\widehat{\sigma}_j^2\right] = \sigma_j^2$.

Proof. b) easily follows from a). Hence we prove a). Consider

$$E\left[\left(\frac{C_{i,k+1}}{C_{i,k}}-\widehat{f}_k\right)^2\middle|\mathcal{B}_k\right]=E\left[\left(\frac{C_{i,k+1}}{C_{i,k}}-f_k\right)^2\middle|\mathcal{B}_k\right] \tag{3.22}$$
$$-2\cdot E\left[\left(\frac{C_{i,k+1}}{C_{i,k}}-f_k\right)\cdot\left(\widehat{f}_k-f_k\right)\middle|\mathcal{B}_k\right]+E\left[\left(\widehat{f}_k-f_k\right)^2\middle|\mathcal{B}_k\right].$$

Hence we calculate the terms on the r.h.s. of the equality above.

$$E\left[\left(\frac{C_{i,k+1}}{C_{i,k}}-f_k\right)^2\middle|\mathcal{B}_k\right]=\mathrm{Var}\left(\frac{C_{i,k+1}}{C_{i,k}}\middle|\mathcal{B}_k\right)=\frac{1}{C_{i,k}}\cdot\sigma_k^2. \tag{3.23}$$

The next term is (using the independence of different accident years)

$$\begin{aligned} E\left[\left(\frac{C_{i,k+1}}{C_{i,k}}-f_k\right)\cdot\left(\widehat{f}_k-f_k\right)\middle|\mathcal{B}_k\right] &= \mathrm{Cov}\left(\frac{C_{i,k+1}}{C_{i,k}},\widehat{f}_k\middle|\mathcal{B}_k\right) \\ &= \frac{C_{i,k}}{\sum_{i=0}^{i^*(k+1)}C_{i,k}}\cdot\mathrm{Var}\left(\frac{C_{i,k+1}}{C_{i,k}}\middle|\mathcal{B}_k\right) \\ &= \frac{\sigma_k^2}{\sum_{i=0}^{i^*(k+1)}C_{i,k}}. \end{aligned} \tag{3.24}$$

Whereas for the last term we obtain

$$E\left[\left(\widehat{f}_k-f_k\right)^2\middle|\mathcal{B}_k\right]=\mathrm{Var}\left(\widehat{f}_k\middle|\mathcal{B}_k\right)=\frac{\sigma_k^2}{\sum_{i=0}^{i^*(k+1)}C_{i,k}}. \tag{3.25}$$

Putting all this together gives

$$E\left[\left(\frac{C_{i,k+1}}{C_{i,k}}-\widehat{f}_k\right)^2\middle|\mathcal{B}_k\right]=\sigma_k^2\cdot\left(\frac{1}{C_{i,k}}-\frac{1}{\sum_{i=0}^{i^*(k+1)}C_{i,k}}\right). \tag{3.26}$$

Hence we have that

$$E\left[\widehat{\sigma}_k^2\middle|\mathcal{B}_k\right]=\frac{1}{i^*(k+1)}\cdot\sum_{i=0}^{i^*(k+1)}C_{i,k}\cdot E\left[\left(\frac{C_{i,k+1}}{C_{i,k}}-\widehat{f}_k\right)^2\middle|\mathcal{B}_k\right]=\sigma_k^2, \tag{3.27}$$

which proves the claim a). This finishes the proof of Lemma 3.5.

□

The following equality plays an important role in the derivation of an estimator for the conditional estimation error

$$E\left[\widehat{f}_k^2\middle|\mathcal{B}_k\right]=\mathrm{Var}\left(\widehat{f}_k\middle|\mathcal{B}_k\right)+f_k^2=\frac{\sigma_k^2}{\sum_{i=0}^{i^*(k+1)}C_{i,k}}+f_k^2. \tag{3.28}$$

In Estimator 2.4 we have already seen how we estimate the ultimate claim $C_{i,J}$, given the information $\mathcal{D}_I$ in the chain-ladder model:

$$\widehat{C_{i,J}}^{CL} = \widehat{E}\left[C_{i,J} \middle| \mathcal{D}_I \right] = C_{i,I-i} \cdot \widehat{f}_{I-i} \cdots \widehat{f}_{J-1}. \tag{3.29}$$

Our goal is to derive an estimate for the conditional mean square error of prediction (conditional MSEP) of $\widehat{C_{i,J}}^{CL}$ for single accident years $i \in \{I-J+1, \ldots, I\}$

$$\begin{aligned} \mathrm{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}\right) &= E\left[\left(\widehat{C_{i,J}}^{CL} - C_{i,J}\right)^2 \middle| \mathcal{D}_I\right] \\ &= \mathrm{Var}\left(C_{i,J}\middle|\mathcal{D}_I\right) + \left(\widehat{C_{i,J}}^{CL} - E\left[C_{i,J}|\mathcal{D}_I\right]\right)^2, \end{aligned} \tag{3.30}$$

and for aggregated accident years we consider

$$\mathrm{msep}_{\sum_i C_{i,J}|\mathcal{D}_I}\left(\sum_{i=I-J+1}^{I} \widehat{C_{i,J}}^{CL}\right) = E\left[\left(\sum_{i=I-J+1}^{I} \widehat{C_{i,J}}^{CL} - \sum_{i=I-J+1}^{I} C_{i,J}\right)^2 \middle| \mathcal{D}_I\right]. \tag{3.31}$$

From (3.30) we see that we need to give an estimate for the process variance and for the estimation error (coming from the fact that f_j is estimated by $\widehat{f}_j$).

3.2.2 Conditional process variance

We consider the first term on the right-hand side of (3.30), which is the conditional process variance. Assume $J > I - i$,

$$\begin{aligned} \mathrm{Var}\left(C_{i,J}\middle|\mathcal{D}_I\right) &= \mathrm{Var}\left(C_{i,J}\middle|C_{i,I-i}\right) \\ &= E\left[\mathrm{Var}\left(C_{i,J}\middle|C_{i,J-1}\right)\middle|C_{i,I-i}\right] + \mathrm{Var}\left(E\left[C_{i,J}\middle|C_{i,J-1}\right]\middle|C_{i,I-i}\right) \\ &= \sigma_{J-1}^2 \cdot E\left[C_{i,J-1}\middle|C_{i,I-i}\right] + f_{J-1}^2 \cdot \mathrm{Var}\left(C_{i,J-1}\middle|C_{i,I-i}\right) \\ &= \sigma_{J-1}^2 \cdot C_{i,I-i} \cdot \prod_{j=I-i}^{J-2} f_j + f_{J-1}^2 \cdot \mathrm{Var}\left(C_{i,J-1}\middle|C_{i,I-i}\right). \end{aligned} \tag{3.32}$$

Hence we obtain a recursive formula for the process variance. If we iterate this procedure, we find that

$$\begin{aligned} \mathrm{Var}\left(C_{i,J}\middle|C_{i,I-i}\right) &= C_{i,I-i} \cdot \sum_{m=I-i}^{J-1} \prod_{n=m+1}^{J-1} f_n^2 \cdot \sigma_m^2 \cdot \prod_{l=I-i}^{m-1} f_l \\ &= \sum_{m=I-i}^{J-1} \prod_{n=m+1}^{J-1} f_n^2 \cdot \sigma_m^2 \cdot E\left[C_{i,m}\middle|C_{i,I-i}\right] \\ &= \left(E\left[C_{i,J}\middle|C_{i,I-i}\right]\right)^2 \cdot \sum_{m=I-i}^{J-1} \frac{\sigma_m^2/f_m^2}{E\left[C_{i,m}\middle|C_{i,I-i}\right]}. \end{aligned} \tag{3.33}$$

This gives the following Lemma:

Lemma 3.6 (Process variance for single accident years) *Under Model Assumptions 3.2 the conditional process variance for the ultimate claim of a single accident year* $i \in \{I-J+1, \ldots, I\}$ *is given by*

$$Var\left(C_{i,J} \middle| \mathcal{D}_I\right) = \left(E\left[C_{i,J} \middle| C_{i,I-i}\right]\right)^2 \cdot \sum_{m=I-i}^{J-1} \frac{\sigma_m^2 / f_m^2}{E\left[C_{i,m} \middle| C_{i,I-i}\right]}. \tag{3.34}$$

Hence we estimate the conditional process variance for a single accident year i by

$$\begin{aligned} \widehat{\mathrm{Var}}\left(C_{i,J} \middle| \mathcal{D}_I\right) &= \widehat{E}\left[\left(C_{i,J} - E\left[C_{i,J} \middle| \mathcal{D}_I\right]\right)^2 \middle| \mathcal{D}_I\right] \\ &= \left(\widehat{C_{i,J}}^{CL}\right)^2 \cdot \sum_{m=I-i}^{J-1} \frac{\widehat{\sigma}_m^2 / \widehat{f}_m^2}{\widehat{C_{i,m}}^{CL}}. \end{aligned} \tag{3.35}$$

The estimator for the conditional process variance can be rewritten in a recursive form. We obtain for $i \in \{I-J+1, \ldots, I\}$

$$\widehat{\mathrm{Var}}\left(C_{i,J} \middle| \mathcal{D}_I\right) = \widehat{\mathrm{Var}}\left(C_{i,J-1} \middle| \mathcal{D}_I\right) \cdot \widehat{f}_{J-1}^2 + \widehat{\sigma}_{J-1}^2 \cdot \widehat{C_{i,J-1}}^{CL}, \tag{3.36}$$

where $\widehat{\mathrm{Var}}\left(C_{i,I-i} \middle| \mathcal{D}_I\right) = 0$.

Because different accident years are independent, we estimate the conditional process variance for aggregated accident years by

$$\widehat{\mathrm{Var}}\left(\sum_{i=I-J+1}^{I} C_{i,J} \middle| \mathcal{D}_I\right) = \sum_{i=I-J+1}^{I} \widehat{\mathrm{Var}}\left(C_{i,J} \middle| \mathcal{D}_I\right). \tag{3.37}$$

Example 3.7 (Chain-ladder method)

We come back to our example in Table 2.2 (see Example 2.7). Since we do not have enough data (i.e. we don't have $I > J$) we are not able to estimate the last variance parameter σ_{J-1}^2 with the estimator $\widehat{\sigma}_{J-1}^2$ (cf. (3.12)). There is an extensive literature about estimation of tail factors and variance estimates. We do not further discuss this here, but we simply choose the extrapolation chosen by Mack [49]:

$$\widehat{\sigma}_{J-1}^2 = \min\left\{\widehat{\sigma}_{J-2}^4 / \widehat{\sigma}_{J-3}^2 ; \widehat{\sigma}_{J-3}^2 ; \widehat{\sigma}_{J-2}^2\right\} \tag{3.38}$$

as estimate for σ_{J-1}^2. This estimate is motivated by the observation that the series $\sigma_0, \ldots, \sigma_{J-2}$ is usually decreasing (cf. Table 3.1). This gives the estimated conditional process standard deviations in Table 3.2.

We define the estimated conditional variational coefficient for accident year i relative to the estimated CL reserves as follows:

$$\mathrm{Vco}_i = \widehat{\mathrm{Vco}}\left(C_{i,J} \middle| \mathcal{D}_I\right) = \frac{\widehat{\mathrm{Var}}\left(C_{i,J} \middle| \mathcal{D}_I\right)^{1/2}}{\widehat{C_{i,J}}^{CL} - C_{i,I-i}}. \tag{3.39}$$

If we take this variational coefficient as a measure for the uncertainty, we see that the uncertainty of the total CL reserves is about 7%.

	0	1	2	3	4	5	6	7	8
1	1.6257	1.0926	1.0197	1.0192	1.0057	1.0060	1.0013	1.0010	1.0014
2	1.5115	1.0754	1.0147	1.0065	1.0035	1.0050	1.0011	1.0011	
3	1.4747	1.0916	1.0260	1.0147	1.0062	1.0051	1.0008		
4	1.4577	1.0845	1.0206	1.0141	1.0092	1.0045			
5	1.4750	1.0767	1.0298	1.0244	1.0109				
6	1.4573	1.0635	1.0255	1.0107					
7	1.5166	1.0663	1.0249						
8	1.4614	1.0683							
9	1.4457								
10									
$\widehat{f}_j$	1.4925	1.0778	1.0229	1.0148	1.0070	1.0051	1.0011	1.0010	1.0014
$\widehat{\sigma}_j$	135.253	33.803	15.760	19.847	9.336	2.001	0.823	0.219	0.059

Table 3.1: Observed historical individual chain-ladder factors $F_{i,j+1}$, estimated chain-ladder factors $\widehat{f}_j$ and estimated standard deviations $\widehat{\sigma}_j$

i	$C_{i,I-i}$	$\widehat{C_{i,J}}^{CL}$	CL reserves	$\widehat{\text{Var}}\left(C_{i,J}\middle\vert\mathcal{D}_I\right)^{1/2}$	Vco_i
0	11'148'124	11'148'124	0		
1	10'648'192	10'663'318	15'126	191	1.3%
2	10'635'751	10'662'008	26'257	742	2.8%
3	9'724'068	9'758'606	34'538	2'669	7.7%
4	9'786'916	9'872'218	85'302	6'832	8.0%
5	9'935'753	10'092'247	156'494	30'478	19.5%
6	9'282'022	9'568'143	286'121	68'212	23.8%
7	8'256'211	8'705'378	449'167	80'077	17.8%
8	7'648'729	8'691'971	1'043'242	126'960	12.2%
9	5'675'568	9'626'383	3'950'815	389'783	9.9%
Total			6'047'061	424'379	7.0%

Table 3.2: Estimated chain-ladder reserves and estimated conditional process standard deviations

3.2.3 Estimation error for single accident years

Next we need to derive an estimate for the conditional parameter/estimation error, i.e. we want to get an estimate for the accuracy of our chain-ladder factor estimates $\widehat{f}_j$. The parameter error for a single accident year in the chain-ladder estimate is given by (see (3.30), (2.3) and (2.8))

$$\begin{aligned} \left(\widehat{C_{i,J}}^{CL} - E\left[C_{i,J}\middle\vert\,\mathcal{D}_I\right]\right)^2 &= C_{i,I-i}^2 \cdot \left(\widehat{f}_{I-i}\cdot\ldots\cdot\widehat{f}_{J-1} - f_{I-i}\cdot\ldots\cdot f_{J-1}\right)^2 \\ &= C_{i,I-i}^2 \cdot \left(\prod_{j=I-i}^{J-1}\widehat{f}_j^2 + \prod_{j=I-i}^{J-1} f_j^2 - 2\cdot\prod_{j=I-i}^{J-1}\widehat{f}_j\cdot f_j\right). \end{aligned} \tag{3.40}$$

Hence we would like to calculate (3.40). Observe that the realizations of the estimators $\widehat{f}_{I-i},\ldots,\widehat{f}_{J-1}$ are known at time I, but the "true" chain-ladder factors $f_{I-i},\ldots,f_{J-1}$ are unknown. Hence (3.40) can not be calculated explicitly. In order to determine the conditional estimation error we will analyze how much the possible chain-ladder factors $\widehat{f}_j$ fluctuate around f_j. We measure these volatilities of the estimates $\widehat{f}_j$ by means of resampled observations for $\widehat{f}_j$.

There are different approaches to resample these values: conditional ones and unconditional ones, see Buchwalder et al. [13]. For the explanation of these different approaches we fix accident year $i \in \{I-J+1, \ldots, I\}$. Then we see from the right-hand side of (3.40) that the main difficulty in the determination of the volatility in the estimates comes from the calculation of the squares of the estimated chain-ladder factors.
Therefore, we focus for the moment on these squares, i.e. we need to resample the following product of squared estimates

$$\widehat{f}_{I-i}^2 \cdot \ldots \cdot \widehat{f}_{J-1}^2, \tag{3.41}$$

the treatment of the last term in (3.40) is then straightforward.
To be able to distinguish the different resample approaches we define by

$$\mathcal{D}_{I,i}^O = \{C_{i,j} \in \mathcal{D}_I;\ j > I-i\} \subset \mathcal{D}_I \tag{3.42}$$

the upper right corner of the observations $\mathcal{D}_I$ with respect to development year $j = I - i + 1$.

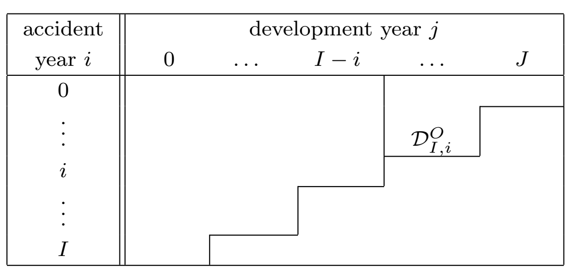


Table 3.3: The upper right corner $\mathcal{D}_{I,i}^O$

For the following explanation observe that $\widehat{f}_j$ is $\mathcal{B}_{j+1}$-measurable.

Approach 1 (Unconditional resampling in $\mathcal{D}_{I,i}^O$). In this approach one calculates the expectation

$$E\left[\left.\widehat{f}_{I-i}^2 \cdot \ldots \cdot \widehat{f}_{J-1}^2 \right| \mathcal{B}_{I-i}\right]. \tag{3.43}$$

This is the complete averaging over the multidimensional distribution after time $I-i$. Since $\mathcal{D}_{I,i}^O \cap \mathcal{B}_{I-i} = \emptyset$ holds true the value (3.43) does not depend on the observations in $\mathcal{D}_{I,i}^O$. I.e. the observed realizations in the upper corner $\mathcal{D}_{I,i}^O$ have no influence on the estimation of the parameter error. Therefore we call this the unconditional version because it gives the average/expected estimation error (independent of the observations in $\mathcal{D}_{I,i}^O$).

Approach 2 (Partial conditional resampling in $\mathcal{D}_{I,i}^O$). In this approach one calculates the value

$$\widehat{f}_{I-i}^2 \cdot \ldots \cdot \widehat{f}_{J-2}^2 \cdot E\left[\left.\widehat{f}_{J-1}^2\right| \mathcal{B}_{J-1}\right]. \tag{3.44}$$

In this version the averaging is only done partially. However, $\mathcal{D}_{I,i}^O \cap \mathcal{B}_{J-1} \neq \emptyset$ holds. I.e. the value (3.44) depends on the observations in $\mathcal{D}_{I,i}^O$. If one decouples the problem of resampling in a smart way, one can even choose the position $j \in \{I-i, \ldots, J-1\}$ at which one wants to do the partial resampling.

Approach 3 (Conditional resampling in $\mathcal{D}_{I,i}^O$). Calculate the value

$$E\left[\left.\widehat{f}_{I-i}^2\right| \mathcal{B}_{I-i}\right] \cdot E\left[\left.\widehat{f}_{I-i+1}^2\right| \mathcal{B}_{I-i+1}\right] \cdot \ldots \cdot E\left[\left.\widehat{f}_{J-1}^2\right| \mathcal{B}_{J-1}\right]. \tag{3.45}$$

Unlike the approach (3.44) the averaging is now done in every position $j \in \{I-i, \ldots, J-1\}$ on the conditional structure. Since $\mathcal{D}_{I,i}^O \cap \mathcal{B}_j \neq \emptyset$ if $j > I-i$ the observed realizations in $\mathcal{D}_{I,i}^O$ have a direct influence on the estimate and (3.45) depends on the observations in $\mathcal{D}_{I,i}^O$. In contrast to (3.43) the averaging is only done over the conditional distributions and not over the multidimensional distribution after $I-i$. Therefore we call this the conditional version. From a numerical point of view it is important to note that Approach 3 allows for a multiplicative structure of the measure of volatility (see Figure 3.1).

Concluding, this means that we consider different probability measures for the resampling, conditional and unconditional ones. Observe that the estimated chain-ladder factors $\widehat{f}_j$ are functions of $(C_{i,j+1})_{i=0,\ldots,I-j-1}$ and $(C_{i,j})_{i=0,\ldots,I-j-1}$, i.e.

$$\widehat{f}_j = \widehat{f}_j\left((C_{i,j+1})_{i=0,\ldots,I-j-1}, (C_{i,j})_{i=0,\ldots,I-j-1}\right) = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}}. \tag{3.46}$$

In the conditional resampling the denominator serves as a fixed volume measure, whereas in the unconditional resampling the denominator is also resampled.
Since our time series $(C_{k,j})_j$ is a Markov chain we can write its probability distribution (with the help of stochastic kernels K_j) as follows:

$$\begin{aligned} dP_k&(x_0, \ldots, x_J) \qquad (3.47)\\ &= K_0(dx_0) \cdot K_1(x_0, dx_1) \cdot K_2(x_0, x_1, dx_2) \cdots K_J(x_0, \ldots, x_{J-1}, dx_J) \\ &= K_0(dx_0) \cdot K_1(x_0, dx_1) \cdot K_2(x_1, dx_2) \cdots K_J(x_{J-1}, dx_J). \end{aligned}$$

In Approach 1 one considers a complete resampling on $\mathcal{D}_{I,i}^{O}$, i.e. one looks, given $\mathcal{B}_{I-i}$, at the measures

$$\begin{aligned} dP\left((x_{k,j})_{k,j} \mid \mathcal{B}_{I-i}\right) &= \prod_{k<i} dP_k\left(x_{k,I-i+1}, \ldots, x_{k,I-k} \middle| C_{k,I-i} = x_{k,I-i}\right) \\ &= \prod_{k<i} K_{I-i+1}\left(x_{k,I-i}, dx_{k,I-i+1}\right) \cdots K_{I-k}\left(x_{k,I-k-1}, dx_{k,I-k}\right), \end{aligned} \tag{3.48}$$

for the resampling of the estimated chain-ladder factors

$$\prod_{j \geq I-i} \widehat{f}_j = \prod_{j \geq I-i} \widehat{f}_j \left((x_{i,j+1})_{i=0,\ldots,I-j-1}, (x_{i,j})_{i=0,\ldots,I-j-1} \right). \tag{3.49}$$

In Approach 3 we always keep fixed the set of actual observations $C_{i,j}$ and we only resample the next step in the time series, i.e. given $\mathcal{D}_I$ we consider the measures (see also Figure 3.1)

$$dP^*_{\mathcal{D}_I}\left((x_{k,j})_{k,j}\right) = \prod_{k<i} K_{I-i+1}\left(C_{k,I-i}, dx_{k,I-i+1}\right) \cdots K_{I-k}\left(C_{k,I-k-1}, dx_{k,I-k}\right), \tag{3.50}$$

for the resampling of

$$\prod_{j \geq I-i} \widehat{f}_j = \prod_{j \geq I-i} \widehat{f}_j \left((x_{i,j+1})_{i=0,\ldots,I-j-1}, (C_{i,j})_{i=0,\ldots,I-j-1} \right). \tag{3.51}$$

Hence in this context $C_{i,j}$ serves as a volume measure for the resampling of $C_{i,j+1}$. In Approach 1 this volume measure is also resampled, whereas in Approach 3 it is kept fixed.

Observe. The question, as to which approach should be chosen, is not a mathematical one and has lead to extensive discussions in the actuarial community (see Buchwalder et al. [11], Mack et al. [52], Gisler [29] and Venter [78]). It depends on the circumstances of the questions as to which approach should be used for a specific practical problem. Henceforth, only the practitioner can choose the appropriate approach for his problems and questions.

Approach 1 (Unconditional resampling)

In the unconditional approach we have (due to the uncorrelatedness of the chain-ladder factors) that

$$\begin{aligned} E&\left[\left.\left(\widehat{C_{i,J}}^{CL}-E\left[C_{i,J}\right|\mathcal{D}_I\right]\right)^2\right|\mathcal{B}_{I-i}\right] \\ &= C_{i,I-i}^2\cdot E\left[\left.\prod_{j=I-i}^{J-1}\widehat{f}_j^2+\prod_{j=I-i}^{J-1}f_j^2-2\cdot\prod_{j=I-i}^{J-1}\widehat{f}_j\cdot f_j\right|\mathcal{B}_{I-i}\right] \\ &= C_{i,I-i}^2\cdot\left(E\left[\left.\prod_{j=I-i}^{J-1}\widehat{f}_j^2\right|\mathcal{B}_{I-i}\right]-\prod_{j=I-i}^{J-1}f_j^2\right). \end{aligned} \tag{3.52}$$

Hence, to give an estimate for the estimation error with the unconditional version, we need to calculate the expectation in the last term of (3.52) (as described in Approach 1). This would be easy, if the estimated chain-ladder factors $\widehat{f}_j$ were independent. But they are **only** uncorrelated, see Lemma 2.5 and the following lemma (for a similar statement see also Mack et al. [52]):

Lemma 3.8 *Under Model Assumptions 3.2 the squares of two successive chain-ladder estimators $\widehat{f}_{j-1}$ and $\widehat{f}_j$ are, given $\mathcal{B}_{j-1}$, negatively correlated, i.e.*

$$Cov\left(\left.\widehat{f}_{j-1}^2,\widehat{f}_j^2\right|\mathcal{B}_{j-1}\right)<0 \tag{3.53}$$

for $1\le j\le J-1$.

Proof. Observe that $\widehat{f}_{j-1}$ is $\mathcal{B}_j$-measurable. We define

$$S_j=\sum_{i=0}^{i^*(j+1)}C_{i,j}. \tag{3.54}$$

Hence, we have that

$$\begin{aligned} \text{Cov}&\left(\left.\widehat{f}_{j-1}^2,\widehat{f}_j^2\right|\mathcal{B}_{j-1}\right) \\ &= E\left[\left.\text{Cov}\left(\left.\widehat{f}_{j-1}^2,\widehat{f}_j^2\right|\mathcal{B}_j\right)\right|\mathcal{B}_{j-1}\right]+\text{Cov}\left(\left.E\left[\left.\widehat{f}_{j-1}^2\right|\mathcal{B}_j\right],E\left[\left.\widehat{f}_j^2\right|\mathcal{B}_j\right]\right|\mathcal{B}_{j-1}\right) \\ &= \text{Cov}\left(\left.\widehat{f}_{j-1}^2,\frac{\sigma_j^2}{S_j}+f_j^2\right|\mathcal{B}_{j-1}\right) \\ &= \frac{\sigma_j^2}{\left(S_{j-1}\right)^2}\cdot\text{Cov}\left(\left.\left(\sum_{i=0}^{i^*(j)}C_{i,j}\right)^2,\frac{1}{S_j}\right|\mathcal{B}_{j-1}\right). \end{aligned} \tag{3.55}$$

Moreover, using

$$\left(\sum_{i=0}^{i^*(j)}C_{i,j}\right)^2=S_j^2+2\cdot S_j\cdot C_{I-j,j}+C_{I-j,j}^2, \tag{3.56}$$

the independence of different accident years and $E\left[C_{I-j,j}\middle|\mathcal{B}_{j-1}\right] = f_{j-1}\cdot C_{I-j,j-1}$ leads to

$$\begin{aligned}
&\mathrm{Cov}\left(\widehat{f}_{j-1}^2, \widehat{f}_j^2\middle|\mathcal{B}_{j-1}\right) \qquad (3.57)\\
&\quad= \frac{\sigma_j^2}{\left(S_{j-1}\right)^2}\cdot\left[\mathrm{Cov}\left(S_j^2, \frac{1}{S_j}\middle|\mathcal{B}_{j-1}\right) + 2\cdot f_{j-1}\cdot C_{I-j,j-1}\cdot\mathrm{Cov}\left(S_j, \frac{1}{S_j}\middle|\mathcal{B}_{j-1}\right)\right].
\end{aligned}$$

Finally, we need to calculate both covariance terms on the right-hand side of (3.57). Using Jensen's inequality we obtain for $\alpha = 1, 2$

$$\begin{aligned}
\mathrm{Cov}\left(S_j^\alpha, \frac{1}{S_j}\middle|\mathcal{B}_{j-1}\right) &= E\left[S_j^{\alpha-1}\middle|\mathcal{B}_{j-1}\right] - E\left[S_j^\alpha\middle|\mathcal{B}_{j-1}\right]\cdot E\left[S_j^{-1}\middle|\mathcal{B}_{j-1}\right] \qquad (3.58)\\
&< E\left[S_j^{\alpha-1}\middle|\mathcal{B}_{j-1}\right] - E\left[S_j\middle|\mathcal{B}_{j-1}\right]^\alpha\cdot E\left[S_j\middle|\mathcal{B}_{j-1}\right]^{-1} = 0,
\end{aligned}$$

Jensen's inequality is strict because we have assumed strictly positive variances $\sigma_{j-1}^2 > 0$, which implies that S_j is not deterministic at time $j-1$. This finishes the proof of Lemma 3.8.

□

Lemma 3.8 implies that the term

$$E\left[\prod_{j=I-i}^{J-1}\widehat{f}_j^2\middle|\mathcal{B}_{I-i}\right] \qquad (3.59)$$

can not easily be calculated. Hence from this point of view Approach 1 is not a promising route for finding a closed formula for the estimation error.

Approach 3 (conditional resampling)

In Approach 3 we explicitly resample the observed chain-ladder factors $\widehat{f}_j$. To do the resampling we introduce stronger model assumptions. This is done with a time series model. Such time series models for the chain-ladder method can be found in several papers in the literature see e.g. Murphy [55], Barnett-Zehnwirth [7] or Buchwalder et al. [13].

Model Assumptions 3.9 (Time series model)

- Different accident years i are independent.
- There exist constants $f_j > 0$, $\sigma_j > 0$ and random variables $\varepsilon_{i,j+1}$ such that for all $i \in \{0, \ldots, I\}$ and $j \in \{0, \ldots, J-1\}$ we have that

  $$C_{i,j+1} = f_j\cdot C_{i,j} + \sigma_j\cdot\sqrt{C_{i,j}}\cdot\varepsilon_{i,j+1}, \qquad (3.60)$$

  with conditionally, given $\mathcal{B}_0$, $\varepsilon_{i,j+1}$ are independent with $E\left[\varepsilon_{i,j+1}\middle|\mathcal{B}_0\right] = 0$, $E\left[\varepsilon_{i,j+1}^2\middle|\mathcal{B}_0\right] = 1$ and $P\left[C_{i,j+1} > 0\middle|\mathcal{B}_0\right] = 1$ for all $i \in \{0, \ldots, I\}$ and $j \in \{0, \ldots, J-1\}$.

□

Remarks 3.10

- The time series model defines an auto-regressive process. It is particularly useful for the derivation of the estimation error and reflects the mechanism of generating sets of "other possible" observations.

- The random variables $\varepsilon_{i,j+1}$ are defined conditionally, given $\mathcal{B}_0$, in order to ensure that the cumulative payments $C_{i,j+1}$ stay positive, $P\left[\cdot|\,\mathcal{B}_0\right]$-a.s.

- It is easy to show that Model Assumptions 3.9 imply the Assumptions 3.2 of the classical stochastic chain-ladder model of Mack [49].

- The definition of the time series model in Buchwalder et al. [11] is slightly different. The difference lies in the fact that here we assume a.s. positivity of $C_{i,j}$. This could also be done with the help of conditional assumptions, i.e. the theory would also run through if we would assume that

$$P\left[C_{i,j+1} > 0 \middle|\, C_{i,j}\right] = 1, \tag{3.61}$$

  for all i and j.

In the sequel we use Approach 3, i.e. we do conditional resampling in the time series model. We therefore resample the observations for $\widehat{f}_{I-i}, \ldots, \widehat{f}_{J-1}$, given the upper trapezoid $\mathcal{D}_I$. Thereby we take into account that, given $\mathcal{D}_I$, the observations for $\widehat{f}_j$ could have been different from the observed values. To account for this source of uncertainty we proceed as usual in statistics: Given $\mathcal{D}_I$, we generate for $i \in \{0, \ldots, I\}$ and $j \in \{0, \ldots, J-1\}$ a set of "new" observations $\widetilde{C}_{i,j+1}$ by the formula

$$\widetilde{C}_{i,j+1} = f_j \cdot C_{i,j} + \sigma_j \cdot \sqrt{C_{i,j}} \cdot \widetilde{\varepsilon}_{i,j+1}, \tag{3.62}$$

where $\sigma_j > 0$ and $\widetilde{\varepsilon}_{i,j+1}$, $\varepsilon_{i,j+1}$ are independent and identically distributed given $\mathcal{B}_0$ (cf. Model Assumptions 3.9). This means that $C_{i,j}$ acts as a fixed volume measure and we resample $\widetilde{C}_{i,j+1} \stackrel{(d)}{=} C_{i,j+1}$, given $\mathcal{B}_j$. This means in the language of stochastic kernels that we consider the distributions $K_{j+1}\left(C_{i,j}, dx_{j+1}\right)$ (see (3.50)).

Remark. We have chosen a different notation ($\widetilde{C}_{i,j+1}$ vs. $C_{i,j+1}$) to clearly illustrate that we resample on the conditional structure, i.e. $\widetilde{C}_{i,j+1}$ are random variables and $C_{i,j}$ are (deterministic) volumes, given $\mathcal{B}_j$.

In the spirit of Approach 3 (cf. (3.45)) we resample the observations for $\widehat{f}_j$ by only resampling the observations of development year $j+1$. Together with the resampling assumption (3.62) this leads to the following resampled representation for the estimates of the development factors

$$\widehat{f}_j = \frac{\sum_{i=0}^{i^*(j+1)} \widetilde{C}_{i,j+1}}{\sum_{i=0}^{i^*(j+1)} C_{i,j}} = f_j + \frac{\sigma_j}{S_j} \cdot \sum_{i=0}^{i^*(j+1)} \sqrt{C_{i,j}} \cdot \widetilde{\varepsilon}_{i,j+1} \qquad (0 \le j \le J-1), \tag{3.63}$$

where

$$S_j = \sum_{i=0}^{i^*(j+1)} C_{i,j}. \tag{3.64}$$

As in (3.50) we denote the probability measure of these resampled chain-ladder estimates by $P^*_{\mathcal{D}_I}$.

These resampled estimates of the development factors have, given $\mathcal{B}_j$, the same distribution as the original estimated chain-ladder factors. Unlike the observations $\{C_{i,j}; i+j \le I\}$ the observations $\{\widetilde{C}_{i,j}; i+j \le I\}$ and also the resampled estimates are random variables given $\mathcal{D}_I$. Furthermore the observations $C_{i,j}$ and the random variables $\widetilde{\varepsilon}_{i,j}$ are independent, given $\mathcal{B}_0 \subset \mathcal{D}_I$. This and (3.63) shows that

1) the estimators $\widehat{f}_0, \ldots, \widehat{f}_{J-1}$ are conditionally independent w.r.t. $P^*_{\mathcal{D}_I}$,

2) $E^*_{\mathcal{D}_I}\left[\widehat{f}_j\right] = f_j$ for $0 \le j \le J-1$ and

3) $E^*_{\mathcal{D}_I}\left[\left(\widehat{f}_j\right)^2\right] = f_j^2 + \frac{\sigma_j^2}{S_j}$ for $0 \le j \le J-1$.

Figure 3.1 illustrates the conditional resampling for two different possible observations $\mathcal{D}_I^{(1)}$ and $\mathcal{D}_I^{(2)}$ of the original data $\mathcal{D}_I$, which would give the two different chain-ladder estimates $\widehat{C}_{i,J}^{(1)}$ and $\widehat{C}_{i,J}^{(2)}$ for $E\left[C_{i,J} \middle| \mathcal{D}_I\right]$.

Therefore in Approach 3 we estimate the estimation error by (using 1)-3))

$$\begin{aligned} & E^*_{\mathcal{D}_I}\left[C_{i,I-i}^2 \cdot \left(\widehat{f}_{I-i} \cdot \ldots \cdot \widehat{f}_{J-1} - f_{I-i} \cdot \ldots \cdot f_{J-1}\right)^2\right] \\ & = C_{i,I-i}^2 \cdot \mathrm{Var}_{P^*_{\mathcal{D}_I}}\left(\widehat{f}_{I-i} \cdot \ldots \cdot \widehat{f}_{J-1}\right) \\ & = C_{i,I-i}^2 \cdot \left(\prod_{l=I-i}^{J-1} E^*_{\mathcal{D}_I}\left[\left(\widehat{f}_l\right)^2\right] - \prod_{l=I-i}^{J-1} f_l^2\right) \\ & = C_{i,I-i}^2 \cdot \left(\prod_{l=I-i}^{J-1} \left(f_l^2 + \frac{\sigma_l^2}{S_l}\right) - \prod_{l=I-i}^{J-1} f_l^2\right). \end{aligned} \tag{3.65}$$

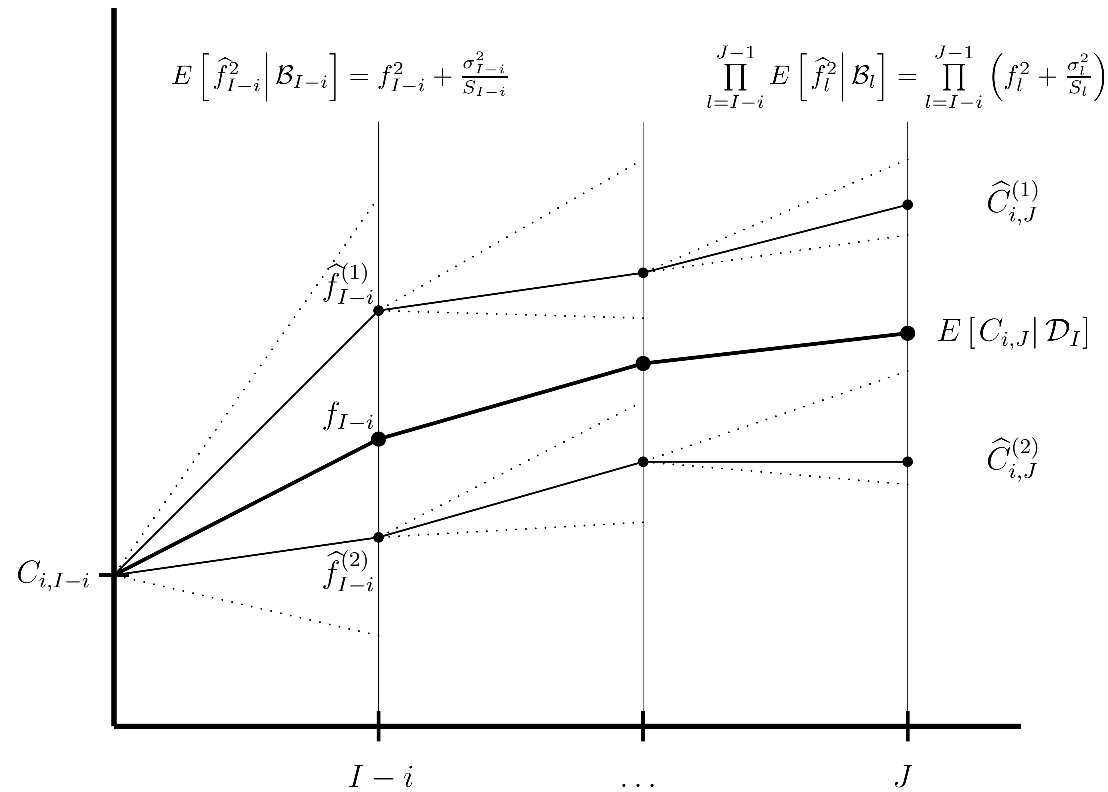


Figure 3.1: Conditional resampling in $\mathcal{D}^O_{I,i}$ (Approach 3)

Observe that this calculation is exact, the estimation has been done at the point where we have decided to use Approach 3 for the estimation error, i.e. the estimate was done choosing the conditional probability measure $P^*_{\mathcal{D}_I}$.

Next, we replace the parameters $\sigma^2_{I-i},\dots,\sigma^2_{J-1}$ and $f_{I-i},\dots,f_{J-1}$ with their estimators, and we obtain the following estimator for the conditional estimation error of accident year $i\in\{I-J+1,\dots,I\}$

$$\begin{aligned}\widehat{\mathrm{Var}}\left(\widehat{C_{i,J}}^{CL}\middle|\mathcal{D}_I\right) &= \widehat{E}^*_{\mathcal{D}_I}\left[\left(\widehat{C_{i,J}}^{CL}-E\left[C_{i,J}\middle|\mathcal{D}_I\right]\right)^2\right] \qquad (3.66)\\ &= {}_1C^2_{i,I-i}\cdot\left(\prod_{l=I-i}^{J-1}\left(\widehat{f}^2_l+\frac{\widehat{\sigma}^2_l}{S_l}\right)-\prod_{l=I-i}^{J-1}\widehat{f}^2_l\right).\end{aligned}$$

The estimator for the conditional estimation error can be written in a recursive form. We obtain for $i \in \{I-J+1, \dots, I\}$

$$
\begin{aligned}
&\widehat{\mathrm{Var}}\left(\left.\widehat{C_{i,J}}^{CL}\right|\mathcal{D}_I\right) \\
&\quad = \widehat{\mathrm{Var}}\left(\left.\widehat{C_{i,J-1}}^{CL}\right|\mathcal{D}_I\right)\cdot \widehat{f}_{J-1}^2 + C_{i,I-i}^2 \cdot \frac{\widehat{\sigma}_{J-1}^2}{S_{J-1}} \cdot \prod_{l=I-i}^{J-2}\left(\widehat{f}_l^2 + \frac{\widehat{\sigma}_l^2}{S_l}\right) \qquad (3.67)\\
&\quad = \widehat{\mathrm{Var}}\left(\left.\widehat{C_{i,J-1}}^{CL}\right|\mathcal{D}_I\right)\cdot \left(\widehat{f}_{J-1}^2 + \frac{\widehat{\sigma}_{J-1}^2}{S_{J-1}}\right) + C_{i,I-i}^2 \cdot \frac{\widehat{\sigma}_{J-1}^2}{S_{J-1}} \cdot \prod_{l=I-i}^{J-2}\widehat{f}_l^2,
\end{aligned}
$$

where $\widehat{\mathrm{Var}}\left(\left.\widehat{C_{i,I-i}}^{CL}\right|\mathcal{D}_I\right) = 0$.

Estimator 3.11 (MSEP for single accident years, conditional version)
Under Model Assumptions 3.9 we have the following estimator for the conditional mean square of prediction of the ultimate claim of a single accident year $i \in \{I-J+1, \dots, I\}$

$$
\begin{aligned}
&\widehat{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}\right) = \widehat{E}\left[\left.\left(\widehat{C_{i,J}}^{CL} - C_{i,J}\right)^2\right|\mathcal{D}_I\right]\\
&= \underbrace{\left(\widehat{C_{i,J}}^{CL}\right)^2 \cdot \sum_{l=I-i}^{J-1}\frac{\widehat{\sigma}_l^2/\widehat{f}_l^2}{\widehat{C_{i,l}}^{CL}}}_{\text{process variance}} + \underbrace{C_{i,I-i}^2\cdot\left(\prod_{l=I-i}^{J-1}\left(\widehat{f}_l^2+\frac{\widehat{\sigma}_l^2}{S_l}\right) - \prod_{l=I-i}^{J-1}\widehat{f}_l^2\right)}_{\text{estimation error}}. \qquad (3.68)
\end{aligned}
$$

We can rewrite (3.68) as follows

$$
\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}\right) = \left(\widehat{C_{i,J}}^{CL}\right)^2 \cdot \left(\sum_{l=I-i}^{J-1}\frac{\widehat{\sigma}_l^2/\widehat{f}_l^2}{\widehat{C_{i,l}}^{CL}} + \prod_{l=I-i}^{J-1}\left(\frac{\widehat{\sigma}_l^2/\widehat{f}_l^2}{S_l}+1\right) - 1\right). \qquad (3.69)
$$

We could also do a linear approximation to the estimation error:

$$
\prod_{l=I-i}^{J-1}\left(\widehat{f}_l^2 + \frac{\widehat{\sigma}_l^2}{S_l}\right) - \prod_{l=I-i}^{J-1}\widehat{f}_l^2 \approx \prod_{l=I-i}^{J-1}\widehat{f}_l^2 \cdot \sum_{l=I-i}^{J-1}\frac{\widehat{\sigma}_l^2/\widehat{f}_l^2}{S_l}. \qquad (3.70)
$$

Observe that in fact the right-hand side of (3.70) is a lower bound for the left-hand side. This immediately gives the following estimate:

Estimator 3.12 (MSEP for single accident years)
Under Model Assumptions 3.9 we have the following estimator for the conditional mean square error of prediction of the ultimate claim of a single accident year $i \in \{I-J+1, \dots, I\}$

$$
\widehat{\widehat{msep}}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}\right) = \left(\widehat{C_{i,J}}^{CL}\right)^2 \cdot \sum_{l=I-i}^{J-1}\frac{\widehat{\sigma}_l^2}{\widehat{f}_l^2}\cdot\left(\frac{1}{\widehat{C_{i,l}}^{CL}} + \frac{1}{S_l}\right). \qquad (3.71)
$$

The Mack [49] approach

Mack [49] even gives a different approach to the estimation of the estimation error. Introduce for $j \in \{I-i, \ldots, J-1\}$

$$T_j = \widehat{f}_{I-i} \cdots \widehat{f}_{j-1} \cdot \left(f_j - \widehat{f}_j\right) \cdot f_{j+1} \cdots f_{J-1}. \tag{3.72}$$

Observe that

$$\left(\widehat{f}_{I-i} \cdot \ldots \cdot \widehat{f}_{J-1} - f_{I-i} \cdot \ldots \cdot f_{J-1}\right)^2 = \left(\sum_{j=I-i}^{J-1} T_j\right)^2. \tag{3.73}$$

This implies that (see (3.40))

$$\left(\widehat{C_{i,J}}^{CL} - E\left[C_{i,J}\middle|\mathcal{D}_I\right]\right)^2 = C_{i,I-i}^2 \cdot \left(\sum_{j=I-i}^{J-1} T_j^2 + 2 \cdot \sum_{I-i \le j < k \le J-1} T_j \cdot T_k\right). \tag{3.74}$$

Each term in the sums on the right-hand side of the equality above is now estimated by a slightly modified version of Approach 2: We estimate $T_j \cdot T_k$ for $j < k$ by

$$\begin{aligned} & E\left[T_j \cdot T_k\middle|\mathcal{B}_k\right] \\ = \;& \widehat{f}_{I-i}^2 \cdots \widehat{f}_{j-1}^2 \cdot \left\{\left(f_j - \widehat{f}_j\right) \cdot \widehat{f}_j\right\} \cdot \left\{f_{j+1} \cdot \widehat{f}_{j+1}\right\} \cdot \ldots \cdot \left\{f_{k-1} \cdot \widehat{f}_{k-1}\right\} \\ & \cdot \left\{f_k \cdot E\left[f_k - \widehat{f}_k\middle|\mathcal{B}_k\right]\right\} \cdot f_{k+1}^2 \cdot \ldots \cdot f_{J-1}^2 \\ = \;& 0, \end{aligned} \tag{3.75}$$

and T_j^2 is estimated by

$$\begin{aligned} E\left[T_j^2\middle|\mathcal{B}_j\right] &= \widehat{f}_{I-i}^2 \cdots \widehat{f}_{j-1}^2 \cdot E\left[\left(f_j - \widehat{f}_j\right)^2\middle|\mathcal{B}_j\right] \cdot f_{j+1}^2 \cdot \ldots \cdot f_{J-1}^2 \\ &= \widehat{f}_{I-i}^2 \cdots \widehat{f}_{j-1}^2 \cdot \frac{\sigma_j^2}{S_j} \cdot f_{j+1}^2 \cdot \ldots \cdot f_{J-1}^2. \end{aligned} \tag{3.76}$$

Hence (3.40) is estimated by

$$C_{i,I-i}^2 \cdot \sum_{j=I-i}^{J-1} \widehat{f}_{I-i}^2 \cdots \widehat{f}_{j-1}^2 \cdot \frac{\sigma_j^2}{S_j} \cdot f_{j+1}^2 \cdot \ldots \cdot f_{J-1}^2. \tag{3.77}$$

If we now replace the unknown parameters σ_j^2 and f_j by its estimates we exactly obtain the estimate $\widehat{\text{msep}}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}\right)$ for the conditional estimation error presented in Estimator 3.12.

Remarks 3.13

- We see that the Mack estimate for the conditional estimation error (also presented in Estimator 3.12) is a linear approximation and lower bound to the estimate coming from Approach 3.

- The difference comes from the fact that Mack [49] decouples the estimation error in an appropriate way (with the help of the terms T_j) and then applies a partial conditional resampling to each of the terms in the decoupling.

- The Time Series Model 3.9 has slightly stronger assumptions than the weighted average development (WAD) factor model studied in Murphy [55], Model IV. To obtain the crucial recursive formula for the conditional estimation error (Theorem 3 in Appendix C of [55]) Murphy assumes independence for the estimators of the chain-ladder factors. However, this assumption is inconsistent with the model assumptions since the chain-ladder factors indeed are uncorrelated (see Lemma 2.5c)) but the squares of two successive chain-ladder estimators are negatively correlated as we can see from Lemma 3.8. The point is that by his assumptions Murphy [55] gets a multiplicative structure of the measure of volatility. In Approach 3 we get the multiplicative structure by the choice of the conditional resampling (probability measure $P^*_{\mathcal{D}_I}$ for the measure of the (conditional) volatility of the chain-ladder estimator (see discussion in Section 3.2.3). This means, in Approach 3 we do not assume that the estimated chain-ladder factors are independent. Henceforth, since in both estimators a multiplicative structure is used it turns out that the recursive estimator (3.67) for the conditional estimation error is exactly the estimator presented in Theorem 3 of Murphy [55] (see also Appendix B in Barnett-Zehnwirth [7]).

Example 3.7 revisited

We come back to our example in Table 2.2. This gives the following error estimates:

From Tables 3.4 and 3.5 we see that the differences in the estimates for the conditional estimation error coming from the linear approximation (Mack formula) are negligible. In all examples we have looked at we came to this conclusion.

3.2.4 Conditional MSEP in the chain-ladder model for aggregated accident years

Consider two different accident years $i < l$. From the model assumptions we know that the ultimate losses $C_{i,J}$ and $C_{l,J}$ are independent. Nevertheless we have to be

i	$\widehat{C_{i,J}}^{CL}$	CL reserves	$\widehat{\text{Var}}\left(C_{i,J}\vert\mathcal{D}_I\right)^{1/2}$		$\widehat{\text{Var}}\left(\widehat{C_{i,J}}^{CL}\vert\mathcal{D}_I\right)^{1/2}$		$\widehat{\text{msep}}_{C_{i,J}\vert\mathcal{D}_I}(\widehat{C_{i,J}}^{CL})^{1/2}$	
0	11'148'124							
1	10'663'318	15'126	191	1.3%	187	1.2%	267	1.8%
2	10'662'008	26'257	742	2.8%	535	2.0%	914	3.5%
3	9'758'606	34'538	2'669	7.7%	1'493	4.3%	3'058	8.9%
4	9'872'218	85'302	6'832	8.0%	3'392	4.0%	7'628	8.9%
5	10'092'247	156'494	30'478	19.5%	13'517	8.6%	33'341	21.3%
6	9'568'143	286'121	68'212	23.8%	27'286	9.5%	73'467	25.7%
7	8'705'378	449'167	80'077	17.8%	29'675	6.6%	85'398	19.0%
8	8'691'971	1'043'242	126'960	12.2%	43'903	4.2%	134'337	12.9%
9	9'626'383	3'950'815	389'783	9.9%	129'770	3.3%	410'817	10.4%

Table 3.4: Estimated chain-ladder reserves and error terms according to Estimator 3.11

i	$\widehat{C_{i,J}}^{CL}$	CL reserves	$\widehat{\text{Var}}\left(C_{i,J}\vert\mathcal{D}_I\right)^{1/2}$		$\widehat{\text{Var}}\left(\widehat{C_{i,J}}^{CL}\vert\mathcal{D}_I\right)^{1/2}$		$\widehat{\widehat{\text{msep}}}_{C_{i,J}\vert\mathcal{D}_I}(\widehat{C_{i,J}}^{CL})^{1/2}$	
0	11'148'124							
1	10'663'318	15'126	191	1.3%	187	1.2%	267	1.8%
2	10'662'008	26'257	742	2.8%	535	2.0%	914	3.5%
3	9'758'606	34'538	2'669	7.7%	1'493	4.3%	3'058	8.9%
4	9'872'218	85'302	6'832	8.0%	3'392	4.0%	7'628	8.9%
5	10'092'247	156'494	30'478	19.5%	13'517	8.6%	33'341	21.3%
6	9'568'143	286'121	68'212	23.8%	27'286	9.5%	73'467	25.7%
7	8'705'378	449'167	80'077	17.8%	29'675	6.6%	85'398	19.0%
8	8'691'971	1'043'242	126'960	12.2%	43'903	4.2%	134'337	12.9%
9	9'626'383	3'950'815	389'783	9.9%	129'769	3.3%	410'817	10.4%

Table 3.5: Estimated chain-ladder reserves and error terms according to Estimator 3.12

careful if we aggregate $\widehat{C_{i,J}}^{CL}$ and $\widehat{C_{l,J}}^{CL}$. The estimators are no longer independent since they use the same observations for estimating the age-to-age factors f_j. We have that

$$\begin{aligned}\text{msep}_{C_{i,J}+C_{l,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}+\widehat{C_{i,J}}^{CL}\right) &= E\left[\left.\left(\widehat{C_{i,J}}^{CL}+\widehat{C_{l,J}}^{CL}-\left(C_{i,J}+C_{l,J}\right)\right)^2\right|\mathcal{D}_I\right]\\ &= \text{Var}\left(C_{i,J}+C_{l,J}\middle|\,\mathcal{D}_I\right)+\left(\widehat{C_{i,J}}^{CL}+\widehat{C_{l,J}}^{CL}-E\left[C_{i,J}+C_{l,J}\middle|\,\mathcal{D}_I\right]\right)^2. \end{aligned}\tag{3.78}$$

Using the independence of the different accident years, we obtain for the first term

$$\text{Var}\left(C_{i,J}+C_{l,J}\middle|\,\mathcal{D}_I\right)=\text{Var}\left(C_{i,J}\middle|\,\mathcal{D}_I\right)+\text{Var}\left(C_{l,J}\middle|\,\mathcal{D}_I\right),\tag{3.79}$$

whereas for the second term we obtain

$$\begin{aligned}&\left(\widehat{C_{i,J}}^{CL}+\widehat{C_{l,J}}^{CL}-E\left[C_{i,J}+C_{l,J}\middle|\,\mathcal{D}_I\right]\right)^2\\ &\quad=\left(\widehat{C_{i,J}}^{CL}-E\left[C_{i,J}\middle|\,\mathcal{D}_I\right]\right)^2+\left(\widehat{C_{l,J}}^{CL}-E\left[C_{l,J}\middle|\,\mathcal{D}_I\right]\right)^2\\ &\qquad+2\cdot\left(\widehat{C_{i,J}}^{CL}-E\left[C_{i,J}\middle|\,\mathcal{D}_I\right]\right)\cdot\left(\widehat{C_{l,J}}^{CL}-E\left[C_{l,J}\middle|\,\mathcal{D}_I\right]\right).\end{aligned}\tag{3.80}$$

Hence we have the following decomposition for the conditional prediction error of the sum of two accident years

$$
\begin{aligned}
& E\left[\left(\widehat{C_{i,J}}^{CL}+\widehat{C_{l,J}}^{CL}-(C_{i,J}+C_{l,J})\right)^2\Big|\,\mathcal{D}_I\right] \\
&\quad = E\left[\left(\widehat{C_{i,J}}^{CL}-C_{i,J}\right)^2\Big|\,\mathcal{D}_I\right]+E\left[\left(\widehat{C_{l,J}}^{CL}-C_{l,J}\right)^2\Big|\,\mathcal{D}_I\right] \\
&\qquad +2\cdot\left(\widehat{C_{i,J}}^{CL}-E\left[C_{i,J}|\,\mathcal{D}_I\right]\right)\cdot\left(\widehat{C_{l,J}}^{CL}-E\left[C_{l,J}|\,\mathcal{D}_I\right]\right).
\end{aligned}
\tag{3.81}
$$

Hence we obtain

$$
\begin{aligned}
& \mathrm{msep}_{C_{i,J}+C_{l,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}+\widehat{C_{i,J}}^{CL}\right) \\
&\quad = \mathrm{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}\right)+\mathrm{msep}_{C_{l,J}|\mathcal{D}_I}\left(\widehat{C_{l,J}}^{CL}\right) \\
&\qquad +2\cdot\left(\widehat{C_{i,J}}^{CL}-E\left[C_{i,J}|\,\mathcal{D}_I\right]\right)\cdot\left(\widehat{C_{l,J}}^{CL}-E\left[C_{l,J}|\,\mathcal{D}_I\right]\right).
\end{aligned}
\tag{3.82}
$$

In addition to the conditional mean square error of prediction of single accident years, we need to average similar to (3.40) over the possible values of $\widehat{f}_j$ for the cross-products of the conditional estimation errors of the two accident years:

$$
\begin{aligned}
& \left(\widehat{C_{i,J}}^{CL}-E\left[C_{i,J}|\,\mathcal{D}_I\right]\right)\cdot\left(\widehat{C_{l,J}}^{CL}-E\left[C_{l,J}|\,\mathcal{D}_I\right]\right) \\
&\quad = C_{i,I-i}\cdot\left(\widehat{f}_{I-i}\cdot\ldots\cdot\widehat{f}_{J-1}-f_{I-i}\cdot\ldots\cdot f_{J-1}\right) \\
&\qquad \cdot C_{l,I-l}\cdot\left(\widehat{f}_{I-l}\cdot\ldots\cdot\widehat{f}_{J-1}-f_{I-l}\cdot\ldots\cdot f_{J-1}\right).
\end{aligned}
\tag{3.83}
$$

Now we could have the same discussions about resampling as above. Here we simply use Approach 3 for resampling, i.e. we choose the probability measure $P^*_{\mathcal{D}_I}$. Then we can explicitly calculate these cross-products. As in (3.65) we obtain as estimate for the cross-products

$$
\begin{aligned}
& C_{i,I-i}\cdot C_{l,I-l}\cdot E^*_{\mathcal{D}_I}\left[\left(\prod_{j=I-i}^{J-1}\widehat{f}_j-\prod_{j=I-i}^{J-1}f_j\right)\cdot\left(\prod_{j=I-l}^{J-1}\widehat{f}_j-\prod_{j=I-l}^{J-1}f_j\right)\right] \\
&= C_{i,I-i}\cdot C_{l,I-l}\cdot\mathrm{Cov}_{P^*_{\mathcal{D}_I}}\left(\widehat{f}_{I-i}\cdot\ldots\cdot\widehat{f}_{J-1},\widehat{f}_{I-l}\cdot\ldots\cdot\widehat{f}_{J-1}\right) \\
&= C_{i,I-i}\cdot C_{l,I-l}\cdot f_{I-l}\cdot\ldots\cdot f_{I-i-1}\cdot\mathrm{Var}_{P^*_{\mathcal{D}_I}}\left(\widehat{f}_{I-i}\cdot\ldots\cdot\widehat{f}_{J-1}\right) \\
&= C_{i,I-i}\cdot C_{l,I-l}\cdot f_{I-l}\cdot\ldots\cdot f_{I-i-1}\cdot\left(\prod_{j=I-i}^{J-1}E^*_{\mathcal{D}_I}\left[\left(\widehat{f}_j\right)^2\right]-\prod_{j=I-i}^{J-1}f_j^2\right) \\
&= C_{i,I-i}\cdot E\left[C_{l,I-i}|\,\mathcal{D}_I\right]\cdot\left(\prod_{j=I-i}^{J-1}\left(f_j^2+\frac{\sigma_j^2}{S_j}\right)-\prod_{j=I-i}^{J-1}f_j^2\right).
\end{aligned}
\tag{3.84}
$$

But then the estimation of the covariance term is straightforward from the estimate of a single accident year.

Estimator 3.14 (MSEP aggregated accident years, conditional version)

Under Model Assumptions 3.9 we have the following estimator for the conditional mean square error of prediction of the ultimate claim for aggregated accident years

$$\widehat{msep}_{\sum_i C_{i,J}|\mathcal{D}_I}\left(\sum_{i=I-J+1}^{I}\widehat{C_{i,J}}^{CL}\right)=\widehat{E}\left[\left(\sum_{i=I-J+1}^{I}\widehat{C_{i,J}}^{CL}-\sum_{i=I-J+1}^{I}C_{i,J}\right)^2\middle|\mathcal{D}_I\right]$$

$$=\sum_{i=I-J+1}^{I}\widehat{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}\right)$$

$$+2\cdot\sum_{I-J+1\le i<l\le I}C_{i,I-i}\cdot\widehat{C_{l,I-i}}^{CL}\cdot\left(\prod_{j=I-i}^{J-1}\left(\widehat{f}_j^2+\frac{\widehat{\sigma}_j^2}{S_j}\right)-\prod_{j=I-i}^{J-1}\widehat{f}_j^2\right). \quad (3.85)$$

Remarks 3.15

- The last terms (covariance terms) from the result above can be rewritten as

$$2\cdot\sum_{I-J+1\le i<l\le I}\frac{\widehat{C_{l,I-i}}^{CL}}{C_{i,I-i}}\cdot\widehat{\mathrm{Var}}\left(\widehat{C_{i,J}}^{CL}\middle|\mathcal{D}_I\right), \quad (3.86)$$

  where $\widehat{\mathrm{Var}}\left(\widehat{C_{i,J}}^{CL}\middle|\mathcal{D}_I\right)$ is the conditional estimation error of the single accident year i (see (3.66)). This may be helpful in the implementation since it leads to matrix multiplications.

- We can again do a linear approximation and then we find the estimator presented in Mack [49].

Example 3.7 revisited
We come back to our example in Table 2.2. This gives the error estimates in Table 3.6.

3.3 Analysis of error terms

In this section we further analyze the conditional mean square error of prediction of the chain-ladder method. In fact, we consider three different kinds of error terms: a) conditional process error, b) conditional prediction error, c) conditional estimation error. To analyze these three terms we define a model, which is different from the classical chain-ladder model. It is slightly more complicated than the classical

i	$\widehat{C_{i,J}}^{CL}$	CL reserves	$\widehat{\mathrm{Var}}\left(C_{i,J}\vert\mathcal{D}_I\right)^{1/2}$		$\widehat{\mathrm{Var}}\left(\widehat{C_{i,J}}^{CL}\vert\mathcal{D}_I\right)^{1/2}$		$\widehat{\mathrm{msep}}_{C_{i,J}\vert\mathcal{D}_I}(\widehat{C_{i,J}}^{CL})^{1/2}$	
0	11'148'124							
1	10'663'318	15'126	191	1.3%	187	1.2%	267	1.8%
2	10'662'008	26'257	742	2.8%	535	2.0%	914	3.5%
3	9'758'606	34'538	2'669	7.7%	1'493	4.3%	3'058	8.9%
4	9'872'218	85'302	6'832	8.0%	3'392	4.0%	7'628	8.9%
5	10'092'247	156'494	30'478	19.5%	13'517	8.6%	33'341	21.3%
6	9'568'143	286'121	68'212	23.8%	27'286	9.5%	73'467	25.7%
7	8'705'378	449'167	80'077	17.8%	29'675	6.6%	85'398	19.0%
8	8'691'971	1'043'242	126'960	12.2%	43'903	4.2%	134'337	12.9%
9	9'626'383	3'950'815	389'783	9.9%	129'770	3.3%	410'817	10.4%
Cov. term					116'811		116'811	
Total		6'047'061	424'379	7.0%	185'026	3.1%	462'960	7.7%

Table 3.6: Estimated chain-ladder reserves and error terms (Estimator 3.14)

model but therefore leads to a clear distinction between these error terms. The motivation for a clear distinction between the three error terms is that the sources of these error classes are rather different ones and we believe that in the light of the solvency discussions (see e.g. SST [73], Sandström [67], Buchwalder et al. [11, 14] or Wüthrich [88]) we should clearly distinguish between the different risk factors. In this section we closely follow Wüthrich [90]. For a similar Bayesian approach we also refer to Gisler [29].

3.3.1 Classical chain-ladder model

The observed individual development factors were defined by (see also (3.19))

$$F_{i,j} = \frac{C_{i,j}}{C_{i,j-1}}, \tag{3.87}$$

then we have with Model Assumptions 3.2 that

$$E\left[F_{i,j}\middle|\, C_{i,j-1}\right] = f_{j-1} \qquad \text{and} \qquad \mathrm{Var}\left(F_{i,j}\middle|\, C_{i,j-1}\right) = \frac{\sigma_{j-1}^2}{C_{i,j-1}}. \tag{3.88}$$

The conditional variational coefficients of the development factors $F_{i,j}$ are given by

$$\mathrm{Vco}\left(F_{i,j}\middle|\, C_{i,j-1}\right) = \mathrm{Vco}\left(C_{i,j}\middle|\, C_{i,j-1}\right) = \frac{\sigma_{j-1}}{f_{j-1}} \cdot C_{i,j-1}^{-1/2} \longrightarrow 0, \qquad \text{as } C_{i,j-1} \to \infty. \tag{3.89}$$

Hence for increasing volume the conditional variational coefficients of $F_{i,j}$ converge to zero! It is exactly this property (3.89) which is crucial in risk management. If we assume that risk is defined through these variational coefficients, it means that the risk completely disappears for very large portfolios (law of large numbers). But we all know that this is not the case in practice. There are always external factors, which influence a portfolio and which are not diversifiable, e.g. if jurisdiction changes it is not helpful to have a large portfolio, etc. Also the experiences in recent years have shown that we have to be very careful about external factors and parameter errors since they can not be diversified. Therefore, in almost all developments of new solvency guidelines and requirements one pays a lot of attention to these risks. The goal here is to define a chain-ladder model, which reflects also this kind of risk class.

3.3.2 Enhanced chain-ladder model

The approach in this section modifies (3.89) as follows. We assume that there exist constants $a_0^2, a_1^2, \ldots \geq 0$ such that for all $1 \leq j \leq J$ we have that

$$\text{Vco}^2\left(F_{i,j}\middle|\, C_{i,j-1}\right) = \frac{\sigma_{j-1}^2}{f_{j-1}^2} \cdot C_{i,j-1}^{-1} + a_{j-1}^2. \tag{3.90}$$

Hence

$$\text{Vco}^2\left(F_{i,j}\middle|\, C_{i,j-1}\right) > \lim_{C_{i,j-1}\to\infty} \text{Vco}^2\left(F_{i,j}\middle|\, C_{i,j-1}\right) = a_{j-1}^2, \tag{3.91}$$

which is now bounded from below by a_{j-1}^2. This implies that we replace the chain-ladder condition on the variance by

$$\text{Var}\left(C_{i,j}\middle|\, C_{i,j-1}\right) = \sigma_{j-1}^2 \cdot C_{i,j-1} + a_{j-1}^2 \cdot f_{j-1}^2 \cdot C_{i,j-1}^2. \tag{3.92}$$

This means that we add a quadratic term to ensure that the variational coefficient does not disappear when the volume is going to infinity.

As above, we define the chain-ladder consistent time series model. This time series model gives an algorithm how we should simulate additional observations. This algorithm will be used for the calculation of the estimation error.

Model Assumptions 3.16 (Enhanced time series model)

- Different accident years i are independent.
- There exist constants $f_j > 0$, $\sigma_j^2 > 0$, $a_j^2 \geq 0$ and random variables $\varepsilon_{i,j+1}$ such that for all $i \in \{0, \ldots, I\}$ and $j \in \{0, \ldots, J-1\}$ we have that

$$C_{i,j+1} = f_j \cdot C_{i,j} + \left(\sigma_j^2 + a_j^2 \cdot f_j^2 \cdot C_{i,j}\right)^{1/2} \cdot \sqrt{C_{i,j}} \cdot \varepsilon_{i,j+1}, \tag{3.93}$$

  with conditionally, given $\mathcal{B}_0$, $\varepsilon_{i,j+1}$ are independent with $E\left[\varepsilon_{i,j+1}\middle|\, \mathcal{B}_0\right] = 0$, $E\left[\varepsilon_{i,j+1}^2\middle|\, \mathcal{B}_0\right] = 1$ and $P\left[C_{i,j+1} > 0\middle|\, \mathcal{B}_0\right] = 1$ for all $i \in \{0, \ldots, I\}$ and $j \in \{0, \ldots, J-1\}$.

□

Remark. See Remarks 3.10.

Lemma 3.17 *Model 3.16 satisfies Model Assumptions 3.2 with* (3.11) *replaced by* (3.92).

I.e. the model satisfies the chain-ladder assumptions with modified variance function. For $a_j = 0$ we obtain the Time Series Version 3.9.

3.3.3 Interpretation

In this subsection we give an interpretation to the variance term (3.92). Alternatively, we could use a model with latent variables $\Theta_{i,j}$. This is similar to the Bayesian approaches such as used in Gisler [29] saying that the "true" chain-ladder factors f_j are themselves random variables (depending on external/latent factors).

(A1) Conditionally, given $\Theta_{i,j}$, we have

$$E\left[C_{i,j+1}\middle|\,\Theta_{i,j}, C_{i,j}\right] = f_j(\Theta_{i,j}) \cdot C_{i,j}, \tag{3.94}$$

$$\mathrm{Var}\left(C_{i,j+1}\middle|\,\Theta_{i,j}, C_{i,j}\right) = \sigma_j^2(\Theta_{i,j}) \cdot C_{i,j}. \tag{3.95}$$

(A2) $\Theta_{i,j}$ are independent with

$$E\left[f_j(\Theta_{i,j})\middle|\,C_{i,j}\right] = f_j, \tag{3.96}$$

$$\mathrm{Var}\left(f_j(\Theta_{i,j})\middle|\,C_{i,j}\right) = a_j^2 \cdot f_j^2, \tag{3.97}$$

$$E\left[\sigma_j^2(\Theta_{i,j})\middle|\,C_{i,j}\right] = \sigma_j^2. \tag{3.98}$$

□

Remark. The variables $F_{i,j} = C_{i,j+1}/C_{i,j}$ satisfy the Bühlmann-Straub model assumptions (see Bühlmann-Gisler [18] and Section 4.3 below).

For the variance term we obtain

$$\begin{aligned} \mathrm{Var}\left(C_{i,j+1}\middle|\,C_{i,j}\right) &= E\left[\mathrm{Var}\left(C_{i,j+1}\middle|\,\Theta_{i,j}, C_{i,j}\right)\middle|\,C_{i,j}\right] \qquad (3.99)\\ &\quad +\mathrm{Var}\left(E\left[C_{i,j+1}\middle|\,\Theta_{i,j}, C_{i,j}\right]\middle|\,C_{i,j}\right)\\ &= \sigma_j^2 \cdot C_{i,j} + a_j^2 \cdot f_j^2 \cdot C_{i,j}^2. \qquad (3.100)\end{aligned}$$

Moreover we see that

$$\mathrm{Vco}\left(f_j(\Theta_{i,j})\middle|\,C_{i,j}\right) = a_j. \tag{3.101}$$

Hence we introduce the following terminology:

a) Conditional process error / Conditional process variance. The conditional process error corresponds to the term

$$\sigma_j^2 \cdot C_{i,j} \tag{3.102}$$

and reflects the fact that $C_{i,j+1}$ are random variables which have to be predicted. For increasing volume $C_{i,j}$ the variational coefficient of this term disappears.

b) Conditional Prediction error. The conditional prediction error corresponds to the term

$$a_j^2 \cdot f_j^2 \cdot C_{i,j}^2 \tag{3.103}$$

and reflects the fact that we have to predict the future development factors $f_j(\Theta_{i,j})$. These future development factors underlay also some uncertainty, and hence may be modelled stochastically (Bayesian point of view). The Mack formula and the Estimator 3.14 for the conditional mean square error of prediction does not consider this kind of risk.

c) Conditional estimation error. There is a third kind of risk, namely the risk which comes from the fact that we have to estimate the true parameters f_j in (3.96) from the data. This error term will be called conditional estimation error. It is also considered in the Mack model and in Estimator 3.14. For the derivation of an estimate for this term we will use Approach 3, page 50. This derivation will use the time series definition of the chain-ladder method.

3.3.4 Chain-ladder estimator in the enhanced model

Under Model Assumptions 3.16 we have that

$$F_{i,j+1} = f_j + \left(\sigma_j^2 \cdot C_{i,j}^{-1} + a_j^2 \cdot f_j^2\right)^{1/2} \cdot \varepsilon_{i,j+1}, \tag{3.104}$$

with

$$E\left[F_{i,j+1} \middle| C_{i,j}\right] = f_j \quad \text{and} \quad \operatorname{Var}\left(F_{i,j+1} \middle| C_{i,j}\right) = \sigma_j^2 \cdot C_{i,j}^{-1} + a_j^2 \cdot f_j^2. \tag{3.105}$$

This immediately gives the following lemma:

Lemma 3.18 *Under Model Assumptions 3.16 we have for $i > I - J$ that*

$$E\left[C_{i,J} \middle| \mathcal{D}_I\right] = E\left[C_{i,J} \middle| C_{i,I-i}\right] = C_{i,I-i} \cdot \prod_{j=I-i}^{J-1} f_j. \tag{3.106}$$

Proof. See proof of Lemma 2.3.

□

Remark. As soon as we know the chain-ladder factors f_j we can calculate the expected conditional ultimate $C_{i,J}$, given the information $\mathcal{D}_I$. Of course, in general, the chain-ladder factors f_j are not known and need to be estimated from the data.

3.3.5 Conditional process and prediction errors

We derive now the recursive formula for the conditional process and prediction error: Under Model Assumptions 3.16 we have for the ultimate claim $C_{i,J}$ of accident year $i > I - J$ that

$$\begin{aligned} \mathrm{Var}\left(C_{i,J} \middle| \mathcal{D}_I\right) &= \mathrm{Var}\left(C_{i,J} \middle| C_{i,I-i}\right) \\ &= E\left[\mathrm{Var}\left(C_{i,J} \middle| C_{i,J-1}\right) \middle| C_{i,I-i}\right] + \mathrm{Var}\left(E\left[C_{i,J} \middle| C_{i,J-1}\right] \middle| C_{i,I-i}\right). \end{aligned} \tag{3.107}$$

For the first term on the right-hand side of (3.107) we obtain under Model Assumptions 3.16 that

$$\begin{aligned} & E\left[\mathrm{Var}\left(C_{i,J} \middle| C_{i,J-1}\right) \middle| C_{i,I-i}\right] \\ &\quad = E\left[\sigma_{J-1}^2 \cdot C_{i,J-1} + a_{J-1}^2 \cdot f_{J-1}^2 \cdot C_{i,J-1}^2 \middle| C_{i,I-i}\right] \\ &\quad = \sigma_{J-1}^2 \cdot \prod_{j=I-i}^{J-2} f_j \cdot C_{i,I-i} + a_{J-1}^2 \cdot f_{J-1}^2 \cdot \left(\mathrm{Var}\left(C_{i,J-1} \middle| \mathcal{D}_I\right) + E\left[C_{i,J-1} \middle| C_{i,I-i}\right]^2\right) \\ &\quad = C_{i,I-i}^2 \cdot \left(\frac{\sigma_{J-1}^2}{C_{i,I-i}} \cdot \prod_{j=I-i}^{J-2} f_j + a_{J-1}^2 \cdot \prod_{j=I-i}^{J-1} f_j^2\right) + a_{J-1}^2 \cdot f_{J-1}^2 \cdot \mathrm{Var}\left(C_{i,J-1} \middle| \mathcal{D}_I\right). \end{aligned} \tag{3.108}$$

For the second term on the right-hand side of (3.107) we obtain under Model Assumptions 3.16

$$\begin{aligned} \mathrm{Var}\left(E\left[C_{i,J} \middle| C_{i,J-1}\right] \middle| C_{i,I-i}\right) &= \mathrm{Var}\left(f_{J-1} \cdot C_{i,J-1} \middle| C_{i,I-i}\right) \\ &= f_{J-1}^2 \cdot \mathrm{Var}\left(C_{i,J-1} \middle| \mathcal{D}_I\right). \end{aligned} \tag{3.109}$$

This leads to the following recursive formula (compare this to (3.32))

$$\begin{aligned} \mathrm{Var}\left(C_{i,J} \middle| \mathcal{D}_I\right) &= C_{i,I-i}^2 \cdot \left(\frac{\sigma_{J-1}^2}{C_{i,I-i}} \cdot \prod_{j=I-i}^{J-2} f_j + a_{J-1}^2 \cdot \prod_{j=I-i}^{J-1} f_j^2\right) \\ &\quad + (1 + a_{J-1}^2) \cdot f_{J-1}^2 \cdot \mathrm{Var}\left(C_{i,J-1} \middle| \mathcal{D}_I\right). \end{aligned} \tag{3.110}$$

For $a_{J-1}^2 = 0$ it coincides with the formula given in (3.32).
This gives the following lemma:

Lemma 3.19 (Process/prediction errors for single accident years) *Under Model Assumptions 3.16 the conditional process variance and prediction errors for the ultimate claim of a single accident year $i \in \{I-j+1, \ldots, I\}$ are given by*

$$\begin{aligned} &Var(C_{i,J} | \mathcal{D}_I) \\ &= C_{i,I-i}^2 \cdot \left[\sum_{m=I-i}^{J-1} \prod_{n=m+1}^{J-1} (1+a_n^2) \cdot f_n^2 \cdot \left(\frac{\sigma_m^2}{C_{i,I-i}} \cdot \prod_{j=I-i}^{m-1} f_j + a_m^2 \cdot \prod_{j=I-i}^{m} f_j^2 \right) \right] \\ &= E\left[C_{i,J} | \mathcal{D}_I \right]^2 \cdot \left[\sum_{m=I-i}^{J-1} \left(\frac{\sigma_m^2 / f_m^2}{E\left[C_{i,m} | \mathcal{D}_I \right]} + a_m^2 \right) \cdot \prod_{n=m+1}^{J-1} (1+a_n^2) \right]. \end{aligned} \tag{3.111}$$

Lemma 3.19 implies that the conditional variational coefficient of the ultimate $C_{i,J}$ is given by

$$\mathrm{Vco}\left(C_{i,J} | \mathcal{D}_I \right) = \left[\sum_{m=I-i}^{J-1} \left(\frac{\sigma_m^2 / f_m^2}{E\left[C_{i,m} | \mathcal{D}_I \right]} + a_m^2 \right) \cdot \prod_{n=m+1}^{J-1} (1+a_n^2) \right]^{1/2}. \tag{3.112}$$

Henceforth we see that the **conditional prediction error** of $C_{i,J}$ corresponds to (the conditional process error disappears for infinitely large volume $C_{i,I-i}$)

$$\lim_{C_{i,I-i} \to \infty} \mathrm{Vco}\left(C_{i,J} | \mathcal{D}_I \right) = \left[\sum_{m=I-i}^{J-1} a_m^2 \cdot \prod_{n=m+1}^{J-1} (1+a_n^2) \right]^{1/2}, \tag{3.113}$$

and the conditional variational coefficient for the **conditional process error** of $C_{i,J}$ is given by

$$\left[\sum_{m=I-i}^{J-1} \left(\frac{\sigma_m^2 / f_m^2}{E\left[C_{i,m} | \mathcal{D}_I \right]} \right) \cdot \prod_{n=m+1}^{J-1} (1+a_n^2) \right]^{1/2}. \tag{3.114}$$

3.3.6 Chain-ladder factors and conditional estimation error

The conditional estimation error comes from the fact that we have to estimate the f_j from the data.

Estimation Approach 1

From Lemma 3.4 we obtain the following lemma:

Lemma 3.20 *Under Model Assumptions 3.16, the estimator*

$$\widehat{F}_j = \frac{\sum_{i=0}^{i^*(j+1)} \frac{C_{i,j}}{\sigma_j^2 + a_j^2 \cdot f_j^2 \cdot C_{i,j}} \cdot F_{i,j+1}}{\sum_{i=0}^{i^*(j+1)} \frac{C_{i,j}}{\sigma_j^2 + a_j^2 \cdot f_j^2 \cdot C_{i,j}}} = \frac{\sum_{i=0}^{i^*(j+1)} \frac{C_{i,j+1}}{\sigma_j^2 + a_j^2 \cdot f_j^2 \cdot C_{i,j}}}{\sum_{i=0}^{i^*(j+1)} \frac{C_{i,j}}{\sigma_j^2 + a_j^2 \cdot f_j^2 \cdot C_{i,j}}}. \tag{3.115}$$

is the $\mathcal{B}_{j+1}$-measurable unbiased estimator for f_j, which has minimal conditional variance among all linear combinations of the unbiased estimators $(F_{i,j+1})_{0\le i\le i^*(j+1)}$ for f_j, conditioned on $\mathcal{B}_j$, i.e.

$$Var(\widehat{F}_j|\mathcal{B}_j) = \min_{\alpha_i\in\mathbb{R}} Var\left(\sum_{i=0}^{i^*(j+1)} \alpha_i \cdot F_{i,j+1}\middle|\mathcal{B}_j\right). \tag{3.116}$$

The conditional variance is given by

$$Var\left(\widehat{F}_j\middle|\mathcal{B}_j\right) = \left(\sum_{i=0}^{i^*(j+1)} \frac{C_{i,j}}{\sigma_j^2 + a_j^2 \cdot f_j^2 \cdot C_{i,j}}\right)^{-1}. \tag{3.117}$$

Proof. From (3.105) we see that $F_{i,j+1}$ is an unbiased estimator for f_j, conditioned on $\mathcal{B}_j$, with

$$E\left[F_{i,j+1}\middle|\mathcal{B}_j\right] = E\left[F_{i,j+1}\middle|C_{i,j}\right] = f_j, \tag{3.118}$$

$$\mathrm{Var}\left(F_{i,j+1}\middle|\mathcal{B}_j\right) = \mathrm{Var}\left(F_{i,j+1}\middle|C_{i,j}\right) = \sigma_j^2 \cdot C_{i,j}^{-1} + a_j^2 \cdot f_j^2. \tag{3.119}$$

Hence the proof follows from Lemma 3.4.

□

Remark. For $a_j = 0$ we obtain the classical chain-ladder estimators (2.7). Moreover, observe that for calculating the estimate $\widehat{F}_j$ one needs to know the parameter f_j, a_j and σ_j (see (3.115)). Of course this contradicts the fact that we need to estimate f_j. One way out of this dilemma is to use an estimate for f_j which is not optimal, i.e. has larger variance.
Let us (in Estimation Approach 1) assume that we can calculate (3.115).

Estimator 3.21 (Chain-ladder estimator, enhanced time series model)
The CL estimator for $E\left[C_{i,j}\middle|\mathcal{D}_I\right]$ in the Enhanced Model 3.16 is given by

$$\widehat{C_{i,j}}^{(CL,2)} = \widehat{E}\left[C_{i,j}\middle|\mathcal{D}_I\right] = C_{i,I-i} \cdot \prod_{l=I-i}^{j-1} \widehat{F}_l. \tag{3.120}$$

for $i + j > I$.

We obtain the following lemma for the estimators in the enhanced time series model:

Lemma 3.22 *Under Assumptions 3.16 we have:*

a) *$\widehat{F}_j$ is, given $\mathcal{B}_j$, an unbiased estimator for f_j, i.e. $E\left[\widehat{F}_j\middle|\mathcal{B}_j\right] = f_j$,* *b)* $\widehat{F}_j$ *is (unconditionally) unbiased for* f_j, *i.e.* $E\left[\widehat{F}_j\right] = f_j$,

c) $\widehat{F}_0, \ldots, \widehat{F}_{J-1}$ *are uncorrelated, i.e.* $E\left[\widehat{F}_0 \cdot \ldots \cdot \widehat{F}_{J-1}\right] = \prod_{j=0}^{J-1} E\left[\widehat{F}_j\right]$,

d) $\widehat{C_{i,J}}^{(CL,2)}$ *is, given* $C_{i,I-i}$, *an unbiased estimator for* $E\left[C_{i,J}\middle|\mathcal{D}_I\right]$, *i.e.* $E\left[\widehat{C_{i,J}}^{(CL,2)}\middle| C_{I-i}\right] = E\left[C_{i,J}\middle|\mathcal{D}_I\right]$ *and*

e) $\widehat{C_{i,J}}^{(CL,2)}$ *is (unconditionally) unbiased for* $E\left[C_{i,J}\right]$, *i.e.* $E\left[\widehat{C_{i,J}}^{(CL,2)}\right] = E\left[C_{i,J}\right]$.

Proof. See proof of Lemma 2.5.

□

Single accident years

In the sequel of this subsection we assume that the parameters in (4.62) are known to calculate $\widehat{F}_j$.

Our goal is to estimate the conditional mean square error of prediction (conditional MSEP) as in the classical chain-ladder model

$$\begin{aligned} \mathrm{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{(CL,2)}\right) &= E\left[\left(C_{i,J} - \widehat{C_{i,J}}^{(CL,2)}\right)^2\middle|\mathcal{D}_I\right] \\ &= \mathrm{Var}\left(C_{i,J}\middle|\mathcal{D}_I\right) + \left(E\left[C_{i,J}\middle|\mathcal{D}_I\right] - \widehat{C_{i,J}}^{(CL,2)}\right)^2. \end{aligned} \tag{3.121}$$

The first term is exactly the conditional process variance and the conditional prediction error obtained in Lemma 3.19, the second term is the conditional estimation error. It is given by

$$\left(E\left[C_{i,J}\middle|\mathcal{D}_I\right] - \widehat{C_{i,J}}^{(CL,2)}\right)^2 = C_{i,I-i}^2 \cdot \left(\prod_{j=I-i}^{J-1} f_j - \prod_{j=I-i}^{J-1} \widehat{F}_j\right)^2. \tag{3.122}$$

Observe that

$$\begin{aligned} \widehat{F}_j &= \frac{\sum_{i=0}^{i^*(j+1)} \frac{C_{i,j}}{\sigma_j^2 + a_j^2 \cdot f_j^2 \cdot C_{i,j}} \cdot F_{i,j+1}}{\sum_{i=0}^{i^*(j+1)} \frac{C_{i,j}}{\sigma_j^2 + a_j^2 \cdot f_j^2 \cdot C_{i,j}}} \\ &= f_j + \frac{1}{\sum_{i=0}^{i^*(j+1)} \frac{C_{i,j}}{\sigma_j^2 + a_j^2 \cdot f_j^2 \cdot C_{i,j}}} \sum_{i=0}^{i^*(j+1)} \left(\frac{C_{i,j}}{\sigma_j^2 + a_j^2 \cdot f_j^2 \cdot C_{i,j}}\right)^{1/2} \cdot \varepsilon_{i,j+1}. \end{aligned} \tag{3.123}$$

Hence $\widehat{F}_j$ consists of a constant f_j and a stochastic error term (see also Lemma 3.20). In order to determine the conditional estimation error we now proceed as in Section 3.2.3 for the Time Series Model 3.9. This means that we use Approach 3 (conditional resampling in $\mathcal{D}_{I,i}^O$, page 50) to estimate the fluctuations of the estimators $\widehat{F}_0, \ldots, \widehat{F}_{J-1}$ around the chain-ladder factors $f_0, \ldots, f_{J-1}$, i.e. to get an estimate for (3.122).

We therefore (conditionally) resample the observations $\widehat{F}_0, \ldots, \widehat{F}_{J-1}$, given $\mathcal{D}_I$, and use the resampled estimates to calculate an estimate for the conditional estimation error. For these resampled observations we again use the notation $P^*_{\mathcal{D}_I}$ for the conditional measure (for a more detailed discussion we refer to Section 3.2.3). Moreover, under $P^*_{\mathcal{D}_I}$, the random variables $\widehat{F}_j$ are independent with

$$E^*_{\mathcal{D}_I}\left[\widehat{F}_j\right] = f_j \quad \text{and} \quad E^*_{\mathcal{D}_I}\left[\left(\widehat{F}_j\right)^2\right] = f_j^2 + \left(\sum_{i=0}^{i^*(j+1)} \frac{C_{i,j}}{\sigma_j^2 + a_j^2 \cdot f_j^2 \cdot C_{i,j}}\right)^{-1} \tag{3.124}$$

(cf. Section 3.2.3, Approach 3). This means that the conditional estimation error (3.122) is estimated by

$$\begin{aligned} E^*_{\mathcal{D}_I}\left[C_{i,I-i}^2 \cdot \left(\prod_{j=I-i}^{J-1} f_j - \prod_{j=I-i}^{J-1} \widehat{F}_j\right)^2\right] &= C_{i,I-i}^2 \cdot \mathrm{Var}_{P^*_{\mathcal{D}_I}}\left(\prod_{j=I-i}^{J-1} \widehat{F}_j\right) \\ &= C_{i,I-i}^2 \cdot \left(\prod_{j=I-i}^{J-1} E^*_{\mathcal{D}_I}\left[\left(\widehat{F}_j\right)^2\right] - \prod_{j=I-i}^{J-1} f_j^2\right) \\ &= C_{i,I-i}^2 \cdot \prod_{j=I-i}^{J-1} f_j^2 \cdot \left[\prod_{j=I-i}^{J-1}\left(\left(\sum_{k=0}^{i^*(j+1)} \frac{C_{k,j}}{\frac{\sigma_j^2}{f_j^2} + a_j^2 \cdot C_{k,j}}\right)^{-1} + 1\right) - 1\right]. \end{aligned} \tag{3.125}$$

Finally, if we do a linear approximation to (3.125) we obtain

$$\begin{aligned} E^*_{\mathcal{D}_I}\left[C_{i,I-i}^2 \cdot \left(\prod_{j=I-i}^{J-1} f_j - \prod_{j=I-i}^{J-1} \widehat{F}_j\right)^2\right] &= C_{i,I-i}^2 \cdot \mathrm{Var}_{P^*_{\mathcal{D}_I}}\left(\prod_{j=I-i}^{J-1} \widehat{F}_j\right) \\ &\approx C_{i,I-i}^2 \cdot \prod_{j=I-i}^{J-1} f_j^2 \cdot \sum_{j=I-i}^{J-1}\left(\sum_{k=0}^{i^*(j+1)} \frac{C_{k,j}}{\frac{\sigma_j^2}{f_j^2} + a_j^2 \cdot C_{k,j}}\right)^{-1}. \end{aligned} \tag{3.126}$$

For $a_j = 0$ this is exactly the conditional estimation error in the Mack Model 3.2. For increasing number of observations (accident years i) this error term goes to zero.

If we use the linear approximation (3.126) and if we replace the parameters in (3.111) and (3.126) by their estimators (cf. Section 3.3.7) we obtain the following estimator for the conditional mean square error of prediction (for the time being we assume that σ_j^2 and a_j^2 are known).

Estimator 3.23 (MSEP for single accident years)

Under Model Assumptions 3.16 we have the following estimator for the conditional mean square error of prediction for the ultimate claim of a single accident year $i \in \{I-J+1, \ldots, I\}$

$$\widehat{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{(CL,2)}\right) = \left(\widehat{C_{i,J}}^{(CL,2)}\right)^2 \cdot \sum_{j=I-i}^{J-1}\left[\left(\frac{\sigma_j^2}{\widehat{F_j}^2 \cdot \widehat{C_{i,j}}^{(CL,2)}} + a_j^2\right) \cdot \prod_{n=j+1}^{J-1}(1+a_n^2) + \left(\sum_{k=0}^{i^*(j+1)} \frac{C_{k,j}}{\frac{\sigma_j^2}{\widehat{F_j}^2} + a_j^2 \cdot C_{k,j}}\right)^{-1}\right]. \tag{3.127}$$

Aggregated accident years

Consider two different accident years $k < i$. From our assumptions we know that the ultimate losses $C_{k,J}$ and $C_{i,J}$ are independent. Nevertheless we have to be careful if we aggregate $\widehat{C_{k,J}}^{(CL,2)}$ and $\widehat{C_{i,J}}^{(CL,2)}$. The estimators are no longer independent since they use the same observations for estimating the chain-ladder factors f_j.

$$\begin{aligned} & E\left[\left(\widehat{C_{k,J}}^{(CL,2)} + \widehat{C_{i,J}}^{(CL,2)} - (C_{k,J}+C_{i,J})\right)^2 \middle| \mathcal{D}_I\right] \\ & \quad = \mathrm{Var}\left(C_{k,J}+C_{i,J}|\, \mathcal{D}_I\right) + \left(\widehat{C_{k,J}}^{(CL,2)} + \widehat{C_{i,J}}^{(CL,2)} - E\left[C_{k,J}+C_{i,J}|\, \mathcal{D}_I\right]\right)^2. \end{aligned} \tag{3.128}$$

Using the independence of the different accident years, we obtain for the first term

$$\mathrm{Var}\left(C_{k,J}+C_{i,J}|\, \mathcal{D}_I\right) = \mathrm{Var}\left(C_{k,J}|\, \mathcal{D}_I\right) + \mathrm{Var}\left(C_{i,J}|\, \mathcal{D}_I\right). \tag{3.129}$$

This term is exactly the conditional process and prediction error from Lemma 3.19. For the second term (3.128) we obtain

$$\begin{aligned} & \left(\widehat{C_{k,J}}^{(CL,2)} + \widehat{C_{i,J}}^{(CL,2)} - E\left[C_{k,J}+C_{i,J}|\, \mathcal{D}_I\right]\right)^2 \\ & \quad = \left(\widehat{C_{k,J}}^{(CL,2)} - E\left[C_{k,J}|\, \mathcal{D}_I\right]\right)^2 + \left(\widehat{C_{i,J}}^{(CL,2)} - E\left[C_{i,J}|\, \mathcal{D}_I\right]\right)^2 \\ & \qquad + 2 \cdot \left(\widehat{C_{k,J}}^{(CL,2)} - E\left[C_{k,J}|\, \mathcal{D}_I\right]\right) \cdot \left(\widehat{C_{i,J}}^{(CL,2)} - E\left[C_{i,J}|\, \mathcal{D}_I\right]\right). \end{aligned} \tag{3.130}$$

Hence we have the following decomposition for the conditional mean square error of prediction error of the sum of two accident years

$$\begin{aligned} & E\left[\left(\widehat{C_{k,J}}^{(CL,2)} + \widehat{C_{i,J}}^{(CL,2)} - (C_{k,J}+C_{i,J})\right)^2 \middle| \mathcal{D}_I\right] \\ & \quad = E\left[\left(\widehat{C_{k,J}}^{(CL,2)} - C_{k,J}\right)^2 \middle| \mathcal{D}_I\right] + E\left[\left(\widehat{C_{i,J}}^{(CL,2)} - C_{i,J}\right)^2 \middle| \mathcal{D}_I\right] \\ & \qquad + 2 \cdot \left(\widehat{C_{k,J}}^{(CL,2)} - E\left[C_{k,J}|\, \mathcal{D}_I\right]\right) \cdot \left(\widehat{C_{i,J}}^{(CL,2)} - E\left[C_{i,J}|\, \mathcal{D}_I\right]\right). \end{aligned} \tag{3.131}$$

In addition to the conditional MSEP of single accident years (see Estimator 3.23), we need to average the covariance terms over the possible values of $\widehat{F}_j$ similar to (3.122):

$$\left(\widehat{C_{k,J}}^{(CL,2)} - E\left[C_{k,J}\middle|\mathcal{D}_I\right]\right)\cdot\left(\widehat{C_{i,J}}^{(CL,2)} - E\left[C_{i,J}\middle|\mathcal{D}_I\right]\right)$$
$$= C_{k,I-k}\cdot\left(\prod_{l=I-k}^{J-1}\widehat{F}_l - \prod_{l=I-k}^{J-1} f_l\right)\cdot C_{i,I-i}\cdot\left(\prod_{l=I-i}^{J-1}\widehat{F}_l - \prod_{l=I-i}^{J-1} f_l\right). \tag{3.132}$$

As in (3.125), using Approach 3, we obtain for the covariance term (3.132)

$$E^*_{\mathcal{D}_I}\left[C_{k,I-k}\cdot C_{i,I-i}\cdot\left(\prod_{j=I-k}^{J-1}\widehat{F}_j - \prod_{j=I-k}^{J-1} f_j\right)\cdot\left(\prod_{j=I-i}^{J-1}\widehat{F}_j - \prod_{j=I-i}^{J-1} f_j\right)\right]$$
$$= C_{k,I-k}\cdot C_{i,I-i}\cdot\prod_{j=I-i}^{I-k-1} f_j\cdot\left(\prod_{j=I-k}^{J-1} E^*_{\mathcal{D}_I}\left[\left(\widehat{F}_j\right)^2\right] - \prod_{j=I-k}^{J-1} f_j^2\right) \tag{3.133}$$
$$= C_{k,I-k}\cdot C_{i,I-i}\cdot\prod_{j=I-i}^{I-k-1} f_j\cdot\prod_{j=I-k}^{J-1} f_j^2$$
$$\cdot\left[\prod_{j=I-k}^{J-1}\left(\left(\sum_{m=0}^{i^*(j+1)}\frac{C_{m,j}}{\frac{\sigma_j^2}{f_j^2}+a_j^2\cdot C_{m,j}}\right)^{-1}+1\right)-1\right].$$

If we do the same linear approximation as in (3.126) the estimation of the covariance term is straightforward from (3.117).

Estimator 3.24 (MSEP for aggregated accident years)
Under Model Assumptions 3.16 we have the following estimator for the conditional mean square error of prediction of the ultimate claim for aggregated accident years

$$\widehat{msep}_{\sum_i C_{i,J}|\mathcal{D}_I}\left(\sum_{I-J+1}^{I}\widehat{C_{i,J}}^{(CL,2)}\right) = \sum_{i=I-J+1}^{I}\widehat{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{(CL,2)}\right) \tag{3.134}$$
$$+\,2\sum_{I-J+1\le k<i\le I}\widehat{C_{k,J}}^{(CL,2)}\cdot\widehat{C_{i,J}}^{(CL,2)}\cdot\sum_{j=I-k}^{J-1}\left(\sum_{m=0}^{i^*(j+1)}\frac{C_{m,j}}{\frac{\sigma_j^2}{\widehat{F}_j^2}+a_j^2\cdot C_{m,j}}\right)^{-1}.$$

Estimation Approach 2

In the derivation of the estimate $\widehat{F}_j$, see (4.62), we have seen that we face the problem, that the parameters need already be known in order to estimate them.

We could also use a different (unbiased) estimator. We define

$$\widehat{F_j}^{(0)} = \frac{\sum_{i=0}^{i^*(j+1)} C_{i,j} \cdot F_{i,j+1}}{\sum_{i=0}^{i^*(j+1)} C_{i,j}} = \frac{\sum_{i=0}^{i^*(j+1)} C_{i,j+1}}{\sum_{i=0}^{i^*(j+1)} C_{i,j}}. \tag{3.135}$$

$\widehat{F_j}^{(0)} = \widehat{f_j}$ is the classical chain-ladder estimator in the Mack Model 3.2. It is optimal under the Mack variance condition, but it is not optimal under our variance condition (3.90). Observe that

$$\begin{aligned}
\operatorname{Var}\left(\widehat{F_j}^{(0)} \middle| \mathcal{B}_j\right) &= \frac{1}{\left(\sum_{i=0}^{i^*(j+1)} C_{i,j}\right)^2} \sum_{i=0}^{i^*(j+1)} \operatorname{Var}\left(C_{i,j+1} \middle| C_{i,j}\right) \qquad (3.136)\\
&= \frac{\sum_{i=0}^{i^*(j+1)} \sigma_j^2 \cdot C_{i,j-1} + a_j^2 \cdot f_j^2 \cdot C_{i,j}^2}{\left(\sum_{i=0}^{i^*(j+1)} C_{i,j}\right)^2} \\
&= \frac{\sigma_j^2}{\sum_{i=0}^{i^*(j+1)} C_{i,j}} + \frac{a_j^2 \cdot f_j^2 \cdot \sum_{i=0}^{i^*(j+1)} C_{i,j}^2}{\left(\sum_{i=0}^{i^*(j+1)} C_{i,j}\right)^2}.
\end{aligned}$$

This immediately gives the following corollary:

Corollary 3.25 *Under Model Assumptions 3.16 we have for $i > I - J$ that*

$$C_{i,I-i} \cdot \prod_{j=I-i}^{J-1} \widehat{F_j}^{(0)} \tag{3.137}$$

defines a conditionally, given $C_{i,I-i}$, unbiased estimator for $E\left[C_{i,J} \middle| \mathcal{D}_I\right]$. The process variance and the prediction error is given by Lemma 3.19.

For the estimation error of a single accident year in Approach 3 we obtain the estimate

$$C_{i,I-i}^2 \cdot \left[\prod_{j=I-i}^{J-1} \left(\frac{\sigma_j^2}{\sum_{i=0}^{i^*(j+1)} C_{i,j}} + \frac{a_j^2 \cdot f_j^2 \cdot \sum_{i=0}^{i^*(j+1)} C_{i,j}^2}{\left(\sum_{i=0}^{i^*(j+1)} C_{i,j}\right)^2} + f_j^2 \right) - \prod_{j=I-i}^{J-1} f_j^2 \right]. \tag{3.138}$$

This expression is of course larger than the one obtained in (3.125).

3.3.7 Parameter estimation

We need to estimate three families of parameters f_j, σ_j and a_j. For Estimation Approach 1 the estimation of f_j is given in (3.115), which gives only an implicit expression for the estimation of f_j, since the chain-ladder factors appear also in the weights. Therefore we propose an iterative estimation in Approach 1 (on the other hand there is no difficulty in Estimation Approach 2).

Estimation of a_j. The sequence a_j can usually not be estimated from the data, unless we have a very large portfolio ($C_{i,j} \to \infty$), such that the conditional process error disappears. Hence a_j can only be obtained if we have data from the whole insurance market. This kind of considerations have been done for the determination of the parameters for prediction errors in the Swiss Solvency Test (see e.g. Tables 6.4.4 and 6.4.7 in [73]). Unfortunately, the tables only give an overall estimate for the conditional prediction error, not a sequence a_j (e.g. the variational coefficient of the overall error (similar to (3.101)) for motor third party liability claims reserves is 3.5%).

We reconstruct a_j with the help of (3.113). Define for $j = 0, \ldots, J-1$

$$V_j^2 = \sum_{m=j-1}^{J-1} a_m^2 \cdot \prod_{n=m+1}^{J-1} (1 + a_n^2). \tag{3.139}$$

Hence a_{j-1} can be determined recursively from $V_j^2 - V_{j+1}^2$:

$$a_{j-1}^2 = \left(V_j^2 - V_{j+1}^2\right) \prod_{n=j}^{J-1} (1 + a_n^2)^{-1}. \tag{3.140}$$

Henceforth, we can estimate a_{j-1} as soon as we have an estimate for V_j. V_j corresponds to

$$V_j = \lim_{C_{i,j-1} \to \infty} \mathrm{Vco}\left(C_{i,J} \middle| C_{i,j-1} \right) \tag{3.141}$$

(cf. (3.113)). Hence we need to estimate the conditional prediction error of $C_{i,J}$, given the observation $C_{i,j-1}$. Since we do not really have a good idea/guess about the conditional variational coefficient in (3.141) we express the conditional variational coefficient in terms of reserves

$$\begin{aligned}
\operatorname{Vco}\left(C_{i,J}\middle|\, C_{i,j-1}\right) &= \frac{\operatorname{Var}\left(C_{i,J}\middle|\, C_{i,j-1}\right)^{1/2}}{E\left[C_{i,J}\middle|\, C_{i,j-1}\right]} \qquad (3.142)\\
&= \frac{\operatorname{Var}\left(C_{i,J}-C_{i,j-1}\middle|\, C_{i,j-1}\right)^{1/2}}{E\left[C_{i,J}-C_{i,j-1}\middle|\, C_{i,j-1}\right]} \cdot \frac{E\left[C_{i,J}-C_{i,j-1}\middle|\, C_{i,j-1}\right]}{E\left[C_{i,J}\middle|\, C_{i,j-1}\right]}\\
&= \operatorname{Vco}\left(C_{i,J}-C_{i,j-1}\middle|\, C_{i,j-1}\right) \cdot \frac{\prod_{l=j-1}^{J-1} f_l - 1}{\prod_{l=j-1}^{J-1} f_l}.
\end{aligned}$$

In our examples we assume that the conditional variational coefficient for the conditional prediction error of the reserves $C_{i,J}-C_{i,j-1}$ is constant equal to r and we set

$$\widehat{V}_j = r \cdot \frac{\prod_{l=j-1}^{J-1} \widehat{F}_l^{(0)} - 1}{\prod_{l=j-1}^{J-1} \widehat{F}_l^{(0)}}. \qquad (3.143)$$

This immediately gives an estimate $\widehat{a}_j$ for the conditional prediction error a_j.

Estimation of σ_j. σ_j^2 is estimated iteratively from the data. A tedious calculation on conditional expectation gives

$$\begin{aligned}
&\frac{1}{i^*(j+1)} \sum_{i=0}^{i^*(j+1)} C_{i,j} \cdot E\left[\left(F_{i,j+1} - \widehat{F}_j^{(0)}\right)^2 \middle|\, \mathcal{B}_j\right] \qquad (3.144)\\
&= \sigma_j^2 + \frac{a_j^2 \cdot f_j^2}{i^*(j+1)} \left(\sum_{i=0}^{i^*(j+1)} C_{i,j} - \frac{\sum_{i=0}^{i^*(j+1)} C_{i,j}^2}{\sum_{i=0}^{i^*(j+1)} C_{i,j}} \right).
\end{aligned}$$

Hence we get the following iteration for the estimation of σ_j^2: For $k \geq 1$

$$\begin{aligned}
\widehat{\sigma_j^2}^{(k)} &= \frac{1}{i^*(j+1)} \sum_{i=0}^{i^*(j+1)} C_{i,j} \left(F_{i,j+1} - \widehat{F}_j^{(0)}\right)^2 \qquad (3.145)\\
&\quad - \frac{\widehat{a}_j^{\,2} \cdot \left(\widehat{F}_j^{(k-1)}\right)^2}{i^*(j+1)} \left(\sum_{i=0}^{i^*(j+1)} C_{i,j} - \frac{\sum_{i=0}^{i^*(j+1)} C_{i,j}^2}{\sum_{i=0}^{i^*(j+1)} C_{i,j}} \right).
\end{aligned}$$

If $\widehat{\sigma_j^2}^{(k)}$ becomes negative, it is set to 0, i.e. we only have a conditional prediction error and the conditional process error is equal to zero (the volume is sufficiently large, such that the conditional process error disappears).

Estimation of $\widehat{F}_j$. The estimators $\widehat{F}_j$ are then iteratively determined via (3.115). For $k \geq 1$

$$\widehat{F}_j^{(k)} = \frac{\sum_{i=0}^{i^*(j+1)} \frac{C_{i,j+1}}{\widehat{\sigma_j^2}^{(k)} + \widehat{a}_j^{\,2} \cdot \left(\widehat{F}_j^{(k-1)}\right)^2 \cdot C_{i,j}}}{\sum_{i=0}^{i^*(j+1)} \frac{C_{i,j}}{\widehat{\sigma_j^2}^{(k)} + \widehat{a}_j^{\,2} \cdot \left(\widehat{F}_j^{(k-1)}\right)^2 \cdot C_{i,j}}}. \tag{3.146}$$

Remarks 3.26

- In all examples we have looked at we have observed very fast convergence of $\widehat{\sigma_j^2}^{(k)}$ and $\widehat{F}_j^{(k)}$ in the sense that we have not observed any changes in the ultimates after three iterations for the $\widehat{F}_j$.

- To determine σ_j^2 we could also choose a different unbiased estimator

$$1 = \frac{1}{i^*(j+1)} \cdot \sum_{i=0}^{i^*(j+1)} \frac{C_{i,j}}{\sigma_j^2 + a_j^2 \cdot f_j^2 \cdot C_{i,j}} \cdot E\left[\left(F_{i,j+1} - \widehat{F_j} \right)^2 \middle| \mathcal{B}_j \right]. \tag{3.147}$$

  The difficulty with (3.147) is that it again leads to an implicit expression for $\widehat{\sigma_j^2}$.

- The formula for the MSEP, Estimator 3.24, was derived under the assumption that the underlying model parameters f_j, σ_j and a_j are known, if we replace these parameters by their estimates (as it is described via the iteration in this section) we obtain additional sources for the estimation errors! However, since calculations get too tedious (or even impossible) we omit further derivations of the MSEP and take Estimator 3.24 as a first approximation.

We close this section with an example:

Example 3.27 (MSEP in the enhanced chain-ladder model)

We choose two portfolios: Portfolio A and Portfolio B. Both are of similar type (i.e. consider the same line of business), moreover Portfolio B is contained in Portfolio A.

Portfolio A

	0	1	2	3	4	5	6	7	8	9	10
0	111'551	154'622	156'159	156'759	157'583	158'666	160'448	160'552	160'568	160'617	160'621
1	116'163	171'449	175'502	176'533	176'989	177'269	178'488	178'556	178'620	178'621	178'644
2	127'615	189'682	193'823	196'324	198'632	200'299	202'740	203'848	204'168	205'560	205'562
3	147'659	217'342	220'123	222'731	222'916	223'320	223'447	223'566	227'103	227'127	227'276
4	157'495	212'770	219'680	220'978	221'276	223'724	223'743	223'765	223'669	223'601	223'558
5	154'969	213'352	219'201	220'469	222'751	223'958	224'005	224'030	223'975	224'048	224'036
6	152'833	209'969	214'692	220'040	223'467	223'754	223'752	223'593	223'585	223'688	223'697
7	144'223	207'644	212'443	214'108	214'661	214'610	214'564	214'484	214'459	214'459	
8	145'612	209'604	214'161	215'982	217'962	220'783	221'078	221'614	221'616		
9	196'695	282'621	288'676	290'036	292'206	294'531	294'671	294'705			
10	181'381	260'308	266'497	269'130	269'404	269'691	269'720				
11	177'168	263'130	268'848	270'787	271'624	271'688					
12	156'505	230'607	237'102	244'847	245'940						
13	157'839	239'723	261'213	264'755							
14	159'429	233'309	239'800								
15	169'990	246'019									
16	173'377										

Table 3.7: Observed cumulative payments $C_{i,j}$ in Portfolio A

	0	1	2	3	4	5	6	7	8	9
$\widehat{f}_j$	1.4416	1.0278	1.0112	1.0057	1.0048	1.0025	1.0008	1.0020	1.0010	1.0001
$\widehat{\sigma}_j$	18.3478	8.7551	3.9082	2.2050	2.1491	2.0887	0.8302	2.4751	1.0757	0.1280

Table 3.8: Chain-ladder parameters in Mack's Model 3.2 for Portfolio A

This leads in Mack's Model 3.2 to the following reserves:

i	CL reserves	$\widehat{\text{msep}}_{C_{i,J}\mid\mathcal{D}_I}(\widehat{C_{i,J}}^{CL})^{1/2}$		$\widehat{\text{Var}}\left(C_{i,J}\mid\mathcal{D}_I\right)^{1/2}$		$\widehat{\text{Var}}\left(\widehat{C_{i,J}}^{CL}\mid\mathcal{D}_I\right)^{1/2}$	
7	20	64	322.0%	59	300.4%	23	115.8%
8	231	543	235.2%	510	220.8%	187	80.9%
9	898	1'582	176.1%	1'468	163.4%	589	65.5%
10	1'044	1'573	150.7%	1'470	140.9%	560	53.7%
11	1'731	1'957	113.1%	1'838	106.2%	674	38.9%
12	2'747	2'169	79.0%	2'055	74.8%	693	25.2%
13	4'487	2'563	57.1%	2'426	54.1%	826	18.4%
14	6'803	3'169	46.6%	3'030	44.5%	928	13.6%
15	14'025	5'663	40.4%	5'443	38.8%	1'564	11.2%
16	90'809	10'121	11.1%	9'762	10.8%	2'669	2.9%
Total	122'795	13'941	11.4%	12'336	10.0%	6'495	5.3%

Table 3.9: Reserves and conditional MSEP in Mack's Model 3.2 for Portfolio A

We compare these results now to the estimates in the Model 3.16: We set $r = 5\%$ and obtain the parameter estimates given below.

Remark. In practice a_j can only be determined with the help of external know how and market data. Therefore, e.g. for solvency purposes, a_j should be determined a priori by the regulator. It answers the question "how good can an actuarial estimate at most be?".

	0	1	2	3	4	5	6	7	8	9	10
$\widehat{F_j}^{(0)} = \widehat{f}_j$	1.4416	1.0278	1.0112	1.0057	1.0048	1.0025	1.0008	1.0020	1.0010	1.0001	
$\widehat{V_j}$	1.7187%	0.2697%	0.1379%	0.0833%	0.0552%	0.0316%	0.0193%	0.0152%	0.0052%	0.0005%	0.0000%
$\widehat{a}_j$	1.6974%	0.2317%	0.1099%	0.0624%	0.0453%	0.0251%	0.0119%	0.0143%	0.0052%	0.0005%	0.0000%

Table 3.10: Estimated $\widehat{a}_j$ and $\widehat{V}_j$ in Model 3.16

	0	1	2	3	4	5	6	7	8	9
$\widehat{F_j}^{(1)}$	1.44152	1.02784	1.01123	1.00572	1.00477	1.00249	1.00082	1.00200	1.00095	1.00009
$\widehat{F_j}^{(2)}$	1.44152	1.02784	1.01123	1.00572	1.00477	1.00249	1.00082	1.00200	1.00095	1.00009
$\widehat{F_j}^{(3)}$	1.44152	1.02784	1.01123	1.00572	1.00477	1.00249	1.00082	1.00200	1.00095	1.00009
$\widehat{\sigma_j}^{(1)}$	15.82901	8.68855	3.87516	2.18642	2.13924	2.08568	0.82851	2.47435	1.07546	0.12802
$\widehat{\sigma_j}^{(2)}$	15.82926	8.68856	3.87516	2.18642	2.13924	2.08568	0.82851	2.47435	1.07546	0.12802
$\widehat{\sigma_j}^{(3)}$	15.82926	8.68856	3.87516	2.18642	2.13924	2.08568	0.82851	2.47435	1.07546	0.12802

Table 3.11: Estimated parameters in Model 3.16 for Portfolio A

Already after 3 iterations the parameters have sufficiently converged such that the reserves are stable.

i	CL reserves	$\widehat{\text{msep}}_{C_{i,J}\|\mathcal{D}_I}(\widehat{C_{i,J}}^{(CL,2)})^{1/2}$		$\widehat{\text{Var}}\left(C_{i,J}\|\mathcal{D}_I\right)^{1/2}$		process error$^{1/2}$		prediction error$^{1/2}$		$\widehat{\text{Var}}\left(\widehat{C_{i,J}}^{(CL,2)}\|\mathcal{D}_I\right)^{1/2}$	
7	20	64	321.9%	59	300.4%	59	300.4%	1	5.0%	23	115.8%
8	231	543	235.1%	510	220.8%	510	220.7%	12	5.0%	187	80.9%
9	898	1'581	176.0%	1'468	163.4%	1'467	163.3%	45	5.0%	589	65.5%
10	1'044	1'573	150.7%	1'470	140.8%	1'469	140.7%	52	5.0%	560	53.7%
11	1'731	1'956	113.0%	1'836	106.1%	1'834	106.0%	87	5.0%	674	38.9%
12	2'747	2'165	78.8%	2'051	74.7%	2'046	74.5%	137	5.0%	693	25.2%
13	4'489	2'556	56.9%	2'418	53.9%	2'408	53.6%	224	5.0%	826	18.4%
14	6'804	3'153	46.3%	3'013	44.3%	2'994	44.0%	340	5.0%	928	13.6%
15	14'024	5'627	40.1%	5'405	38.5%	5'360	38.2%	701	5.0%	1'565	11.2%
16	90'796	9'244	10.2%	8'844	9.7%	7'590	8.4%	4'540	5.0%	2'688	3.0%
Total	122'784	13'298	10.8%	11'598	9.4%	10'640	8.7%	4'615	3.8%	6'504	5.3%

Table 3.12: Reserves and conditional MSEP in Model 3.16 for Portfolio A

Comment. The resulting reserves are almost the same in the Mack Model 3.2 and in Model 3.16. We obtain now both, a conditional process error and a conditional prediction error term. The sum of these two terms has about the same size as the conditional process error in Mack's method. This comes from the fact that we use the same data to estimate the parameters. But the error term in the enhanced chain-ladder model is now bounded from below by the conditional prediction error, whereas the conditional process error in the Mack model converges to zero for increasing volume.

Portfolio B

We choose now a second portfolio B, which includes similar business as our example given in Table 3.7 (Portfolio A). In fact, Portfolio B is a sub-portfolio of Portfolio A given in Table 3.13 containing exactly the same line of business. Therefore we assume that the conditional prediction errors are the same as in Table 3.12.

	0	1	2	3	4	5	6	7	8	9	10
0	53'095	73'067	74'548	75'076	75'894	76'128	77'904	78'008	78'022	78'071	78'075
1	59'183	87'679	89'303	90'033	90'058	90'303	91'454	91'472	91'482	91'483	91'494
2	64'640	95'734	97'648	99'429	100'462	101'683	103'549	104'642	104'917	105'560	105'560
3	72'150	105'349	106'546	106'919	106'934	107'144	107'170	107'225	107'232	107'232	107'232
4	76'272	105'630	108'406	108'677	108'838	110'140	110'110	110'111	110'155	110'155	110'110
5	75'469	105'987	108'779	109'093	111'366	111'390	111'422	111'448	111'367	111'369	111'369
6	78'835	108'835	111'455	116'231	117'896	118'161	118'157	117'940	117'940	117'972	117'974
7	70'780	98'753	101'347	102'624	102'629	102'587	102'545	102'500	102'474	102'474	
8	73'311	101'911	103'657	104'516	105'297	107'749	107'911	107'949	107'949		
9	102'741	144'167	147'211	147'777	149'506	149'753	149'865	149'899			
10	97'797	143'742	147'683	149'575	149'710	149'857	149'890				
11	98'682	147'042	151'029	151'960	152'645	152'682					
12	86'067	126'032	129'969	131'858	131'972						
13	87'013	131'721	150'062	152'883							
14	83'678	124'048	128'322								
15	90'415	129'970									
16	86'382										

Table 3.13: Observed cumulative payments $C_{i,j}$ Portfolio B

	0	1	2	3	4	5	6	7	8	9
$\widehat{F_j}^{(1)}$	1.43999	1.03310	1.01168	1.00632	1.00463	1.00415	1.00102	1.00026	1.00088	0.99996
$\widehat{F_j}^{(2)}$	1.43999	1.03310	1.01168	1.00632	1.00463	1.00415	1.00102	1.00026	1.00088	0.99996
$\widehat{F_j}^{(3)}$	1.43999	1.03310	1.01168	1.00632	1.00463	1.00415	1.00102	1.00026	1.00088	0.99996
$\widehat{\sigma_j}^{(1)}$	11.16524	10.86582	3.52827	2.24269	2.35370	2.63740	1.10600	0.30292	0.69058	0.05593
$\widehat{\sigma_j}^{(2)}$	11.16676	10.86582	3.52827	2.24269	2.35370	2.63740	1.10600	0.30292	0.69058	0.05593
$\widehat{\sigma_j}^{(3)}$	11.16676	10.86582	3.52827	2.24269	2.35370	2.63740	1.10600	0.30292	0.69058	0.05593

Table 3.14: Estimated parameters in Model 3.16 for Portfolio B

i	CL reserves	$\widehat{\mathrm{msep}}_{C_{i,J}\mid\mathcal{D}_I}(\widehat{C_{i,J}}^{(CL,2)})^{1/2}$		$\widehat{\mathrm{Var}}\left(C_{i,J}\mid\mathcal{D}_I\right)^{1/2}$		process error$^{1/2}$		prediction error$^{1/2}$		$\widehat{\mathrm{Var}}\left(\widehat{C_{i,J}}^{(CL,2)}\mid\mathcal{D}_I\right)^{1/2}$	
7	-4	19	-485.1%	18	-453.9%	18	-453.8%	0	-12.0%	7	-171.1%
8	91	242	265.6%	228	249.8%	228	249.7%	6	6.2%	82	90.5%
9	166	318	191.6%	293	176.4%	292	175.8%	23	13.7%	124	74.7%
10	320	557	174.4%	519	162.5%	518	162.2%	29	9.1%	202	63.3%
11	961	1'232	128.3%	1'159	120.6%	1'158	120.5%	49	5.1%	419	43.7%
12	1'445	1'453	100.6%	1'381	95.6%	1'379	95.4%	74	5.1%	452	31.3%
13	2'650	1'835	69.2%	1'734	65.4%	1'729	65.3%	130	4.9%	599	22.6%
14	3'749	2'144	57.2%	2'051	54.7%	2'043	54.5%	182	4.9%	625	16.7%
15	8'224	4'726	57.5%	4'545	55.3%	4'530	55.1%	373	4.5%	1'295	15.7%
16	45'878	5'885	12.8%	5'652	12.3%	5'175	11.3%	2'273	5.0%	1'639	3.6%
Total	63'480	8'933	14.1%	7'967	12.6%	7'623	12.0%	2'316	3.6%	4'041	6.4%

Table 3.15: Reserves and conditional MSEP in Model 3.16 for Portfolio B

Comments.

- The error terms between portfolio A and portfolio B are now directly comparable: The conditional prediction errors are the same for both portfolios. The conditional process error decreases now from portfolio B to portfolio A by about factor $\sqrt{2}$, since portfolio A has about twice the size of portfolio B. The conditional estimation error decreases from portfolio B to portfolio A since in portfolio A we have more data to estimate the parameters.

- The conditional prediction errors in portfolio A and portfolio B slightly differ since we choose different development factors $\widehat{F}_j$ and since the relative weights $C_{i,I-i}$ between the accident years i differ in portfolio A and portfolio B.

- A more conservative model would be to assume total dependence for the conditional prediction errors between the accident years, i.e. then we would not observe any diversification of the conditional prediction error between the accident years.

Chapter 4

Bayesian models

4.1 Introduction to credibility claims reserving methods

In the broadest sense, Bayesian methods for claims reserving can be considered as methods in which one combines expert knowledge or existing a priori information with observations resulting in an estimate for the ultimate claim. In the simplest case this a priori knowledge/information is given e.g. by a single value like an a priori estimate for the ultimate claim or for the average loss ratio (see this section and the following section). However, in a strict sense the a priori knowledge/information in Bayesian methods for claims reserving is given by an a priori distribution of a random quantity such as the ultimate claim or a risk parameter. The Bayesian inference is then understood to be the process of combining the a priori distribution of the random quantity with the data given in the upper trapezoid via Bayes' theorem. In this manner it is sometimes possible to obtain an analytic expression for the a posteriori distribution of the ultimate claim that reflects the change in the uncertainty due to the observations. The a posteriori expectation of the ultimate claim is then called the "Bayesian estimator" for the ultimate claim and minimizes the quadratic loss in the class of all estimators which are square integrable functions of the observations (see Section 4.2). In cases where we are not able to explicitly calculate the a posteriori expectation of the ultimate we restrict the search of the best estimator to the smaller class of estimators, which are linear functions of the observations (see Sections 4.3, 4.4 and 4.5).

4.1.1 Benktander-Hovinen method

This method goes back to Benktander [8] and Hovinen [37]. They have developed independently a method which leads to the same total estimated loss amount. Choose $i > I - J$. Assume we have an a priori estimate μ_i for $E[C_{i,J}]$ and that the claims development pattern $(\beta_j)_{0 \le j \le J}$ with $E[C_{i,j}] = \mu_i \cdot \beta_j$ is known. Since the Bornhuetter-Ferguson method completely ignores the observations $C_{i,I-i}$ on the last observed diagonal and the chain-ladder method completely ignores the a priori estimate μ_i at hand, one could consider a credibility mixture of these two methods (see (2.23)-(2.24)): For $c \in [0,1]$ we define the following credibility mixture

$$S_i(c) = c \cdot \widehat{C_{i,J}}^{CL} + (1-c) \cdot \mu_i \tag{4.1}$$

for $I - J + 1 \le i \le I$, where $\widehat{C_{i,J}}^{CL}$ is the chain-ladder estimate for the ultimate claim. The parameter c should increase with developing $C_{i,I-i}$ since we have more information in $C_{i,j}$ with increasing time. Benktander [8] proposed to choose $c = \beta_{I-i}$. This leads to the following estimator:

Estimator 4.1 (Benktander-Hovinen estimator) *The BH estimator is given by*

$$\widehat{C_{i,J}}^{BH} = C_{i,I-i} + (1-\beta_{I-i}) \cdot \left(\beta_{I-i} \cdot \widehat{C_{i,J}}^{CL} + (1-\beta_{I-i}) \cdot \mu_i \right) \tag{4.2}$$

for $I - J + 1 \le i \le I$.

Observe that we could again identify the claims development pattern $(\beta_j)_{0 \le j \le J}$ with the chain-ladder factors $(f_j)_{0 \le j < J}$. This can be done if we use Model Assumptions 2.9 for the Bornhuetter-Ferguson motivation, see also (2.22). Henceforth we identify in the sequel of this section

$$\beta_j = \frac{1}{\prod_{k=j}^{J-1} f_k}. \tag{4.3}$$

Since the development pattern β_j is known we also have (using (4.3)) known chain-ladder factors, which implies that we set

$$f_j = \widehat{f}_j \tag{4.4}$$

for $0 \le j \le J - 1$. Then the BH estimator can be written in the form

$$\begin{aligned} \widehat{C_{i,J}}^{BH} &= \beta_{I-i} \cdot \widehat{C_{i,J}}^{CL} + (1-\beta_{I-i}) \cdot \widehat{C_{i,J}}^{BF} && (4.5) \\ &= C_{i,I-i} + (1-\beta_{I-i}) \cdot \widehat{C_{i,J}}^{BF} && (4.6) \end{aligned}$$

(cf. (2.18) and (2.24)).

Remarks 4.2

- Equation (4.6) shows that the Benktander-Hovinen estimator can be seen as an iterated Bornhuetter-Ferguson estimator using the BF estimate as new a priori estimate.
- The following lemma shows that the weighting β_{I-i} is not a fixe point of our iteration since we have to evaluate the BH estimate at $1-(1-\beta_{I-i})^2$.

Lemma 4.3 *We have that*

$$\widehat{C_{i,J}}^{BH} \quad = \quad S_i\left(1-(1-\beta_{I-i})^2\right) \tag{4.7}$$

for $I-J+1 \le i \le I$.

Proof. It holds that

$$\begin{aligned}\widehat{C_{i,J}}^{BH} &= C_{i,I-i} + (1-\beta_{I-i})\cdot\left(\beta_{I-i}\cdot\widehat{C_{i,J}}^{CL} + (1-\beta_{I-i})\cdot\mu_i\right)\\ &= \beta_{I-i}\cdot\widehat{C_{i,J}}^{CL} + \left(\beta_{I-i}-\beta_{I-i}^2\right)\cdot\widehat{C_{i,J}}^{CL} + (1-\beta_{I-i})^2\cdot\mu_i \qquad (4.8)\\ &= \left(1-(1-\beta_{I-i})^2\right)\cdot\widehat{C_{i,J}}^{CL} + (1-\beta_{I-i})^2\cdot\mu_i = S_i\left(1-(1-\beta_{I-i})^2\right).\end{aligned}$$

This finishes the proof of the lemma.

□

Example 4.4 (Benktander-Hovinen method)

We revisit the data set given in Examples 2.7 and 2.11. We see that the Benktander-

						estimated reserves		
i	$C_{i,I-i}$	μ_i	β_{I-i}	$\widehat{C_{i,J}}^{CL}$	$\widehat{C_{i,J}}^{BH}$	CL	BH	BF
0	11'148'124	11'653'101	100.0%	11'148'124	11'148'124			
1	10'648'192	11'367'306	99.9%	10'663'318	10'663'319	15'126	15'127	16'124
2	10'635'751	10'962'965	99.8%	10'662'008	10'662'010	26'257	26'259	26'998
3	9'724'068	10'616'762	99.6%	9'758'606	9'758'617	34'538	34'549	37'575
4	9'786'916	11'044'881	99.1%	9'872'218	9'872'305	85'302	85'389	95'434
5	9'935'753	11'480'700	98.4%	10'092'247	10'092'581	156'494	156'828	178'024
6	9'282'022	11'413'572	97.0%	9'568'143	9'569'793	286'121	287'771	341'305
7	8'256'211	11'126'527	94.8%	8'705'378	8'711'824	449'167	455'612	574'089
8	7'648'729	10'986'548	88.0%	8'691'971	8'725'026	1'043'242	1'076'297	1'318'646
9	5'675'568	11'618'437	59.0%	9'626'383	9'961'926	3'950'815	4'286'358	4'768'384
Total						6'047'061	6'424'190	7'356'580

Table 4.1: Claims reserves from the Benktander-Hovinen method

Hovinen reserves are in between the chain-ladder reserves and the Bornhuetter-Ferguson reserves. They are closer to the chain-ladder reserves because β_{I-i} is larger than 50% for all accident years $i \in \{0,\dots,I\}$.

The next theorem is due to Mack [51]. It says that if we further iterate the BF method, we arrive at the chain-ladder reserve:

Theorem 4.5 (Mack [51]) *Choose $\widehat{C}^{(0)} = \mu_i$ and define for $m \geq 0$*

$$\widehat{C}^{(m+1)} = C_{i,I-i} + (1 - \beta_{I-i}) \cdot \widehat{C}^{(m)}. \tag{4.9}$$

If $\beta_{I-i} > 0$ then

$$\lim_{m \to \infty} \widehat{C}^{(m)} = \widehat{C_{i,J}}^{CL}. \tag{4.10}$$

Proof. For $m \geq 1$ we claim that

$$\widehat{C}^{(m)} = \left(1 - (1 - \beta_{I-i})^m\right) \cdot \widehat{C_{i,J}}^{CL} + (1 - \beta_{I-i})^m \cdot \mu_i. \tag{4.11}$$

The claim is true for $m = 1$ (BF estimator) and for $m = 2$ (BH estimator, see Lemma 4.3). Hence we prove the claim by induction. Induction step $m \to m + 1$:

$$\begin{aligned} \widehat{C}^{(m+1)} &= C_{i,I-i} + (1 - \beta_{I-i}) \cdot \widehat{C}^{(m)} \qquad (4.12) \\ &= C_{i,I-i} + (1 - \beta_{I-i}) \cdot \left(\left(1 - (1 - \beta_{I-i})^m\right) \cdot \widehat{C_{i,J}}^{CL} + (1 - \beta_{I-i})^m \cdot \mu_i \right) \\ &= \beta_{I-i} \cdot \widehat{C_{i,J}}^{CL} + \left((1 - \beta_{I-i}) - (1 - \beta_{I-i})^{m+1}\right) \widehat{C_{i,J}}^{CL} + (1 - \beta_{I-i})^{m+1} \mu_i, \end{aligned}$$

which proves (4.11). But from (4.11) the claim of the theorem immediately follows.

□

Example 4.4, revisited

In view of Theorem 4.5 we have

	$\widehat{C}^{(1)} = \widehat{C_{i,J}}^{BF}$	$\widehat{C}^{(2)} = \widehat{C_{i,J}}^{BH}$	$\widehat{C}^{(3)}$	$\widehat{C}^{(4)}$	$\widehat{C}^{(5)}$	...	$\widehat{C}^{(\infty)} = \widehat{C_{i,J}}^{CL}$
0	11'148'124	11'148'124	11'148'124	11'148'124	11'148'124	...	11'148'124
1	10'664'316	10'663'319	10'663'318	10'663'318	10'663'318		10'663'318
2	10'662'749	10'662'010	10'662'008	10'662'008	10'662'008		10'662'008
3	9'761'643	9'758'617	9'758'606	9'758'606	9'758'606		9'758'606
4	9'882'350	9'872'305	9'872'218	9'872'218	9'872'218	...	9'872'218
5	10'113'777	10'092'581	10'092'252	10'092'247	10'092'247		10'092'247
6	9'623'328	9'569'793	9'568'192	9'568'144	9'568'143		9'568'143
7	8'830'301	8'711'824	8'705'711	8'705'395	8'705'379		8'705'378
8	8'967'375	8'725'026	8'695'938	8'692'447	8'692'028		8'691'971
9	10'443'953	9'961'926	9'764'095	9'682'902	9'649'579	...	9'626'383

Table 4.2: Iteration of the Bornhuetter-Ferguson method

4.1.2 Minimizing quadratic loss functions

We choose $i > I - J$ and define the reserves for accident year i (see also (1.43))

$$R_i = C_{i,J} - C_{i,I-i}. \tag{4.13}$$

Hence under the assumption that the development pattern and the chain-ladder factors are known (and identified by (4.3) under Model Assumptions 2.9) the chain-ladder reserve and the Bornhuetter-Ferguson reserve are given by

$$\widehat{R_i}^{CL} = \widehat{C_{i,J}}^{CL} - C_{i,I-i} = C_{i,I-i} \cdot \left(\prod_{j=I-i}^{J-1} f_j - 1 \right), \tag{4.14}$$

$$\widehat{R_i}^{BF} = \widehat{C_{i,J}}^{BF} - C_{i,I-i} = (1 - \beta_{I-i}) \cdot \mu_i. \tag{4.15}$$

If we mix the chain-ladder and Bornhuetter-Ferguson methods we obtain for the credibility mixture ($c \in [0,1]$)

$$c \cdot \widehat{C_{i,J}}^{CL} + (1-c) \cdot \widehat{C_{i,J}}^{BF} \tag{4.16}$$

the following reserves

$$\begin{aligned} \widehat{R_i}(c) &= c \cdot \widehat{R_i}^{CL} + (1-c) \cdot \widehat{R_i}^{BF} \\ &= (1 - \beta_{I-i}) \cdot \left(c \cdot \widehat{C_{i,J}}^{CL} + (1-c) \cdot \mu_i \right) \\ &= \widehat{C_{i,J}}^{BF} - C_{i,I-i} + c \cdot (\widehat{C_{i,J}}^{CL} - \widehat{C_{i,J}}^{BF}) \end{aligned} \tag{4.17}$$

(see also (4.5)).

Question. Which is the optimality c? Optimal is defined in the sense of minimizing the quadratic error function. This means:

Our goal is to minimize the (unconditional) mean square error of prediction for the reserves $\widehat{R_i}(c)$

$$\mathrm{msep}_{R_i}\left(\widehat{R_i}(c)\right) = E\left[\left(R_i - \widehat{R_i}(c) \right)^2 \right] \tag{4.18}$$

(see also Section 3.1).

In order to do this minimization we need a proper stochastic model definition.

Model Assumptions 4.6

Assume that different accident years i are independent. There exists a sequence $(\beta_j)_{0\le j\le J}$ with $\beta_J = 1$ such that we have for all $j \in \{0,\dots,J\}$

$$E\left[C_{i,j}\right] = \beta_j \cdot E\left[C_{i,J}\right]. \tag{4.19}$$

Moreover, we assume that U_i is a random variable, which is unbiased for $E[C_{i,J}]$, i.e. $E[U_i] = E[C_{i,J}]$, and that U_i is independent of $C_{i,I-i}$ and $C_{i,J}$.

□

Remarks 4.7

- Model Assumptions 4.6 coincide with Model Assumptions 2.9 if we assume that $U_i = \mu_i$ is deterministic.
- Observe that we do not assume that the chain-ladder model is satisfied! The chain-ladder model satisfies Model Assumptions 4.6 but not necessarily vica versa. Assume that f_j is identified with β_j (via (4.3)) and that

  $$\widehat{C_{i,J}}^{CL} = \frac{C_{i,I-i}}{\beta_{I-i}} \qquad \text{and} \qquad \widehat{C_{i,J}}^{BF} = C_{i,I-i} + (1-\beta_{I-i})\cdot U_i. \tag{4.20}$$

  Hence the reserves are given by

  $$\widehat{R_i}(c) = (1-\beta_{I-i})\cdot\left(c\cdot\widehat{C_{i,J}}^{CL} + (1-c)\cdot U_i\right). \tag{4.21}$$

  Under these model assumptions and definitions we minimize

  $$\mathrm{msep}_{R_i}\left(\widehat{R_i}(c)\right) = E\left[\left(R_i - \widehat{R_i}(c)\right)^2\right]. \tag{4.22}$$

- Observe that also if we would assume that the chain-ladder model is satisfied we could not directly compare this situation to the mean square error of prediction calculation in Chapter 3. For the derivation of a MSEP formula for the chain-ladder method we have always assumed that the chain-ladder factors f_j are not known. If they would be known the mean square error of prediction of the chain-ladder reserves simply is given by (see (3.30))

  $$\begin{aligned}\mathrm{msep}_{C_{i,J}}\left(\widehat{C_{i,J}}^{CL}\right) &= E\left[E\left[\left.\left(C_{i,J} - \widehat{C_{i,J}}^{CL}\right)^2\right|\mathcal{D}_I\right]\right]\\ &= E\left[\mathrm{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}\right)\right]\\ &= E\left[\mathrm{Var}\left(\left.C_{i,J}\right|\mathcal{D}_I\right)\right]\\ &= \mathrm{Var}\left(C_{i,J}\right) - \mathrm{Var}\left(E\left[\left.C_{i,J}\right|\mathcal{D}_I\right]\right)\\ &= \mathrm{Var}\left(C_{i,J}\right) - \mathrm{Var}\left(C_{i,I-i}\right)\cdot\prod_{j=I-i}^{J-1} f_j^2.\end{aligned} \tag{4.23}$$

If we calculate (4.17) under Model Assumptions 4.6 and with (4.20) we obtain $E\big[\widehat{R}_i(c)\big] = E[R_i]$ and

$$\begin{aligned} \text{msep}_{R_i}\big(\widehat{R_i}(c)\big) &= \text{Var}\,(R_i) + E\left[\left(E[R_i] - \widehat{R_i}(c)\right)^2\right] \\ &\quad +2 \cdot E\left[\left(R_i - E[R_i]\right) \cdot \left(E[R_i] - \widehat{R_i}(c)\right)\right] \\ &= \text{Var}\,(R_i) + \text{Var}\left(\widehat{R_i}(c)\right) - 2 \cdot \text{Cov}\left(R_i, \widehat{R_i}(c)\right). \end{aligned} \tag{4.24}$$

Then we have the following theorem:

Theorem 4.8 (Mack [51]) *Under Model Assumptions 4.6 and* (4.20) *the optimal credibility factor* c_i^* *which minimizes the (unconditional) mean square error of prediction* (4.22) *is given by*

$$c_i^* = \frac{\beta_{I-i}}{1-\beta_{I-i}} \cdot \frac{Cov(C_{i,I-i}, R_i) + \beta_{I-i}(1-\beta_{I-i})\,Var(U_i)}{Var(C_{i,I-i}) + \beta_{I-i}^2\,Var(U_i)}. \tag{4.25}$$

Proof. We have that

$$\begin{aligned} E\left[\left(\widehat{R_i}(c_i) - R_i\right)^2\right] &= c_i^2 \cdot E\left[\left(\widehat{R_i}^{CL} - \widehat{R_i}^{BF}\right)^2\right] + E\left[\left(R_i - \widehat{R_i}^{BF}\right)^2\right] \\ &\quad -2 \cdot c_i \cdot E\left[\left(\widehat{R_i}^{CL} - \widehat{R_i}^{BF}\right)\left(R_i - \widehat{R_i}^{BF}\right)\right]. \end{aligned} \tag{4.26}$$

Hence the optimal c_i is given by

$$\begin{aligned} c_i^* &= \frac{E\left[\left(\widehat{R_i}^{CL} - \widehat{R_i}^{BF}\right)\left(R_i - \widehat{R_i}^{BF}\right)\right]}{E\left[\left(\widehat{R_i}^{CL} - \widehat{R_i}^{BF}\right)^2\right]} \qquad (4.27) \\ &= \frac{E\left[\left((1/\beta_{I-i} - 1)\cdot C_{i,I-i} - (1-\beta_{I-i})\cdot U_i\right)\left(R_i - (1-\beta_{I-i})\cdot U_i\right)\right]}{E\left[\left((1/\beta_{I-i} - 1)\cdot C_{i,I-i} - (1-\beta_{I-i})\cdot U_i\right)^2\right]} \\ &= \frac{\beta_{I-i}}{1-\beta_{I-i}} \cdot \frac{E\left[\left(C_{i,I-i} - \beta_{I-i}\cdot U_i\right)\left(R_i - (1-\beta_{I-i})\cdot U_i\right)\right]}{E\left[\left(C_{i,I-i} - \beta_{I-i}\cdot U_i\right)^2\right]}. \end{aligned}$$

Since $E[\beta_{I-i} \cdot U_i] = E[C_{i,I-i}]$ and $E[U_i] = E[C_{i,J}]$ we obtain

$$\begin{aligned} c_i^* &= \frac{\beta_{I-i}}{1-\beta_{I-i}} \cdot \frac{\text{Cov}(C_{i,I-i} - \beta_{I-i}\cdot U_i, R_i - (1-\beta_{I-i})\cdot U_i)}{\text{Var}(C_{i,I-i} - \beta_{I-i}\cdot U_i)} \qquad (4.28) \\ &= \frac{\beta_{I-i}}{1-\beta_{I-i}} \cdot \frac{\text{Cov}(C_{i,I-i}, R_i) + \beta_{I-i}\cdot(1-\beta_{I-i})\cdot\text{Var}(U_i)}{\text{Var}(C_{i,I-i}) + \beta_{I-i}^2\cdot\text{Var}(U_i)}. \end{aligned}$$

This finishes the proof.

□

We would like to mention once more that we have not considered the estimation errors in the claims development pattern β_j and f_j, respectively. In this sense Theorem 4.8 is a statement giving optimal credibility weights considering process variance and the uncertainty in the a priori estimate U_i.

Remark. To explicitly calculate c_i^* in Theorem 4.8 we need to specify an explicit stochastic model. We will do this below in Section 4.1.4, and close this subsection for the moment.

4.1.3 Cape-Cod Model

One main deficiency in the chain-ladder model is that the chain-ladder model completely depends on the last observation on the diagonal (see Chain-ladder Estimator 2.4). If this last observation is an outlier, this outlier will be projected to the ultimate claim (using the age-to-age factors). Often in long-tailed lines of business the first observations are not always representative. One possibility to smoothen outliers on the last observed diagonal is to combine BF and CL methods as e.g. in the Benktander-Hovinen method, another possibility is to robustify such observations. This is done in the Cape-Cod method. The Cape-Cod method goes back to Bühlmann [15].

Model Assumptions 4.9 (Cape-Cod method)

There exist parameters $\Pi_0, \ldots, \Pi_I > 0$, $\kappa > 0$ and a claims development pattern $(\beta_j)_{0 \le j \le J}$ with $\beta_J = 1$ such that

$$E[C_{i,j}] = \kappa \cdot \Pi_i \cdot \beta_j \tag{4.29}$$

for all $i = 0, \ldots, I$. Moreover, different accident years i are independent.

□

Observe that the Cape-Cod model assumptions coincide with Model Assumptions 2.9, set $\mu_i = \kappa \cdot \Pi_i$. For the Cape-Cod method we have described these new assumptions, to make clear the parameter Π_i can be interpreted as the premium in year i and κ reflects the average loss ratio. We assume that κ is independent of the accident year i, i.e. the premium level w.r.t. κ is the same for all accident years. Under (4.3) we can for each accident year estimate the loss ratio using the chain-ladder estimate

$$\widehat{\kappa}_i = \frac{\widehat{C_{i,J}}^{CL}}{\Pi_i} = \frac{C_{i,I-i}}{\prod_{j=I-i}^{J-1} f_j^{-1} \cdot \Pi_i} = \frac{C_{i,I-i}}{\beta_{I-i} \cdot \Pi_i}. \tag{4.30}$$

This is an unbiased estimate for κ,

$$E\left[\widehat{\kappa}_i\right] = \frac{1}{\Pi_i} \cdot E\left[\widehat{C_{i,J}}^{CL}\right] = \frac{1}{\Pi_i \cdot \beta_{I-i}} \cdot E\left[C_{i,I-i}\right] = \frac{1}{\Pi_i} \cdot E\left[C_{i,J}\right] = \kappa. \tag{4.31}$$

The robustified overall loss ratio is then estimate by

$$\widehat{\kappa}^{CC} = \sum_{i=0}^{I} \frac{\beta_{I-i} \cdot \Pi_i}{\sum_{k=0}^{I} \beta_{I-k} \cdot \Pi_k} \cdot \widehat{\kappa}_i = \frac{\sum_{i=0}^{I} C_{i,I-i}}{\sum_{i=0}^{I} \beta_{I-i} \cdot \Pi_i}. \tag{4.32}$$

Observe that $\widehat{\kappa}^{CC}$ is an unbiased estimate for κ.
A robustified value for $C_{i,I-i}$ is then found by $(i > I - J)$

$$\widehat{C_{i,I-i}}^{CC} = \widehat{\kappa}^{CC} \cdot \Pi_i \cdot \beta_{I-i}. \tag{4.33}$$

This leads to the Cape-Cod estimator:

Estimator 4.10 (Cape-Cod estimator) *The CC estimator is given by*

$$\widehat{C_{i,J}}^{CC} = C_{i,I-i} - \widehat{C_{i,I-i}}^{CC} + \prod_{j=I-i}^{J-1} f_j \cdot \widehat{C_{i,I-i}}^{CC} \tag{4.34}$$

for $I - J + 1 \leq i \leq I$.

Lemma 4.11 *Under Model Assumptions 4.9 and* (4.3) *the estimator* $\widehat{C_{i,J}}^{CC} - C_{i,I-i}$ *is unbiased for* $E\left[C_{i,J} - C_{i,I-i}\right] = \kappa \cdot \Pi_i \cdot (1 - \beta_{I-i})$.

Proof. Observe that

$$E\left[\widehat{C_{i,I-i}}^{CC}\right] = E\left[\widehat{\kappa}^{CC}\right] \cdot \Pi_i \cdot \beta_{I-i} = \kappa \cdot \Pi_i \cdot \beta_{I-i} = E\left[C_{i,I-i}\right]. \tag{4.35}$$

Moreover we have with (4.3) that

$$\widehat{C_{i,J}}^{CC} - C_{i,I-i} = \widehat{C_{i,I-i}}^{CC} \cdot \left(\prod_{j=I-i}^{J-1} f_j - 1 \right) = \widehat{\kappa}^{CC} \cdot \Pi_i \cdot (1 - \beta_{I-i}). \tag{4.36}$$

This finishes the proof.

□

Remarks 4.12

- The chain-ladder iteration is applied to the robustified diagonal value $\widehat{C_{i,I-i}}^{CC}$, but still we add the difference between original observation $C_{i,I-i}$ and robustified diagonal value in order to calculate the ultimate.

  If we transform the Cap-Code estimator we obtain (see also (4.36))

$$\widehat{C_{i,J}}^{CC} = C_{i,I-i} + (1 - \beta_{I-i}) \cdot \widehat{\kappa}^{CC} \cdot \Pi_i, \tag{4.37}$$

  which is a Bornhuetter-Ferguson type estimate with modified a priori estimate $\widehat{\kappa}^{CC} \cdot \Pi_i$.

- Observe that

$$\mathrm{Var}\left(\widehat{\kappa}_i\right) = \frac{1}{\Pi_i^2 \cdot \beta_{I-i}^2} \cdot \mathrm{Var}\left(C_{i,I-i}\right). \tag{4.38}$$

According to the choice of the variance function of $C_{i,j}$ this may also suggest that the robustification can be done in an other way (with smaller variance), see also Lemma 3.4.

Example 4.13 (Cape-Cod method)

We revisit the data set given in Examples 2.7, 2.11 and 4.4.

	Π_i	$\widehat{\kappa}_i$	$\widehat{C_{i,I-i}}^{CC}$	$\widehat{C_{i,J}}^{CC}$		estimated reserves Cape-Cod	CL	BF
0	15'473'558	72.0%	10'411'192	11'148'124		0	0	0
1	14'882'436	71.7%	9'999'259	10'662'396		14'204	15'126	16'124
2	14'456'039	73.8%	9'702'614	10'659'704		23'953	26'257	26'998
3	14'054'917	69.4%	9'423'208	9'757'538		33'469	34'538	37'575
4	14'525'373	68.0%	9'688'771	9'871'362		84'446	85'302	95'434
5	15'025'923	67.2%	9'953'237	10'092'522		156'769	156'494	178'024
6	14'832'965	64.5%	9'681'735	9'580'464		298'442	286'121	341'305
7	14'550'359	59.8%	9'284'898	8'761'342		505'131	449'167	574'089
8	14'461'781	60.1%	8'562'549	8'816'611		1'167'882	1'043'242	1'318'646
9	15'210'363	63.3%	6'033'871	9'875'801		4'200'233	3'950'815	4'768'384
	$\widehat{\kappa}^{CC}$	67.3%			Total	6'484'530	6'047'061	7'356'580

Table 4.3: Claims reserves from the Cape-Cod method

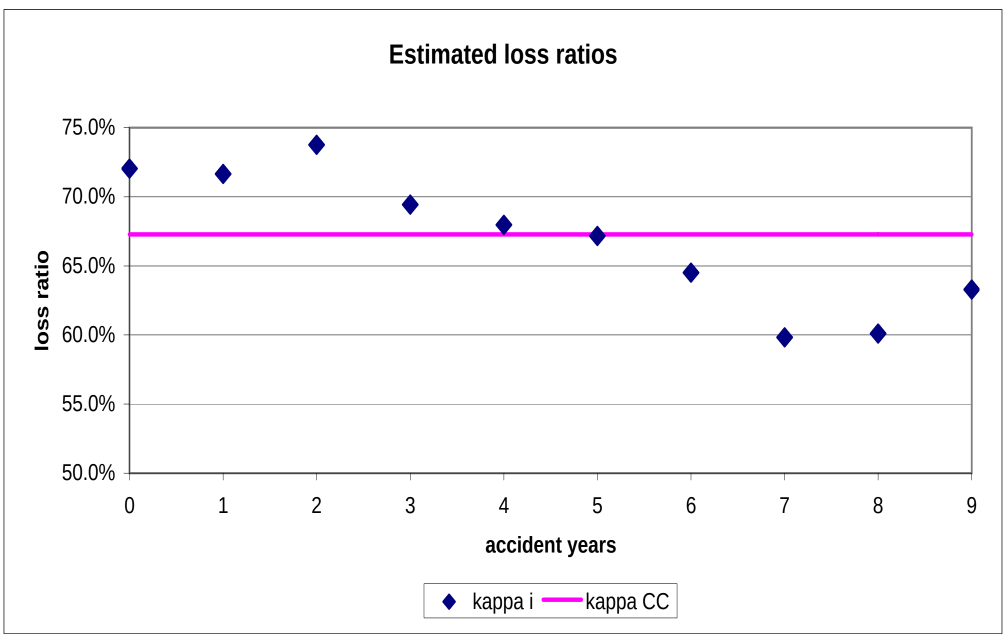


Figure 4.1: Estimated individual loss ratios $\widehat{\kappa}_i$ and estimated Cap-Code loss ratio $\widehat{\kappa}^{CC}$

The example shows the robustified diagonal values $\widehat{C_{i,I-i}}^{CC}$. This leads to the Cape-Cod estimate. The Cape-Cod estimate $\widehat{C_{i,J}}^{CC}$ is smaller than the Bornhuetter-Ferguson estimate $\widehat{C_{i,J}}^{BF}$. This comes from the fact, that the a priori estimates $\widehat{\mu}_i$ used for the Bornhuetter-Ferguson method are rather pessimistic. The loss ratios $\widehat{\mu}_i/\Pi_i$ are all above 75%, whereas the Cape-Cod method gives loss ratios $\widehat{\kappa}_i$, which are all below 75% (see Figure 4.1).
However, as Figure 4.1 shows: We have to be careful with the assumption of constant loss ratios κ. The figure suggests that we have to consider underwriting cycles carefully. In soft markets, loss ratios are rather low (we are able to charge rather high premiums). If there is a keen competition we expect low profit margins. If possible, we should adjust our premium with underwriting cycle information. For this reason one finds in practice modified versions of the Cape-Cod method, e.g. smoothening of the last observed diagonal is only done over neighboring values.

4.1.4 A distributional example to credible claims reserving

To construct the Benktander-Hovinen estimate we have used a credible weighting between the Bornhuetter-Ferguson method and the chain-ladder method. Theorem 4.8 gave a statement for the best weighted average (relative to the quadratic loss function). We now specify an explicit model to apply Theorem 4.8.

Model Assumptions 4.14 (Mack [51])

- Different accident years i are independent.
- There exists a sequence $(\beta_j)_{0\le j\le J}$ with $\beta_J = 1$ such that we have

$$E\left[C_{i,j}\middle|\, C_{i,J}\right] = \beta_j \cdot C_{i,J}, \tag{4.39}$$

$$\mathrm{Var}\left(C_{i,j}\middle|\, C_{i,J}\right) = \beta_j \cdot (1-\beta_j)\cdot \alpha^2(C_{i,J}) \tag{4.40}$$

  for all $i = 0, \ldots, I$ and $j = 0, \ldots, J$.

□

Remarks 4.15

- The spirit of this model is different from the Chain-ladder Model 3.2. In the chain-ladder model we have a "forward" iteration, i.e. we condition on the preceding observation. In the model above we have rather a "backward" iteration, conditioning on the ultimate $C_{i,J}$ we determine intermediate cumulative payment states, i.e. this is simply a refined definition of the development pattern.

- This model can also be viewed as a Bayesian approach, with latent variables which determine the ultimate claim $C_{i,J}$. This will be further discussed below.

- Observe that this model satisfies the Model Assumptions 2.9 with $\mu_i = E[C_{i,J}]$. Moreover $C_{i,j}$ satisfies the assumptions given in Model Assumptions 4.6. The chain-ladder model is in general not satisfied (see also Section 4.2.2, e.g. (4.68) below).

- Observe that the variance condition is such that it converges to zero for $\beta_j \to 1$, i.e. if the expected outstanding payments are low, also the uncertainty is low.

In view of Theorem 4.8 we have the following corollary (use definitions (4.21) and (4.20)):

Corollary 4.16 *Under the assumption that U_i is an unbiased estimator for $E[C_{i,J}]$ which is independent of $C_{i,I-i}$ and $C_{i,J}$ and Model Assumption 4.14 the optimal credibility factor c_i^* which minimizes the (unconditional) mean square error of prediction* (4.22) *is given by*

$$c_i^* = \frac{\beta_{I-i}}{\beta_{I-i} + t_i} \qquad \textit{with} \qquad t_i = \frac{E\left[\alpha^2(C_{i,J})\right]}{Var(U_i) + Var(C_{i,J}) - E\left[\alpha^2(C_{i,J})\right]} \tag{4.41}$$

for $i \in \{I-J+1, \ldots, I\}$.

Proof. From Theorem 4.8 we have

$$c_i^* = \frac{\beta_{I-i}}{1-\beta_{I-i}} \cdot \frac{\mathrm{Cov}(C_{i,I-i}, C_{i,J} - C_{i,I-i}) + \beta_{I-i} \cdot (1-\beta_{I-i}) \cdot \mathrm{Var}(U_i)}{\mathrm{Var}(C_{i,I-i}) + \beta_{I-i}^2 \cdot \mathrm{Var}(U_i)}. \tag{4.42}$$

Now we need to calculate the elements of the equation above. We obtain

$$\begin{aligned} \mathrm{Var}(C_{i,I-i}) &= E\left[\mathrm{Var}(C_{i,I-i}|C_{i,J})\right] + \mathrm{Var}\left(E[C_{i,I-i}|C_{i,J}]\right) \\ &= \beta_{I-i} \cdot (1-\beta_{I-i}) \cdot E\left[\alpha^2(C_{i,J})\right] + \beta_{I-i}^2 \mathrm{Var}\left(C_{i,J}\right), \end{aligned} \tag{4.43}$$

and

$$\mathrm{Cov}(C_{i,I-i}, C_{i,J} - C_{i,I-i}) = \mathrm{Cov}(C_{i,I-i}, C_{i,J}) - \mathrm{Var}(C_{i,I-i}). \tag{4.44}$$

Henceforth we need to calculate

$$\begin{aligned} \mathrm{Cov}(C_{i,I-i}, C_{i,J}) &= E\left[\mathrm{Cov}(C_{i,I-i}, C_{i,J}|C_{i,J})\right] + \mathrm{Cov}\left(E[C_{i,I-i}|C_{i,J}], E[C_{i,J}|C_{i,J}]\right) \\ &= 0 + \mathrm{Cov}\left(\beta_{I-i} \cdot C_{i,J}, C_{i,J}\right) = \beta_{I-i} \cdot \mathrm{Var}(C_{i,J}). \end{aligned} \tag{4.45}$$

This implies that

$$\mathrm{Cov}(C_{i,I-i}, C_{i,J} - C_{i,I-i}) = \beta_{I-i} \cdot \mathrm{Var}(C_{i,J}) - \mathrm{Var}(C_{i,I-i}). \tag{4.46}$$

Hence we obtain

$$\begin{aligned} c_i^* &= \frac{\beta_{I-i}}{1-\beta_{I-i}} \cdot \frac{\beta_{I-i} \cdot \mathrm{Var}(C_{i,J}) - \mathrm{Var}(C_{i,I-i}) + \beta_{I-i} \cdot (1-\beta_{I-i}) \cdot \mathrm{Var}(U_i)}{\mathrm{Var}(C_{i,I-i}) + \beta_{I-i}^2 \cdot \mathrm{Var}(U_i)} \\ &= \frac{\mathrm{Var}(C_{i,J}) - E\left[\alpha^2(C_{i,J})\right] + \mathrm{Var}(U_i)}{\left(\beta_{I-i}^{-1} - 1\right) \cdot E\left[\alpha^2(C_{i,J})\right] + \mathrm{Var}(C_{i,J}) + \mathrm{Var}(U_i)} \\ &= \frac{\mathrm{Var}(C_{i,J}) - E\left[\alpha^2(C_{i,J})\right] + \mathrm{Var}(U_i)}{\beta_{I-i}^{-1} \cdot E\left[\alpha^2(C_{i,J})\right] + \mathrm{Var}(C_{i,J}) - E\left[\alpha^2(C_{i,J})\right] + \mathrm{Var}(U_i)}. \end{aligned} \tag{4.47}$$

This finishes the proof of the corollary.

□

Corollary 4.17 *Under the assumption that U_i is an unbiased estimator for $E[C_{i,J}]$ which is independent of $C_{i,I-i}$ and $C_{i,J}$ and the Model Assumption 4.14 we find the following mean square errors of prediction (see also* (4.21)-(4.22)*):*

$$\begin{aligned} msep_{R_i}\left(\widehat{R_i}(c)\right) &= E\left[\alpha^2(C_{i,J})\right] \cdot \left(\frac{c^2}{\beta_{I-i}} + \frac{1}{1-\beta_{I-i}} + \frac{(1-c)^2}{t_i}\right) \cdot (1-\beta_{I-i})^2, \\ msep_{R_i}\left(\widehat{R_i}(0)\right) &= E\left[\alpha^2(C_{i,J})\right] \cdot \left(\frac{1}{1-\beta_{I-i}} + \frac{1}{t_i}\right) \cdot (1-\beta_{I-i})^2, \\ msep_{R_i}\left(\widehat{R_i}(1)\right) &= E\left[\alpha^2(C_{i,J})\right] \cdot \left(\frac{1}{\beta_{I-i}} + \frac{1}{1-\beta_{I-i}}\right) \cdot (1-\beta_{I-i})^2, \\ msep_{R_i}\left(\widehat{R_i}(c_i^*)\right) &= E\left[\alpha^2(C_{i,J})\right] \cdot \left(\frac{1}{\beta_{I-i}+t_i} + \frac{1}{1-\beta_{I-i}}\right) \cdot (1-\beta_{I-i})^2 \end{aligned} \tag{4.48}$$

for $i \in \{I-J+1, \ldots, I\}$.

Proof. Exercise.

□

Remarks 4.18

- The reserve $\widehat{R_i}(0)$ corresponds to the Bornhuetter-Ferguson reserve $\widehat{R_i}^{BF}$ and $\widehat{R_i}(1)$ corresponds to the chain-ladder reserve $\widehat{R_i}^{CL}$. However, $\mathrm{msep}_{R_i}\left(\widehat{R_i}(1)\right)$ and $\mathrm{msep}_{R_i}\left(\widehat{R_i}^{CL}\right)$ from Section 3 are not comparable since a) we use a completely different model, which leads to different process error and prediction error terms; b) in Corollary 4.17 we do not investigate the estimation error coming from the fact that we have to estimate f_j and β_j.

- From Corollary 4.17 we see that the Bornhuetter-Ferguson estimate in Model 4.14 is better than the chain-ladder estimate as long as

$$t_i > \beta_{I-i}. \tag{4.49}$$

I.e. for years with small loss experience β_{I-i} one should take the BF estimate whereas for older years one should take the CL estimate. Similar estimates can be derived for the BH estimate.

Example 4.19 (Mack model, Model Assumptions 4.14)

An easy distributional example satisfying Model Assumptions 4.14 is the following. Assume that, conditionally given $C_{i,J}$, $C_{i,j}/C_{i,J}$ has a $\text{Beta}\big(\alpha_i \cdot \beta_j, \alpha_i \cdot (1-\beta_j)\big)$-distribution. Hence

$$\begin{aligned} E\left[C_{i,j}\middle| C_{i,J}\right] &= C_{i,J} \cdot E\left[\frac{C_{i,j}}{C_{i.J}}\middle| C_{i,J}\right] = \beta_j \cdot C_{i,J}, & (4.50)\\ \text{Var}\left(C_{i,j}\middle| C_{i,J}\right) &= C_{i,J}^2 \cdot \text{Var}\left(\frac{C_{i,j}}{C_{i,J}}\middle| C_{i,J}\right) = \beta_j \cdot (1-\beta_j) \cdot \frac{C_{i,J}^2}{1+\alpha_i} & (4.51)\end{aligned}$$

for all $i = 0, \ldots, I$ and $j = 0, \ldots, J$.
See appendix, Section B.2.4, for the definition of the Beta distribution and its moments.
We revisit the data set given in Examples 2.7, 2.11 and 4.4. Observe that

$$E\left[\alpha^2(C_{i,J})\right] = \frac{1}{1+\alpha_i} \cdot E\left[C_{i,J}^2\right] = \frac{E\left[C_{i,J}\right]^2}{1+\alpha_i} \cdot \left(\text{Vco}^2\left(C_{i,J}\right) + 1\right). \tag{4.52}$$

As already mentioned before, we assume that the claims development pattern $(\beta_j)_{0 \le j \le J}$ is known. This means that in our estimates there is no estimation error term coming from the claims development parameters. We only have a process variance term and the uncertainty in the estimation of the ultimate U_i. This corresponds to a prediction error term in the language of Section 3.3. We assume that an actuary is able to predict the true a priori estimate with an error of 5%, i.e. $\text{Vco}(U_i) = 5\%$. Hence we assume that

$$\text{Vco}\left(C_{i,J}\right) = \left(\text{Vco}^2(U_i) + r^2\right)^{1/2}, \tag{4.53}$$

where $r = 6\%$ corresponds to the pure process error. This leads to the results in Table 4.4.
We already see from the choices of our parameters α_i, r and $\text{Vco}(U_i)$ that it is rather difficult to apply this method in practice, since we have not estimated these parameters from the data available.

	α_i	$\text{Vco}(U_i)$	r	$\text{Vco}(C_{i,J})$	t_i	c_i^*	estimated reserves			
								CL	BF	$\widehat{R_i}(c_i^*)$
0	600	5.0%	6.0%	7.8%	24.2%	80.5%		0	0	0
1	600	5.0%	6.0%	7.8%	24.2%	80.5%		15'126	16'124	15'320
2	600	5.0%	6.0%	7.8%	24.2%	80.5%		26'257	26'998	26'401
3	600	5.0%	6.0%	7.8%	24.2%	80.5%		34'538	37'575	35'131
4	600	5.0%	6.0%	7.8%	24.2%	80.4%		85'302	95'434	87'288
5	600	5.0%	6.0%	7.8%	24.2%	80.3%		156'494	178'024	160'738
6	600	5.0%	6.0%	7.8%	24.2%	80.1%		286'121	341'305	297'128
7	600	5.0%	6.0%	7.8%	24.2%	79.7%		449'167	574'089	474'538
8	600	5.0%	6.0%	7.8%	24.2%	78.5%		1'043'242	1'318'646	1'102'588
9	600	5.0%	6.0%	7.8%	24.2%	70.9%		3'950'815	4'768'384	4'188'531
							Total	6'047'061	7'356'580	6'387'663

Table 4.4: Claims reserves from Model 4.14

	$E[U_i]$	$E[\alpha^2(C_{i,J})]^{1/2}$	$\text{msep}^{1/2}(\widehat{R_i}(1))$	$\text{msep}^{1/2}(\widehat{R_i}(0))$	$\text{msep}^{1/2}(\widehat{R_i}(c_i^*))$
0	11'653'101	476'788			
1	11'367'306	465'094	17'529	17'568	17'527
2	10'962'965	448'551	22'287	22'373	22'282
3	10'616'762	434'386	25'888	26'031	25'879
4	11'044'881	451'902	42'189	42'751	42'153
5	11'480'700	469'734	58'952	60'340	58'862
6	11'413'572	466'987	81'990	85'604	81'745
7	11'126'527	455'243	106'183	113'911	105'626
8	10'986'548	449'515	166'013	190'514	163'852
9	11'618'437	475'369	396'616	500'223	372'199
total			457'811	560'159	435'814

Table 4.5: Mean square errors of prediction according to Corollary 4.17

The prediction errors are given in Table 4.5.
Observe that these mean square errors of prediction can not be compared to the mean square error of prediction obtained in the chain-ladder method (see Section 3). We do not know whether the model assumptions in this example imply the chain-ladder model assumptions. Moreover, we do not investigate the uncertainties in the parameter estimates, and the choice of the parameters was rather artificial, motivated by expert opinions.

□

4.2 Exact Bayesian models

4.2.1 Motivation

Bayesian methods for claims reserving are methods in which one combines a priori information or expert knowledge with observations in the upper trapezoid $\mathcal{D}_I$. Available information/knowledge is incorporated through an a priori distribution of a random quantity such as the ultimate claim (see Sections 4.2.2 and 4.2.3) or a risk parameter (see Section 4.2.4) which must be modeled by the actuary. This distribution is then connected with the likelihood function via Bayes' theorem. If we use a smart choice for the distribution of the observations and the a priori distribution such as the exponential dispersion family (EDF) and its associate conjugates (see Section 4.2.4), we are able to derive an analytic expression for the a posteriori distribution of the ultimate claim. This means that we can compute the a posteriori expectation $E\left[C_{i,J}|\mathcal{D}_I\right]$ of the ultimate claim $C_{i,J}$ which is called "Bayesian estimator" for the ultimate claim, given the observations $\mathcal{D}_I$. The Bayesian method is called exact since the Bayesian estimator $E\left[C_{i,J}|\mathcal{D}_I\right]$ is optimal in the sense that it minimizes the squared loss function (MSEP) in the class $L^2_{C_{i,J}}(\mathcal{D}_I)$ of all estimators for $C_{i,J}$ which are square integrable functions of the observations in $\mathcal{D}_I$, i.e.

$$E\left[C_{i,J}|\mathcal{D}_I\right] = \underset{Y \in L^2_{C_{i,J}}(\mathcal{D}_I)}{\operatorname{argmin}} E\left[\left.(C_{i,J}-Y)^2\right|\mathcal{D}_I\right]. \tag{4.54}$$

For its conditional mean square error of prediction we have that

$$\operatorname{msep}_{C_{i,J}|\mathcal{D}_I}\left(E\left[C_{i,J}|\mathcal{D}_I\right]\right) = \operatorname{Var}(C_{i,J}|\mathcal{D}_I). \tag{4.55}$$

Of course, if there are unknown parameters in the underlying probabilistic model, we can not explicitly calculate $E\left[C_{i,J}|\mathcal{D}_I\right]$. These parameters need to be estimated by $\mathcal{D}_I$-measurable estimators. Hence we obtain a $\mathcal{D}_I$-measurable estimator $\widehat{E}\left[C_{i,J}|\mathcal{D}_I\right]$ for $E\left[C_{i,J}|\mathcal{D}_I\right]$ (and $C_{i,J}|\mathcal{D}_I$, resp.) which implies that

$$\operatorname{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{E}\left[C_{i,J}|\mathcal{D}_I\right]\right) = \operatorname{Var}(C_{i,J}|\mathcal{D}_I) + \left(\widehat{E}\left[C_{i,J}|\mathcal{D}_I\right] - E\left[C_{i,J}|\mathcal{D}_I\right]\right)^2, \tag{4.56}$$

and now we are in a similar situation as in the chain-ladder model, see (3.30).

We close this section with some remarks: For pricing and tariffication of insurance contracts Bayesian ideas and techniques are well investigated and widely used in practice. For the claims reserving problem Bayesian methods are less used although we believe that they are very useful for answering practical questions (this has e.g. already be mentioned in de Alba [2]).

In the literature exact Bayesian models have been studied e.g. in a series of papers by Verrall [79, 81, 82], de Alba [2, 4], de Alba-Corzo [3], Gogol [30], Haastrup-Arjas [32] Ntzoufras-Dellaportas [60] and the corresponding implementation by Scollnik [70]. Many of these results refer to explicit choices of distributions, e.g. the Poisson-gamma or the (log-)normal-normal cases are considered. Below, we give an approach which suites for rather general distributions (see Section 4.2.4).

4.2.2 Log-normal/Log-normal model

In this section we revisit Model 4.14. We make distributional assumptions on $C_{i,J}$ and $C_{i,j}|C_{i,J}$ which satisfy Model Assumptions 4.14 and hence would allow for the application of Corollary 4.16. However, in this section we don't follow that route: In Corollary 4.16 we have specified a second distribution for an a priori estimate $E[U_i]$ which then led to Corollary 4.16.
Here, we don't use a distribution for the a priori estimate, but we explicitly specify the distribution of $C_{i,J}$. The distributional assumptions will be such that we can determine the exact distribution of $C_{i,J}|C_{i,j}$ according to Bayes' theorem. It figures out that the best estimate for $E\left[C_{i,J}\middle|\, C_{i,I-i}\right]$ is a crediblity mixture between the observation $C_{i,I-i}$ and the a priori mean $E[C_{i,J}]$. Gogol [30] proposed the following model.

Model Assumptions 4.20 (Log-normal/Log-normal model)

- Different accident years i are independent.
- $C_{i,J}$ is log-normally distributed with parameters $\mu^{(i)}$ and σ_i^2 for $i = 0, \ldots, I$.
- Conditionally, given $C_{i,J}$, $C_{i,j}$ has a Log-normal distribution with parameters $\nu_j(C_{i,J})$ and $\tau_j^2(C_{i,J})$ for $i = 0, \ldots, I$ and $j = 0, \ldots, J$.

□

Remark. In appendix, Section B.2.2, we provide the definition of the Log-normal distribution. $\mu^{(i)}$ and σ_i^2 denote the parameters of the Log-normal distribution of $C_{i,J}$, with

$$\mu_i = E[C_{i,J}] = \exp\left\{\mu^{(i)} + \sigma_i^2/2\right\} \tag{4.57}$$

is the a priori mean of $C_{i,J}$.

If $(C_{i,j})_{0\le j\le J}$ also satisfies Model Assumptions 4.14 we have that

$$\begin{aligned} E\left[C_{i,j}\middle|\, C_{i,J}\right] &= \exp\left\{\nu_j + \tau_j^2/2\right\} \stackrel{!}{=} \beta_j \cdot C_{i,J}, \qquad (4.58)\\ \operatorname{Var}\left(C_{i,j}\middle|\, C_{i,J}\right) &= \exp\left\{2\cdot\nu_j + \tau_j^2\right\}\cdot\left(\exp\{\tau_j^2\} - 1\right) \stackrel{!}{=} \beta_j\cdot(1-\beta_j)\cdot\alpha^2(C_{i,J}). \end{aligned}$$

This implies that we have to choose

$$\tau_j^2 = \tau_j^2(C_{i,J}) = \log\left(1 + \frac{1-\beta_j}{\beta_j}\cdot\frac{\alpha^2(C_{i,J})}{C_{i,J}^2}\right), \tag{4.59}$$

$$\nu_j = \nu_j(C_{i,J}) = \log\left(\beta_j\cdot C_{i,J}\right) - \frac{1}{2}\cdot\log\left(1 + \frac{1-\beta_j}{\beta_j}\cdot\frac{\alpha^2(C_{i,J})}{C_{i,J}^2}\right). \tag{4.60}$$

The joint distribution of $(C_{i,j}, C_{i,J})$ is given by

$$\begin{aligned} f_{C_{i,j},C_{i,J}}(x,y) &= f_{C_{i,j}|C_{i,J}}(x|y) \cdot f_{C_{i,J}}(y) \\ &= \frac{1}{(2\pi)^{1/2}\tau_j(y)} \frac{1}{x} \exp\left\{ -\frac{1}{2} \left(\frac{\log(x) - \nu_j(y)}{\tau_j(y)} \right)^2 \right\} \\ &\quad \cdot \frac{1}{(2\pi)^{1/2}\sigma_i} \frac{1}{y} \exp\left\{ -\frac{1}{2} \left(\frac{\log(y) - \mu^{(i)}}{\sigma_i} \right)^2 \right\} \\ &= \frac{1}{2\pi \cdot \sigma_i \cdot \tau_j(y)} \frac{1}{x \cdot y} \exp\left\{ -\frac{1}{2} \left(\frac{\log(x) - \nu_j(y)}{\tau_j(y)} \right)^2 - \frac{1}{2} \left(\frac{\log(y) - \mu^{(i)}}{\sigma_i} \right)^2 \right\}. \end{aligned} \tag{4.61}$$

Lemma 4.21 *The Model Assumptions 4.20 combined with Model Assumptions 4.14 with $\alpha^2(c) = a^2 \cdot c^2$ for some $a \in \mathbb{R}$ satisfies the following equalities*

$$\tau_j^2(c) = \tau_j^2 = \log\left(1 + \frac{1-\beta_j}{\beta_j} \cdot a^2\right), \tag{4.62}$$

$$\nu_j(c) = \log c + \log \beta_j - \tau_j^2/2. \tag{4.63}$$

Moreover, the conditional distribution of $C_{i,J}$ given $C_{i,j}$ is again a Log-normal distribution with updated parameters

$$\mu_{post(i,j)} = \left(1 - \frac{\tau_j^2}{\sigma_i^2 + \tau_j^2}\right) \cdot \left(\tau_j^2/2 + \log(C_{i,j}/\beta_j)\right) + \frac{\tau_j^2}{\sigma_i^2 + \tau_j^2} \cdot \mu^{(i)}, \tag{4.64}$$

$$\sigma^2_{post(i,j)} = \frac{\tau_j^2}{\sigma_i^2 + \tau_j^2} \cdot \sigma_i^2. \tag{4.65}$$

Remarks 4.22

- This example shows a typical Bayesian and credibility result: i) In this example of conjugated distributions we can exactly calculate the a posteriori distribution of the ultimate claim $C_{i,J}$ given the information $C_{i,j}$ (cf. Section 4.2.4 and see also Bühlmann-Gisler [18]). ii) We see that we need to update the parameter $\mu^{(i)}$ by choosing a credibility weighted average of the a priori parameter $\mu^{(i)}$ and the transformed observation $\tau_j^2/2 + \log(C_{i,j}/\beta_j)$, where the credibility weight is given by

$$\alpha_{i,j} = \sigma_i^2/(\sigma_i^2 + \tau_j^2). \tag{4.66}$$

  This implies the updating of the a priori mean of the ultimate claim $C_{i,J}$

$$E[C_{i,J}] = \exp\{\mu^{(i)} + \sigma_i^2/2\} \tag{4.67}$$

to the a posteriori mean of the ultimate claim $C_{i,J}$

$$\begin{aligned} E[C_{i,J}|C_{i,j}] &= \exp\{\mu_{post(i,j)} + \sigma^2_{post(i,j)}/2\} \\ &= \exp\left\{(1-\alpha_{i,j})\cdot\left(\mu^{(i)}+\sigma_i^2/2\right) + \alpha_{i,j}\cdot\left(\log(C_{i,j}/\beta_j)+\tau_j^2/2\right)\right\} \\ &= \exp\left\{(1-\alpha_{i,j})\cdot\left(\mu^{(i)}+\sigma_i^2/2\right) + \alpha_{i,j}\cdot\left(-\log\beta_j+\tau_j^2/2\right)\right\}\cdot C_{i,j}^{\frac{\sigma_i^2}{\sigma_i^2+\tau_j^2}}, \end{aligned} \tag{4.68}$$

see also (4.83) below.

- Observe that this model does in general not satisfy the chain-ladder assumptions (cf. last expression in (4.68)). This has already been mentioned in Remarks 4.15.

- Observe that in the current derivation we only consider one observation $C_{i,j}$. We could also consider the whole sequence of observations $C_{i,0},\ldots,C_{i,j}$ then the a posteriori distribution of $C_{i,J}$ is log-normally distributed with mean

$$\begin{aligned} \mu^*_{post(i,j)} &= \frac{\sum_{k=0}^{j}\frac{\log C_{i,k}-\log\beta_k+\tau_k^2}{\tau_k^2}+\frac{\mu^{(i)}}{\sigma_i^2}}{\sum_{k=0}^{j}\frac{1}{\tau_k^2}+\frac{1}{\sigma_i^2}} \\ &= \alpha^*_{i,j}\cdot\frac{1}{\sum_{k=0}^{j}\frac{1}{\tau_k^2}}\cdot\sum_{k=0}^{j}\frac{\log C_{i,k}-\log\beta_k+\tau_k^2}{\tau_k^2}+\left(1-\alpha^*_{i,j}\right)\cdot\mu^{(i)}, \end{aligned} \tag{4.69}$$

with

$$\alpha^*_{i,j} = \frac{\sum_{k=0}^{j}\frac{1}{\tau_k^2}}{\sum_{k=0}^{j}\frac{1}{\tau_k^2}+\frac{1}{\sigma_i^2}}, \tag{4.70}$$

and variance

$$\sigma^{2,*}_{post(i,j)} = \left[\sum_{k=0}^{j}\frac{1}{\tau_k^2}+\frac{1}{\sigma_i^2}\right]^{-1}. \tag{4.71}$$

Observe that this is again a credibility weighted average between the a priori estimate $\mu^{(i)}$ and the observations $C_{i,0},\ldots,C_{i,j}$. The credibility weights are given by $\alpha^*_{i,j}$. Moreover, observe that this model does not have the Markov property, this is in contrast to our chain-ladder assumptions.

Proof of Lemma 4.21. The equations (4.62)-(4.63) easily follow from (4.59)-(4.60). Hence we only need to calculate the conditional distribution of $C_{i,J}$ given $C_{i,j}$. From (4.61) and (4.63) we see that the joint density of $(C_{i,j}, C_{i,J})$ is given by

$$\begin{aligned} f_{C_{i,j},C_{i,J}}(x,y) &= \frac{1}{2\pi\cdot\sigma_i\cdot\tau_j}\cdot\frac{1}{x\cdot y} \\ &\cdot\exp\left\{-\frac{1}{2}\left(\frac{\log(x)-\log(y)-\log\beta_j+\tau_j^2/2}{\tau_j}\right)^2-\frac{1}{2}\left(\frac{\log(y)-\mu^{(i)}}{\sigma_i}\right)^2\right\}. \end{aligned} \tag{4.72}$$

Now we have that

$$\left(\frac{z-c}{\tau}\right)^2+\left(\frac{z-\mu}{\sigma}\right)^2=\frac{\left(z-\frac{\sigma^2c+\tau^2\mu}{\sigma^2+\tau^2}\right)^2}{\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}}+\frac{(\mu-c)^2}{\sigma^2+\tau^2}. \tag{4.73}$$

This implies that

$$\begin{aligned} f_{C_{i,j},C_{i,J}}(x,y) &= \frac{1}{2\pi\cdot\sigma_i\cdot\tau_j}\cdot\frac{1}{x\cdot y} \\ &\cdot\exp\left\{-\frac{1}{2}\left(\frac{\left(\log y-\frac{\sigma_i^2c(x)+\tau_j^2\mu^{(i)}}{\sigma_i^2+\tau_j^2}\right)^2}{\frac{\sigma_i^2\tau_j^2}{\sigma_i^2+\tau_j^2}}+\frac{\left(\mu^{(i)}-c(x)\right)^2}{\sigma_i^2+\tau_j^2}\right)\right\}, \end{aligned} \tag{4.74}$$

where

$$c(x)=\log(x)-\log\beta_j+\tau_j^2/2. \tag{4.75}$$

From this we see that

$$f_{C_{i,J}|C_{i,j}}(y|x)=\frac{f_{C_{i,j},C_{i,J}}(x,y)}{f_{C_{i,j}}(x)}=\frac{f_{C_{i,j},C_{i,J}}(x,y)}{\int f_{C_{i,j},C_{i,J}}(x,y)dy} \tag{4.76}$$

is the density of a Log-normal distribution with parameters

$$\mu_{post(i,j)} = \frac{\sigma_i^2\cdot c(C_{i,j})+\tau_j^2\cdot\mu^{(i)}}{\sigma_i^2+\tau_j^2}, \tag{4.77}$$

$$\sigma^2_{post(i,j)} = \frac{\sigma_i^2\cdot\tau_j^2}{\sigma_i^2+\tau_j^2}. \tag{4.78}$$

Finally, we rewrite $\mu_{post(i,j)}$ (cf. (4.75)):

$$\mu_{post(i,j)} = \frac{\sigma_i^2\cdot\left(\log(C_{i,j})-\log\beta_j+\tau_j^2/2\right)+\tau_j^2\cdot\mu^{(i)}}{\sigma_i^2+\tau_j^2}. \tag{4.79}$$

This finishes the proof of Lemma 4.21.

□

Estimator 4.23 (Log-normal/Log-normal model, Gogol [30])
Under the assumptions of Lemma 4.21 we have the following estimator for the ultimate claim $E\left[C_{i,J}\middle|\,C_{i,I-i}\right]$

$$\begin{aligned} \widehat{C_{i,J}}^{Go} &= E\left[C_{i,J}\middle|\,C_{i,I-i}\right] \\ &= \exp\left\{\left(1-\frac{\sigma_i^2}{\sigma_i^2+\tau_{I-i}^2}\right)\cdot\left(\mu^{(i)}+\frac{\sigma_i^2}{2}\right)+\frac{\sigma_i^2}{\sigma_i^2+\tau_{I-i}^2}\cdot\left(\log\left(\frac{C_{i,I-i}}{\beta_{I-i}}\right)+\frac{\tau_{I-i}^2}{2}\right)\right\} \end{aligned} \tag{4.80}$$

for $I-J+1\le i\le I$.

Observe that we only condition on the last observation $C_{i,I-i}$, see also Remarks 4.22 on Markov property.

Remark. We could also consider

$$\widehat{C_{i,J}}^{Go,2} = C_{i,I-i} + (1-\beta_{I-i}) \cdot \widehat{C_{i,J}}^{Go}. \tag{4.81}$$

From a practical point of view $\widehat{C_{i,J}}^{Go,2}$ is more useful, if we have an outlier on the diagonal. However, both estimators are not easily obtained in practice, since there are too many parameters which are difficult to estimate.

Example 4.24 (Model Gogol [30], Assumptions of Lemma 4.21)

We revisit the data set given in Example 2.7. For the parameters we do the same choices as in Example 4.19 (see Table 4.4). I.e. we set $\text{Vco}(C_{i,J})$ equal to the value obtained in (4.53). In formula (4.53) this variational coefficient was decomposed into process error and parameter uncertainties, here we only use the overall uncertainty. Moreover we choose $a^2 = \frac{1}{1+\alpha_i}$ with $\alpha_i = 600$. Using (4.62), (4.57) and

$$\sigma_i^2 = \ln\big(\text{Vco}^2(C_{i,J}) + 1\big) \tag{4.82}$$

(cf. appendix, Table B.5) leads to Table 4.6.

	$\mu_i = E[C_{i,J}]$	$\text{Vco}(C_{i,J})$	$\mu^{(i)}$	σ_i	β_{I-i}	a^2	τ_{I-i}
0	11'653'101	7.8%	16.27	7.80%	100.0%	0.17%	0.0%
1	11'367'306	7.8%	16.24	7.80%	99.9%	0.17%	0.2%
2	10'962'965	7.8%	16.21	7.80%	99.8%	0.17%	0.2%
3	10'616'762	7.8%	16.17	7.80%	99.6%	0.17%	0.2%
4	11'044'881	7.8%	16.21	7.80%	99.1%	0.17%	0.4%
5	11'480'700	7.8%	16.25	7.80%	98.4%	0.17%	0.5%
6	11'413'572	7.8%	16.25	7.80%	97.0%	0.17%	0.7%
7	11'126'527	7.8%	16.22	7.80%	94.8%	0.17%	1.0%
8	10'986'548	7.8%	16.21	7.80%	88.0%	0.17%	1.5%
9	11'618'437	7.8%	16.27	7.80%	59.0%	0.17%	3.4%

Table 4.6: Parameter choice for the Log-normal/Log-normal model

We obtain the credibility weights and estimates for the ultimates in Table 4.7. Using (4.66), (4.57) and $\widehat{C_{i,J}}^{CL} = \frac{C_{i,I-i}}{\beta_{I-i}}$ we obtain for Estimator 4.23 the following representation:

$$\begin{aligned} \widehat{C_{i,J}}^{Go} &= \exp\left\{(1-\alpha_{i,I-i}) \cdot \big(\mu^{(i)} + \sigma_i^2/2\big) + \alpha_{i,I-i} \cdot \left(\log\left(\frac{C_{i,I-i}}{\beta_{I-i}}\right) + \tau_{I-i}^2/2\right)\right\} \\ &= \mu_i^{1-\alpha_{i,I-i}} \cdot \exp\left\{\log \widehat{C_{i,J}}^{CL} + \tau_{I-i}^2/2\right\}^{\alpha_{i,I-i}}. \end{aligned} \tag{4.83}$$

	$C_{i,I-i}$	$1-\alpha_{i,I-i}$	$\mu_{post(i,I-i)}$	$\sigma_{post(i,I-i)}$	$\widehat{C_{i,J}}^{Go}$	estimated reserves Go	CL	BF
0	11'148'124	0.0%	16.23	0.00%	11'148'124	0	0	0
1	10'648'192	0.0%	16.18	0.15%	10'663'595	15'403	15'126	16'124
2	10'635'751	0.1%	16.18	0.20%	10'662'230	26'479	26'257	26'998
3	9'724'068	0.1%	16.09	0.24%	9'759'434	35'365	34'538	37'575
4	9'786'916	0.2%	16.11	0.38%	9'874'925	88'009	85'302	95'434
5	9'935'753	0.4%	16.13	0.51%	10'097'962	162'209	156'494	178'024
6	9'282'022	0.8%	16.08	0.71%	9'582'510	300'487	286'121	341'305
7	8'256'211	1.5%	15.98	0.94%	8'737'154	480'942	449'167	574'089
8	7'648'729	3.6%	15.99	1.48%	8'766'487	1'117'758	1'043'242	1'318'646
9	5'675'568	16.0%	16.11	3.12%	9'925'132	4'249'564	3'950'815	4'768'384
						6'476'218	6'047'061	7'356'580

Table 4.7: Estimated reserves in model Lemma 4.21

Hence we obtain a weighted average between the a priori estimate $\mu_i = E[C_{i,J}]$ and the chain-ladder estimate $\widehat{C_{i,J}}^{CL}$ on the log-scale. This leads (together with the bias correction) to multiplicative credibility formula. In Table 4.7 we see that the weights $1-\alpha_{i,I-i}$ given to the a priori mean μ_i are rather low.

For the conditional mean square error of prediction we have

$$\begin{aligned}
\text{msep}_{C_{i,J}|C_{i,I-i}}\left(\widehat{C_{i,J}}^{Go}\right) &= \text{Var}(C_{i,J}|C_{i,I-i}) \\
&= \exp\left\{2\cdot\mu_{post(i,I-i)} + \sigma^2_{post(i,I-i)}\right\}\cdot\left(\exp\left\{\sigma^2_{post(i,I-i)}\right\}-1\right) \qquad (4.84)\\
&= \left(E\left[C_{i,J}\middle|\, C_{i,I-i}\right]\right)^2\cdot\left(\exp\left\{\sigma^2_{post(i,I-i)}\right\}-1\right)\\
&= \left(\widehat{C_{i,J}}^{Go}\right)^2\cdot\left(\exp\left\{\sigma^2_{post(i,I-i)}\right\}-1\right).
\end{aligned}$$

This holds under the assumption that the parameters β_j, $\mu^{(i)}$, σ_i and a^2 are known. Hence it is not directly comparable to the mean square error of prediction obtained from the chain-ladder model, since we have no canonical model for the estimation of these parameters and hence we can not quantify the estimation error.
If we want to compare this mean square error of prediction to the ones obtained in Corollary 4.17 we need to calculate the unconditional version:

$$\begin{aligned}
\text{msep}_{C_{i,J}}\left(\widehat{C_{i,J}}^{Go}\right) &= E\left[\left(C_{i,J}-\widehat{C_{i,J}}^{Go}\right)^2\right]\\
&= E\left[\text{Var}(C_{i,J}|C_{i,I-i})\right] \qquad (4.85)\\
&= E\left[\left(\widehat{C_{i,J}}^{Go}\right)^2\right]\cdot\left(\exp\left\{\sigma^2_{post(i,I-i)}\right\}-1\right).
\end{aligned}$$

Hence we need the distribution of $\widehat{C_{i,J}}^{CL} = C_{i,I-i}/\beta_{I-i}$ (cf. (4.83)). Using (4.74) we obtain

$$f_{C_{i,I-i}}(x) = \int_{\mathbb{R}_+} f_{C_{i,I-i},C_{i,J}}(x,y)\, dy$$
$$= \underbrace{\int_{\mathbb{R}_+} \frac{1}{\sqrt{2\pi}\cdot\frac{\sigma_i\cdot\tau_{I-i}}{\sqrt{\sigma_i^2+\tau_{I-i}^2}}} \cdot \frac{1}{y} \cdot \exp\left\{ -\frac{1}{2} \cdot \frac{\left(\log y - \frac{\sigma_i^2 c(x) + \tau_{I-i}^2 \mu^{(i)}}{\sigma_i^2+\tau_{I-i}^2}\right)^2}{\frac{\sigma_i^2\cdot\tau_{I-i}^2}{\sigma_i^2+\tau_{I-i}^2}} \right\} dy}_{=1} \tag{4.86}$$
$$\cdot \frac{1}{\sqrt{2\pi\cdot(\sigma_i^2+\tau_{I-i}^2)}} \cdot \frac{1}{x} \cdot \exp\left\{ -\frac{1}{2} \cdot \frac{\left(\log(x/\beta_{I-i}) + \frac{\tau_{I-i}^2}{2} - \mu^{(i)}\right)^2}{\sigma_i^2+\tau_{I-i}^2} \right\}.$$

This shows the estimator $\widehat{C_{i,J}}^{CL} = C_{i,I-i}/\beta_{I-i}$ is log-normally distributed with parameters $\mu^{(i)}-\tau_{I-i}^2/2$ and $\sigma_i^2+\tau_{I-i}^2$. Moreover, the multiplicative reproductiveness of the Log-normal distribution implies that for $\gamma > 0$

$$\left(\widehat{C_{i,J}}^{CL}\right)^\gamma \overset{(d)}{\sim} \mathcal{LN}\left(\gamma\cdot\mu^{(i)} - \gamma\cdot\tau_{I-i}^2/2, \gamma^2\cdot\left(\sigma_i^2+\tau_{I-i}^2\right)\right). \tag{4.87}$$

Using (4.83) and (4.57) this leads to

$$\begin{aligned}
&\mathrm{msep}_{C_{i,J}}\left(\widehat{C_{i,J}}^{Go}\right)\\
&\quad = E\left[\left(\widehat{C_{i,J}}^{Go}\right)^2\right]\cdot\left(\exp\left\{\sigma^2_{post(i,I-i)}\right\}-1\right) \qquad (4.88)\\
&\quad = \mu_i^{2\cdot(1-\alpha_{i,I-i})}\cdot\exp\left\{\alpha_{i,I-i}\cdot\tau_{I-i}^2\right\}\cdot\left(\exp\left\{\sigma^2_{post(i,I-i)}\right\}-1\right)\cdot E\left[\left(\widehat{C_{i,J}}^{CL}\right)^{2\cdot\alpha_{i,I-i}}\right]\\
&\quad = \exp\left\{2\cdot\mu^{(i)} + (1-\alpha_{i,I-i})\cdot\sigma_i^2 + \alpha_{i,I-i}\cdot\tau_{I-i}^2\right\}\\
&\qquad\quad \cdot\exp\left\{-\alpha_{i,I-i}\cdot\tau_{I-i}^2 + 2\cdot\alpha_{i,I-i}^2\left(\sigma_i^2+\tau_{I-i}^2\right)\right\}\cdot\left(\exp\left\{\sigma^2_{post(i,I-i)}\right\}-1\right).
\end{aligned}$$

Observe that

$$\alpha_{i,I-i}\cdot\left(\sigma_i^2+\tau_{I-i}^2\right) = \sigma_i^2 \tag{4.89}$$

(cf. (4.66)). This immediately implies the following corollary:

Corollary 4.25 *Under the assumptions of Lemma 4.21 we have*

$$msep_{C_{i,J}}\left(\widehat{C_{i,J}}^{Go}\right) = \exp\left\{2\cdot\mu^{(i)} + (1+\alpha_{i,I-i})\cdot\sigma_i^2\right\}\cdot\left(\exp\left\{\sigma^2_{post(i,I-i)}\right\}-1\right) \tag{4.90}$$

for all $I-J+1 \le i \le I$.

	$\mathrm{msep}^{1/2}_{C_{i,J}\mid C_{i,I-i}}\left(\widehat{C_{i,J}}^{Go}\right)$	$\mathrm{msep}^{1/2}_{C_{i,J}}\left(\widehat{C_{i,J}}^{Go}\right)$	$\mathrm{msep}^{1/2}\left(\widehat{R_i}(c_i^*)\right)$
0			
1	16'391	17'526	17'527
2	21'602	22'279	22'282
3	23'714	25'875	25'879
4	37'561	42'139	42'153
5	51'584	58'825	58'862
6	68'339	81'644	81'745
7	82'516	105'397	105'626
8	129'667	162'982	163'852
9	309'586	363'331	372'199
total	359'869	427'850	435'814

Table 4.8: Mean square errors of prediction under the assumptions of Lemma 4.21 and in Model 4.14

4.2.3 Overdispersed Poisson model with gamma a priori distribution

In the next subsections we will consider a different class of Bayesian models. In Model Assumptions 4.14 we had a distributional assumption on $C_{i,j}$ given the ultimate claim $C_{i,J}$ (which can be seen as a backward iteration). Now we introduce a latent variable Θ_i. Conditioned on Θ_i we will do distributional assumptions on the cumulative sizes $C_{i,j}$ and incremental quantities $X_{i,j}$, respectively. Θ_i describes the risk characteristics of accident year i (e.g. was it a "good" or a "bad" year). $C_{i,J}$ is then a random variable with parameters which depend on Θ_i. In the spirit of the previous chapters Θ_i reflects the prediction uncertainties.

We start with the overdispersed Poisson model. The overdispersed Poisson model differs from the Poisson Model 2.12 in that the variance is not equal to the mean. This model was introduced for claims reserving in a Bayesian context by Verrall [79, 81, 82] and Renshaw-Verrall [65]. Furthermore, the overdispersed Poisson model is also used in a generalized linear model context (see McCullagh-Nelder [53], England-Verrall [25] and references therein, and Chapter 5, below). The overdispersed Poisson model as considered below can be generalized to the exponential dispersion family, this is done in Subsection 4.2.4.

We start with the overdispersed Poisson model with Gamma a priori distribution (cf. Verrall [81, 82]).

Model Assumptions 4.26 (Overdispersed Poisson-gamma model)

There exist random variables Θ_i and $Z_{i,j}$ as well as constants $\phi_i > 0$ and $\gamma_0, \ldots, \gamma_J > 0$ with $\sum_{j=0}^{J} \gamma_j = 1$ such that for all $i \in \{0, \ldots, I\}$ and $j \in \{0, \ldots, J\}$ we have

- conditionally, given Θ_i, the $Z_{i,j}$ are independent and Poisson distributed, and the incremental variables $X_{i,j} = \phi_i \cdot Z_{i,j}$ satisfy

$$E\left[X_{i,j}\middle|\,\Theta_i\right] = \Theta_i \cdot \gamma_j \qquad \text{and} \qquad \mathrm{Var}\left(X_{i,j}\middle|\,\Theta_i\right) = \phi_i \cdot \Theta_i \cdot \gamma_j. \tag{4.91}$$

- The pairs $\big(\Theta_i, (X_{i,0}, \ldots, X_{i,J})\big)$ $(i = 0, \ldots, I)$ are independent and Θ_i is Gamma distributed with shape parameter a_i and scale parameter b_i.

□

Remarks 4.27

- See appendix, Sections B.1.2 and B.2.3 for the definition of the Poisson and Gamma distribution.

- Observe, given Θ_i, the expectation and variance of $Z_{i,j}$ satisfy

$$E\left[Z_{i,j}\middle|\,\Theta_i\right] = \mathrm{Var}\left[Z_{i,j}\middle|\,\Theta_i\right] = \frac{\Theta_i \cdot \gamma_j}{\phi_i}. \tag{4.92}$$

The a priori expectation of the increments $X_{i,j}$ is given by

$$E\left[X_{i,j}\right] = E\big[E\left[X_{i,j}\middle|\,\Theta_i\right]\big] = \gamma_j \cdot E\left[\Theta_i\right] = \gamma_j \cdot \frac{a_i}{b_i}. \tag{4.93}$$

- For the cumulative ultimate claim we obtain

$$C_{i,J} = \phi_i \cdot \sum_{j=0}^{J} Z_{i,j}. \tag{4.94}$$

This implies that conditionally, given Θ_i,

$$\frac{C_{i,J}}{\phi_i} \overset{(d)}{\sim} \mathrm{Poisson}(\Theta_i/\phi_i), \qquad \text{and} \qquad E\left[C_{i,J}\middle|\,\Theta_i\right] = \Theta_i, \tag{4.95}$$

this means that Θ_i plays the role of the (unknown) total expected claim amount of accident year i. The Bayesian approach chosen tells us, how we should combine the a priori expectation $E[C_{i,J}] = a_i/b_i$ and the information $\mathcal{D}_I$.

- This model is sometimes problematic in practical applications. It assumes that we have no negative increments $X_{i,j}$. If we count the number of reported claims this may hold true. However if $X_{i,j}$ denotes incremental payments, we can have negative values. E.g. in motor hull insurance in old development periods one gets more money (via subrogation and repayments of deductibles) than one spends.

- Observe that we have assumed that the claims development pattern γ_j is known.

- Observe that in the overdispersed Poisson model, in general, $C_{i,j}$ is not a natural number. Henceforth, if we work with claims counts with dispersion $\phi_i \neq 1$ there is not really an interpretation for this model.

Lemma 4.28 *Under Model Assumptions 4.26 the a posteriori distribution of* Θ_i*, given* $(X_{i,0}, \ldots, X_{i,j})$*, is a Gamma distribution with updated parameters*

$$a_{i,j}^{post} = a_i + C_{i,j}/\phi_i, \tag{4.96}$$

$$b_{i,j}^{post} = b_i + \sum_{k=0}^{j} \gamma_k/\phi_i = b_i + \beta_j/\phi_i, \tag{4.97}$$

where $\beta_j = \sum_{k=0}^{j} \gamma_k$.

Remarks 4.29

- Since accident years are independent it suffices to consider $(X_{i,0}, \ldots, X_{i,j})$ for the calculation of the a posteriori distribution of Θ_i.

- We assume that a priori all accident years are equal (Θ_i are i.i.d.). After we have a set of observations $\mathcal{D}_I$, we obtain a posteriori risk characteristics which differ according to the observations.

- Model 4.26 belongs to the well-known class of exponential dispersion models with associated conjugates (see e.g. Bühlmann-Gisler [18] in Subsection 2.5.1, and Subsection 4.2.4 below).

- Using Lemma 4.28 we obtain for the a posteriori expectation

$$\begin{aligned} E\left[\Theta_i \middle| \mathcal{D}_I\right] &= \frac{a_{i,I-i}^{post}}{b_{i,I-i}^{post}} = \frac{a_i + C_{i,I-i}/\phi_i}{b_i + \beta_{I-i}/\phi_i} \\ &= \frac{b_i}{b_i + \beta_{I-i}/\phi_i} \cdot \frac{a_i}{b_i} + \left(1 - \frac{b_i}{b_i + \beta_{I-i}/\phi_i}\right) \cdot \frac{C_{i,I-i}}{\beta_{I-i}}, \end{aligned} \tag{4.98}$$

  which is a credibility weighted average between the a priori expectation $E[\Theta_i] = \frac{a_i}{b_i}$ and the observation $\frac{C_{i,I-i}}{\beta_{I-i}}$ (see next section and Bühlmann-Gisler [18] for more detailed discussions).

- In fact we can specify the a posteriori distribution of $(C_{i,J} - C_{i,I-i})/\phi_i$, given $\mathcal{D}_I$. It holds for $k \in \{0, 1, \ldots\}$ that

$$
\begin{aligned}
& P\big((C_{i,J} - C_{i,I-i})/\phi_i = k\big) \\
&\quad = \int_{\mathbb{R}_+} e^{-(1-\beta_{I-i})\cdot\theta} \cdot \frac{((1-\beta_{I-i})\cdot\theta)^k}{k!} \cdot \frac{(b_{i,I-i}^{post})^{a_{i,I-i}^{post}}}{\Gamma(a_{i,I-i}^{post})} \cdot \theta^{a_{i,I-i}^{post}-1} \cdot e^{-b_{i,I-i}^{post}\cdot\theta}\, d\theta \\
&\quad = \frac{\left(b_{i,I-i}^{post}\right)^{a_{i,I-i}^{post}} \cdot (1-\beta_{I-i})^k}{\Gamma\left(a_{i,I-i}^{post}\right)\cdot k!} \cdot \underbrace{\int_{\mathbb{R}_+} \theta^{k+a_{i,I-i}^{post}-1} \cdot e^{(b_{i,I-i}^{post}+1-\beta_{I-i})\cdot\theta}\, d\theta}_{\propto \text{ density of } \Gamma\left(k+a_{i,I-i}^{post},\, b_{i,I-i}^{post}+1-\beta_{I-i}\right)} \\
&\quad = \frac{\left(b_{i,I-i}^{post}\right)^{a_{i,I-i}^{post}} \cdot (1-\beta_{I-i})^k}{\Gamma\left(a_{i,I-i}^{post}\right)\cdot k!} \cdot \frac{\Gamma\left(k+a_{i,I-i}^{post}\right)}{\left(b_{i,I-i}^{post}+1-\beta_{I-i}\right)^{k+a_{i,I-i}^{post}}} \qquad (4.99) \\
&\quad = \frac{\Gamma\left(k+a_{i,I-i}^{post}\right)}{k!\cdot\Gamma\left(a_{i,I-i}^{post}\right)} \cdot \left(\frac{b_{i,I-i}^{post}}{b_{i,I-i}^{post}+1-\beta_{I-i}}\right)^{a_{i,I-i}^{post}} \cdot \left(\frac{1-\beta_{I-i}}{b_{i,I-i}^{post}+1-\beta_{I-i}}\right)^k \\
&\quad = \binom{k+a_{i,I-i}^{post}-1}{k} \cdot \left(\frac{b_{i,I-i}^{post}}{b_{i,I-i}^{post}+1-\beta_{I-i}}\right)^{a_{i,I-i}^{post}} \cdot \left(\frac{1-\beta_{I-i}}{b_{i,I-i}^{post}+1-\beta_{I-i}}\right)^k,
\end{aligned}
$$

  which is a Negative binomial distribution with parameters $r = a_{i,I-i}^{post}$ and $p = b_{i,I-i}^{post}/\big(b_{i,I-i}^{post} + 1 - \beta_{I-i}\big)$ (see appendix Section B.1.3).

Proof. Using (4.92) we obtain for the conditionally density of $(X_{i,0}, \ldots, X_{i,j})$, given Θ_i, that

$$
f_{X_{i,0},\ldots,X_{i,j}|\Theta_i}(x_0, \ldots, x_j|\theta) = \prod_{k=0}^{j} \exp\{-\theta\cdot\gamma_k/\phi_i\} \cdot \frac{(\theta\cdot\gamma_k/\phi_i)^{x_k/\phi_i}}{x_k/\phi_i}. \qquad (4.100)
$$

Hence the joint distribution of Θ_i and $(X_{i,0}, \ldots, X_{i,j})$ is given by

$$
\begin{aligned}
f_{\Theta_i,X_{i,0},\ldots,X_{i,j}}(\theta, x_0, \ldots, x_j) &= f_{X_{i,0},\ldots,X_{i,j}|\Theta_i}(x_0, \ldots, x_j|\theta) \cdot f_{\Theta_i}(\theta) \\
&= \prod_{k=0}^{j} \exp\{-\theta\cdot\gamma_k/\phi_i\} \cdot \frac{(\theta\cdot\gamma_k/\phi_i)^{x_k/\phi_i}}{x_k/\phi_i} \cdot \frac{b_i^{a_i}}{\Gamma(a_i)} \theta^{a_i-1} e^{-b_i\cdot\theta}. \qquad (4.101)
\end{aligned}
$$

This shows that the a posteriori distribution of Θ_i, given $(X_{i,0}, \ldots, X_{i,j})$, is again a Gamma distribution with updated parameters

$$
a_{i,j}^{post} = a_i + C_{i,j}/\phi_i, \qquad (4.102)
$$

$$
b_{i,j}^{post} = b_i + \sum_{k=0}^{j} \gamma_k/\phi_i. \qquad (4.103)
$$

This finishes the proof of the lemma.

□

Using the conditional independence of $X_{i,j}$, given Θ_i, and (4.91) we obtain

$$\begin{aligned}
E\left[C_{i,J}\middle|\,\mathcal{D}_I\right] &= E\left[E\left[C_{i,J}\middle|\,\Theta_i,\mathcal{D}_I\right]\middle|\,\mathcal{D}_I\right]\\
&= E\left[E\left[\sum_{j=0}^{I-i} X_{i,j}\middle|\,\Theta_i,\mathcal{D}_I\right]\middle|\,\mathcal{D}_I\right] + E\left[E\left[\sum_{j=I-i+1}^{J} X_{i,j}\middle|\,\Theta_i,\mathcal{D}_I\right]\middle|\,\mathcal{D}_I\right]\\
&= C_{i,I-i} + E\left[E\left[\sum_{j=I-i+1}^{J} X_{i,j}\middle|\,\Theta_i\right]\middle|\,\mathcal{D}_I\right] \qquad (4.104)\\
&= C_{i,I-i} + (1-\beta_{I-i})\cdot E\left[\Theta_i\middle|\,\mathcal{D}_I\right].
\end{aligned}$$

Together with (4.98) this motivates the following estimator:

Estimator 4.30 (Poisson-gamma model, Verrall [81, 82]) *Under Model Assumptions 4.26 we have the following estimator for the ultimate claim* $E\left[C_{i,J}\middle|\,\mathcal{D}_I\right]$

$$\widehat{C_{i,J}}^{PoiGa} = C_{i,I-i} + (1-\beta_{I-i})\cdot\left[\frac{b_i}{b_i+\frac{\beta_{I-i}}{\phi_i}}\cdot\frac{a_i}{b_i} + \left(1-\frac{b_i}{b_i+\frac{\beta_{I-i}}{\phi_i}}\right)\cdot\frac{C_{i,I-i}}{\beta_{I-i}}\right] \qquad (4.105)$$

for $I-J+1 \le i \le I$.

Example 4.31 (Poisson-gamma model)

We revisit the data set given in Example 2.7. For the a priori parameters we do the same choices as in Example 4.19 (see Table 4.4). Since Θ_i is Gamma distributed with shape parameter a_i and scale parameter b_i we have

$$E\left[\Theta_i\right] = \frac{a_i}{b_i}, \qquad (4.106)$$

$$\mathrm{Vco}\left(\Theta_i\right) = a_i^{-1/2}, \qquad (4.107)$$

and, using (4.91), we obtain

$$\begin{aligned}
\mathrm{Var}\left(C_{i,J}\right) &= E\left[\mathrm{Var}\left(C_{i,J}\middle|\,\Theta_i\right)\right] + \mathrm{Var}\left(E\left[C_{i,J}\middle|\,\Theta_i\right]\right)\\
&= \phi_i\cdot E[\Theta_i] + \mathrm{Var}\left(\Theta_i\right) \qquad (4.108)\\
&= \frac{a_i}{b_i}\cdot\left(\phi_i + b_i^{-1}\right).
\end{aligned}$$

This leads to Table 4.9.

We define

$$\alpha_{i,I-i} = \frac{\beta_{I-i}/\phi_i}{b_i+\beta_{I-i}/\phi_i}, \qquad (4.109)$$

	$E[\Theta_i]$	$\mathrm{Vco}(\Theta_i)$	$\mathrm{Vco}(C_{i,J})$	a_i	b_i	ϕ_i
0	11'653'101	5.00%	7.8%	400	0.00343%	41'951
1	11'367'306	5.00%	7.8%	400	0.00352%	40'922
2	10'962'965	5.00%	7.8%	400	0.00365%	39'467
3	10'616'762	5.00%	7.8%	400	0.00377%	38'220
4	11'044'881	5.00%	7.8%	400	0.00362%	39'762
5	11'480'700	5.00%	7.8%	400	0.00348%	41'331
6	11'413'572	5.00%	7.8%	400	0.00350%	41'089
7	11'126'527	5.00%	7.8%	400	0.00360%	40'055
8	10'986'548	5.00%	7.8%	400	0.00364%	39'552
9	11'618'437	5.00%	7.8%	400	0.00344%	41'826

Table 4.9: Parameter choice for the Poisson-gamma model

which is the credibility weight given to the observation $\frac{C_{i,I-i}}{\beta_{I-i}}$ (cf. (4.98)). The credibility weights and estimates for the ultimates are given in Table 4.10.

Observe that

$$\alpha_{i,I-i} = \frac{\beta_{I-i}}{\beta_{I-i} + \phi_i \cdot b_i} = \frac{\beta_{I-i}}{\beta_{I-i} + \frac{E\left[\mathrm{Var}\left(X_{i,I-i}\middle|\Theta_i\right)\right]}{\gamma_{I-i}\cdot \mathrm{Var}(\Theta_i)}}. \tag{4.110}$$

The term $\frac{E\left[\mathrm{Var}\left(X_{i,I-i}\middle|\Theta_i\right)\right]}{\gamma_{I-i}\cdot \mathrm{Var}(\Theta_i)}$ is the so-called credibility coefficient (see also Remark 4.38).

						estimated reserves		
	$C_{i,I-i}$	$\beta_{i,I-i}$	$\alpha_{i,I-i}$	$\frac{a^{post}_{i,I-i}}{b^{post}_{i,I-i}}$	$\widehat{C_{i,J}}^{PoiGa}$	PoiGa	CL	BF
0	11'148'124	100.0%	41.0%	11'446'143	11'148'124	0	0	0
1	10'648'192	99.9%	40.9%	11'079'028	10'663'907	15'715	15'126	16'124
2	10'635'751	99.8%	40.9%	10'839'802	10'662'446	26'695	26'257	26'998
3	9'724'068	99.6%	40.9%	10'265'794	9'760'401	36'333	34'538	37'575
4	9'786'916	99.1%	40.8%	10'566'741	9'878'219	91'303	85'302	95'434
5	9'935'753	98.4%	40.6%	10'916'902	10'105'034	169'281	156'494	178'024
6	9'282'022	97.0%	40.3%	10'670'762	9'601'115	319'093	286'121	341'305
7	8'256'211	94.8%	39.7%	10'165'120	8'780'696	524'484	449'167	574'089
8	7'648'729	88.0%	37.9%	10'116'206	8'862'913	1'214'184	1'043'242	1'318'646
9	5'675'568	59.0%	29.0%	11'039'755	10'206'452	4'530'884	3'950'815	4'768'384
						6'927'973	6'047'061	7'356'580

Table 4.10: Estimated reserves in the Poisson-gamma model

The conditional mean square error of prediction is given by

$$\begin{aligned}
\mathrm{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{PoiGa}\right) &= E\left[\left(C_{i,J}-\widehat{C_{i,J}}^{PoiGa}\right)^2\middle|\mathcal{D}_I\right]\\
&= E\left[\left(\sum_{j=I-i+1}^{J} X_{i,j}-(1-\beta_{I-i})\cdot E\left[\Theta_i|\,\mathcal{D}_I\right]\right)^2\middle|\mathcal{D}_I\right]\\
&= E\left[\left(\sum_{j=I-i+1}^{J}\left(X_{i,j}-\gamma_j\cdot E\left[\Theta_i|\,\mathcal{D}_I\right]\right)\right)^2\middle|\mathcal{D}_I\right] \qquad (4.111)
\end{aligned}$$

(cf. (4.104)-(4.105)). Since for $j > I-i$

$$\begin{aligned}
E\left[X_{i,j}|\,\mathcal{D}_I\right] &= E\big[E\left[X_{i,j}|\,\Theta_i,\mathcal{D}_I\right]|\,\mathcal{D}_I\big]\\
&= E\big[E\left[X_{i,j}|\,\Theta_i\right]|\,\mathcal{D}_I\big] \qquad (4.112)\\
&= \gamma_j\cdot E\left[\Theta_i|\,\mathcal{D}_I\right],
\end{aligned}$$

we have that

$$\mathrm{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{PoiGa}\right) = \mathrm{Var}\left(\sum_{j=I-i+1}^{J} X_{i,j}\middle|\mathcal{D}_I\right). \qquad (4.113)$$

This last expression can be calculated. We do the complete calculation, but we could also argue with the help of the negative binomial distribution. Using the conditional independence of $X_{i,j}$, given Θ_i, and (4.91) we obtain

$$\begin{aligned}
&\mathrm{Var}\left(\sum_{j=I-i+1}^{J} X_{i,j}\middle|\mathcal{D}_I\right)\\
&= E\left(\mathrm{Var}\left(\sum_{j=I-i+1}^{J} X_{i,j}\middle|\Theta_i\right)\middle|\mathcal{D}_I\right)+\mathrm{Var}\left(E\left(\sum_{j=I-i+1}^{J} X_{i,j}\middle|\Theta_i\right)\middle|\mathcal{D}_I\right)\\
&= E\left(\sum_{j=I-i+1}^{J}\phi_i\cdot\Theta_i\cdot\gamma_j\middle|\mathcal{D}_I\right)+\mathrm{Var}\left(\sum_{j=I-i+1}^{J}\Theta_i\cdot\gamma_j\middle|\mathcal{D}_I\right) \qquad (4.114)\\
&= \phi_i\cdot(1-\beta_{I-i})\cdot E\left[\Theta_i|\,\mathcal{D}_I\right]+(1-\beta_{I-i})^2\cdot\mathrm{Var}\left(\Theta_i|\,\mathcal{D}_I\right).
\end{aligned}$$

With Lemma 4.28 this leads to the following corollary:

Corollary 4.32 *Under Model Assumptions 4.26 the conditional mean square error of prediction is given by*

$$msep_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{PoiGa}\right) = \phi_i\cdot(1-\beta_{I-i})\cdot\frac{a_{i,I-i}^{post}}{b_{i,I-i}^{post}}+(1-\beta_{I-i})^2\cdot\frac{a_{i,I-i}^{post}}{\left(b_{i,I-i}^{post}\right)^2} \qquad (4.115)$$

for $I-J+1\le i\le I$.

Remark. Observe that we have assumed that the parameters a_i, b_i, ϕ_i and γ_j are known. If these need to be estimated we obtain an additional term in the MSEP calculation which corresponds to the parameter estimation error.

The unconditional mean square error of prediction can then easily be calculated. We have

$$\begin{aligned} \mathrm{msep}_{C_{i,J}}\left(\widehat{C_{i,J}}^{PoiGa}\right) &= E\left[\mathrm{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{PoiGa}\right)\right] \\ &= \phi_i \cdot (1-\beta_{I-i}) \cdot \frac{E\left[a_{i,I-i}^{post}\right]}{b_{i,I-i}^{post}} + (1-\beta_{I-i})^2 \cdot \frac{E\left[a_{i,I-i}^{post}\right]}{\left(b_{i,I-i}^{post}\right)^2}, \end{aligned} \tag{4.116}$$

and using $E[C_{i,I-i}] = \beta_{I-i} \cdot \frac{a_i}{b_i}$ (cf. (4.93)) we obtain

$$\mathrm{msep}_{C_{i,J}}\left(\widehat{C_{i,J}}^{PoiGa}\right) = \phi_i \cdot (1-\beta_{I-i}) \cdot \frac{a_i}{b_i} \cdot \frac{1+\phi_i \cdot b_i}{\phi_i \cdot b_i + \beta_{I-i}}. \tag{4.117}$$

Hence we obtain the Table 4.11 for the conditional prediction errors.

	$\mathrm{msep}^{1/2}_{C_{i,J}\|C_{i,I-i}}(\cdot)$		$\mathrm{msep}^{1/2}_{C_{i,J}}(\cdot)$		
	$\widehat{C_{i,J}}^{PoiGa}$	$\widehat{C_{i,J}}^{Go}$	$\widehat{C_{i,J}}^{PoiGa}$	$\widehat{C_{i,J}}^{Go}$	$\widehat{R_i}(c_i^*)$
0					
1	25'367	16'391	25'695	18'832	17'527
2	32'475	21'602	32'659	23'940	22'282
3	37'292	23'714	37'924	27'804	25'879
4	60'359	37'561	61'710	45'276	42'153
5	83'912	51'584	86'052	63'200	58'862
6	115'212	68'339	119'155	87'704	81'745
7	146'500	82'516	153'272	113'195	105'626
8	224'738	129'667	234'207	174'906	163'852
9	477'318	309'586	489'668	388'179	372'199
total	571'707	359'869	588'809	457'739	435'814

Table 4.11: Mean square errors of prediction in the Poisson-gamma model, the Log-normal/Log-normal model and in Model 4.14

We have already seen in Table 4.10 that the Poisson-gamma reserves are closer to the Bornhuetter-Ferguson estimate (this stands in contrast with the other methods presented in this chapter). Table 4.11 shows that the prediction error is substantially larger in the Poisson-gamma model than in the other models (comparable to the estimation error in $\widehat{R_i}(0)$ in Table 4.5). This suggests that in the present case the Poisson-gamma method is not an appropriate method.

4.2.4 Exponential dispersion family with its associate conjugates

In the subsection above we have seen that in the Poisson-gamma model Θ_i has as a posteriori distribution again a Gamma distribution with updated parameters. This indicates, using a smart choice of the distributions we were able to calculate the a posteriori distribution. We generalize the Poisson-gamma model to the Exponential dispersion family (EDF), and we look for its associate conjugates. This are standard models in Bayesian inference, for literature we refer e.g. to Bernardo-Smith [9]. Similar ideas have been applied for tariffication and pricing (see Bühlmann-Gisler [18], Chapter 2), we transform these ideas to the reserving context (see also Wüthrich [89]).

Model Assumptions 4.33 (Exponential dispersion model)

There exists a claims development pattern $(\beta_j)_{0 \le j \le J}$ with $\beta_J = 1$, $\gamma_0 = \beta_0 \neq 0$ and $\gamma_j = \beta_j - \beta_{j-1} \neq 0$ for $j \in \{1, \ldots, J\}$.

- Conditionally, given Θ_i, the $X_{i,j}$ $(0 \le i \le I,\ 0 \le j \le J)$ are independent with

$$\frac{X_{i,j}}{\gamma_j \cdot \mu_i} \stackrel{(d)}{\sim} dF_{i,j}^{(\Theta_i)}(x) = a\left(x, \frac{\sigma^2}{w_{i,j}}\right) \cdot \exp\left\{\frac{x \cdot \Theta_i - b(\Theta_i)}{\sigma^2 / w_{i,j}}\right\} d\nu(x), \quad (4.118)$$

  where ν is a suitable σ-finite measure on $\mathbb{R}$, $b(\cdot)$ is some real-valued twice-differentiable function of Θ_i and $\mu_i > 0$, σ^2 and $w_{i,j} > 0$ are some real-valued constants, and $F_{i,j}^{(\Theta_i)}$ is a probability distribution on $\mathbb{R}$.

- The random vectors $\big(\Theta_i, (X_{i,0}, \ldots, X_{i,J})\big)$ $(i = 0, \ldots, I)$ are independent and Θ_i are real-valued random variables with densities (w.r.t. the Lebesgue measure)

$$u_{\mu,\tau^2}(\theta) = d(\mu, \tau^2) \cdot \exp\left\{\frac{\mu \cdot \theta - b(\theta)}{\tau^2}\right\}, \quad (4.119)$$

  with $\mu \equiv 1$ and $\tau^2 > 0$.

□

Remarks 4.34

- In the following the measure ν is given by the Lebesgue measure or by the counting measure.

- A distribution of the type (4.118) is said to be a (one parametric) Exponential dispersion family (EDF). The class of (one parametric) Exponential dispersion families covers a large class of families of distributions, e.g. the families of the Poisson, Bernoulli, Gamma, Normal and Inverse-gaussian distributions (cf. Bühlmann-Gisler [18], Section 2.5).

- The first assumption implies that the scaled sizes $Y_{i,j} = X_{i,j}/(\gamma_j \cdot \mu_i)$ have, given Θ_i, a distribution $F_{i,j}^{(\Theta_i)}$ which belongs to the EDF. A priori they are all the same, which is described by the fact that μ and τ^2 do not depend on i.

- For the time being we assume that all parameters of the underlying distributions are known, $w_{i,j}$ is a known volume measure which will be further specified below.

- For the moment we could also concentrate on a single accident year i, i.e. we only need the Model Assumptions 4.33 for a fixed accident year i.

- A pair of distributions given by (4.118) and (4.119) is said to be a (one parametric) Exponential dispersion family with associated conjugates. Examples are (see Bühlmann-Gisler [18], Section 2.5): Poisson-gamma case (see also Verrall [81, 82] and Subsection 4.2.3), Binomial-beta case, Gamma-gamma case or Normal-normal case.

We have the following lemma:

Lemma 4.35 (Associate Conjugate) *Under Model Assumptions 4.33 the conditional distribution of Θ_i, given $X_{i,0}, \ldots, X_{i,j}$, has the density $u_{\mu^{(i)}_{post,j}, \tau^2_{post,j}}(\cdot)$ with the a posteriori parameters*

$$\tau^2_{post,j} = \sigma^2 \cdot \left[\frac{\sigma^2}{\tau^2} + \sum_{k=0}^{j} w_{i,k} \right]^{-1}, \tag{4.120}$$

$$\mu^{(i)}_{post,j} = \frac{\tau^2_{post,j}}{\sigma^2} \cdot \left[\frac{\sigma^2}{\tau^2} \cdot 1 + \sum_{k=0}^{j} w_{i,k} \cdot \bar{Y}_i^{(j)} \right], \tag{4.121}$$

where

$$\bar{Y}_i^{(j)} = \sum_{k=0}^{j} \frac{w_{i,k}}{\sum_{l=0}^{j} w_{i,l}} \cdot \frac{X_{i,k}}{\gamma_k \cdot \mu_i}. \tag{4.122}$$

Proof. Define $Y_{i,j} = X_{i,j}/(\gamma_j \cdot \mu_i)$. The joint distribution of $(\Theta_i, Y_{i,0}, \dots, Y_{i,j})$ is given by

$$\begin{aligned} f_{\Theta_i, Y_{i,0},\dots,Y_{i,j}}(\theta, y_0, \dots, y_j) &= f_{Y_{i,0},\dots,Y_{i,j}|\Theta_i}(y_0, \dots, y_j|\theta) \cdot u_{1,\tau^2}(\theta) \\ &= d(1,\tau^2) \cdot \exp\left\{ \frac{1 \cdot \theta - b(\theta)}{\tau^2} \right\} \qquad (4.123) \\ &\quad \cdot \prod_{k=0}^{j} a\left(y_k, \frac{\sigma^2}{w_{i,k}} \right) \cdot \exp\left\{ \frac{y_k \cdot \theta - b(\theta)}{\sigma^2 / w_{i,k}} \right\}. \end{aligned}$$

Hence the conditional distribution of Θ_i, given $X_{i,0}, \dots, X_{i,j}$, is proportional to

$$\exp\left\{ \theta \cdot \left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{w_{i,k}}{\sigma^2} \cdot \frac{X_{i,k}}{\gamma_k \cdot \mu_i} \right] - b(\theta) \cdot \left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{w_{i,k}}{\sigma^2} \right] \right\}. \qquad (4.124)$$

This finishes the proof of the lemma.

□

Remarks 4.36

- Lemma 4.35 states that the distribution defined by density (4.119) is a conjugated distribution to the distribution given by (4.118). This means that the a posteriori distribution of Θ_i, given $X_{i,0}, \dots, X_{i,j}$, is again of the type (4.119) with updated parameters $\tau^2_{post,j}$ and $\mu^{(i)}_{post,j}$.

- From Lemma 4.35 we can calculate the distribution of $(Y_{i,I-i+1}, \dots Y_{i,J})$, given $\mathcal{D}_I$. First we remark that different accident years are independent, hence we can restrict ourselves to the observations $Y_{i,0}, \dots, Y_{i,I-i}$, then we have that the a posteriori distribution is given by

$$\int \prod_{j=I-i+1}^{J} F^{(\theta)}_{i,j}(y_j) \cdot u_{\mu^{(i)}_{post,I-i}, \tau^2_{post,I-i}}(\theta)\, d\theta. \qquad (4.125)$$

  In the Poisson-gamma case this is a negative binomial distribution. Observe that for the EDF with its associate conjugates we can determine the explicit distributions, not only estimates for the first two moments.

Theorem 4.37 *Under the Model Assumptions 4.33 we have for $i = 0, \dots, I$ and $j = 0, \dots, J$:*
1. The conditional moments of the standardized observations $\frac{X_{i,j}}{\gamma_j \cdot \mu_i}$ are given by

$$\mu(\Theta_i) \stackrel{def.}{=} E\left[\left. \frac{X_{i,j}}{\gamma_j \cdot \mu_i} \right| \Theta_i \right] = b'(\Theta_i), \qquad (4.126)$$

$$Var\left(\left. \frac{X_{i,j}}{\gamma_j \cdot \mu_i} \right| \Theta_i \right) = \frac{\sigma^2 \cdot b''(\Theta_i)}{w_{i,j}}. \qquad (4.127)$$

2. If $\exp\left\{\left(\mu_i \cdot \theta - b(\theta)\right)/\tau^2\right\}$ *disappears on the boundary of* Θ_i *for all* μ_i, τ^2 *then*

$$E\left[X_{i,j}\right] = \gamma_j \cdot E\left[\mu(\Theta_i)\right] \;\; = \;\; \gamma_j \cdot \mu_i, \tag{4.128}$$

$$E\left[\mu(\Theta_i)\middle|\, X_{i,0}, \ldots, X_{i,j}\right] = \alpha_{i,j} \cdot \bar{Y}_i^{(j)} + (1 - \alpha_{i,j}) \cdot 1, \tag{4.129}$$

where

$$\alpha_{i,j} = \frac{\sum_{k=0}^{j} w_{i,k}}{\sum_{k=0}^{j} w_{i,k} + \sigma^2/\tau^2}. \tag{4.130}$$

Proof. See Lemma 5 below, Theorem 2.20 in Bühlmann-Gisler [18] or Bernardo-Smith [9].

□

Remarks 4.38

- In Model Assumptions 4.33 and in Theorem 4.37 we study the standardized version for the observations $X_{i,j}$. If μ_i is equal for all i, the standardization is not necessary. If they are not equal, the standardized version is straightforward for comparisons between accident years.

- Theorem 4.37 says, that the a posteriori mean of $\mu(\Theta_i)$, given the observations $X_{i,0}, \ldots, X_{i,j}$, is a credibility weighted average between the a priori mean $E[\mu(\Theta_i)] = 1$ and the weighted average $\bar{Y}_i^{(j)}$ of the standardized observations. The larger the individual variation σ^2 the smaller the credibility weight $\alpha_{i,j}$; the larger the collective variability τ^2 the larger the credibility weight $\alpha_{i,j}$. For a detailed discussion on the credibility coefficient

$$\kappa = \sigma^2/\tau^2 \tag{4.131}$$

we refer to Bühlmann-Gisler [18].

Estimator 4.39 *Under Model Assumptions 4.33 we have the following estimators for the increments* $E\left[X_{i,I-i+k}\middle|\,\mathcal{D}_I\right]$ *and the ultimate claims* $E\left[C_{i,J}\middle|\,\mathcal{D}_I\right]$

$$\widehat{X_{i,I-i+k}}^{EDF} = \gamma_{I-i+k} \cdot \mu_i \cdot \widetilde{\mu(\Theta_i)}, \tag{4.132}$$

$$\widehat{C_{i,J}}^{EDF} = C_{i,I-i} + (1 - \beta_{I-i}) \cdot \mu_i \cdot \widetilde{\mu(\Theta_i)} \tag{4.133}$$

for $I - J + 1 \le i \le I$ *and* $k \in \{1, \ldots, J - I + i\}$, *where*

$$\widetilde{\mu(\Theta_i)} = E\left[\mu(\Theta_i)\middle|\,\mathcal{D}_I\right] = \alpha_{i,I-i} \cdot \bar{Y}_i^{(I-i)} + (1 - \alpha_{i,I-i}) \cdot 1. \tag{4.134}$$

Theorem 4.40 (Bayesian estimator) *Under Model Assumptions 4.33 the estimators* $\widetilde{\mu(\Theta_i)}$, $\widehat{X_{i,j+k}}^{EDF}$ *and* $\widehat{C_{i,J}}^{EDF}$ *are* $\mathcal{D}_I$*-measurable and minimize the conditional mean square errors* $msep_{\mu(\Theta_i)|\mathcal{D}_I}(\cdot)$, $msep_{X_{i,j+k}|\mathcal{D}_I}(\cdot)$ *and* $msep_{C_{i,J}|\mathcal{D}_I}(\cdot)$*, respectively, for* $I-J+1 \le i \le I$*. I.e. these estimators are Bayesian w.r.t.* $\mathcal{D}_I$ *and minimize the quadratic loss function (*$L^2(P)$*-norm).*

Proof. The $\mathcal{D}_I$-measurability is clear. But then the claim for $\widetilde{\mu(\Theta_i)}$ is clear, since the conditional expectation minimizes the mean square error given $\mathcal{D}_I$ (see Theorem 2.5 in [18]). Due to our independence assumptions we have

$$E\left[X_{i,I-i+k}|\mathcal{D}_I\right] = E\left[E\left[X_{i,I-i+k}|\Theta_i\right]|\mathcal{D}_I\right] = \gamma_{I-i+k}\cdot\mu_i\cdot\widetilde{\mu(\Theta_i)}, \quad (4.135)$$

$$E\left[C_{i,J}|\mathcal{D}_I\right] = C_{i,I-i} + (1-\beta_{I-i})\cdot\mu_i\cdot\widetilde{\mu(\Theta_i)}, \quad (4.136)$$

which finishes the proof of the theorem.

□

Explicit choice of weights.

W.l.o.g. we may and will assume that

$$m_b = E\left[b''(\Theta_i)\right] = 1. \quad (4.137)$$

Otherwise we simply multiply σ^2 and τ^2 by m_b, which in our context of EDF with associate conjugates leads to the same model with $b(\theta)$ replaced by $b_{(1)}(\theta) = m_b \cdot b(\theta/m_b)$. This rescaled model has then

$$\mathrm{Var}\left(\left.\frac{X_{i,j}}{\gamma_j\cdot\mu_i}\right|\Theta_i\right) = \frac{m_b\cdot\sigma^2\cdot b''_{(1)}(\Theta_i)}{w_{i,j}}, \quad \text{with} \quad E\left[b''_{(1)}(\Theta_i)\right]=1, \quad (4.138)$$

$$\mathrm{Var}\left(b'_{(1)}(\Theta_i)\right) = m_b\cdot\tau^2. \quad (4.139)$$

Since both, σ^2 and τ^2 are multiplied by m_b, the credibility weights $\alpha_{i,j}$ do not change under this transformation. Hence we assume (4.137) for the rest of this work.

In Section 4.2.4 we have not specified the weights $w_{i,j}$. In Mack [47] there is a discussion choosing appropriate weights (Assumption (A4$^\alpha$) in Mack [47]). In fact we could choose a design matrix $\omega_{i,j}$ which gives a whole family of models. We do not further discuss this here, we will do a canonical choice (which is favoured in many applications) that has the nice side effect that we obtain a natural mixture between the chain-ladder estimate and the Bornhuetter-Ferguson estimate.

Model Assumptions 4.41

In addition to Model Assumptions 4.33 and (4.137) we assume that there exists $\delta \geq 0$ with $w_{i,j} = \gamma_j \cdot \mu_i^\delta$ for all $i = 0, \dots, I$ and $j = 0, \dots, J$ and that $\exp\left\{\left(\mu_0 \cdot \theta - b(\theta)\right)/\tau^2\right\}$ disappears on the boundary of Θ_i for all μ_0 and τ^2.

□

Hence, we have $\sum_{k=0}^{j} w_{i,k} = \beta_j \cdot \mu_i^\delta$. This immediately implies:

Corollary 4.42 *Under Model Assumptions 4.41 we have for all $i = 0, \dots, I$ that*

$$\widetilde{\mu(\Theta_i)} = \alpha_{i,I-i} \cdot \frac{C_{i,I-i}}{\beta_{I-i} \cdot \mu_i} + (1 - \alpha_{i,I-i}) \cdot 1, \tag{4.140}$$

$$\textit{where} \quad \alpha_{i,I-i} = \frac{\beta_{I-i}}{\beta_{I-i} + \frac{\sigma^2}{\mu_i^\delta} \cdot \tau^{-2}}. \tag{4.141}$$

Remark. Compare the weight $\alpha_{i,I-i}$ from (4.141) to $\alpha_{i,I-i}$ from (4.110):
In the notation of Subsection 4.2.3 (see (4.110)) we have

$$\kappa_i = \phi_i \cdot b_i = \frac{E\left[\operatorname{Var}\left(X_{i,I-i} \middle| \Theta_i\right)\right]}{\gamma_{I-i} \cdot \operatorname{Var}(\Theta_i)} \tag{4.142}$$

and in the notation of this subsection we have

$$\kappa_i = \frac{\sigma^2/\mu_i^\delta}{\tau^2} = \frac{E\left[\operatorname{Var}\left(\frac{X_{i,I-i}}{\mu_i} \middle| \Theta_i\right)\right]}{\gamma_{I-i} \cdot \operatorname{Var}\big(\mu(\Theta_i)\big)}. \tag{4.143}$$

This shows that the estimators $\widehat{C_{i,J}}^{PoiGa}$ and $\widehat{C_{i,J}}^{EDF}$ give the same estimated reserve (the Poisson-gamma model is an example for the Exponential dispersion family with associate conjugates).

Example 4.43 (Exponential dispersion model with associate conjugate)

We revisit the data set given in Example 2.7. For the a priori parameters we do the same choices as in Example 4.19 (see Table 4.4).
Observe that the credibility weight of the reserves does not depend on the choice of $\delta \geq 0$ for given $\operatorname{Vco}(C_{i,J})$: Using the conditional independence of the increments $X_{i,j}$, given Θ_i, and (4.126), (4.127) and (4.137) leads to

$$\begin{aligned}
\operatorname{Var}(C_{i,J}) &= E\left[\operatorname{Var}\left(\sum_{j=0}^{J} X_{i,j} \middle| \Theta_i\right)\right] + \operatorname{Var}\left(E\left[\sum_{j=0}^{J} X_{i,j} \middle| \Theta_i\right]\right) \\
&= E\left[\sum_{j=0}^{J} \gamma_j^2 \cdot \mu_i^2 \cdot \operatorname{Var}\left(\frac{X_{i,j}}{\gamma_j \cdot \mu_i} \middle| \Theta_i\right)\right] + \operatorname{Var}\left(\sum_{j=0}^{J} \gamma_j \cdot \mu_i \cdot E\left[\frac{X_{i,j}}{\gamma_j \cdot \mu_i} \middle| \Theta_i\right]\right) \\
&= \sum_{j=0}^{J} \gamma_j^2 \cdot \mu_i^2 \cdot \frac{\sigma^2}{\omega_{i,j}} + \mu_i^2 \cdot \operatorname{Var}\big(b'(\Theta_i)\big) \qquad (4.144) \\
&= \frac{\mu_i^2}{\mu_i^\delta} \cdot \sigma^2 + \mu_i^2 \cdot \tau^2.
\end{aligned}$$

Hence, we have that

$$\mathrm{Vco}^2(C_{i,J}) = \frac{\sigma^2}{\mu_i^\delta} + \tau^2. \tag{4.145}$$

This implies that

$$\alpha_{i,I-i} = \frac{\beta_{I-i}}{\beta_{I-i} + \frac{\sigma^2}{\mu_i^\delta} \cdot \tau^{-2}} = \frac{\beta_{I-i}}{\beta_{I-i} + \frac{\mathrm{Vco}^2(C_{i,J})}{\tau^2} - 1}. \tag{4.146}$$

For simplicity we have chosen $\delta = 0$ in Table 4.12, which implies that $\kappa_i \equiv \kappa$.

	τ	σ	κ	$\alpha_{i,I-i}$	$\widetilde{\mu(\Theta_i)}$	reserves EDF
0	5.00%	6.00%	1.4400	41.0%	0.9822	0
1	5.00%	6.00%	1.4400	40.9%	0.9746	15'715
2	5.00%	6.00%	1.4400	40.9%	0.9888	26'695
3	5.00%	6.00%	1.4400	40.9%	0.9669	36'333
4	5.00%	6.00%	1.4400	40.8%	0.9567	91'303
5	5.00%	6.00%	1.4400	40.6%	0.9509	169'281
6	5.00%	6.00%	1.4400	40.3%	0.9349	319'093
7	5.00%	6.00%	1.4400	39.7%	0.9136	524'484
8	5.00%	6.00%	1.4400	37.9%	0.9208	1'214'184
9	5.00%	6.00%	1.4400	29.0%	0.9502	4'530'884
						6'927'973

Table 4.12: Estimated reserves in the Exponential dispersion model with associate conjugate

The estimates in Table 4.10 and Table 4.12 lead to the same result.
Moreover, we see that the Bayesian estimate $\widetilde{\mu(\Theta_i)}$ is below 1 for all accident years i (see Table 4.12). This suggests (once more) that the choices of the a priori estimates μ_i for the ultimate claims were too conservative.

Conclusion 1. Corollary 4.42 implies that the estimator $\widehat{C_{i,J}}^{EDF}$ gives the optimal mixture between the Bornhuetter-Ferguson and the chain-ladder estimates in the EDF with associate conjugate: Assume that β_j and f_j are identified by (4.3) and set $\widehat{C}_{i,J}^{CL} = C_{i,I-i}/\beta_{I-i}$. Then we have that

$$\begin{aligned} \widehat{C_{i,J}}^{EDF} &= C_{i,I-i} + (1-\beta_{I-i}) \cdot \left[\alpha_{i,I-i} \cdot \frac{C_{i,I-i}}{\beta_{I-i}} + (1-\alpha_{i,I-i}) \cdot \mu_i \right] \\ &= C_{i,I-i} + (1-\beta_{I-i}) \cdot \left[\alpha_{i,I-i} \cdot \widehat{C_{i,J}}^{CL} + (1-\alpha_{i,I-i}) \cdot \mu_i \right] \\ &= C_{i,I-i} + (1-\beta_{I-i}) \cdot S_i\left(\alpha_{i,I-i}\right), \end{aligned} \tag{4.147}$$

where $S_i(\cdot)$ is the function defined in (4.1). Hence we have the mixture

$$\widehat{C_{i,J}}^{EDF} = \alpha_{i,I-i} \cdot \widehat{C_{i,J}}^{CL} + (1-\alpha_{i,I-i}) \cdot \widehat{C_{i,J}}^{BF} \tag{4.148}$$

between the CL estimate and the BF estimate. Moreover it minimizes the conditional MSEP in the Exponential dispersion model with associate conjugate. Observe that

$$\alpha_{i,I-i} = \frac{\beta_{I-i}}{\beta_{I-i} + \kappa_i}, \tag{4.149}$$

where the credibility coefficient was defined in (4.143). If we choose $\kappa_i = 0$ we obtain the chain-ladder estimate and if we choose $\kappa_i = \infty$ we obtain the Bornhuetter-Ferguson reserve.

Conclusion 2. Using (4.135) we find for all $I - i \leq j < J$ that

$$\begin{aligned} E\left[C_{i,j+1} | C_{i,0}, \ldots, C_{i,I-i}\right] \\ &= C_{i,I-i} + E\left[\left.\sum_{l=I-i+1}^{j+1} X_{i,l} \right| C_{i,0}, \ldots, C_{i,I-i}\right] \\ &= C_{i,I-i} + \sum_{l=I-i+1}^{j+1} \gamma_l \cdot \mu_i \cdot \widetilde{\mu(\Theta_i)} \\ &= \left(1 + \frac{\beta_{j+1} - \beta_{I-i}}{\beta_{I-i}} \cdot \alpha_{i,I-i}\right) \cdot C_{i,I-i} + (\beta_{j+1} - \beta_{I-i}) \cdot (1 - \alpha_{i,I-i}) \cdot \mu_i. \end{aligned} \tag{4.150}$$

In the 2nd step we explicitly use, that we have an exact Bayesian estimator. (4.150) does not hold true in the Bühlmann-Straub model (see Section 4.3 below). Formula (4.150) suggests that the EDF with associate conjugate is a "linear mixture" of the chain-ladder model and the Bornhuetter-Ferguson model. If we choose the credibility coefficient $\kappa_i = 0$, we obtain

$$E\left[C_{i,j+1} | C_{i,0}, \ldots, C_{i,j}\right] = \left(1 + \frac{\beta_{j+1} - \beta_j}{\beta_j}\right) \cdot C_{i,j} = f_j \cdot C_{i,j}, \tag{4.151}$$

if we assume (4.3). This is exactly the chain-ladder assumption (2.1). If we choose $\kappa_i = \infty$ then $\alpha_{i,I-i} = 0$ and we

$$E\left[C_{i,J} | C_{i,0}, \ldots, C_{i,I-i}\right] = C_{i,I-i} + (1 - \beta_{I-i}) \cdot \mu_i, \tag{4.152}$$

which is Model 2.8 that we have used to motivate the Bornhuetter-Ferguson estimate $\widehat{C_{i,J}}^{BF}$.

Under Model Assumptions 4.41 we obtain for the conditional mean square error of prediction

$$\mathrm{msep}_{\mu(\Theta_i)|\mathcal{D}_I}\left(\widetilde{\mu(\Theta_i)}\right) = E\left[\left.\left(\widetilde{\mu(\Theta_i)} - \mu(\Theta_i)\right)^2 \right| \mathcal{D}_I\right] = \mathrm{Var}(\mu(\Theta_i)|\mathcal{D}_I), \tag{4.153}$$

and hence we have that

$$\mathrm{msep}_{\mu(\Theta_i)}\left(\widetilde{\mu(\Theta_i)}\right) = E\big[\mathrm{Var}(\mu(\Theta_i)|\mathcal{D}_I)\big]. \tag{4.154}$$

If we plug in the estimator (4.140) we obtain

$$\begin{aligned}
&\mathrm{msep}_{\mu(\Theta_i)}\left(\widetilde{\mu(\Theta_i)}\right)\\
&= E\left[\left(\alpha_{i,I-i}\cdot\frac{C_{i,I-i}}{\beta_{I-i}\cdot\mu_i} + (1-\alpha_{i,I-i})\cdot 1 - \mu(\Theta_i)\right)^2\right]\\
&= E\left[\left(\alpha_{i,I-i}\cdot\left(\frac{C_{i,I-i}}{\beta_{I-i}\cdot\mu_i} - \mu(\Theta_i)\right) - (1-\alpha_{i,I-i})\cdot\big(\mu(\Theta_i)-1\big)\right)^2\right] \qquad (4.155)\\
&= (\alpha_{i,I-i})^2\cdot E\left[\mathrm{Var}\left(\left.\frac{C_{i,I-i}}{\beta_{I-i}\cdot\mu_i}\right|\Theta_i\right)\right] + (1-\alpha_{i,I-i})^2\cdot\mathrm{Var}\left(\mu(\Theta_i)\right)\\
&= (\alpha_{i,I-i})^2\cdot\frac{\sigma^2}{\beta_{I-i}\cdot\mu_i^\delta} + (1-\alpha_{i,I-i})^2\cdot\tau^2\\
&= (1-\alpha_{i,I-i})\cdot\tau^2,
\end{aligned}$$

where in the last step we have used $\tau^2\cdot(1-\alpha_{i,I-i}) = \alpha_{i,I-i}\cdot\frac{\sigma^2}{\mu_i^\delta\cdot\beta_{I-i}}$ (cf. (4.141)). From this we derive the unconditional mean square error of prediction for the estimate of $C_{i,J}$:

$$\begin{aligned}
\mathrm{msep}_{C_{i,J}}\left(\widehat{C_{i,J}}^{EDF}\right) &= E\left[\left((1-\beta_{I-i})\cdot\mu_i\cdot\widetilde{\mu(\Theta_i)} - (C_{i,J}-C_{i,I-i})\right)^2\right] \qquad (4.156)\\
&= \mu_i^2\cdot E\left[\left((1-\beta_{I-i})\left(\widetilde{\mu(\Theta_i)} - \mu(\Theta_i) + \mu(\Theta_i)\right) - \frac{C_{i,J}-C_{i,I-i}}{\mu_i}\right)^2\right]\\
&= \mu_i^2\cdot(1-\beta_{I-i})^2\cdot\mathrm{msep}_{\mu(\Theta_i)}\left(\widetilde{\mu(\Theta_i)}\right) + \sum_{k=I-i+1}^{J} E\big[\mathrm{Var}\left(X_{i,k}|\,\Theta_i\right)\big]\\
&= \mu_i^2\cdot\left[(1-\beta_{I-i})^2\cdot(1-\alpha_{i,I-i})\cdot\tau^2 + (1-\beta_{I-i})\cdot\sigma^2/\mu_i^\delta\right].
\end{aligned}$$

Moreover, if we set $\delta = 0$ we obtain

$$\mathrm{msep}_{C_{i,J}}\left(\widehat{C_{i,J}}^{EDF}\right) = \sigma^2\cdot(1-\beta_{I-i})\cdot\mu_i\cdot\frac{1+\frac{\sigma^2}{\mu_i\cdot\tau^2}}{\beta_{i-i}+\frac{\sigma^2}{\mu_i\cdot\tau}}. \tag{4.157}$$

This is the same value as for the Poisson-gamma case, see (4.116) and Table 4.11. For the conditional mean square error of prediction for the estimate of $C_{i,J}$, one needs to calculate

$$\mathrm{Var}\big(\mu(\Theta_i)|\mathcal{D}_I\big) = \mathrm{Var}\big(b'(\Theta_i)|\mathcal{D}_I\big), \tag{4.158}$$

where Θ_i, given $\mathcal{D}_I$, has a posteriori distribution $u_{\mu^{(i)}_{post,j},\tau^2_{post,j}}(\cdot)$ given by Lemma 4.35. We omit its further calculation.

Remarks on parameter estimation. So far we have always assumed that μ_i, γ_j, σ^2 and τ^2 are known. Under these assumptions we have calculated the Bayesian estimator which was optimal in the sense that it minimizes the MSEP. If the parameters are not known, the problem becomes substantially more difficult and in general one loses the optimality results.

Estimation of γ_j. At the moment we do not have a canonical way, how the claims development pattern should be estimated. In practice one often chooses the chain-ladder estimate $\widehat{\beta}_j^{(CL)}$ provided in (2.25) and then sets

$$\widehat{\gamma}_j^{(CL)} = \widehat{\beta}_j^{(CL)} - \widehat{\beta}_{j-1}^{(CL)}. \tag{4.159}$$

At the current stage we can not say anything about the optimality of this estimator. However, observe that for the Poisson-gamma model this estimator is natural in the sense that it coincides with the MLE estimator provided in the Poisson model (see Corollary 2.18). For more on this topic we refer to Subsection 4.2.5.

Estimation of μ_i. Usually, one takes a plan value, a budget value or the value used for the premium calculation (as in the BF method).

Estimation of σ^2 and τ^2. For known γ_j and μ_i one can give unbiased estimators for these variance parameters. For the moment we omit its formulation, because in Section 4.3 we see that the Exponential dispersion model with its associate conjugates satisfies the assumptions of the Bühlmann-Straub model. Hence we can take the same estimators as in the Bühlmann-Straub model and these are provided in Subsection 4.3.1.

4.2.5 Poisson-gamma case, revisited

In Model Assumptions 4.26 and 4.33 we have assumed that the claims development pattern γ_j is known. Of course, in general this is not the case and in practice one usually uses estimate (4.159) for the claims development pattern. In Verrall [82] this is called the "plug-in" estimate (which leads to the CL and BF mixture). However, in a full Bayesian approach one should also estimate this paremater in a Bayesian way (since usually it is not known). This means that we should also give an a priori distribution to the claims development pattern. For simplicity, we only treat the Poisson-gamma case (which was also considered in Verrall [82]). We have the following assumptions

Model Assumptions 4.44 (Poisson-gamma model)

There exist positive random vectors $\boldsymbol{\Theta} = (\Theta_i)_i$ and $\gamma = (\gamma_j)_j$ with $\sum_{j=0}^{J} \gamma_j = 1$ such that for all $i \in \{0, \dots, I\}$ and $j \in \{0, \dots, J\}$ we have

- conditionally, given $\boldsymbol{\Theta}$ and γ, the $X_{i,j}$ are independent and Poisson distributed with mean $\Theta_i \cdot \gamma_j$.
- $\boldsymbol{\Theta}$ and γ are independent and Θ_i are independent Gamma distributed with shape parameter a_i and scale parameter b_i, and γ is f_γ distributed.

□

As before, we can calculate the joint distribution of $\{X_{i,j}, i + j \leq I\}$, $\boldsymbol{\Theta}$ and γ, which is given by

$$f\left((x_{i,j})_{i+j\leq I}, \boldsymbol{\Theta}, \gamma\right) = \prod_{i+j\leq I} e^{-\Theta_i \cdot \gamma_j} \cdot \frac{(\Theta_i \cdot \gamma_j)^{x_{i,j}}}{x_{i,j}!} \prod_{i=0}^{I} f_{a_i, b_i}(\Theta_i) \cdot f_\gamma(\gamma). \tag{4.160}$$

The posterior distribution $(\boldsymbol{\Theta}, \gamma)$ given the observations $\mathcal{D}_I$ is proportional to

$$\prod_{i=0}^{I} f_{a^{post}_{i,I-i}, b^{post}_{I-i}}(\Theta_i) \cdot \prod_{j=0}^{(I-i)\wedge J} \gamma_j^{x_{i,j}} \cdot f_\gamma(\gamma), \tag{4.161}$$

with

$$a^{post}_{i,j} = a_i + \sum_{k=0}^{j\wedge J} X_{i,j} \qquad \text{and} \qquad b^{post}_{i,j} = b_i + \sum_{k=0}^{j\wedge J} \gamma_k, \tag{4.162}$$

see also Lemma 4.28. From this we immediately see, that one can not calculate analytically the posterior distribution of $(\boldsymbol{\Theta}, \gamma)$ given the observations $\mathcal{D}_I$, but this also implies that we can not easily calculate the conditional distribution of X_{k+l}, $k + l > I$, given the observations $\mathcal{D}_I$. Hence these Bayesian models can only be implemented with the help of numerical simulations, e.g. the Markov Chain Monte Carlo (MCMC) approach. The implementation using a simulation-based MCMC is discussed in de Alba [2, 4] and Scollnik [70].

4.3 Bühlmann-Straub Credibility Model

In the last section we have seen an exact Bayesian approach to the claims reserving problem. The Bayesian estimator

$$\widetilde{\mu(\Theta_i)} = E\left[\mu(\Theta_i) \middle| X_{i,0}, \dots, X_{i,I-i}\right] \tag{4.163}$$

is the best estimator for $\mu(\Theta_i)$ in the class of all estimators which are square integrable functions of the observations $X_{i,0}, \ldots, X_{i,I-i}$. The crucial point in the calculation was that – from the EDF with its associate conjugates – we were able to explicitly calculate the a posteriori distribution of $\mu(\Theta_i)$. Moreover, the parameters of the a posteriori distribution and the Bayesian estimator were linear in the observations. However, in most of the Bayesian models we are not in the situation where we are able to calculate the a posteriori distribution, and therefore the Bayesian estimator cannot be expressed in a closed analytical form. I.e. in general the Bayesian estimator does not meet the practical requirements of simplicity and intuitiveness and can only be calculated by numerical procedures such as Markov Chain Monte Carlo methods (MCMC methods).
In cases where we are not able to derive the Bayesian estimator we restrict the class of possible estimators to a smaller class, which are linear functions of the observations $X_{i,0}, \ldots, X_{i,I-i}$. This means that we try to get an estimator which minimizes the quadratic loss function among all estimators which are linear combinations of the observations $X_{i,0}, \ldots, X_{i,I-i}$. The result will be an estimator which is practicable and intuitive by definition. This approach is well-known in actuarial science as credibility theory and since "best" is also to be understood in the Bayesian sense credibility estimators are linear Bayes estimators (see Bühlmann-Gisler [18]).
In claims reserving the credibility theory was used e.g. by De Vylder [84], Neuhaus [56] and Mack [51] in the Bühlmann-Straub context.

In the sequel we always assume that the incremental loss development pattern $(\gamma_j)_{j=0,\ldots,J}$ given by

$$\gamma_0 = \beta_0 \quad \text{and} \quad \gamma_j = \beta_j - \beta_{j-1} \text{ for } j = 1, \ldots, J \tag{4.164}$$

is known as in the previous sections on Bayesian estimates.

Model Assumptions 4.45 (Bühlmann-Straub model [18])

- Conditionally, given Θ_i, the increments $X_{i,0}, \ldots, X_{i,J}$ are independent with

$$E\left[X_{i,j}/\gamma_j \middle| \Theta_i\right] = \mu(\Theta_i), \tag{4.165}$$

$$\mathrm{Var}\left(X_{i,j}/\gamma_j \middle| \Theta_i\right) = \sigma^2(\Theta_i)/\gamma_j \tag{4.166}$$

  for all $i = 0, \ldots, I$ and $j = 0, \ldots, J$.

- The pairs $(\Theta_i, \mathbf{X}_i)$ $(i = 0, \ldots, I)$ are independent, and the Θ_i are independent and identically distributed.

For the cumulative claim amount we obtain

$$E\left[C_{i,j}|\Theta_i\right] = \beta_j \cdot \mu(\Theta_i), \tag{4.167}$$

$$\mathrm{Var}\left(C_{i,j}|\Theta_i\right) = \beta_j \cdot \sigma^2(\Theta_i). \tag{4.168}$$

The latter equation shows that this model is different from Model 4.14. The term $(1-\beta_j)\cdot\alpha^2(C_{i,J})$ is replaced by $\sigma^2(\Theta_i)$. On the other hand the Bühlmann-Straub model is very much in the spirit of the EDF with its associate conjugates. The parameter Θ_i plays the role of the underlying risk characteristics, i.e. the parameter Θ_i is unknown and tells us whether we have a good or bad accident year. For a more detailed explanation in the framework of tariffication and pricing we refer to Bühlmann-Gisler [18].

In linear credibility theory one looks for an estimate $\widehat{\mu(\Theta_i)}$ of $\mu(\Theta_i)$ which minimizes the quadratic loss function among all estimators which are linear in the observations $X_{i,j}$ (see also [18], Definition 3.8). I.e. one has to solve the optimization problem

$$\widehat{\mu(\Theta_i)}^{cred} = \mathrm{argmin}_{\widetilde{\mu}\in L(\mathbf{X},1)} E\left[\left(\mu(\Theta)-\widetilde{\mu}\right)^2\right], \tag{4.169}$$

where

$$L(\mathbf{X},1) = \left\{\widetilde{\mu};\ \widetilde{\mu} = a_{i,0} + \sum_{i=0}^{I}\sum_{j=0}^{(I-i)\wedge J} a_{i,j}\cdot X_{i,j} \quad \text{with } a_{i,j}\in\mathbb{R}\right\}. \tag{4.170}$$

Remarks 4.46

- Observe that the credibility estimator $\widehat{\mu(\Theta_i)}^{cred}$ is linear in the observations $X_{i,j}$ by definition. We could also allow for general real-valued, square integrable functions of the observations $X_{i,j}$. In that case we obtain simply the Bayesian estimator since the conditional a posteriori expectation minimizes the quadratic loss function among all estimators which are a square integrable function of the observations.

- Credibility estimators can also be constructed using Hilbert space theory. Indeed (4.169) asks for a minimization in an L^2-sense, which corresponds to orthogonal projections in Hilbert spaces. For more on this topic we refer to Bühlmann-Gisler [18].

We define the structural parameters

$$\mu_0 = E\left[\mu(\Theta_i)\right], \tag{4.171}$$

$$\sigma^2 = E\left[\sigma^2(\Theta_i)\right], \tag{4.172}$$

$$\tau^2 = \mathrm{Var}\left(\mu(\Theta_i)\right). \tag{4.173}$$

Theorem 4.47 (inhomogeneous Bühlmann-Straub estimator)
Under Model Assumptions 4.45 the optimal linear inhomogeneous estimator of $\mu(\Theta_i)$, given the observations $\mathcal{D}_I$, is given by

$$\widehat{\mu(\Theta_i)}^{cred} = \alpha_i \cdot Y_i + (1 - \alpha_i) \cdot \mu_0 \tag{4.174}$$

for $I - J + 1 \le i \le I$, where

$$\alpha_i = \frac{\beta_{I-i}}{\beta_{I-i} + \sigma^2/\tau^2}, \tag{4.175}$$

$$Y_i = \sum_{j=0}^{(I-i)\wedge J} \frac{\gamma_j}{\beta_{I-i}} \cdot \frac{X_{i,j}}{\gamma_j} = \frac{C_{i,(I-i)\wedge J}}{\beta_{I-i}}. \tag{4.176}$$

In credibility theory the a priori mean μ_0 can also be estimated from the data. This leads to the homogeneous credibility estimator.

Theorem 4.48 (homogeneous Bühlmann-Straub estimator)
Under Model Assumptions 4.45 the optimal linear homogeneous estimator of $\mu(\Theta_i)$ given the observations $\mathcal{D}_I$ is given by

$$\widehat{\mu(\Theta_i)}^{hom} = \alpha_i \cdot Y_i + (1 - \alpha_i) \cdot \widehat{\mu}_0 \tag{4.177}$$

for $I - J + 1 \le i \le I$, where α_i and Y_i are given in Theorem 4.47 and

$$\widehat{\mu}_0 = \sum_{i=0}^{I} \frac{\alpha_i}{\alpha_\bullet} \cdot Y_i, \qquad \textit{with} \qquad \alpha_\bullet = \sum_{i=0}^{I} \alpha_i. \tag{4.178}$$

Proof of Theorem 4.47 and Theorem 4.48. We refer to Theorems 4.2 and 4.4 in Bühlmann-Gisler [18].

□

Remarks 4.49

- If the a priori mean μ_0 is known we choose the inhomogeneous credibility estimator $\widehat{\mu(\Theta_i)}^{cred}$ from Theorem 4.47. This estimator minimizes the quadratic loss function given in (4.169) among all estimators given in (4.170).

  If the a priori mean μ_0 is unknown, we estimate its value also from the data. This is done by switching to the homogeneous credibility estimator $\widehat{\mu(\Theta_i)}^{hom}$ given in Theorem 4.48. The crucial part is that we have to slightly change the set of possible estimators given in (4.170) towards

$$L_e(\mathbf{X}) = \left\{ \widetilde{\mu}\ ;\ \widetilde{\mu} = \sum_{i=0}^{I} \sum_{j=0}^{(I-i)\wedge J} a_{i,j} \cdot X_{i,j} \text{ with } a_{i,j} \in \mathbb{R},\ E\left[\widetilde{\mu}\right] = \mu_0 \right\}. \tag{4.179}$$

The homogeneous credibility estimator minimizes the quadratic loss function among all estimators from the set $L_e(\mathbf{X})$, i.e.

$$\widehat{\mu(\Theta_i)}^{hom} = \text{argmin}_{\widetilde{\mu} \in L_e(\mathbf{X})} E\left[(\mu(\Theta) - \widetilde{\mu})^2 \right]. \tag{4.180}$$

- The crucial point in the credibility estimators in (4.174) and (4.177) is that we take a weighted average Y_i between the individual observations of accident year i and the a priori mean μ_0 and its estimator $\widehat{\mu}_0$, respectively. Observe that the weighted average Y_i only depends on the observations of accident year i. This is a consequence of the independence assumption between the accident years. However, the estimator $\widehat{\mu}_0$ uses the observations of all accident years since the a priori mean μ_0 holds for all accident years. The credibility weight $\alpha_i \in [0, 1]$ for the weighted average of the individual observations Y_i becomes small when the expected fluctuations within the accidents years σ^2 are large and becomes large if the fluctuations between the accident years τ^2 are large.

- The estimator (4.174) is exactly the same as the one from the exponential dispersion model with associate conjugates (Corollary 4.42) if we assume that all a priori means μ_i are equal and $\delta = 0$.

- Since the inhomogeneous estimator $\widehat{\mu(\Theta_i)}^{cred}$ contains a constant it is automatically an unbiased estimator for the a priori mean μ_0. In contrast to $\widehat{\mu(\Theta_i)}^{cred}$ the homogeneous $\widehat{\mu(\Theta_i)}^{hom}$ is unbiased for μ_0 by definition.

- The weights γ_j in the model assumptions could be replaced by weights $\gamma_{i,j}$, then the Bühlmann-Straub result still holds true. Indeed, one could choose a design matrix $\gamma_{i,j} = \Gamma_i(j)$ to apply the Bühlmann-Straub model (see Taylor [75] and Mack [47]) and the variance condition is then replaced by

$$\text{Var}\left(X_{i,j}/\gamma_{j,i} | \Theta_i \right) = \frac{\sigma^2(\Theta_i)}{V_i \cdot \gamma_{j,i}^{\delta}}, \tag{4.181}$$

where $V_i > 0$ is an appropriate measure for the volume and $\delta > 0$. $\delta = 1$ is the model favoured by Mack [47], whereas De Vylder [84] has chosen $\delta = 2$. For $\delta = 0$ we obtain a condition which is independent of j (credibility model of Bühlmann, see [18]).

Different a priori means μ_i. If $X_{i,j}/\gamma_j$ has different a priori means μ_i for different accident years i, we modify the Bühlmann-Straub assumptions (4.165)-(4.166) to

$$E\left[\left. \frac{X_{i,j}}{\gamma_j \cdot \mu_i} \right| \Theta_i \right] = \mu(\Theta_i), \tag{4.182}$$

$$\text{Var}\left(\left. \frac{X_{i,j}}{\gamma_j \cdot \mu_i} \right| \Theta_i \right) = \frac{\sigma^2(\Theta_i)}{\gamma_j \cdot \mu_i^{\delta}}, \tag{4.183}$$

for an appropriate choice $\delta \geq 0$. In this case we have $E[\mu(\Theta_i)] = 1$ and the inhomogeneous and homogeneous credibility estimator are given by

$$\widehat{\mu(\Theta_i)}^{cred} = \alpha_i \cdot Y_i + (1-\alpha_i) \cdot 1, \tag{4.184}$$

and

$$\widehat{\mu(\Theta_i)}^{hom} = \alpha_i \cdot Y_i + (1-\alpha_i) \cdot \widehat{\mu}_0, \tag{4.185}$$

respectively, where

$$Y_i = \frac{C_{i,I-i\wedge J}}{\mu_i \cdot \beta_{I-i}}, \quad \alpha_i = \frac{\beta_{I-i}}{\beta_{I-i}+\kappa_i} \quad \text{with } \kappa_i = \frac{\sigma^2}{\mu_i^\delta \cdot \tau^2}. \tag{4.186}$$

Observe that this gives now completely the same estimator as in the exponential dispersion family with its associate conjugates (see Corollary 4.42).
This immediately gives the following estimators:

Estimator 4.50 (Bühlmann-Straub credibility reserving estimator)
In the Bühlmann-Straub model 4.45 with generalized assumptions (4.182)-(4.183) *we have the following estimators*

$$\widehat{C_{i,J}}^{cred} = \widehat{E}\left[C_{i,J}\middle|\mathcal{D}_I\right] = C_{i,I-i} + (1-\beta_{I-i}) \cdot \mu_i \cdot \widehat{\mu(\Theta_i)}^{cred}, \tag{4.187}$$

$$\widehat{C_{i,J}}^{hom} = \widehat{E}\left[C_{i,J}\middle|\mathcal{D}_I\right] = C_{i,I-i} + (1-\beta_{I-i}) \cdot \mu_i \cdot \widehat{\mu(\Theta_i)}^{hom} \tag{4.188}$$

for $I-J+1 \leq i \leq I$.

Lemma 4.51 *In the Bühlmann-Straub model 4.45 the quadratic losses for the credibility estimators are given by*

$$E\left[\left(\widehat{\mu(\Theta_i)}^{cred} - \mu(\Theta_i)\right)^2\right] = \tau^2 \cdot (1-\alpha_i), \tag{4.189}$$

$$E\left[\left(\widehat{\mu(\Theta_i)}^{hom} - \mu(\Theta_i)\right)^2\right] = \tau^2 \cdot (1-\alpha_i) \cdot \left(1 + \frac{1-\alpha_i}{\alpha_\bullet}\right) \tag{4.190}$$

for $I-J+1 \leq i \leq I$.

Proof. We refer to Theorems 4.3 and 4.6 in Bühlmann-Gisler [18].

□

Corollary 4.52 *In the Bühlmann-Straub model 4.45 with generalized assumptions* (4.182)-(4.183) *the mean square errors of prediction of the inhomogeneous and homogeneous credibility reserving estimator are given by*

$$msep_{C_{i,J}}\left(\widehat{C_{i,J}}^{cred}\right) = \mu_i^2 \cdot \left[(1-\beta_{I-i}) \cdot \sigma^2/\mu_i^\delta + (1-\beta_{I-i})^2 \cdot \tau^2 \cdot (1-\alpha_i)\right]. \tag{4.191}$$

and

$$msep_{C_{i,J}}\left(\widehat{C_{i,J}}^{hom}\right) = msep_{C_{i,J}}\left(\widehat{C_{i,J}}^{cred}\right) + \mu_i^2 \cdot (1-\beta_{I-i})^2 \cdot \tau^2 \cdot \frac{(1-\alpha_i)^2}{\alpha_\bullet}, \quad (4.192)$$

respectively, for $I - J + 1 \leq i \leq I$.

Remarks 4.53

- The first term on the right-hand side of the above equalities stands again for the process error whereas the second terms stand for the parameter/prediction errors (how good can an actuary predict the mean). Observe again, that we assume that the incremental loss development pattern $(\gamma_j)_{j=0,\dots,J}$ is known, and hence we do not estimate the estimation error in the claims development pattern.

- Observe the MSEP formula for the credibility estimator conicides with the one for the exponential dispersion family, see (4.156).

Proof. We separate the mean square error of prediction as follows

$$\mathrm{msep}_{C_{i,J}}\left(\widehat{C_{i,J}}^{cred}\right) = E\left[\left((1-\beta_{I-i}) \cdot \mu_i \cdot \widehat{\mu(\Theta_i)}^{cred} - (C_{i,J} - C_{i,I-i})\right)^2\right]. \quad (4.193)$$

Conditionally, given $\boldsymbol{\Theta} = (\Theta_0, \dots, \Theta_I)$, we have that the increments $X_{i,j}$ are independent. But this immediately implies that the expression in (4.193) is equal to

$$E\left[E\left[(1-\beta_{I-i})^2 \cdot \mu_i^2 \cdot \left(\widehat{\mu(\Theta_i)}^{cred} - \mu(\Theta_i)\right)^2 \middle| \boldsymbol{\Theta}\right]\right] \quad (4.194)$$

$$+ E\left[E\left[\left((1-\beta_{I-i}) \cdot \mu_i \cdot \mu(\Theta_i) - (C_{i,J} - C_{i,I-i})\right)^2 \middle| \boldsymbol{\Theta}\right]\right]$$

$$= (1-\beta_{I-i})^2 \cdot \mu_i^2 \cdot \mathrm{msep}_{\mu(\Theta_i)}\left(\widehat{\mu(\Theta_i)}^{cred}\right) + E\left[\mathrm{Var}\left(C_{i,J} - C_{i,I-i} \middle| \boldsymbol{\Theta}\right)\right].$$

But then the claim follows from Lemma 4.51 and

$$\mathrm{Var}\left(C_{i,J} - C_{i,I-i} \middle| \boldsymbol{\Theta}\right) = (1-\beta_{I-i}) \cdot \mu_i^{2-\delta} \cdot \sigma^2(\Theta_i). \quad (4.195)$$

□

4.3.1 Parameter estimation

So far (in the example) the choice of the variance parameters was rather artificial. In this subsection we provide estimators for σ^2 and τ^2. In practical applications it is often convenient to eliminate outliers for the estimation of σ^2 and τ^2, since the estimators are often not very robust.

Before we start with the parameter estimations we would like to mention that in this section essentially the same remarks apply as the ones mentioned on page 125. We need to estimate γ_j, σ^2 and τ^2. For the weights γ_j we proceed as in (4.159) Estimate the claims development pattern β_j from (2.25). The incremental losse development pattern γ_j is then estimated by (4.164).

We define

$$S_i = \frac{1}{(I-i)\wedge J} \sum_{j=0}^{(I-i)\wedge J} \gamma_j \cdot \left(\frac{X_{i,j}}{\gamma_j} - Y_i \right)^2. \tag{4.196}$$

Then S_i is an unbiased estimator for σ^2 (see [18], (4.22)). Hence σ^2 is estimated by the following unbiased estimator

$$\widehat{\sigma^2} = \frac{1}{I} \sum_{i=0}^{I-1} S_i. \tag{4.197}$$

For the estimation of τ^2 we define

$$T = \sum_{i=0}^{I} \frac{\beta_{I-i}}{\sum_i \beta_{I-i}} \cdot \left(Y_i - \overline{Y} \right)^2, \tag{4.198}$$

where

$$\overline{Y} = \frac{\sum_i \beta_{I-i} \cdot Y_i}{\sum_i \beta_{I-i}} = \frac{\sum_i C_{i,(I-i)\wedge J}}{\sum_i \beta_{I-i}}. \tag{4.199}$$

Then an unbiased estimator for τ^2 is given by (see [18], (4.26))

$$\widehat{\tau^2} = c \cdot \left\{ T - \frac{I \cdot \sigma^2}{\sum_i \beta_{I-i}} \right\}, \tag{4.200}$$

with

$$c = \left(\sum_{i=0}^{I} \frac{\beta_{I-i}}{\sum_i \beta_{I-i}} \cdot \left(1 - \frac{\beta_{I-i}}{\sum_i \beta_{I-i}} \right) \right)^{-1}. \tag{4.201}$$

If $\widehat{\tau^2}$ is negative it is set to zero.

If we work with different μ_i we have to slightly change the estimators (see Bühlmann-Gisler [18], Section 4.8).

Example 4.54 (Bühlmann-Straub model, constant μ_i)

We revisit the data given in Example 2.7. We recall that we have set $\mathrm{Vco}\big(\mu(\Theta_i)\big) = 5\%$ and $\mathrm{Vco}(C_{i,J}) = 7.8\%$, using external know how only (see Tables 4.9 and 4.4). For this example we assume that all a priori expectations μ_i are equal and we use the homogeneous credibility estimator. We have the following observations, where the incremental claims development pattern γ_j is estimated via the chain-ladder method.

	0	1	2	3	4	5	6	7	8	9
0	5'946'975	9'668'212	10'563'929	10'771'690	10'978'394	11'040'518	11'106'331	11'121'181	11'132'310	11'148'124
1	6'346'756	9'593'162	10'316'383	10'468'180	10'536'004	10'572'608	10'625'360	10'636'546	10'648'192	
2	6'269'090	9'245'313	10'092'366	10'355'134	10'507'837	10'573'282	10'626'827	10'635'751		
3	5'863'015	8'546'239	9'268'771	9'459'424	9'592'399	9'680'740	9'724'068			
4	5'778'885	8'524'114	9'178'009	9'451'404	9'681'692	9'786'916				
5	6'184'793	9'013'132	9'585'897	9'830'796	9'935'753					
6	5'600'184	8'493'391	9'056'505	9'282'022						
7	5'288'066	7'728'169	8'256'211							
8	5'290'793	7'648'729								
9	5'675'568									
$\widehat{f}_j$	1.4925	1.0778	1.0229	1.0148	1.0070	1.0051	1.0011	1.0010	1.0014	

Table 4.13: Observed historical cumulative payments $C_{i,j}$ and estimated chain-ladder factors $\widehat{f}_j$, see Table 2.2

	1	2	3	4	5	6	7	8	9	10
0	10'086'719	12'814'544	13'090'078	9'577'303	14'357'308	9'048'371	12'901'245	13'793'367	10'658'637	11'148'124
1	10'764'791	11'179'404	10'569'215	6'997'504	4'710'946	5'331'290	10'340'861	10'390'677	11'152'804	
2	10'633'061	10'248'997	12'378'890	12'113'052	10'606'498	9'531'934	10'496'344	8'289'406		
3	9'944'313	9'240'019	10'559'131	8'788'679	9'236'259	12'866'767	8'493'606			
4	9'801'620	9'453'540	9'556'060	12'602'942	15'995'382	15'325'912				
5	10'490'085	9'739'738	8'370'426	11'289'335	7'290'128					
6	9'498'524	9'963'120	8'229'388	10'395'847						
7	8'969'136	8'402'801	7'716'853							
8	8'973'762	8'119'848								
9	9'626'383									
γ_j	59.0%	29.0%	6.8%	2.2%	1.4%	0.7%	0.5%	0.1%	0.1%	0.1%

Table 4.14: Observed scaled incremental payments $X_{i,j}/\gamma_j$ and estimated incremental claims development pattern $\widehat{\gamma}_j$

Hence we obtain the following estimators:

$$c = 1.11316, \tag{4.202}$$

$$\overline{Y} = 9'911'975, \tag{4.203}$$

$$\widehat{\sigma} = 337'289, \tag{4.204}$$

$$\widehat{\tau} = 734'887, \tag{4.205}$$

$$\widehat{\mu_0} = 9'885'584. \tag{4.206}$$

This leads with $\widehat{\kappa} = \frac{\widehat{\sigma^2}}{\widehat{\tau^2}} = 21.1\%$, $\widehat{\text{Vco}}\big(\mu(\Theta_i)\big) = \frac{\widehat{\tau}}{\widehat{\mu_0}} = 7.4\%$ and $\widehat{\text{Vco}}(C_{i,J}) = \frac{\left(\widehat{\sigma^2}+\widehat{\tau^2}\right)^{1/2}}{\widehat{\mu_0}} = 8.2\%$ to the following reserves:

	α_i	$\widehat{C_{i,J}}^{cred}$	estimated reserves CL	estimated reserves hom. cred.
0	82.6%	11'148'124	0	0
1	82.6%	10'663'125	15'126	14'934
2	82.6%	10'661'675	26'257	25'924
3	82.5%	9'758'685	34'538	34'616
4	82.5%	9'872'238	85'302	85'322
5	82.4%	10'091'682	156'494	155'929
6	82.2%	9'569'836	286'121	287'814
7	81.8%	8'716'445	449'167	460'234
8	80.7%	8'719'642	1'043'242	1'070'913
9	73.7%	9'654'386	3'950'815	3'978'818
			6'047'061	6'114'503

Table 4.15: Estimated reserves in the homogeneous Bühlmann-Straub model (constant μ_i)

We see that the estimates are close to the chain-ladder method. This comes from the fact that the credibility weights are rather big: Since $\widehat{\kappa}$ is rather small compared to β_{I-i} we obtain credibility weights which are all larger than 70%.
For the mean square errors of prediction we obtain the values in Table 4.16.

Example 4.55 (Bühlmann-Straub model, varying μ_i)

We revisit the data set given in Example 2.7 and Example 4.55. This time we assume that an a priori differentiation μ_i is given by Table 4.6 (a priori mean for Bornhuetter-Ferguson method). We apply the scaled model (4.182)-(4.183) for $\delta = 0, 1, 2$ and obtain the reserves in Table 4.17.
We see that the estimates for different δ's do not differ too much, and they are still close to the chain-ladder method. However, they differ from the estimates for the constant μ_i case (see Table 4.15).
For the estimated variational coefficient we have for $\delta = 0, 1, 2$

$$\widehat{\text{Vco}}\big(\mu(\Theta_i)\big) \approx 6.8\%. \tag{4.207}$$

	$\operatorname{msep}^{1/2}_{C_{i,J}}\left(\widehat{C_{i,J}}^{cred}\right)$	$\operatorname{msep}^{1/2}_{C_{i,J}}\left(\widehat{C_{i,J}}^{hom}\right)$
0	0	0
1	12'711	12'711
2	16'755	16'755
3	20'095	20'096
4	31'465	31'467
5	42'272	42'278
6	59'060	59'076
7	78'301	78'339
8	123'114	123'259
9	265'775	267'229
total	314'699	315'998

Table 4.16: Mean square error of prediction in the Bühlmann-Straub model (constant μ_i)

	credibility weights α_i			reserves	credibility reserves		
	$\delta = 0$	$\delta = 1$	$\delta = 2$	CL	$\delta = 0$	$\delta = 1$	$\delta = 2$
0	80.2%	80.6%	81.1%	0	0	0	0
1	80.1%	80.2%	80.3%	15'126	14'943	14'944	14'944
2	80.1%	79.6%	79.1%	26'257	25'766	25'753	25'740
3	80.1%	79.1%	78.0%	34'538	34'253	34'238	34'222
4	80.0%	79.7%	79.3%	85'302	85'056	85'051	85'046
5	79.9%	80.2%	80.4%	156'494	156'562	156'561	156'559
6	79.7%	79.8%	80.0%	286'121	289'078	289'056	289'035
7	79.3%	79.0%	78.8%	449'167	460'871	461'021	461'180
8	78.1%	77.6%	77.0%	1'043'242	1'069'227	1'069'815	1'070'427
9	70.4%	71.0%	71.5%	3'950'815	4'024'687	4'023'270	4'021'903
total				6'047'061	6'160'443	6'159'709	6'159'056
$\widehat{\mu_0}$	0.8810	0.8809	0.8809				

Table 4.17: Estimated reserves in the homogeneous Bühlmann-Straub model (varying μ_i)

This describes the accuracy of the estimate of the "true" expected mean by the actuary. Observe that we have chosen 5% in Example 4.54.

Moreover, we see (once more) that the a priori estimate μ_i seems to be rather pessimistic, since $\widehat{\mu_0}$ is substantially smaller than 1 (for all δ).

For the mean square error of prediction we obtain the values in Table 4.18.

4.4 Multidimensional credibility models

In Section 4.3 we have assumed that the incremental payments have the following form

$$E\left[X_{i,j}|\Theta_i\right] = \gamma_j \cdot \mu(\Theta_i). \tag{4.208}$$

	$\text{msep}^{1/2}_{C_{i,J}}\left(\widehat{C_{i,J}}^{hom}\right)$		
	$\delta = 0$	$\delta = 1$	$\delta = 2$
1	12'835	12'771	12'711
2	16'317	16'532	16'755
3	18'952	19'511	20'094
4	30'871	31'161	31'464
5	43'110	42'682	42'272
6	59'876	59'456	59'059
7	77'383	77'819	78'282
8	120'119	121'536	123'008
9	273'931	269'926	266'054
total	320'377	317'540	314'889

Table 4.18: Mean square error of prediction in the Bühlmann-Straub model (varying μ_i)

The constant γ_j denotes the payment ratio in period j. If we rewrite this in vector form we obtain

$$E\left[\mathbf{X}_i|\Theta_i\right] = \gamma \cdot \mu(\Theta_i), \tag{4.209}$$

where $\mathbf{X}_i = (X_{i,0}, \ldots, X_{i,J})'$ and $\gamma = (\gamma_0, \ldots, \gamma_J)'$.
We see that the stochastic terms $\mu(\Theta_i)$ can only act as a scalar. Sometimes we would like to have more flexibility, i.e. we replace $\mu(\Theta_i)$ by a vector. This leads to a generalization of the Bühlmann-Straub model.

4.4.1 Hachemeister regression model

Model Assumptions 4.56 (Hachemeister regression model [31])

- There exist p-dimensional design vectors $\gamma_j(i) = \left(\gamma_{j,1}(i), \ldots, \gamma_{j,p}(i)\right)'$ and vectors $\mu(\Theta_i) = \left(\mu_1(\Theta_i), \ldots, \mu_p(\Theta_i)\right)'$ ($p \leq J+1$) such that we have

$$\begin{aligned} E\left[X_{i,j}|\Theta_i\right] &= \gamma_j(i)' \cdot \mu(\Theta_i), && (4.210)\\ \text{Cov}\left(X_{i,j}, X_{i,k}|\Theta_i\right) &= \Sigma_{j,k,i}(\Theta_i) && (4.211)\end{aligned}$$

  for all $i \in \{0, \ldots, I\}$ and $j \in \{0, \ldots, J\}$.

- The $(J+1)\times p$ matrix $\Gamma_i = \left(\gamma_0(i), \ldots, \gamma_J(i)\right)'$ has rank p and the components $\mu_1(\Theta_i), \ldots, \mu_p(\Theta_i)$ of $\mu(\Theta_i)$ are linearly independent.

- The pairs $(\Theta_i, \mathbf{X}_i)$ ($i = 0, \ldots, I$) are independent, and the Θ_i are independent and identically distributed.

Remarks 4.57

- We are now in the credibility regression case, see Bühlmann-Gisler [18], Section 8.3, where $\mu(\Theta_i) = \big(\mu_1(\Theta_i), \ldots, \mu_p(\Theta_i)\big)'$ is a p-dimensional vector, which we would like to estimate.

- Γ_i is a known $(J+1) \times p$ design matrix.

We define the following parameters

$$\begin{aligned} \mu &= E\big[\mu(\Theta_i)\big], &(4.212)\\ S_{j,k,i} &= E\big[\Sigma_{j,k,i}(\Theta_i)\big], &(4.213)\\ T &= \mathrm{Cov}\big(\mu(\Theta), \mu(\Theta)\big), &(4.214)\\ S_i &= \left(S_{j,k,i}\right)_{j,k=0,\cdots,J} &(4.215)\end{aligned}$$

for $i \in \{0, \ldots, I\}$ and $j, k \in \{0, \ldots, J\}$. Hence T is a $p \times p$ covariance matrix for the variability between the different accident years and S_i is a $(J+1) \times (J+1)$ matrix that describes the variability within the accident year i. An important special case for S_i is given by

$$S_i = \sigma^2 \cdot W_i^{-1} = \sigma^2 \cdot \mathrm{diag}\big(w_{i,0}^{-1}, \ldots, w_{i,J}^{-1}\big), \tag{4.216}$$

for appropriate weights $w_{i,j} > 0$ and a scalar $\sigma^2 > 0$.

Theorem 4.58 (Hachemeister estimator)
Under Model Assumptions 4.56 the optimal linear inhomogeneous estimator for $\mu(\Theta_i)$ is given by

$$\widehat{\mu(\Theta_i)}^{cred} = A_i \cdot \mathbf{B}_i + (1 - A_i) \cdot \mu, \tag{4.217}$$

with

$$\begin{aligned} A_i &= T \cdot \left(T + \left(\Gamma_i^{[I-i]\prime} S_i^{-1} \Gamma_i^{[I-i]}\right)^{-1}\right)^{-1}, &(4.218)\\ \mathbf{B}_i &= \left(\Gamma_i^{[I-i]\prime} S_i^{-1} \Gamma_i^{[I-i]}\right)^{-1} \cdot \Gamma_i^{[I-i]\prime} S_i^{-1} \cdot \mathbf{X}_i^{[I-i]}, &(4.219)\end{aligned}$$

where

$$\begin{aligned} \Gamma_i^{[I-i]} &= \big(\gamma_0(i), \ldots, \gamma_{(I-i)\wedge J}(i), \mathbf{0}, \ldots, \mathbf{0}\big)' &(4.220)\\ \mathbf{X}_i^{[I-i]} &= (X_{i,0}, \ldots, X_{i,(I-i)\wedge J}, 0, \ldots, 0)' &(4.221)\end{aligned}$$

for $I - J + 1 \le i \le I$ with $p \le I - i + 1$. The quadratic loss matrix for the credibility estimator is given by

$$E\left[\left(\widehat{\mu(\Theta_i)}^{cred} - \mu(\Theta_i)\right) \cdot \left(\widehat{\mu(\Theta_i)}^{cred} - \mu(\Theta_i)\right)'\right] = (1 - A_i) \cdot T. \tag{4.222}$$

Proof. See Theorem 8.7 in Bühlmann-Gisler [18].

□

We have the following corollary:

Corollary 4.59 (Standard Regression) *Under Model Assumption 4.56 with S_i given by* (4.216) *we have*

$$A_i = T \cdot \left(T + \sigma^2 \cdot \left(\Gamma_i^{[I-i]\prime} W_i \Gamma_i^{[I-i]} \right)^{-1} \right)^{-1}, \tag{4.223}$$

$$\mathbf{B}_i = \left(\Gamma_i^{[I-i]\prime} W_i \Gamma_i^{[I-i]} \right)^{-1} \cdot \Gamma_i^{[I-i]\prime} W_i \cdot \mathbf{X}_i^{[I-i]} \tag{4.224}$$

for $I - J + 1 \leq i \leq I$ with $p \leq I - i + 1$.

This leads to the following reserving estimator:

Estimator 4.60 (Hachemeister credibility reserving estimator) *In the Hachemeister Regression Model 4.56 the estimator is given by*

$$\widehat{C_{i,J}}^{cred} = C_{i,I-i} + \sum_{j=I-i+1}^{J} \gamma_j(i)' \cdot \widehat{\mu(\Theta_i)}^{cred} \tag{4.225}$$

for $I - J + 1 \leq i \leq I$ with $p \leq I - i + 1$.

Remarks 4.61

- If μ is not known, then (4.217) can be replaced by the homogeneous credibility estimator for $\mu(\Theta_i)$ using

$$\widehat{\mu} = \left(\sum_{i=0}^{I} A_i \right)^{-1} \cdot \sum_{i=0}^{I} A_i \cdot \mathbf{B}_i. \tag{4.226}$$

  In that case the right-hand side of (4.222) needs to be replaced by

$$(1 - A_i) \cdot T \cdot \left(1 + \left(\sum_{i=0}^{I} A_i' \right)^{-1} \cdot (1 - A_i') \right). \tag{4.227}$$

- Term (4.219) gives the formula for the data compression (see also Theorem 8.6 in Bühlmann-Gisler [18]). We already see from this that for $p > 1$ we have some difficulties with considering the youngest years since the dimension of μ is larger than the available number of observations if $p > I - i + 1$. Observe that

$$E\left[\mathbf{B}_i \middle| \Theta_i\right] = \mu(\Theta_i), \tag{4.228}$$

$$E\left[\left(\mathbf{B}_i - \mu(\Theta_i) \right) \cdot \left(\mathbf{B}_i - \mu(\Theta_i) \right)' \right] = \left(\Gamma_i^{[I-i]\prime} S_i^{-1} \Gamma_i^{[I-i]} \right)^{-1}. \tag{4.229}$$

- **Choices of the design matrix Γ_i.** There are various possibilities to choose the design matrix Γ_i. One possibility which is used is the so-called Hoerl curve (see De Jong-Zehnwirth [42] and Zehnwirth [92]), set $p = 3$ and

$$\gamma_j(i) = (1, \log(j+1), j)'. \tag{4.230}$$

- **Parameter estimation.** It is rather difficult to get good parameter estimations in this model for $p > 1$. If we assume that the covariance matrix $\left(\Sigma_{j,k,i}(\Theta_i)\right)_{j,k=0,\ldots,J}$ is diagonal with mean S_i given by (4.216), we can estimate S_i with the help of the one-dimensional Bühlmann-Straub model (see Subsection 4.3.1). An unbiased estimator for the covariance matrix T is given by

$$\widehat{T} = \frac{1}{I-p} \sum_{i=0}^{I-p} E\left[\left(\mathbf{B}_i - \overline{\mathbf{B}}\right) \cdot \left(\mathbf{B}_i - \overline{\mathbf{B}}\right)'\right] - \frac{1}{I-p+1} \sum_{i=0}^{I-p} \left(\Gamma_i^{[I-i]'} S_i^{-1} \Gamma_i^{[I-i]}\right)^{-1}, \tag{4.231}$$

with

$$\overline{\mathbf{B}} = \frac{1}{I-p+1} \sum_{i=0}^{I-p} \mathbf{B}_i. \tag{4.232}$$

- **Examples.** In all examples we have looked at it was rather difficult to obtain reasonable estimates for the claims reserves. This has various reasons: 1) There is not an obvious choice for a good design matrix Γ_i. In our examples the Hoerl curve has not well-behaved. 2) The estimation of the structural parameters S_i and T are always difficult. Moreover they are not robust against outliers. 3) Already, slight perturbations of the data had a large effect on the resulting reserves. For all these reasons we do not give a real data example, i.e. the Hachemeister model is very interesting from a theoretical point of view, from a practical point of view it is rather difficult to apply it to real data.

4.4.2 Other credibility models

In the Bühlmann-Straub credibility model we had a deterministic cashflow pattern γ_j and we have estimated the exposure $\mu(\Theta_i)$ of the accident years. We could also exchange the role of these two parameters

Model Assumptions 4.62

There exist scalars μ_i $(i = 0, \ldots, I)$ such that

- conditionally, given Θ_i, we have for all $j \in \{0, \dots, J\}$

$$E[X_{i,j}|\Theta_i] = \gamma_j(\Theta_i) \cdot \mu_i. \tag{4.233}$$

- The pairs $(\Theta_i, \mathbf{X}_i)$ $(i = 0, \dots, I)$ are independent, and the Θ_i are independent and identically distributed.

Remarks 4.63

- Now the whole vector $\gamma(\Theta_i) = \big(\gamma_0(\Theta_i), \dots, \gamma_J(\Theta_i)\big)'$ is a random drawing with

$$E\big[\gamma_j(\Theta_i)\big] = \gamma_j, \tag{4.234}$$
$$\mathrm{Cov}\big(\gamma_j(\Theta_i), \gamma_k(\Theta_i)\big) = T_{j,k}, \tag{4.235}$$
$$\mathrm{Cov}\left(X_{i,j}, X_{i,k} \middle| \Theta_i\right) = \Sigma_{j,k,i}(\Theta_i). \tag{4.236}$$

- The difficulty in this model is that we have observations $X_{i,0}, \dots, X_{i,I-i}$ for $\gamma_0(\Theta_i), \dots, \gamma_{I-i}(\Theta_i)$ and we need to estimate $\gamma_{I-i+1}(\Theta_i), \dots, \gamma_J(\Theta_i)$. This is slightly different from classical one-dimensional credibility applications. From this it is clear that a crucial role is played by the covariance structures, which projects past observations to the future.

- For general covariance structures it is difficult to give nice formulas. Special cases were studied by Jewell [38] and Hesselager-Witting [35]. Hesselager-Witting [35] assume that the vectors

$$(\gamma_0(\Theta_i), \dots, \gamma_J(\Theta_i)) \tag{4.237}$$

  are i.i.d. Dirichlet distributed with parameters $a_0, \dots, a_J$. Define $a = \sum_{j=0}^J a_j$ then we have (see Hesselager-Witting [35], formula (3))

$$E\big[\gamma_j(\Theta_i)\big] = \gamma_j = a_j/a, \tag{4.238}$$
$$\mathrm{Cov}\big(\gamma_j(\Theta_i), \gamma_k(\Theta_i)\big) = T_{j,k} = \frac{1}{1+a} \cdot \left(1_{j=k} \cdot \gamma_j - \gamma_j \cdot \gamma_k\right). \tag{4.239}$$

  If we then choose a specific form for the covariance structure $\Sigma_{j,k,i}(\Theta_i)$ we can work out a credibility formula for the expected ultimate claim.

Of course there is a large variety of other credibility models, such as e.g. hierarchical credibility models Hesselager [36]. We do not further discuss them here.

4.5 Kalman filter

Kalman filters are an enhancement of credibility models. We will treat only the one-dimensional case, since already in the multivariate credibility context we have seen that it becomes difficult to go to higher dimensions.

Kalman filters are evolutionary credibility models. If we take e.g. the Bühlmann-Straub model then it is assumed that Θ_i $(i = 0, \ldots, I)$ are independent and identically distributed (see Model 4.45). If we go back to Example 4.54 we obtain the following picture for the observations $Y_0, \ldots, Y_I$ and the estimate $\widehat{\mu_0}$ for the a priori mean μ_0 (cf. (4.176) and (4.178), respectively):

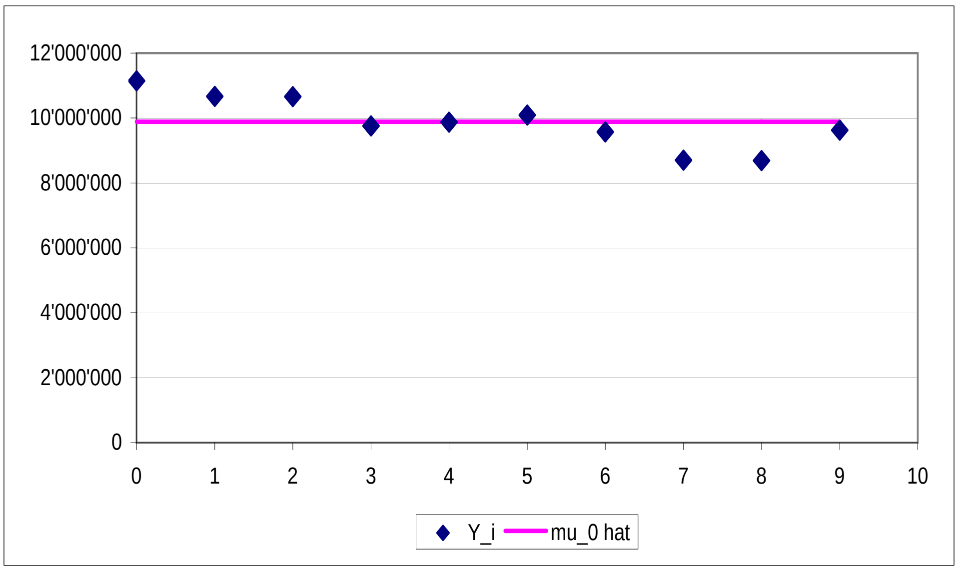


Figure 4.2: Observations Y_i and estimate $\widehat{\mu_0}$

From Figure 4.2 it is not obvious that $\boldsymbol{\Theta} = (\Theta_0, \Theta_1, \ldots)$ is a process of identically distributed random variables. We could also have underwriting cycles which would rather suggest, that neighboring Θ_i's are dependent. Hence, we assume that $\boldsymbol{\Theta} = (\Theta_0, \Theta_1, \ldots)$ is a stochastic process of random variables which are not necessarily independent and identically distributed.

Model Assumptions 4.64 (Kalman filter)

- $\boldsymbol{\Theta} = (\Theta_0, \Theta_1, \ldots)$ is a stochastic process.
- Conditionally, given $\boldsymbol{\Theta}$, the increments $X_{i,j}$ are independent with for all i, j

$$E\left[X_{i,j}/\gamma_j \middle| \boldsymbol{\Theta}\right] = \mu(\Theta_i), \tag{4.240}$$

$$\operatorname{Cov}\left(X_{i,j}/\gamma_j, X_{k,l}/\gamma_l \middle| \boldsymbol{\Theta}\right) = 1_{\{i=k, j=l\}} \cdot \sigma^2(\Theta_i)/\gamma_j. \tag{4.241}$$

- $\big(\mu(\Theta_i)\big)_{i\ge 0}$ is a martingale.

□

Remarks 4.65

- The assumption (4.241) can be relaxed in the sense that we only need in the average (over $\mathbf{\Theta}$) conditional uncorrelatedness. Assumption (4.241) implies that we obtain an updating procedure which is recursive.

- The martingale assumption implies that we have uncorrelated centered increments $\mu(\Theta_{i+1}) - \mu(\Theta_i)$ (see also (1.25)),

$$E\left[\mu(\Theta_{i+1})\right|\mu(\Theta_0),\ldots,\mu(\Theta_i)] = \mu(\Theta_i). \tag{4.242}$$

  In Hilbert space language this reads as follows: The projection of $\mu(\Theta_{i+1})$ onto the subspace of all square integrable functions of $\mu(\Theta_0),\ldots,\mu(\Theta_i)$ is simply $\mu(\Theta_i)$, i.e. the process $\big(\mu(\Theta_i)\big)_{i\ge 0}$ has centered orthogonal increments. This last assumption could be generalized to linear transformations (see Corollary 9.5 in Bühlmann-Gisler [18]).

We introduce the following notations (the notation is motivated by the usual terminology from state space models, see e.g. Abraham-Ledolter [1]):

$$\begin{aligned}
\mathbf{Y}_i &= (X_{i,0}/\gamma_0,\ldots,X_{i,I-i}/\gamma_{I-i}), && (4.243)\\
\mu_{i|i-1} &= \operatorname{argmin}_{\widetilde{\mu}\in L(\mathbf{Y}_0,\ldots,\mathbf{Y}_{i-1},1)} E\left[\left(\mu(\Theta_i)-\widetilde{\mu}\right)^2\right], && (4.244)\\
\mu_{i|i} &= \operatorname{argmin}_{\widetilde{\mu}\in L(\mathbf{Y}_0,\ldots,\mathbf{Y}_{i},1)} E\left[\left(\mu(\Theta_i)-\widetilde{\mu}\right)^2\right] && (4.245)
\end{aligned}$$

(cf. (4.170)). $\mu_{i|i-1}$ is the best linear forecast for $\mu(\Theta_i)$ based on the information $\mathbf{Y}_0,\ldots,\mathbf{Y}_{i-1}$. Whereas $\mu_{i|i}$ is the best linear forecast for $\mu(\Theta_i)$ which is also based on $\mathbf{Y}_i$. Hence there are two updating procedures: 1) updating from $\mu_{i|i-1}$ to $\mu_{i|i}$ on the basis of the newest observation $\mathbf{Y}_i$ and 2) updating from $\mu_{i|i}$ to $\mu_{i+1|i}$ due to the parameter movement from $\mu(\Theta_i)$ to $\mu(\Theta_{i+1})$.
We define the following structural parameters

$$\begin{aligned}
\sigma^2 &= E\left[\sigma^2(\Theta_i)\right], && (4.246)\\
\delta_i^2 &= \operatorname{Var}\big(\mu(\Theta_i)-\mu(\Theta_{i-1})\big), && (4.247)\\
q_{i|i-1} &= E\left[\big(\mu_{i,i-1}-\mu(\Theta_i)\big)^2\right], && (4.248)\\
q_{i|i} &= E\left[\big(\mu_{i,i}-\mu(\Theta_i)\big)^2\right]. && (4.249)
\end{aligned}$$

Theorem 4.66 (Kalman filter recursion formula, Theorem 9.6 in [18])
Under Model Assumptions 4.64 we have

1. *Anchoring ($i = 0$)*

$$\mu_{0|-1} = \mu_0 = E\left[\mu(\Theta_0)\right] \quad \text{and} \quad q_{0|-1} = \tau_0^2 = Var\big(\mu(\Theta_0)\big). \tag{4.250}$$

2. *Recursion ($i \geq 0$)*

   (a) *Observation update:*

$$\mu_{i|i} = \alpha_i \cdot Y_i + (1-\alpha_i) \cdot \mu_{i|i-1}, \tag{4.251}$$
$$q_{i|i} = (1-\alpha_i) \cdot q_{i|i-1}, \tag{4.252}$$
$$\tag{4.253}$$

   with

$$\alpha_i = \frac{\beta_{I-i}}{\beta_{I-i} + \sigma^2/q_{i|i-1}}, \tag{4.254}$$
$$Y_i = \sum_{j=0}^{(I-i)\wedge J} \frac{\gamma_j}{\beta_{I-i}} \frac{X_{i,j}}{\gamma_j} = \frac{C_{i,(I-i)\wedge J}}{\beta_{I-i}}. \tag{4.255}$$

   (b) *Parameter update:*

$$\mu_{i+1|i} = \mu_{i|i} \quad \text{and} \quad q_{i+1|i} = q_{i|i} + \delta_{i+1}^2. \tag{4.256}$$

Proof. For the proof we refer to Theorem 9.6 in Bühlmann-Gisler [18].

□

This leads to the following reserving estimator:

Estimator 4.67 (Kalman filter reserving estimator)
In the Kalman filter model 4.64 the estimator is given by

$$\widehat{C_{i,J}}^{Ka} = \widehat{E}\left[C_{i,J}\middle|\,\mathcal{D}_I\right] = C_{i,I-i} + (1-\beta_{I-i}) \cdot \mu_{i|i} \tag{4.257}$$

for $I - J + 1 \leq i \leq I$.

Remarks 4.68

- In practice we face two difficulties: 1) We need to estimate all the parameters. 2) We need "good" estimates for the starting values μ_0 and τ_0^2 for the iteration.
- Parameter estimation: For the estimation of σ^2 we choose $\widehat{\sigma^2}$ as in the Bühlmann-Straub model (see (4.197)). The estimation of δ_i^2 is less straightforward, in fact we need to define a special case of the Model Assumptions 4.64.

Model Assumptions 4.69 (Gerber-Jones [28])

- Model Assumptions 4.64 hold.
- There exists a sequence $(\Delta_i)_{i\geq 1}$ of independent random variables with $E[\Delta_i] = 0$ and $\mathrm{Var}(\Delta_i) = \delta^2$ such that

$$\mu(\Theta_i) = \mu(\Theta_{i-1}) + \Delta_i \tag{4.258}$$

  for all $i \geq 1$.
- $\mu(\Theta_0)$ and Δ_i are independent for all $i \geq 1$.

□

Remark. In this model holds $\delta_i = \mathrm{Var}\big(\mu(\Theta_i) - \mu(\Theta_{i-1})\big) = \mathrm{Var}(\Delta_i) = \delta^2$.

Let us first calculate the variances and covariances of Y_i defined in (4.255).

$$\begin{aligned}
\mathrm{Var}(Y_i) &= \mathrm{Var}\big(E\left[Y_i\middle|\,\mathbf{\Theta}\right]\big) + E\big[\mathrm{Var}\left(Y_i\middle|\,\mathbf{\Theta}\right)\big] \\
&= \mathrm{Var}\big(\mu(\Theta_i)\big) + E\left[\sum_{j=0}^{(I-i)\wedge J} \frac{\gamma_j^2}{\beta_{I-i}^2} \cdot \mathrm{Var}\left(\frac{X_{i,j}}{\gamma_j}\middle|\,\mathbf{\Theta}\right)\right] \\
&= \mathrm{Var}\big(\mu(\Theta_0)\big) + i\cdot\delta^2 + \frac{1}{\beta_{I-i}}\cdot\sigma^2.
\end{aligned} \tag{4.259}$$

Assume that $i > l$

$$\begin{aligned}
\mathrm{Cov}(Y_i, Y_l) &= \mathrm{Cov}\big(E\left[Y_i\middle|\,\mathbf{\Theta}\right], E\left[Y_l\middle|\,\mathbf{\Theta}\right]\big) + E\big[\mathrm{Cov}\left(Y_i, Y_l\middle|\,\mathbf{\Theta}\right)\big] \\
&= \mathrm{Cov}\big(\mu(\Theta_i), \mu(\Theta_l)\big) \\
&= \mathrm{Cov}\left(\mu(\Theta_l) + \sum_{k=l+1}^{i}\Delta_k, \mu(\Theta_l)\right) \\
&= \mathrm{Var}\big(\mu(\Theta_0)\big) + l\cdot\delta^2.
\end{aligned} \tag{4.260}$$

We define $\overline{Y}$ as in (4.199) with $\beta_\bullet = \sum_{i=0}^I \beta_{I-i}$. Hence

$$\begin{aligned}
\sum_{i=0}^{I} \frac{\beta_{I-i}}{\beta_\bullet}\cdot E\left[\left(Y_i - \overline{Y}\right)^2\right] &= \sum_{i=0}^{I} \frac{\beta_{I-i}}{\beta_\bullet}\cdot \mathrm{Var}\left(Y_i - \frac{\sum_{i=0}^I \beta_i\cdot Y_i}{\beta_\bullet}\right) \\
&= \sum_{i=0}^{I} \frac{\beta_{I-i}}{\beta_\bullet}\cdot\mathrm{Var}(Y_i) - \sum_{k,l=0}^{I}\frac{\beta_{I-k}\cdot\beta_{I-l}}{\beta_\bullet^2}\cdot\mathrm{Cov}(Y_k, Y_l) \\
&= \frac{(I+1)\cdot\sigma^2}{\beta_\bullet} + \delta^2\cdot\sum_{i=0}^{I}\left(i\cdot\frac{\beta_{I-i}}{\beta_\bullet} - \sum_{k=0}^{I}\min\{i,k\}\cdot\frac{\beta_{I-k}\cdot\beta_{I-i}}{\beta_\bullet^2}\right) \\
&= \frac{(I+1)\cdot\sigma^2}{\beta_\bullet} + \delta^2\cdot\sum_{i=0}^{I}\sum_{k=0}^{i-1}(i-k)\cdot\frac{\beta_{I-k}\cdot\beta_{I-i}}{\beta_\bullet^2}.
\end{aligned} \tag{4.261}$$

This motivates the following unbiased estimator for δ^2 (see also (4.198)):

$$\begin{aligned}\widehat{\delta^2} &= \left(\sum_{i=0}^{I}\sum_{k=0}^{i-1}(i-k)\cdot\frac{\beta_{I-k}\cdot\beta_{I-i}}{\beta_\bullet^2}\right)^{-1}\cdot\left(\sum_{i=0}^{I}\frac{\beta_{I-i}}{\beta_\bullet}\cdot\left(Y_i-\overline{Y}\right)^2-\frac{(I+1)\cdot\sigma^2}{\beta_\bullet}\right)\\ &= c^*\cdot\left(T-\frac{(I+1)\cdot\sigma^2}{\beta_\bullet}\right), \qquad (4.262)\end{aligned}$$

with

$$c^* = \left(\sum_{i,k=0}^{I}\max\{i-k,0\}\cdot\frac{\beta_{I-i}\cdot\beta_{I-k}}{\beta_\bullet^2}\right)^{-1}. \qquad (4.263)$$

Observe that expression (4.262) is similar to the estimator of τ^2 in the Bühlmann-Straub model (4.200). The difference lies in the constant.

Example 4.70 (Kalman filter)

We revisit the Example 4.54.
We have the following parameters and estimates

$$\begin{aligned} c^* &= 0.62943, & (4.264)\\ \overline{Y} &= 9'911'975, & (4.265)\\ \widehat{\sigma} &= 337'289, & (4.266)\\ \widehat{\delta} &= 545'637. & (4.267)\end{aligned}$$

We start the iteration with the estimates

$$\widehat{\mu}_0 = 9'885'584 \qquad \text{and} \qquad \widehat{\tau}_0 = \widehat{\delta} = 545'637 \qquad (4.268)$$

(see also (4.178)).

	$\mu_{i\|i-1}$	$q_{i\|i-1}^{1/2}$	α_i	Y_i	$\mu_{i\|i}$	$q_{i\|i}^{1/2}$	$\mu_{i+1\|i}$	$q_{i+1\|i}^{1/2}$
0	9’885’584	545’637	72.4%	11’148’123	10’799’066	286’899	10’799’066	616’466
1	10’799’066	616’466	76.9%	10’663’316	10’694’625	296’057	10’694’625	620’781
2	10’694’625	620’781	77.2%	10’662’005	10’669’454	296’651	10’669’454	621’064
3	10’669’454	621’064	77.2%	9’758’602	9’966’628	296’805	9’966’628	621’138
4	9’966’628	621’138	77.1%	9’872’213	9’893’857	297’401	9’893’857	621’423
5	9’893’857	621’423	77.0%	10’092’241	10’046’550	298’230	10’046’550	621’820
6	10’046’550	621’820	76.7%	9’568’136	9’679’468	299’967	9’679’468	622’655
7	9’679’468	622’655	76.4%	8’705’370	8’935’539	302’670	8’935’539	623’962
8	8’935’539	623’962	75.1%	8’691’961	8’752’681	311’533	8’752’681	628’309
9	8’752’681	628’309	67.2%	9’626’366	9’339’528	360’009	9’339’528	653’702

Table 4.19: Iteration in the Kalman filter

We see that in this example the credibility weights are smaller compared to the Bühlmann-Straub model (see Table 4.15). However, they are still rather high, which means that the a priori value $\mu_{i|i-1}$ will move rather closely with the observations Y_{i-1}. Hence we are now able to model dependent time-series, where the a priori value incorporates the past observed loss ratios: see Figure 4.3.

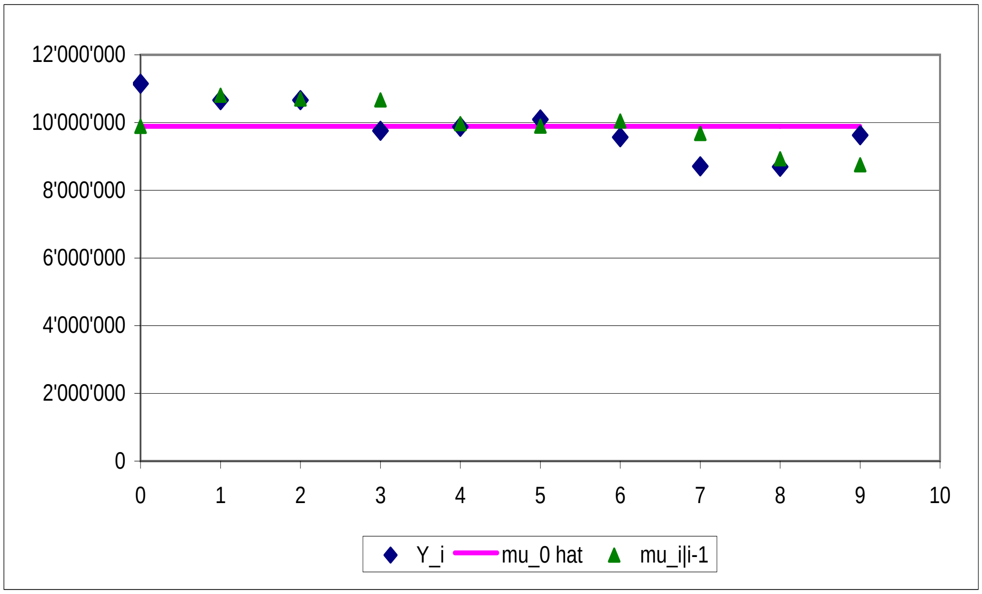


Figure 4.3: Observations Y_i, estimate $\widehat{\mu_0}$ and estimates $\mu_{i|i-1}$

	$\widehat{C_{i,J}}^{Ka}$	estimated reserves CL	hom. cred.	Kalman
0	11'148'123	0	0	0
1	10'663'360	15'126	14'934	15'170
2	10'662'023	26'257	25'924	26'275
3	9'759'339	34'538	34'616	35'274
4	9'872'400	85'302	85'322	85'489
5	10'091'532	156'494	155'929	155'785
6	9'571'465	286'121	287'814	289'450
7	8'717'246	449'167	460'234	461'042
8	8'699'249	1'043'242	1'070'913	1'050'529
9	9'508'643	3'950'815	3'978'818	3'833'085
total		6'047'061	6'114'503	5'952'100

Table 4.20: chain-ladder reserves, homogeneous Bühlmann-Straub reserves and Kalman filter reserves

Chapter 5

Outlook

Several topics on stochastic claims reserving methods need to be added to the current version of this manuscript: e.g.

- explicit distributional models and methods, such as the Log-normal model or Tweedie's compound Poisson model
- generalized linear model methods
- bootstrapping methods
- multivariate methods
- Munich chain-ladder method
- etc.

Appendix A

Unallocated loss adjustment expenses

A.1 Motivation

In this section we describe the "New York"-method for the estimation of unallocated loss adjustment expenses (ULAE). The "New York"-method for estimating ULAE is, unfortunately, only poorly documented in the literature (e.g. as footnotes in Feldblum [26] and Foundation CAS [19]).

In non-life insurance there are usually two different kinds of claims handling costs, external ones and internal ones. External costs like costs for external lawyers or for an external expertise etc. are usually allocated to single claims and are therefore contained in the usual claims payments and loss development figures. These payments are called allocated loss adjustment expenses (ALAE). Typically, internal loss adjustment expenses (income of claims handling department, maintenance of claims handling system, etc.) are not contained in the claims figures and therefore have to be estimated separately. These internal costs can usually not be allocated to single claims. We call these costs unallocated loss adjustment expenses (ULAE). From a regulatory point of view, we should also build reserves for these costs/expenses because they are part of the claims handling process which guarantees that an insurance company is able to meet all its obligations. I.e. ULAE reserves should guarantee the smooth run off of the old insurance liabilities without "pay-as-you-go" from new business/premium for the internal claims handling processes.

A.2 Pure claims payments

Usually, claims development figures only consist of "pure" claims payments not containing ULAE charges. They are usually studied in loss development triangles or trapezoids as above (see Section 1.3).

In this section we denote by $X_{i,j}^{(pure)}$ the "pure" incremental payments for accident year i $(0 \le i \le I)$ in development year j $(0 \le j \le J)$. "Pure" always means, that these quantities do not contain ULAE (this is exactly the quantity studied in Section 1.3). The cumulative pure payments for accident year i after development period j are denoted by (see (1.41))

$$C_{i,j}^{(pure)} = \sum_{k=0}^{j} X_{i,k}{}^{(pure)}. \tag{A.1}$$

We assume that $X_{i,j}^{(pure)} = 0$ for all $j > J$, i.e. the ultimate pure cumulative loss is given by $C_{i,J}^{(pure)}$.

We have observations for $\mathcal{D}_I = \{X_{i,j}^{(pure)};\ 0 \le i \le I \text{ and } 0 \le j \le \min\{J, I-i\}\}$ and the complement of $\mathcal{D}_I$ needs to be predicted.

For the New York-method we also need a second type of development trapezoids, namely a "reporting" trapezoid: For accident year i, $Z_{i,j}^{(pure)}$ denotes the pure cumulative ultimate claim amount for all those claims, which are reported up to (and including) development year j. Hence $\left(Z_{i,0}^{(pure)}, Z_{i,1}^{(pure)}, \ldots\right)$ with $Z_{i,J}^{(pure)} = C_{i,J}^{(pure)}$ describes, how the pure ultimate claim $C_{i,J}^{(pure)}$ is reported over time at the insurance company. Of course, this reporting pattern is much more delicate, because sizes which are reported in the upper set $\widetilde{\mathcal{D}}_I = \{Z_{i,j}^{(pure)};\ 0 \le i \le I \text{ and } 0 \le j \le \min\{J, I-i\}\}$ are still developping, since usually it takes quite some time between the reporting and the final settlement of a claim. In general, the final value for $Z_{i,j}^{(pure)}$ is only known at time $i + J$.

Remark: Since the New York-method is an algorithm based on deterministic numbers, we assume that all our variables are deterministic. Stochastic variables are replaced by their "best estimate" for its conditional mean at time I. We think that for the current presentation (to explain the New York-method) it is not helpful to work in a stochastic framework.

A.3 ULAE charges

The cumulative ULAE payments for accident year i until development period j are denoted by $C_{i,j}^{(ULAE)}$. And finally, the total cumulative payments (pure and ULAE) are denoted by

$$C_{i,j} = C_{i,j}^{(pure)} + C_{i,j}^{(ULAE)}. \tag{A.2}$$

The cumulative ULAE payments $C_{i,j}^{(ULAE)}$ and the incremental ULAE charges

$$X_{i,j}^{(ULAE)} = C_{i,j}^{(ULAE)} - C_{i,j-1}^{(ULAE)} \tag{A.3}$$

need to be estimated: The main difficulty is that for each accounting year $t \leq I$ we usually have only one aggregated observation

$$X_t^{(ULAE)} = \sum_{\substack{i+j=t \\ 0 \leq j \leq J}} X_{i,j}^{(ULAE)} \qquad \text{(sum over } t\text{-diagonal)}. \tag{A.4}$$

I.e. ULAE payments are usually not available for single accident years but rather we have a position "Total ULAE Expenses" for each accounting year t (in general ULAE charges are contained in the position "Administrative Expenses" in the annual profit-and-loss statement).
Hence, for the estimation of future ULAE payments we need first to define an appropriate model in order to split the aggregated observations $X_t^{(ULAE)}$ into the different accident years $X_{i,j}^{(ULAE)}$.

A.4 New York-method

The New York-method assumes that one part of the ULAE charge is proportional to the claims registration (denote this proportion by $r \in [0, 1]$) and the other part is proportional to the settlement (payments) of the claims (proportion $1 - r$).

Assumption A.1 *We assume that there are two development patterns* $(\gamma_j)_{j=0,\dots,J}$ *and* $(\delta_j)_{j=0,\dots,J}$ *with* $\gamma_j \geq 0$, $\delta_j \geq 0$, *for all* j, *and* $\sum_{j=0}^J \gamma_j = \sum_{j=0}^J \delta_j = 1$ *such that (***cashflow or payout pattern***)*

$$X_{i,j}^{(pure)} = \gamma_j \cdot C_{i,J}^{(pure)} \tag{A.5}$$

*and (***reporting pattern***)*

$$Z_{i,j}^{(pure)} = \sum_{l=0}^{j} \delta_l \cdot C_{i,J}^{(pure)} \tag{A.6}$$

for all i *and* j.

Remarks:

- Equation (A.5) describes, how the pure ultimate claim $C_{i,J}^{(pure)}$ is paid over time. In fact γ_j gives the cashflow pattern for the pure ultimate claim $C_{i,J}^{(pure)}$. We propose that γ_j is estimated by the classical chain-ladder factors f_j, see (3.12)

$$\widehat{\gamma}_j^{CL} = \frac{1}{\widehat{f}_j \cdots \widehat{f}_{J-1}} \cdot \left(1 - \frac{1}{\widehat{f}_{j-1}}\right). \tag{A.7}$$

- The estimation of the claims reporting pattern δ_j in (A.6) is more delicate. As we have seen there are not many claims reserving methods which give a reporting pattern δ_j. Such a pattern can only be obtained if one separates the claims estimates for reported claims and IBNyR claims (incurred but not yet reported).

Model Assumptions A.2 *Assume that there exists $r \in [0,1]$ such that the incremental ULAE payments satisfy for all i and all j*

$$X_{i,j}^{(ULAE)} = \big(r \cdot \delta_j + (1-r) \cdot \gamma_j\big) \cdot C_{i,J}^{(ULAE)}. \tag{A.8}$$

Henceforth, we assume that one part (r) of the ULAE charge is proportional to the reporting pattern (one has loss adjustment expenses at the registration of the claim), and the other part $(1-r)$ of the ULAE charge is proportional to the claims settlement (measured by the payout pattern).

Definition A.3 (Paid-to-paid ratio) *We define for all t*

$$\pi_t = \frac{X_t^{(ULAE)}}{X_t^{(pure)}} = \frac{\sum_{\substack{i+j=t \\ 0 \le j \le J}} X_{i,j}^{(ULAE)}}{\sum_{\substack{i+j=t \\ 0 \le j \le J}} X_{i,j}^{(pure)}}. \tag{A.9}$$

The paid-to-paid ratio measures the ULAE payments relative to the pure claim payments in each accounting year t.

Lemma A.4 *Assume there exists $\pi > 0$ such that for all accident years i we have*

$$\frac{C_{i,J}^{(ULAE)}}{C_{i,J}^{(pure)}} = \pi. \tag{A.10}$$

Under Assumption A.1 and Model A.2 we have for all accounting years t

$$\pi_t = \pi, \tag{A.11}$$

whenever $C_{i,J}^{(pure)}$ is constant in i.

Proof of Lemma A.4. We have

$$\pi_t = \frac{\sum_{\substack{i+j=t \\ 0\le j\le J}} X_{i,j}^{(ULAE)}}{\sum_{\substack{i+j=t \\ 0\le j\le J}} X_{i,j}^{(pure)}} = \frac{\sum_{j=0}^{J}\big(r\cdot\delta_j + (1-r)\cdot\gamma_j\big)\cdot C_{t-j,J}^{(ULAE)}}{\sum_{j=0}^{J}\gamma_j\cdot C_{t-j,J}^{(pure)}}$$

$$= \pi\cdot\frac{\sum_{j=0}^{J}\big(r\cdot\delta_j + (1-r)\cdot\gamma_j\big)\cdot C_{t-j,J}^{(pure)}}{\sum_{j=0}^{J}\gamma_j\cdot C_{t-j,J}^{(pure)}} = \pi. \tag{A.12}$$

This finishes the proof.

□

We define the following split of the claims reserves for accident year i at time j:

$$R_{i,j}^{(pure)} = \sum_{l>j} X_{i,l}^{(pure)} = \sum_{l>j}\gamma_l\cdot C_{i,J}^{(pure)} \quad \text{(total pure claims reserves)},$$

$$R_{i,j}^{(IBNyR)} = \sum_{l>j}\delta_l\cdot C_{i,J}^{(pure)} \quad \text{(IBNyR reserves, incurred but not yet reported)},$$

$$R_{i,j}^{(rep)} = R_{i,j}^{(pure)} - R_{i,j}^{(IBNyR)} \quad \text{(reserves for reported claims)}.$$

Estimator A.5 (New York-method) *Under the assumptions of Lemma A.4 we can predict π using the observations π_t (accounting year data). The reserves for ULAE charges for accident year i after development year j, $R_{i,j}^{(ULAE)} = \sum_{l>j} X_{i,l}^{(ULAE)}$, are estimated by*

$$\begin{aligned}\widehat{R}_{i,j}^{(ULAE)} &= \pi\cdot r\cdot R_{i,j}^{(IBNyR)} + \pi\cdot(1-r)\cdot R_{i,j}^{(pure)} \\ &= \pi\cdot R_{i,j}^{(IBNyR)} + \pi\cdot(1-r)\cdot R_{i,j}^{(rep)}.\end{aligned} \tag{A.13}$$

Explanation of Result A.5.

We have under the assumptions of Lemma A.4 for all i, j that

$$\begin{aligned}R_{i,j}^{(ULAE)} &= \sum_{l>j}\big(r\cdot\delta_l + (1-r)\cdot\gamma_l\big)\cdot C_{i,J}^{(ULAE)} \qquad (A.14)\\ &= \pi\cdot\sum_{l>j}\big(r\cdot\delta_l + (1-r)\cdot\gamma_l\big) C_{i,J}^{(pure)} \\ &= \pi\cdot r\cdot R_{i,j}^{(IBNyR)} + \pi\cdot(1-r)\cdot R_{i,j}^{(pure)}.\end{aligned}$$

Remarks:

- In practice one assumes the stationarity condition $\pi_t = \pi$ for all t. This implies that π can be estimated from the accounting data of the annual profit-and-loss statements. Pure claims payments are directly contained in the profit-and-loss statements, whereas ULAE payments are often contained in the administrative expenses. Hence one needs to divide this position into further subpositions (e.g. with the help of an activity-based cost allocation split).

- Result A.5 gives an easy formula for estimating ULAE reserves. If we are interested into the total ULAE reserves after accounting year t we simply have

$$\widehat{R}_t^{(ULAE)} = \sum_{i+j=t} \widehat{R}_{i,j}^{(ULAE)} = \pi \cdot \sum_{i+j=t} R_{i,j}^{(IBNyR)} + \pi \cdot (1-r) \cdot \sum_{i+j=t} R_{i,j}^{(rep)}, \quad (A.15)$$

  i.e. all we need to know is, how to split of total pure claims reserves into reserves for IBNyR claims and reserves for reported claims.

- The assumptions for the New York-method are rather restrictive in the sense that the pure cumulative ultimate claim $C_{i,J}^{(pure)}$ must be constant in k (see Lemma A.4). Otherwise the paid-to-paid ratio π_t for accounting years is not the same as the ratio $C_{i,J}^{(ULAE)}/C_{i,J}^{(pure)}$ even if the latter is assumed to be constant. Of course in practice the assumption of equal pure cumulative ultimate claim is never fulfilled. If we relax this condition we obtain the following lemma.

Lemma A.6 *Assume there exists $\pi > 0$ such that for all accident years i we have*

$$\frac{C_{i,J}^{(ULAE)}}{C_{i,J}^{(pure)}} = \pi \cdot \left(r \cdot \frac{\overline{\delta}}{\overline{\gamma}} + (1-r) \right)^{-1}, \quad (A.16)$$

with

$$\overline{\gamma} = \frac{\sum_{j=0}^{J} \gamma_j \cdot C_{t-j,J}^{(pure)}}{\sum_{j=0}^{J} C_{t-j,J}^{(pure)}} \qquad \text{and} \qquad \overline{\delta} = \frac{\sum_{j=0}^{J} \delta_j \cdot C_{t-j,J}^{(pure)}}{\sum_{j=0}^{J} C_{t-j,J}^{(pure)}}. \quad (A.17)$$

Under Assumption A.1 and Model A.2 we have for all accounting years t

$$\pi_t = \pi. \quad (A.18)$$

Proof of Lemma A.6. As in Lemma A.4 we obtain

$$\pi_t = \pi \cdot \left(r \cdot \frac{\overline{\delta}}{\overline{\gamma}} + (1-r) \right)^{-1} \cdot \frac{\sum_{j=0}^{J} \left(r \cdot \delta_j + (1-r) \cdot \gamma_j \right) \cdot C_{t-j,J}^{(pure)}}{\sum_{j=0}^{J} \gamma_j \cdot C_{t-j,J}^{(pure)}} = \pi. \quad (A.19)$$

This finishes the proof.

□

Remarks:

- If all pure cumulative ultimates are equal then $\overline{\gamma} = \overline{\delta} = 1/(J+1)$ (apply Lemma A.4).

- Assume that there exists a constant $i^{(p)} > 0$ such that for all $i \geq 0$ we have $C_{i+1,J}^{(pure)} = (1+i^{(p)}) \cdot C_{i,J}^{(pure)}$, i.e. constant growth $i^{(p)}$. If we blindly apply (A.11) of Lemma A.4 (i.e. we do not apply the correction factor in (A.16)) and estimate the incremental ULAE payments by (A.13) and (A.15) we obtain

$$\begin{aligned} \sum_{i+j=t} \widehat{X}_{i,j}^{(ULAE)} &= \pi \cdot \sum_{j=0}^{J} \big(r \cdot \delta_j + (1-r) \cdot \gamma_j\big) \cdot C_{t-j,J}^{(pure)} \\ &= \frac{X_t^{(ULAE)}}{X_t^{(pure)}} \cdot \sum_{j=0}^{J} \big(r \cdot \delta_j + (1-r) \cdot \gamma_j\big) \cdot C_{t-j,J}^{(pure)} \qquad \text{(A.20)} \\ &= \sum_{i+j=t} X_{i,j}^{(ULAE)} \cdot \left(r \cdot \frac{\overline{\delta}}{\overline{\gamma}} + (1-r) \right) \\ &= \sum_{i+j=t} X_{i,j}^{(ULAE)} \cdot \left(r \cdot \frac{\sum_{j=0}^{J} \delta_j \cdot \big(1+i^{(p)}\big)^{J-j}}{\sum_{j=0}^{J} \gamma_j \cdot \big(1+i^{(p)}\big)^{J-j}} + (1-r) \right) \\ &> \sum_{i+j=t} X_{i,j}^{(ULAE)}, \end{aligned}$$

  where the last inequality in general holds true for $i^{(p)} > 0$, since usually $(\delta_j)_j$ is more concentrated than $(\gamma_j)_j$, i.e. we usually have $J > 1$ and

$$\sum_{l=0}^{j} \delta_l > \sum_{l=0}^{j} \gamma_l \quad \text{for } j = 0, \ldots, J-1. \qquad \text{(A.21)}$$

  This comes from the fact that the claims are reported before they are paid. I.e. if we blindly apply the New York-method for constant positive growth then the ULAE reserves are too high (for constant negative growth we obtain the opposite sign). This implies that we have always a positive loss experience on ULAE reserves for constant positive growth.

A.5 Example

We assume that the observations for π_t are generated by i.i.d. random variables $\frac{X_t^{(ULAE)}}{X_t^{(pure)}}$. Hence we can estimate π from this sequence. Assume $\pi = 10\%$. Moreover $i^{(p)} = 0$ and set $r = 50\%$ (this is the usual choice, also done in the SST [73]).

Moreover we assume that we have the following reporting and cash flow patterns ($J = 4$):

$$\begin{aligned} (\beta_0, \ldots, \beta_4) &= (90\%, 10\%, 0\%, 0\%, 0\%), && \text{(A.22)} \\ (\alpha_0, \ldots, \alpha_4) &= (30\%, 20\%, 20\%, 20\%, 10\%). && \text{(A.23)} \end{aligned}$$

Assume that $C_{i,J}^{(pure)} = 1'000$. Then the ULAE reserves for accident year i are given by

$$\left(\widehat{R}_{i,-1}^{(ULAE)}, \ldots, \widehat{R}_{i,3}^{(ULAE)} \right) = (100, 40, 25, 15, 5), \tag{A.24}$$

which implies for the estimated incremental ULAE payments

$$\left(\widehat{X}_{i,0}^{(ULAE)}, \ldots, \widehat{X}_{i,4}^{(ULAE)} \right) = (60, 15, 10, 10, 5). \tag{A.25}$$

Hence for the total estimated payments $\widehat{X}_{i,j} = X_{i,j}^{(pure)} + \widehat{X}_{i,j}^{(ULAE)}$ we have

$$\left(\widehat{X}_{i,0}, \ldots, \widehat{X}_{i,4} \right) = (360, 215, 210, 210, 105). \tag{A.26}$$

Appendix B

Distributions

B.1 Discrete distributions

B.1.1 Binomial distribution

For $n \in \mathbb{N}$ and $p \in (0,1)$ the Binomial distribution $\text{Bin}(n,p)$ is defined to be the discrete distribution with probability function

$$f_{n,p}(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \tag{B.1}$$

for all $x \in \{0, \ldots, n\}$.

$\text{E}(X)$	$\text{Var}(X)$	$\text{Vco}(X)$
$n \cdot p$	$n \cdot p \cdot (1-p)$	$\sqrt{\frac{1-p}{n \cdot p}}$

Table B.1: Expectation, variance and variational coefficient of a $\text{Bin}(n,p)$-distributed random variable X

B.1.2 Poisson distribution

For $\lambda \in (0, \infty)$ the Poisson distribution $\text{Poisson}(\lambda)$ is defined to be the discrete distribution with probability function

$$f_\lambda(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!} \tag{B.2}$$

for all $x \in \mathbb{N}_0$.

E(X)	Var(X)	Vco(X)
λ	λ	$\frac{1}{\sqrt{\lambda}}$

Table B.2: Expectation, variance and variational coefficient of a Poisson(λ)-distributed random variable X

B.1.3 Negative binomial bistribution

For $r \in (0, \infty)$ and $p \in (0, 1)$ the Negative binomial distribution NB(r, p) is defined to be the discrete distribution with probability function

$$f_{r,p}(x) = \binom{r + x - 1}{x} \cdot p^r \cdot (1 - p)^x \tag{B.3}$$

for all $x \in \mathbb{N}_0$.
For $\alpha \in \mathbb{R}$ and $n \in \mathbb{N}_0$, the generalized binomial coefficient is defined to be

$$\binom{\alpha}{n} = \frac{\alpha \cdot (\alpha - 1) \cdot \ldots \cdot (\alpha - n + 1)}{n!} = \prod_{k=1}^{n} \frac{\alpha - k + 1}{k}. \tag{B.4}$$

E(X)	Var(X)	Vko(X)
$r \cdot \frac{1-p}{p}$	$r \cdot \frac{1-p}{p^2}$	$\frac{1}{\sqrt{r \cdot (1-p)}}$

Table B.3: Expectation, variance and variational coefficient of a NB(r, p)-distributed random variable X

B.2 Continuous distributions

B.2.1 Normal distribution

For $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ the Normal distribution $\mathcal{N}(\mu, \sigma^2)$ is defined to be the continuous distribution with density

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot \exp\left(-\frac{(x - \mu)^2}{2 \cdot \sigma^2}\right) \cdot 1_{\mathbb{R}}(x). \tag{B.5}$$

B.2.2 Log-normal distribution

For $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ the Log-normal distribution $\mathcal{LN}(\mu, \sigma^2)$ is defined to be the continuous distribution with density

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2 \cdot x}} \cdot \exp\left(-\frac{(\ln x - \mu)^2}{2 \cdot \sigma^2}\right) \cdot 1_{(0,\infty)}(x). \tag{B.6}$$

E(X)	Var(X)	Vco(X)
μ	σ^2	$\frac{\sigma}{\mu}$

Table B.4: Expectation, variance and variational coefficient of a $\mathcal{N}(\mu, \sigma^2)$-distributed random variable X

E(X)	Var(X)	Vco(X)
$e^{\mu+\frac{\sigma^2}{2}}$	$e^{2\cdot\mu+\sigma^2} \cdot \left(e^{\sigma^2} - 1\right)$	$\sqrt{e^{\sigma^2} - 1}$

Table B.5: Expectation, variance and variational coefficient of a $\mathcal{LN}(\mu, \sigma^2)$-distributed random variable X

B.2.3 Gamma distribution

For $\gamma, c \in (0, \infty)$ the Gamma distribution $\Gamma(\gamma, c)$ is defined to be the continuous distribution with density

$$f_{\gamma,c}(x) = \frac{c^\gamma}{\Gamma(\gamma)} \cdot x^{\gamma-1} \cdot e^{-c\cdot x} \cdot 1_{(0,1)}(x). \tag{B.7}$$

The map $\Gamma : (0, \infty) \to (0, \infty)$ given by

$$\Gamma(\gamma) = \int_0^\infty u^{\gamma-1} \cdot e^{-u}\, du \tag{B.8}$$

is called the Gamma function. The parameters γ and c are called shape and scale respectively.
The Gamma function has the following properties

1) $\Gamma(1) = 1$.

2) $\Gamma(1/2) = \sqrt{\pi}$.

3) $\Gamma(\gamma + 1) = \gamma \cdot \Gamma(\gamma)$.

E(X)	Var(X)	Vco(X)
$\frac{\gamma}{c}$	$\frac{\gamma}{c^2}$	$\frac{1}{\sqrt{\gamma}}$

Table B.6: Expectation, variance and variational coefficient of a $\Gamma(\gamma, c)$-distributed random variable X

B.2.4 Beta distribution

For $a, b \in (0, \infty)$ the Beta distribution $\mathrm{Beta}(a, b)$ is defined to be the continuous distribution with density

$$f_{a,b}(x) = \frac{1}{\mathrm{B}(a,b)} \cdot x^{a-1} \cdot (1-x)^{b-1} \cdot 1_{(0,1)}(x). \tag{B.9}$$

The map $\mathrm{B} : (0, \infty) \times (0, \infty) \to (0, \infty)$ given by

$$\mathrm{B}(a,b) = \int_0^1 u^{a-1} \cdot (1-u)^{b-1} \, du = \frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a+b)} \tag{B.10}$$

is called the Beta function.

$\mathrm{E}(X)$	$\mathrm{Var}(X)$	$\mathrm{Vco}(X)$
$\frac{a}{a+b}$	$\frac{a \cdot b}{(a+b)^2 \cdot (a+b+1)}$	$\sqrt{\frac{b}{a \cdot (a+b+1)}}$

Table B.7: Expectation, variance and variational coefficient of a $\mathrm{Beta}(a, b)$-distributed random variable X

Bibliography

[1] Abraham, B., Ledolter, J. (1983), Statistical Methods for Forecasting. John Wiley and Sons, NY.

[2] Alba de, E. (2002), Bayesian estimation of outstanding claim reserves. North American Act. J. 6/4, 1-20.

[3] Alba de, E., Corzo, M.A.R. (2006), Bayesian claims reserving when there are negative values in the runoff triangle. Actuarial Research Clearing House, Jan 01, 2006.

[4] Alba de, E. (2006), Claims reserving when there are negative values in the runoff triangle: Bayesian analysis using the three-parameter log-normal distribution. North American Act. J. 10/3, 45-59.

[5] Arjas, E. (1989), The claims reserving problem in non-life insurance: some structural ideas. ASTIN Bulletin 19/2, 139-152.

[6] Barnett, G., Zehnwirth, B. (1998), Best estimate reserves. CAS Forum, 1-54.

[7] Barnett, G., Zehnwirth, B. (2000), Best estimates for reserves. Proc. CAS, Vol. LXXXII, 245-321.

[8] Benktander, G. (1976), An approach to credibility in calculating IBNR for casualty excess reinsurance. The Actuarial Review, April 1976, p.7.

[9] Bernardo, J.M., Smith, A.F.M. (1994), Bayesian Theory. John Wiley and Sons, NY.

[10] Bornhuetter, R.L., Ferguson, R.E. (1972), The actuary and IBNR. Proc. CAS, Vol. LIX, 181-195.

[11] Buchwalder, M., Bühlmann H., Merz, M., Wüthrich, M.V. (2005), Legal valuation portfolio in non-life insurance. Conference paper, presented at the 36th International ASTIN Colloquium, 4-7 September 2005, ETH Zürich. www.astin2005.ch

[12] Buchwalder, M., Bühlmann H., Merz, M., Wüthrich, M.V. (2006), Estimation of unallocated loss adjustment expenses. Bulletin SAA 2006/1, 43-53.

[13] Buchwalder, M., Bühlmann H., Merz, M., Wüthrich, M.V. (2006), The mean square error of prediction in the chain ladder reserving method (Mack and Murphy revisited). To appear in ASTIN Bulletin 36/2.

[14] Buchwalder, M., Bühlmann H., Merz, M., Wüthrich, M.V. (2006), Valuation portfolio in non-life insurance. Preprint

[15] Bühlmann, H. (1983), Chain ladder, Cape Cod and complementary loss ratio. International Summer School 1983, unpublished.

[16] Bühlmann, H. (1992), Stochastic discounting. Insurance: Math. Econom. 11, 113-127.

[17] Bühlmann, H. (1995), Life insurance with stochastic interest rates. In: Financial Risk in Insurance, G. Ottaviani (Ed.), Springer.

[18] Bühlmann, H., Gisler, A. (2005), A Course in Credibility Theory and its Applications. Springer Universitext.

[19] Casualty Actuarial Society (CAS) (1990). Foundations of Casualty Actuarial Science, fourth edition.

[20] Clark, D.R. (2003), LDF curve-fitting and stochastic reserving: a maximum likelihood approach. CAS Forum (Fall), 41-92.

[21] Efron, B. (1979), Bootstrap methods: another look at the jackknife. Ann. Statist. 7/1, 1–26.

[22] Efron, B., Tibshirani, R.J. (1995), An Introduction to the Bootstrap. Chapman & Hall, NY.

[23] England, P.D., Verrall, R.J. (1999), Analytic and bootstrap estimates of prediction errors in claims reserving. Insurance: Math. Econom. 25, 281-293.

[24] England, P.D., Verrall, R.J. (2001), A flexible framework for stochastic claims reserving. Proc. CAS, Vol. LXXXIII, 1-18.

[25] England, P.D., Verrall, R.J. (2002), Stochastic claims reserving in general insurance. British Act. J. 8/3, 443-518.

[26] Feldblum, S. (2002), Completing and using schedule P. CAS Forum, 353-590.

[27] Finney, D.J. (1941), On the distribution of a variate whose algorithm is normally distributed. JRSS Suppl. 7, 155-161.

[28] Gerber, H.U., Jones, D.A. (1975), Credibility formulas of the updating type. In: Credibility: Theory and Applications, P.M. Kahn (Ed.), Academic Press, NY.

[29] Gisler, A. (2006), The estimation error in the chain-ladder reserving method: a Bayesian approach. To appear in ASTIN Bulletin 36/2.

[30] Gogol, D. (1993), Using expected loss ratios in reserving. Insurance: Math. Econom. 12, 297-299.

[31] Hachemeister, C.A. (1975), Credibility for regression models with application to trend. In: Credibility: Theory and Applications, P.M. Kahn (Ed.), Academic Press, NY.

[32] Haastrup, S., Arjas, E. (1996), Claims reserving in continuous time; a nonparametric Bayesian approach. ASTIN Bulletin 26/2, 139-164.

[33] Herbst, T. (1999), An application of randomly truncated data models in reserving IBNR claims. Insurance: Math. Econom. 25, 123-131.

[34] Hertig, J. (1985), A statistical approach to the IBNR-reserves in marine insurance. ASTIN Bulletin 15, 171-183.

[35] Hesselager, O., Witting, T. (1988), A credibility model with random fluctutations in delay probabilities for the prediction of IBNR claims. ASTIN Bulletin 18, 79-90.

[36] Hesselager, O. (1991), Prediction of outstanding claims: A hierarchical credibility approach. Scand. Act. J. 1991, 25-47.

[37] Hovinen, E. (1981), Additive and continuous IBNR. ASTIN Colloquium Loen, Norway.

[38] Jewell, W.S. (1976), Two classes of covariance matrices giving simple linear forecasts. Scand. Act. J. 1976, 15-29.

[39] Jewell, W.S. (1989), Predicting IBNyR events and delays. ASTIN Bulletin 19/1, 25-55.

[40] Jones, A.R., Copeman, P.J., Gibson, E.R., Line, N.J.S., Lowe, J.A., Martin, P., Matthews, P.N., Powell, D.S. (2006), A change agenda for reserving. Report of the general insurance reserving issues taskforce (GRIT). Institute of Actuaries and Faculty of Actuaries.

[41] Jong de, E. (2006), Forecasting runoff triangles. North American Act. J. 10/2, 28-38.

[42] Jong de, P., Zehnwirth, B. (1983), Claims reserving, state-space models and the Kalman filter. J.I.A. 110, 157-182.

[43] Jørgensen, B., de Souza, M.C.P. (1994), Fitting Tweedie's compound Poisson model to insurance claims data. Scand. Act. J. 1994, 69-93.

[44] Kremer, E. (1982), IBNR claims and the two way model of ANOVA. Scand. Act. J. 1982, 47-55.

[45] Larsen, C.R. (2005), A dynamic claims reserving model. Conference paper, presented at the 36th International ASTIN Colloquium, 4-7 September 2005, ETH Zürich. www.astin2005.ch

[46] Lyons, G., Forster, W., Kedney, P., Warren, R., Wilkinson, H. (2002), Claims reserving working party paper. General Insurance Convention 2002. http://www.actuaries.org.uk

[47] Mack, T. (1990), Improved estimation of IBNR claims by credibility. Insurance: Math. Econom. 9, 51-57.

[48] Mack, T. (1991), A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. ASTIN Bulletin 21/1, 93-109.

[49] Mack, T. (1993), Distribution-free calculation of the standard error of chain ladder reserve estimates. ASTIN Bulletin 23/2, 213-225.

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

[51] Mack, T. (2000), Credible claims reserves: The Benktander method. ASTIN Bulletin 30/2, 333-347.

[52] Mack, T., Quarg, G., Braun, C. (2006), The mean square error of prediction in the chain ladder reserving method - a comment. To appear in ASTIN Bulletin 36/2.

[53] McCullagh, P., Nelder, J.A. (1989), Generalized Linear Models. 2nd edition, Chapman & Hall, London.

[54] Merz, M., Wüthrich, M.V. (2006), A credibility approach to the Munich chain-ladder method. To appear in Blätter DGVFM.

[55] Murphy, D.M. (1994), Unbiased loss development factors. Proc. CAS, Vol. LXXXI, 154-222.

[56] Neuhaus, W. (1992), Another pragmatic loss reserving method or Bornhuetter/Ferguson revisited. Scand. Act. J. 1992, 151-162.

[57] Neuhaus, W. (2004), On the estimation of outstanding claims. Conference paper, presented at the 35th International ASTIN Colloquium 2004, Bergen, Norway.

[58] Norberg, R. (1993), Prediction of outstanding liabilities in non-life insurance. ASTIN Bulletin 23/1, 95-115.

[59] Norberg, R. (1999), Prediction of outstanding liabilities II. Model variations and extensions. ASTIN Bulletin 29/1, 5-25.

[60] Ntzoufras, I., Dellaportas, P. (2002), Bayesian modelling of outstanding liabilities incorporating claim count uncertainty. North American Act. J. 6/1, 113-128.

[61] Partrat, C., Pey, N., Schilling, J. (2005), Delta method and reserving. Conference paper, presented at the 36th International ASTIN Colloquium, 4-7 September 2005, ETH Zürich. www.astin2005.ch

[62] Quarg, G., Mack, T. (2004), Munich Chain Ladder. Blätter DGVFM, Band XXVI, 597-630.

[63] Radtke, M., Schmidt, K.D. (2004), Handbuch zur Schadenreservierung. Verlag Versicherungswirtschaft, Karlsruhe.

[64] Renshaw, A.E. (1994), Claims reserving by joint modelling. Actuarial research paper no. 72, Department of Actuarial Sciences and Statistics, City University, London.

[65] Renshaw, A.E., Verrall, R.J. (1998), A stochastic model underlying the chain ladder technique. British Act. J. 4/4, 903-923.

[66] Ross, S.M. (1985), Introdution to Probability Models, 3rd edition. Academic Press, Orlando Florida.

[67] Sandström, A. (2006), Solvency: Models, Assessment and Regulation. Chapman and Hall, CRC.

[68] Schmidt, K.D., Schaus, A. (1996), An extension of Mack's model for the chain-ladder method. ASTIN Bulletin 26, 247-262.

[69] Schnieper, R. (1991), Separating true IBNR and IBNER claims. ASTIN Bulletin 21, 111-127.

[70] Scollnik, D.P.M. (2002), Implementation of four models for outstanding liabilities in WinBUGS: A discussion of a paper by Ntzoufras and Dellaportas. North American Act. J. 6/1, 113-128.

[71] Smyth, G.K., Jørgensen, B. (2002), Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. ASTIN Bulletin 32, 143-157.

[72] Srivastava, V.K., Giles, D.E. (1987), Seemingly Unrelated Regression Equation Models: Estimation and Inference. Marcel Dekker, NY.

[73] Swiss Solvency Test (2005), BPV SST Technisches Dokument, Version 22.Juni 2005. Available under www.sav-ausbildung.ch

[74] Taylor, G. (1987), Regression models in claims analysis I: theory. Proc. CAS, Vol. XLVI, 354-383.

[75] Taylor, G. (2000), Loss Reserving: An Actuarial Perspective. Kluwer Academic Publishers.

[76] Taylor, G., McGuire, G. (2005), Synchronous bootstrapping of seemingly unrelated regressions. Conference paper, presented at the 36th International ASTIN Colloquium, 4-7 September 2005, ETH Zürich. www.astin2005.ch

[77] Venter, G.G. (1998), Testing the assumptions of age-to-age factors. Proc. CAS, Vol. LXXXV, 807-847.

[78] Venter, G.G. (2006), Discussion of MSEP in the CLRM (MMR). To appear in ASTIN Bulletin 36/2.

[79] Verrall, R.J. (1990), Bayesian and empirical Bayes estimation for the chain ladder model. ASTIN Bulletin 20/2, 217-238.

[80] Verrall, R.J. (1991), On the estimation of reserves from loglinear models. Insurance: Math. Econom. 10, 75-80.

[81] Verrall, R.J. (2000), An investigation into stochastic claims reserving models and the chain-ladder technique. Insurance: Math. Econom. 26, 91-99.

[82] Verrall, R.J. (2004), A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving. North American Act. J. 8/3, 67-89.

[83] Verdier, B., Klinger, A. (2005), JAB Chain: A model-based calculation of paid and incurred loss development factors. Conference paper, presented at the 36th International ASTIN Colloquium, 4-7 September 2005, ETH Zürich. www.astin2005.ch

[84] Vylder De, F. (1982), Estimation of IBNR claims by credibility theory. Insurance: Math. Econom. 1, 35-40.

[85] Vylder De, F., Goovaerts, M.J. (1979), Proceedings of the first meeting of the contact group "Actuarial Sciences". KU Leuven, nl. 7904B wettelijk report: D/1979/23761/5.

[86] Wright, T.S. (1990), A stochastic method for claims reserving in general insurance. J.I.A. 117, 677-731.

[87] Wüthrich, M.V. (2003), Claims reserving using Tweedie's compound Poisson model. ASTIN Bulletin 33/2, 331-346.

[88] Wüthrich, M.V. (2006), Premium liability risks: modeling small claims. Bulletin SAA 2006/1, 27-38.

[89] Wüthrich, M.V. (2006), Using a Bayesian approach for claims reserving. Accepted for publication in CAS Journal.

[90] Wüthrich, M.V. (2006), Prediction error in the chain ladder method. Preprint.

[91] Wüthrich, M.V., Bühlmann, H., Furrer, H. (2006), Lecture notes on market consistent actuarial valuation. ETH Zürich, Summer Term 2006.

[92] Zehnwirth, B. (1998), ICRFS-Plus 8.3 Manual. Insureware Pty Ltd. St. Kilda, Australia.

[93] Zellner, A. (1962), An efficient method of estimating seemingly unrelated regressions and test for aggregated bias. J. American Stat. Assoc. 57, 346-368.