EDUCATION COMMITTEE

OF THE

SOCIETY OF ACTUARIES

ADVANCED SHORT-TERM ACTUARIAL MATHEMATICS STUDY NOTE

OUTSTANDING CLAIMS RESERVES
(Version 1.3)

by

Mary R. Hardy, PhD, FIA, FSA, CERA
University of Waterloo, Ontario

ASTAM-21-23 **Printed in the USA**

Notes:

Version 1.2: Corrects an error in the limits of the equation for the a parameter estimator under the Bühlmann-Straub model, in Section 3.4, and a typo in equation (4.12).

Version 1.3: Corrects an error in the variance of the combined T statistic in Section 3.2. I am grateful to Howard Mahler for this correction.

Preface

This note has been written to support students preparing for the Advanced Short Term Actuarial Mathematics (ASTAM) exam of the Society of Actuaries.

My objective in writing this note is to present both deterministic and some stochastic methods and models for estimating outstanding claims liabilities, in a form suitable for both classroom teaching and self-study. The mathematics is intended to balance rigour and accessibility; the idea is that understanding the mathematics provides a more solid foundation for understanding, applying, and adapting the models, compared with simply presenting formulas or algorithms.

I owe a very large debt to Mario Wüthrich and Michael Merz, whose excellent text '*Stochastic Claims Reserving Methods in Insurance*' (Wüthrich and Merz, 2008) was a key resource for Chapters 3, 4 and 5. I highly recommend this text to any readers interested in developing their understanding of advanced techniques in outstanding claims estimation. Other sources worth exploring include Hindley (2017), England and Verrall (2002), and Mack (1993). The R package 'ChainLadder' (Gesmann *et al.*, 2022) is also invaluable, and is used extensively in Chapters 4 and 5.

I am very grateful to Carlos Araiza Iturria for his diligent and insightful comments on early drafts, and to Barry Koklefsky, Scott Lennox, Stuart Klugman and Chris Groendyke for many helpful suggestions.

Chapter 1

Introduction

Outstanding Claims Reserves (OCR) represent the provision made by an insurer for losses that have been incurred (that is, an insured loss event has occurred), but which have not yet been settled. Typical reasons for the delay between the loss event and the settlement of the claim include:

Delays in reporting The time between the occurrence and reporting of a loss event varies significantly by the line of business, and by the type of loss. In auto insurance, loss events are typically reported relatively quickly, although the severity of the loss may not be known for some time. In medical malpractice insurance there may be a long delay between the occurrence and reporting of a loss event, as the injury or loss may not be apparent for some time.

Claims processing delays Once a loss is reported, the insurer must ascertain the details of the event, assess the extent to which the loss is covered under the insurance, wait for estimates of the severity of the loss, and investigate potential recovery from other parties (subrogation) or from the residual value of insured property acquired by the insurer after a claim is settled (salvage). There will be delays involved in waiting for quotes to repair or replace property, and, typically, longer delays involved in assessing losses associated with physical injury, as it may take some time for the long-term cost implications of an injury to be apparent. Individual claims may be settled in a single lump sum, but for complex claims involving injury there may be a series of interim payments before the claim is finally settled and closed. Salvage and subrogation can result in negative incremental payments for individual claims.

Legal proceedings Legal disputes typically involve disagreements about the appropriate amount of indemnification for a loss, or disagreements about the allocation of blame between different parties. Legal proceedings are typically long and expensive, and significantly extend the period between the loss event and claim payment.

Short-tail insurance refers to lines for which losses are typically reported promptly, with the longest claim delays no more than around five year from the loss event. Property damage in auto or home insurance, and health insurance, are all short-tailed lines. For example, around 40% of claims for auto insurance are settled within the same calendar year as the loss event itself. **Long-tail insurance** refers to insurance where, for a significant proportion of claims, it may take well over five years to settle all the claims arising in a single year. Claims involving injury and liability tend to be long term, because it is more complex to determine and agree on the amount of loss for personal injury, and because there are more likely to be lengthy legal disputes regarding liability. Long-tail business may involve delays in reporting losses, delays in paying claims, or both. Medical malpractice is a prime example of long-tail business, with only around 1% of ultimate losses settled within the same calendar year as the loss (Enz and Holzheu, 2004).

The outstanding claims reserve may be subdivided into claims that are **reported but not (fully) settled** (RBNS), and claims that have been **incurred but not (yet) reported**, called IBNR or **pure IBNR**.[1] Most of the methods that we consider in this note provide estimates of the total outstanding claims, with IBNR implicitly included.

Separate to the outstanding claims reserve is the **Unearned Premium Reserve** (UPR). This represents the provision made for future potential claims. As premiums are paid in advance, at each year end the **earned premium** is the part relating to the period of insurance from the premium payment date to the year end, and the **unearned premium** is the part relating to future cover. The UPR is (typically) the sum of the unearned premium. Usually, premiums paid annually are assigned to each relevant calendar year pro rata. For example, consider a premium of 1,200 paid for 1-year of insurance cover on 1 November 2022. The period of cover in 2022 is 2 months, and in 2023 is 10 months. So the earned premium in 2022 would be $2/12 \times 1,200 = 200$, and the unearned premium reserve at 31 Dec 2022 would be $10/12 \times 1,200 = 1,000$. this would be added to the earned premium in 2023.

[1] The term IBNR is now often used (though not in this note) to refer to all outstanding claims.

OCR Categories

Case-based estimation Under case-based estimation, each loss is individually reviewed and an estimate made of the likely settlement amount. The estimate of outstanding claims would be updated from time to time, as the insurer obtains new information. This method is typically used for very large claims, and for lines of business where coverage and losses tend to be very idiosyncratic, such as marine insurance. It requires skilled expertise, and is inefficient for lines of business with large volumes of reasonably heterogeneous claims. IBNR losses must be estimated separately under this approach.

Expected loss ratio The expected loss ratio method assumes an average cost of claims arising from accidents occurring in each year, based on the earned premiums during the year. For example, if the insurer expects the portfolio to experience a loss ratio (including allocated loss adjustment expenses) of, say, 85%, then the expected cost of claims incurred in year Y would be $0.85 \times$ Earned Premiums in Year Y. The claims outstanding at each subsequent year end, in respect of loss events that occurred in year Y, would be the expected total cost of claims incurred in year Y, minus the sum of the payments made by the year end in respect of those events. This method is clearly very simplistic, and has the significant disadvantage that the outstanding claims reserve will be negative if the initial estimated loss ratio is exceeded. The earned premiums are used here as an exposure measure, and other measures could be used, suitable to the underlying risk. Using premiums is particularly problematic when premiums have been set too low, as the solvency risk is intensified by having both premiums and outstanding claims reserves set below the true expected values of the relevant liabilities.

Aggregate run-off triangle methods The run-off triangle, which is described in more detail below, is a summary of the aggregate claims payment information for all recent loss events, organised by the year that the losses occurred, and by the period between the loss event and the claim payment dates. Run-off triangle methods use past claims settlement patterns to project future payments for open claims. The most common run-off triangle methodology in practice is the chain ladder, which is a central focus of this note.

Credibility methods Credibility approaches to outstanding claims reserve evaluation use a weighted average of two estimates for outstanding claims; one estimate is derived from the run-off triangle, and the other may be an exogenous estimate, for example, based on the expected loss ratio.

Frequency-severity methods Frequency-severity methods model the number and severity of outstanding claims separately, generally starting from the run-off triangle data. Depending on the approach and data, IBNR reserves may need to be separately evaluated.

Parametric models The insurer may model incremental or cumulative settled claims using a parametric approach. Often, this is based on a generalized linear model.

In this note we focus on data-driven aggregate claims estimation approaches which use the run-off triangle data to project future payments. These methods are most appropriate when the volume of claims is high, claims are reasonably homogeneous, and where settlement patterns are reasonably stable. The starting point for many of these methods is the deterministic chain ladder algorithm. It has many advantages, being intuitive and easy to apply, but it is limited by the fact that without a probabilistic framework, there is no measure of how volatile the outstanding claims may be, or how much uncertainty is associated with the estimated liability. In Chapters 3 and 4 we present extensions of the chain ladder method that place it within a stochastic framework, where the deterministic chain ladder estimate is interpreted as the mean of a random variable representing the outstanding claims. The stochastic framework allows us to examine and test the assumptions that are implicit in the deterministic approach, but are explicit in the stochastic frame, and also to quantify the uncertainty in the estimates, and potentially calculate other metrics of interest, such as risk measures or confidence intervals.

Chapter 2

Run-off triangles

There are three time variables relevant to the settlement patterns and amounts of insurance claims.

> The first is the **accident year (AY)** – that is, the calendar year in which the accident or loss occurred. We group all losses incurred in the same year together, as a single cohort of claims. In the mathematical formulas, we will generally indicate the accident year with i (occasionally l).
>
> The second is the **development year (DY)**, which indicates the payment year, relative to the AY. We will use j (occasionally k) for the DY, so that payments made in the same calendar year as the loss occurred are assigned to $j = 0$; payments made in the following calendar year are assigned to $j = 1$, and so on.
>
> The third is the **calendar year** of individual payments. This is determined by adding the AY and the DY together, so a payment made in DY $j = 3$, in settlement of a loss incurred in AY $i = 2$, say, is a payment made in year $i+j = 5$, measured from accident year 0.

Run-off triangle methods for estimating outstanding claims use past settlement patterns to project future settlement patterns for accident years that are not closed (an AY is closed when all claims are settled). The current and past data are presented as a triangle of claims payment data, separated by accident year and development year. An example is presented in Table 2.1. The values in the top table are the incremental claim payments, that is, amounts paid in each development year, for each accident year cohort. The values in the bottom table are the cumulative payments, that is, the total payments made by the end of each development year, separately for each accident year. The triangles summarize the payments made up to the end of 2020 from the 2011 to 2020 accident year cohorts. In the mathematical formulas, we will generally label the first AY in the triangle as $i = 0$, and measure all payments from that date. The anti-diagonals[1] of the run-off triangle relate to a single calendar year, so that the leading anti-diagonal in the incremental run-off triangle in Table 2.1 represents the payments made in 2020, in respect of accidents occurring between 2011 and 2020.

In the illustrative triangles used in this note, we assume that AY 0 is closed – that is, that all claims are fully settled by the end of the latest calendar year – and that AY 1 is the first open accident year. This means that we have the same number of development and accident years. If AY 0 is not closed, then we would need to project losses beyond the final available development year, using a **tail factor**; see Hindley (2017) for details. The assumption that AY 1 is open is not necessary, and does not (significantly) impact any of the formulas or results.

Other data can be analysed in the run-off triangle format. For example, we may consider a triangle of the number of claims reported, to estimate the IBNR reserve (which is implicitly incorporated in the settlement numbers in Table 2.1). Or we may use case estimates of reported claims, so that changes from year to year represent newly reported claims, and revised estimates of previously reported claims. In this paper we do not consider these variants. We will assume all run-off triangles are based on claim payments, which may be negative through salvage or subrogation. We also assume that all claim expenses that are to be included in the outstanding claims reserve are included in the amounts reported in the run-off triangle.

2.1 Notation

Let $i = 0, 1, .., I$ represent successive accident year cohorts. In Table 2.1, we set 2011 as $i = 0$, which is assumed to be closed, and the latest accident year, 2020, represents the tenth and latest cohort of accident years, and the ninth cohort of open accident years, so $I = 9$. From here on, when we refer to AY i, we mean the calendar year associated with the $(i+1)$th accident year.

Let $j = 0, 1, .., J$ represent the development year. In Table 2.1, we have $J = 9$. If AY 1 is the first open accident year, then we have $I = J$; if we have several closed years at the top of the table then we have $J < I$.

Let $X_{i,j}$ denote the incremental claim payments in respect of losses incurred in AY i, paid during DY j. When $i + j \leq I$, this figure is known. For $i + j > I$, this value is to be

[1] An anti-diagonal of a matrix is a set of elements for which the sum of row and column indices is the same, that is, a lower left to upper right diagonal.

Accident		Development Year, j									
Year	i	0	1	2	3	4	5	6	7	8	9
2011	0	4360	2516	625	207	118	39	51	20	11	3
2012	1	3996	2578	449	134	49	31	31	20	4	
2013	2	3840	1738	655	175	96	40	29	14		
2014	3	5108	1757	680	216	114	69	16			
2015	4	4585	1532	414	189	80	87				
2016	5	5767	2164	410	193	86					
2017	6	5550	2540	458	252						
2018	7	6525	2828	562							
2019	8	6620	2544								
2020	9	7014									

Incremental claim payments, $X_{i,j}$

Accident		Development Year, j									
Year	i	0	1	2	3	4	5	6	7	8	9
2011	0	4360	6876	7501	7708	7826	7865	7916	7936	7947	7950
2012	1	3996	6574	7023	7157	7206	7237	7268	7288	7292	
2013	2	3840	5578	6233	6408	6504	6544	6573	6587		
2014	3	5108	6865	7545	7761	7875	7944	7960			
2015	4	4585	6117	6531	6720	6800	6887				
2016	5	5767	7931	8341	8534	8620					
2017	6	5550	8090	8548	8800						
2018	7	6525	9353	9915							
2019	8	6620	9164								
2020	9	7014									

Cumulative claim payments, $C_{i,j}$

Table 2.1: Data for worked examples: incremental and cumulative claim payments up to the end of 2020, arising from losses incurred between 2011 and 2020.

estimated.

Let $C_{i,j}$ denote the cumulative claims paid in respect of losses incurred in AY i, up to the end of DY j. That is

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

Again, these are known when $i+j \leq I$.

Let $\mathcal{D}_I$ represent the data in the cumulative run-off triangle, that is

$$\mathcal{D}_I = \{C_{i,j} : i+j \leq I\}$$

The objective of run-off triangle methods is to use $\mathcal{D}_I$ to estimate future $X_{i,j}$ or $C_{i,j}$, for $j \leq J$ and for $i + j > I$.

Let $\mathcal{B}_k$ denote the subset of $\mathcal{D}_I$ that only contains data from development years $0,1,\cdots,k$, so

$$\mathcal{B}_k = \{C_{i,j} : i+j \leq I,\ j \leq k\}$$

$\mathcal{B}_k$ uses the first $k+1$ columns of $\mathcal{D}_I$, corresponding to development years from 0 to $k \leq J$. We use $\mathcal{B}_k$ in Chapters 3 and 4, where we often condition on the available information up to DY k, to evaluate conditional moments of $C_{i,k+1} \big| C_{i,k}$.

2.2 The deterministic chain ladder approach

The outstanding claims reserve is the sum of the estimated incremental claim payments for all future development years, over all open accident years. For the data in Table 2.1, if we assume that claims are completely run off by the end of DY 9, then for AY 2012, we need to project the payments in 2021, for AY 2013 we need to project payments in 2021 and 2022, and so in, down to AY 2020, for which we have 9 years of development to project. If the future payments are not discounted, or adjusted for inflation, then it is sufficient to estimate the ultimate cumulative payments made in respect of each accident year. The estimated outstanding claims for each accident year can then be found by subtracting the cumulative claims paid to date from the estimated ultimate cumulative amount. If we are discounting, or inflation adjusting, then we need to project the incremental payments made in each future development year, for each accident year, to apply the right discounting or inflation adjustment.

The intuition behind the chain ladder approach is that we assume that the ratio of cumulative claims in successive development years is reasonably stable, so we can use the known ratios of cumulative payments in successive DYs, calculated from the cumulative claims run-off triangle, to project the cumulative payments for subsequent accident years. The objective is to complete the lower triangle in Table 2.1. Specifically, we assume that, on average, cumulative claims follow a settlement pattern across development years based on **development factors** (also called **link factors**) denoted f_j, such that, deterministically,

$$C_{i,j+1} = C_{i,j}\ f_j \quad i = 0, 1, ...I, \quad j = 0, 1, ...J-1 \tag{2.1}$$

We estimate the development factors using $\mathcal{D}_I$. The individual historical development factors for each AY and DY are

$$f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}} \quad i+j \leq I-1$$

Values of $f_{i,j}$ for the data in Table 2.1 are given in Table 2.2.

AY	DY, j									
i	0	1	2	3	4	5	6	7	8	9
0	1.5771	1.0909	1.0276	1.0153	1.0050	1.0065	1.0025	1.0014	1.0004	–
1	1.6451	1.0683	1.0191	1.0068	1.0043	1.0043	1.0028	1.0005		
2	1.4526	1.1174	1.0281	1.0150	1.0062	1.0044	1.0021			
3	1.3440	1.0991	1.0286	1.0147	1.0088	1.0020				
4	1.3341	1.0677	1.0289	1.0119	1.0128					
5	1.3752	1.0517	1.0231	1.0101						
6	1.4577	1.0566	1.0295							
7	1.4334	1.0601								
8	1.3843									
9	–									

Table 2.2: Individual development factors, $f_{i,j}$, for the data in Table 2.1.

To estimate a single development factor for each development year, we could simply average the individual accident year factors, but this ignores the variability in the cumulative claims by accident year. A more efficient approach is to weight the development factors by the available

$C_{i,j}$ values, giving estimated development factors for each DY, $j = 1, 2, ..., J-1$,

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

Example 2.1. Calculate the development factors, $\hat{f}_0$, $\hat{f}_1$, ..., $\hat{f}_8$ from the data in Table 2.1.

Solution We start with $\hat{f}_8$, which is an estimate of the ratio of cumulative claims in DY 9 to cumulative claims in DY 8. We only have one source for this ratio within the $\mathcal{D}_I$ triangle, from the AY $i = 0$ row of data, so

$$\hat{f}_8 = \frac{7950}{7947} = f_{0,8} = 1.00038$$

For $\hat{f}_7$, we are interested in the development from $j = 7$ to $j = 8$, for which we have data from two cohorts, $i = 0$ and $i = 1$. The development factor is therefore

$$\hat{f}_7 = \frac{7947 + 7292}{7936 + 7288} = \frac{7936 \times f_{0,7} + 7288 \times f_{1,7}}{7936 + 7288} = 1.00099$$

Similarly

$$\hat{f}_6 = \frac{7936 + 7288 + 6587}{7916 + 7268 + 6573} = 1.00248$$

$$\hat{f}_5 = \frac{7916 + \cdots + 7960}{7865 + \cdots + 7944} = 1.00429$$

$$\vdots$$

$$\hat{f}_0 = \frac{6876 + \cdots + 9164}{4360 + \cdots + 6620} = 1.43574$$

$\square$

Given a run-off triangle $\mathcal{D}_I$, the cumulative claims paid to date for each accident year are represented by the anti-diagonal values, $C_{i,I-i}$ for $i = 0, \cdots, I$. Projected cumulative claims for

AY i, DY $j \geq I - i + 1$ are constructed as follows:

$$\begin{aligned}
&\widehat{C}_{i,I-i+1} = C_{i,I-i}\ \hat{f}_{I-i} \\
&\widehat{C}_{i,I-i+2} = \widehat{C}_{i,I-i+1}\ \hat{f}_{I-i+1} = C_{i,I-i}\ \hat{f}_{I-i}\ \hat{f}_{I-i+1} \\
&\vdots \\
&\widehat{C}_{i,J} = \widehat{C}_{i,J-1}\ \hat{f}_{J-1} = C_{i,I-i}\ \hat{f}_{I-i}\ \hat{f}_{I-i+1}\ \cdots\ \hat{f}_{J-1}
\end{aligned}$$

The estimated future incremental claims for AY i and DY $j \geq I-i+1$, are given by

$$\widehat{X}_{i,j} = \widehat{C}_{i,j} - \widehat{C}_{i,j-1} = C_{i,I-i}\left(\hat{f}_{I-i}\ \cdots\ \hat{f}_{j-2}\right)\left(\hat{f}_{j-1} - 1\right) \tag{2.2}$$

Let $\widehat{R}_i$ denote the estimated claims outstanding in respect of losses arising in AY i. Assuming that claims are fully run off by the end of DY J, and also assuming that payments are not discounted, we have

$$\widehat{R}_i = \widehat{C}_{i,J} - C_{i,I-i} = C_{i,I-i}\Bigg(\left(\hat{f}_{I-i}\ \hat{f}_{I-i+1}\ \cdots\ \hat{f}_{J-1}\right) - 1\Bigg) \tag{2.3}$$

$$\text{Let}\ \ \hat{\lambda}_j = \hat{f}_j\ \hat{f}_{j+1}\ \cdots\ \hat{f}_{J-1} \tag{2.4}$$

$$\implies \widehat{R}_i = C_{i,I-i}\left(\hat{\lambda}_{I-i} - 1\right) \tag{2.5}$$

The estimate at the end of year I of the aggregate outstanding claims from all open accident years is therefore

$$\widehat{R} = \sum_{i=0}^{I} \widehat{R}_i$$

If the outstanding claims reserve uses discounting, then we need to work with the projected incremental payments. We may assume that claims are paid, on average $\frac{1}{2}$-way through the relevant calendar year, in which case, assuming a constant annual effective interest rate r, we have

$$\widehat{R}_i \ = \sum_{j=I-i+1}^{J} \widehat{X}_{i,j}(1+r)^{-(i+j-I-0.5)} \tag{2.6}$$

AY	DY, j										
i	0	1	2	3	4	5	6	7	8	9	$\widehat{R}_i$
0	4360	6876	7501	7708	7826	7865	7916	7936	7947	7950	0
1	3996	6574	7023	7157	7206	7237	7268	7288	7292	7295	3
2	3840	5578	6233	6408	6504	6544	6573	6587	6593	6596	9
3	5108	6865	7545	7761	7875	7944	7960	7980	7988	7991	31
4	4585	6117	6531	6720	6800	6887	6917	6934	6941	6943	56
5	5767	7931	8341	8534	8620	8683	8721	8742	8751	8754	134
6	5550	8090	8548	8800	8908	8973	9012	9034	9043	9047	247
7	6525	9353	9915	10177	10302	10377	10422	10448	10458	10462	547
8	6620	9164	9843	10103	10227	10302	10346	10372	10382	10386	1222
9	7014	10070	10817	11102	11238	11321	11370	11398	11409	11413	4399

AY	DY, j										Discounted
i	0	1	2	3	4	5	6	7	8	9	$\widehat{R}_i$
0	4360	2516	625	207	118	39	51	20	11	3	0
1	3996	2578	449	134	49	31	31	20	4	3	3
2	3840	1738	655	175	96	40	29	14	6	2	9
3	5108	1757	680	216	114	69	16	20	8	3	29
4	4585	1532	414	189	80	87	30	17	7	3	53
5	5767	2164	410	193	86	63	37	22	9	3	126
6	5550	2540	458	252	108	65	39	22	9	3	229
7	6525	2828	562	262	125	76	45	26	10	4	508
8	6620	2544	679	260	124	75	44	26	10	4	1143
9	7014	3056	746	286	136	83	49	28	11	4	4179

Table 2.3: Projected cumulative claims (top) and projected incremental claims (bottom), with estimated undiscounted outstanding claims reserves (top) and estimated discounted claims (bottom), using an effective interest rate of 5% p.y., for the data from Table 2.1.

Example 2.2. Calculate the OCR using the chain ladder method, for the data in Table 2.1. Assume (i) no discounting, and (ii) payments are discounted at 5% per year effective.

Solution

(i) The estimated development factors and λs are

j	0	1	2	3	4	5	6	7	8
$\hat{f}_j$	1.43574	1.07411	1.02641	1.01226	1.00735	1.00429	1.00248	1.00099	1.00038
$\hat{\lambda}_j$	1.62722	1.13337	1.05516	1.02801	1.01556	1.00816	1.00385	1.00136	1.00038

The cumulative claims development matrix is the top table in Table 2.3. The upper left triangle shows the past cumulative claims, the lower right shows the projected future claims. The outstanding claims for each AY are found by subtracting the current value of cumulative claims (just above the line in the matrix) from the projected ultimate cumulative claims at $j = 9$. The total of the outstanding claims is estimated at $\sum_{i=0}^{9} \widehat{R}_i = 6648$.

(ii) The undiscounted incremental claims are given in the lower table of Table 2.3. The payments are discounted using equation (2.6). The total outstanding claims reserve is 6277. □

Exercise 2.1.

In Table 2.4 you are given cumulative run-off information for two different lines of business, a short-tail line (top) and a long-tail line (bottom).

(a) Use the chain ladder method, without discounting, to estimate the outstanding claims reserve in each case.

(b) Comment on the difference between the two lines, and between these settlement patterns and the one underlying Table 2.1.

Solution

(a) The undiscounted chain ladder estimate of outstanding claims for the short-tail example is 56,955.

The undiscounted chain ladder estimate of outstanding claims for the long tail example is 37,914.

(b) Table 2.5 shows the settlement patterns for all three run-off tables. We summarize the settlement patterns using two different functions of the development factors; $\hat{\beta}_j$ is the

Accident Year	Development year j				
i	0	1	2	3	4
0	9502	25827	37275	44083	44490
1	8138	26292	37496	42114	
2	9802	25563	37257		
3	9498	25266			
4	9072				

Accident Year	Development year j										
i	0	1	2	3	4	5	6	7	8	9	10
0	65	276	797	1626	3093	4412	4890	5153	5335	5360	5365
1	46	405	1039	2194	3448	4746	5298	5563	5681	5706	
2	73	388	1017	2588	4213	5088	5969	6210	6409		
3	95	401	1030	2186	4042	5520	6287	6638			
4	72	502	1146	2614	4402	5713	6397				
5	97	472	1251	2273	3909	5420					
6	93	1134	2578	4141	5739						
7	77	585	1247	3213							
8	75	1288	2143								
9	84	568									
10	109										

Table 2.4: Short-tail and long-tail run-off triangles

estimated proportion of ultimate claims paid by the end of DY j, and $\hat{\gamma}_j$ is the estimated proportion of ultimate claims paid during DY j, where

$$\hat{\beta}_j = \frac{1}{\hat{\lambda}_j},\ j = 0, 1, \cdots, J-1; \quad \hat{\beta}_J = 1.0$$
$$\hat{\gamma}_j = \hat{\beta}_j - \hat{\beta}_{j-1},\ \ j = 1, 2, \cdots, J; \quad \hat{\gamma}_0 = \hat{\beta}_0.$$

Comparing the short tail run-off to the Table 2.1, we see that although the development period is shorter for the short-tail data, the payments in the earliest development year are lower, indicating that the payment or reporting delays are slightly longer for the short-tail business than for the Table 2.1 business. Only 21% of claims are covered in the first year under the short-tail development, compared with 61% under the Table 2.1 development. The largest $\hat{\gamma}_j$ occurs in DY 1 for the short tail, indicating that the modal time to payment is around 1 year, and this is also the median. Less than 1% of claims are paid after the third DY. For the Table 2.1 data, the median and modal time to payment is less than 1-year, and around 1.5% of claims are paid after DY 4.

The long-tail development pattern is very different. Only 1% of claims are paid within the calendar year of the loss event. The median and modal payment times are in DY 4, and nearly 40% of claims are paid after DY 4.

	Development Year j										
	0	1	2	3	4	5	6	7	8	9	10
	Short-Tail										
$\hat{\beta}_j$	0.214	0.596	0.860	0.991	1.000						
$\hat{\gamma}_j$	0.214	0.382	0.264	0.131	0.009						
	Table 2.1										
$\hat{\beta}_j$	0.615	0.882	0.948	0.973	0.985	0.992	0.996	0.999	1.000	1.000	
$\hat{\gamma}_j$	0.615	0.268	0.065	0.025	0.012	0.007	0.004	0.002	0.001	0.000	
	Long-Tail										
$\hat{\beta}_j$	0.010	0.080	0.180	0.371	0.608	0.813	0.920	0.966	0.995	0.999	1.000
$\hat{\gamma}_j$	0.010	0.070	0.100	0.191	0.237	0.205	0.107	0.046	0.028	0.005	0.001

Table 2.5: Development patterns for the Table 2.1 and Table 2.4 run-off data.

2.3 Inflation-adjusted chain ladder

As the outstanding claims data covers several years, we need to consider the impact of inflation on historical and projected figures. If inflation rates are relatively steady, there may be no need to make any adjustments, as the run-off triangle based methods described in this note will implicitly allow for future inflation to continue at similar rates to past inflation. If inflation rates have been very variable over the period covered by the run-off triangle, or if they are expected to be significantly different in the future compared with the recent past, then we should adjust the historical data to eliminate the impact of past inflation, and we should also adjust the projected claim payments to make allowance for future inflation.

We adjust the past data using an index of inflation, Q_t, say. For $t \leq I$, Q_t is known; it will be calculated by the insurer using internal or industry-wide information about claims inflation between time t and time I. For example, if g_k denotes the historic inflation rate from mid-year k to mid-year k+1, then we can set $Q_0 = 1.0$ (as this is an index, the starting value is arbitrary), and $Q_{k+1} = Q_k(1 + g_k)$. The inflation-adjusted incremental run-off triangle data, $X^{\mathrm{a}}_{i,j}$, say, where $i+j \leq I$, would then be calculated from the original incremental data as

$$X^{\mathrm{a}}_{i,j} = X_{i,j}\, \frac{Q_I}{Q_{i+j}}$$

This gives a triangle of inflation-adjusted incremental claims, now expressed in terms of year I values. The inflation-adjusted cumulative claims in the run-off triangle are $C^{\mathrm{a}}_{i,j} = \sum_{k=0}^{j} X^{\mathrm{a}}_{i,k}$.

So for the inflation-adjusted chain ladder, we run the chain ladder algorithm using the inflation adjusted cumulative claims run-off triangle, generating inflation-adjusted estimates for future cumulative claims, $\widehat{C}^{\mathrm{a}}_{i,j}$, where $i+j \geq I$. These are the estimated cumulative claims for AY i and DY j, expressed in year I values. To project to year $i+j$ values, we need to extract the projected incremental claims, and then re-adjust for future inflation, using the projected, extrapolated inflation index, $\widetilde{Q}_t$, for $t > I$, which is calculated using assumed future inflation rates. That is, for $i+j > I$:

$$\begin{aligned}
\widehat{X}^{\mathrm{a}}_{i,j} &= \widehat{C}^{\mathrm{a}}_{i,j} - \widehat{C}^{\mathrm{a}}_{i,j-1} \\
\widehat{X}_{i,j} &= \widehat{X}^{\mathrm{a}}_{i,j} \frac{\widetilde{Q}_{i+j}}{Q_I} \\
\widehat{R}_i &= \sum_{k=I-i+1}^{J} \widehat{X}_{i,j}
\end{aligned}$$

Exercise 2.2. In Table 2.6 you are given a cumulative run-off triangle and historic and assumed future claim inflation rates, where the inflation rates given are assumed to run from mid-year t to mid-year $t+1$.

(i) Calculate the claims inflation index, Q_k using the inflation rates, for $k = 0, 1, ..., 7$. Set $Q_0 = 100.0$.

(ii) Calculate the inflation adjusted outstanding claims reserve.

	Development Year				
AY	0	1	2	3	4
0	168,830	442,760	1,062,807	1,311,257	1,333,517
1	177,540	436,618	873,088	1,013,083	
2	203,860	499,301	1,027,061		
3	215,988	405,472			
4	191,753				

Year t	Inflation Rate (past)	Year t	Inflation Rate (assumed)
0	2.5%	4	4.0%
1	3.0%	5	4.6%
2	3.5%	6	4.0%
3	3.5%	7	3.5%

Table 2.6: Data for inflation-adjusted chain ladder exercise.

Solution

(i)

t	Q_t	t	Q_t
0	100.0	4	113.1
1	102.5	5	117.6
2	105.6	6	123.0
3	109.3	7	127.9

(ii) $\widehat{R} = 1,926,174.$

Chapter 3

$\widehat{C}_{i,J}$ as an expected value

3.1 The chain ladder method

In this section we take a more probabilistic perspective on outstanding claims estimation, by assuming that the estimated ultimate cumulative losses, $\widehat{C}_{i,J}$, are expected values of the random variable $C_{i,J}$. Unconditionally, (that is, assuming that we have no claims data) the $C_{i,j}$ are random variables, and hence so are the $\hat{f}_j$, because they are functions of the $C_{i,j}$. Conditional on $\mathcal{D}_I$, $C_{i,j}$ are known for (i,j) where $i = 0, 1, ..., I$, and $i + j \leq I$, and $f_{i,j}$ are known for (i,j) such that $i = 0, 1, ..., I-1$, and $i+j \leq I-1$.

We reinterpret the deterministic chain ladder relationships as expected value relationships in the following assumptions.

CL Assumption (1) There exist development factors f_j such that

$$\mathrm{E}[C_{i,j+1}|\mathcal{D}_{i+j}] = \mathrm{E}[C_{i,j+1}|C_{i,j}] = C_{i,j}\, f_j.$$

This is just re-stating the chain ladder approach in terms of expected values, where the development factors are unknown underlying parameters.

CL Assumption (2) $C_{i,j}$ and $C_{l,k}$ are independent for $i \neq l$, and for all j, k.

That is, we assume independence of losses for different accident years.

Given these two assumptions, we can prove the following results.

Theorem 1.

(a) $\hat{f}_j$ is an unbiased estimator of f_j.

(b) $\mathrm{E}\left[\hat{f}_0\,\hat{f}_1\,\cdots\,\hat{f}_j\right] = f_0\,f_1\,\cdots\,f_j$ for $j = 0, 1, \cdots, J-1$.

(c) $\widehat{C}_{i,J}\big|\mathcal{D}_I$ is an unbiased estimator of $C_{i,J}\big|\mathcal{D}_I$.

Proof

(a) We use iterated expectation for $\mathrm{E}[\hat{f}_j]$, conditioning on $\mathcal{B}_j$, which contains all the run-off data up to DY j.

$$\begin{aligned}
\mathrm{E}\left[\hat{f}_j\right] &= \mathrm{E}\Big[\mathrm{E}\left[\hat{f}_j|\mathcal{B}_j\right]\Big] \\
\mathrm{E}\left[\hat{f}_j|\mathcal{B}_j\right] &= \frac{\mathrm{E}\left[C_{0,j+1} + C_{1,j+1} + \cdots + C_{I-j,j+1}\big|C_{0,j}, C_{1,j}, \cdots, C_{I-j,j}\right]}{C_{0,j} + C_{1,j} + \cdots + C_{I-j,j}}. \\
&= \frac{f_j\left(C_{0,j} + C_{1,j} + \cdots + C_{I-j,j}\right)}{C_{0,j} + C_{1,j} + \cdots + C_{I-j,j}} \qquad \text{(by assumption (1))} \\
&= f_j \\
\mathrm{E}\left[\hat{f}_j\right] &= \mathrm{E}\Big[\mathrm{E}\left[\hat{f}_j|\mathcal{B}_j\right]\Big] = \mathrm{E}[f_j] = f_j \quad \text{as required.}
\end{aligned}$$

(b) We will prove this by induction, again using iterated expectation. To start, from (a) we have that $\mathrm{E}[\hat{f}_0] = f_0$. Assume that the result is true for the product of development factors up to k, i.e. that

$$\mathrm{E}\left[\hat{f}_0\,\hat{f}_1\,\cdots\,\hat{f}_k\right] = f_0\,f_1\,\cdots\,f_k \tag{3.1}$$

Now consider $\mathrm{E}\left[\hat{f}_0\,\hat{f}_1\,\cdots\hat{f}_{k+1}\right]$. We will condition on all the $C_{i,j}$ values involved in the calculation of $\hat{f}_0$, $\hat{f}_1$, $\cdots$, $\hat{f}_k$, i.e. on $\mathcal{B}_{k+1}$. Based on this conditioning, the only unknown $C_{i,j}$ values in $\hat{f}_0\,\hat{f}_1\cdots\hat{f}_{k+1}$ are those in the numerator of $\hat{f}_{k+1}$, i.e.

$C_{0,k+2}$, $C_{1,k+2} \ldots C_{I-(k+2),k+2}$, so that

$$\mathrm{E}\left[\hat{f}_0\,\hat{f}_1 \cdots \hat{f}_k|\mathcal{B}_{k+1}\right] = \hat{f}_0\,\hat{f}_1 \,\cdots\, \hat{f}_k$$

$$\text{and}\quad \mathrm{E}\left[\hat{f}_0\,\hat{f}_1 \cdots \hat{f}_{k+1}|\mathcal{B}_{k+1}\right] = \left(\hat{f}_0\,\hat{f}_1 \,\cdots\, \hat{f}_k\right)\mathrm{E}\left[\hat{f}_{k+1}|\mathcal{B}_{k+1}\right] \tag{3.2}$$

$$= \left(\hat{f}_0\,\hat{f}_1 \,\cdots\, \hat{f}_k\right) f_{k+1} \qquad \text{from (a) above} \tag{3.3}$$

$$\text{Hence}\quad \mathrm{E}\left[\hat{f}_0\,\hat{f}_1 \cdots \hat{f}_{k+1}\right] = \mathrm{E}\Big[\mathrm{E}\left[\hat{f}_0\,\hat{f}_1 \,\cdots\, \hat{f}_{k+1}|\mathcal{B}_{k+1}\right]\Big]$$

$$= \mathrm{E}\left[\left(\hat{f}_0\,\hat{f}_1 \,\cdots\, \hat{f}_k\right) f_{k+1}\right] \quad \text{(from (3.3))}$$

$$= f_0\,f_1 \,\cdots\, f_k\,f_{k+1} \quad \text{by the inductive assumption in (3.1).}$$

Note that we can use the set $\{C_{0,k}, C_{1,k}, \cdots, C_{I-k-1,k+1}\} \subseteq \mathcal{B}_{k+1}$ in place of $\mathcal{B}_{k+1}$ in the conditioning in (3.2), as these are the only elements in $\mathcal{B}_{k+1}$ that impact $\hat{f}_{k+1}$.

It is easy to show that the product of any sequence of $\hat{f}_j$ has a similar property - in particular,

$$\mathrm{E}\left[\hat{f}_j\,\hat{f}_{j+1}\,\cdots\,\hat{f}_{J-1}\right] = \mathrm{E}\left[\hat{\lambda}_j\right] = f_j\,f_{j+1}\,\cdots, f_{J-1} = \lambda_j. \tag{3.4}$$

(c) $$\mathrm{E}\left[\widehat{C}_{i,J}|\mathcal{D}_I\right] = \mathrm{E}\left[C_{i,I-i}\,\hat{\lambda}_{I-i}|\mathcal{D}_I\right] = C_{i,I-i}\mathrm{E}\left[\hat{\lambda}_{I-i}|\mathcal{D}_I\right] = C_{i,I-i}\,\lambda_{I-i} = \mathrm{E}\left[C_{i,J}|C_{i,I-i}\right]$$

□

Exercise 3.1. Show that $\hat{f}_j$ and $\hat{f}_l$ are uncorrelated for $j \neq l$.

3.2 Testing the chain ladder assumptions

Correlated development factors

The chain ladder assumptions imply uncorrelated (but not independent) development factors. If a test of the development factors shows that they are correlated, then the chain ladder assumptions do not hold.

Example 3.1.

(a) For the data in Table 2.1, calculate the correlations between vectors of development factors for successive development years, separately, up to DY 5 and DY6. That is, find the correlations between

(1) $\left(f_{0,0},\, f_{1,0},\, f_{2,0}, \cdots, f_{6,0},\, f_{7,0}\right)^{\mathrm{T}}$ and $\left(f_{0,1},\, f_{1,1},\, f_{2,1}, \cdots, f_{6,1},\, f_{7,1}\right)^{\mathrm{T}}$

(2) $\left(f_{0,1},\, f_{1,1},\, f_{2,1}, \cdots, f_{6,1}\right)^{\mathrm{T}}$ and $\left(f_{0,2},\, f_{1,2},\, f_{2,2}, \cdots, f_{6,2}\right)^{\mathrm{T}}$

$\vdots$

(6) $\left(f_{0,5},\, f_{1,5},\, f_{2,5}\right)^{\mathrm{T}}$ and $\left(f_{0,6},\, f_{1,6},\, f_{2,6}\right)^{\mathrm{T}}$

Test each correlation for significance, using the test statistic, T_j, defined as

$$T_j = r_j \sqrt{\frac{n_j - 2}{1 - r_j^2}} \tag{3.5}$$

where $n_j = I - 1 - j$ is the sample size, and r_j is the sample correlation coefficient. Under the null hypothesis, that the correlation is not significantly different from 0, we have that, approximately, T_j has a Student's t distribution, with $\nu_j = n_j - 2$ degrees of freedom.

(b) Repeat (a), but use Spearman's rank correlation (that is, calculate the correlation between the relative ranks of each value in each vector, rather than of the values themselves). Do the rank correlations give the same message as the Pearson correlations? Which do you think is more reliable here?

Solution

(a) The calculations are easily done in Excel. The results are given below. In the table, r_j represents the correlation of the pairs of development factors $(f_{i,j},\, f_{i,j+1})$ for $i = 0, \cdots, I - (j+1)$. We only compare columns up to $(f_{i,5},\, f_{i,6})$, as we need at least 3 data points to test significance.

Development Year, j	0-1	1-2	2-3	3-4	4-5	5-6
Pearson Correlation r_j	0.0523	0.3447	0.8929	0.1498	-0.7889	0.0974
Test statistic, T_j	0.1283	0.8211	3.9654	0.2625	-1.8154	0.0979
ν_j	6	5	4	3	2	1
p-value	0.902	0.449	**0.017**	0.810	0.211	0.938

One set of development factors, from $j = 2$ to $j = 3$, shows significant correlation, with a p value of 1.7%.

(b)

Development Year	0-1	1-2	2-3	3-4	4-5	5-6
Spearman Correlation	0.1190	-0.0357	0.4286	0.0000	-0.4000	-0.5000
Test statistic, T_j	0.2937	-0.0799	0.9487	0.0000	-0.6172	-0.5774
ν_j	6	5	4	3	2	1
p-value	0.779	0.939	0.397	1.000	0.600	0.667

When we work with the ranks, rather than the values of the development factors, we find no significant correlations between adjacent sets of development factors. The t-test for the Pearson correlation is suitable when the underlying random variables have the same variance. That is not part of the chain ladder assumptions, and indeed, is unlikely to be true, as later development factors are based on more data (in Chapter 4 we will formalize the assumption of non-constant variance). The Spearman test is more robust to outliers, and eliminates the problem of non-constant variance, but is not directly testing for zero correlation. That is, two random variables may have zero Spearman correlation, but non-zero covariance, which is what we are testing for. In practice, zero Spearman correlation of the development factors is a strong indication that the covariances are close to zero, so the Spearman correlation test is considered more robust than the Pearson. It is also worth noting, though, that with such small sample sizes, the power of these tests is small.

We can combine the test statistics into a single statistic, to give a (rather rough) overall assessment of rank correlation. We will only use statistics with $\nu_j \geq 3$, i.e. $j \leq I-6$, because the variance of the t distribution is infinite for $\nu_j \leq 2$, and we will need the statistic to have finite variance. Under the null hypothesis, we have that for each j, $\mathrm{E}[T_j] = 0$ and $\mathrm{Var}[T_j] = \frac{\nu_j}{\nu_j - 2}$, for $\nu_j \geq 3$. We construct the combined statistic as a weighted average of the T_j's, using the inverse variances as weights.

That is, let $w_j = (\nu_j - 2)/\nu_j = \left(\mathrm{Var}\left[T_j\right]\right)^{-1}$ denote the weights, then $T = \dfrac{\sum_{j=0}^{I-6} T_j\, w_j}{\sum_{j=0}^{I-6} w_j}$.

Under the null hypotheses of zero rank correlation, the T_j have zero covariance, so we have

$\mathrm{E}[T] = 0$ and

$$\mathrm{Var}[T] = \frac{\sum_{j=0}^{I-6} \mathrm{Var}\left[T_j\right] w_j^2}{\left(\sum_{j=0}^{I-6} w_j\right)^2} = \frac{\sum_{j=0}^{I-6} w_j}{\left(\sum_{j=0}^{I-6} w_j\right)^2} = \frac{1}{\sum_{j=0}^{I-6} w_j}$$
$$= \frac{1}{\sum_{j=0}^{I-6} (\nu_j - 2)/\nu_j}$$

Assuming further that under the null hypothesis, approximately, $T \sim \mathrm{N}\big(0, \mathrm{Var}[T]\big)$, we can determine a p-value for the significance of T, based on the null hypothesis of zero correlation. In the example above, using Pearson correlations, the weighted average test statistic using T_0 to T_3 is $T = 1.26$; the standard deviation under the null hypothesis is (approximately) 0.69, giving a p value of 0.07, indicating no significant evidence of correlation at the 5% level.[1]

Calendar Year Effects

The other chain ladder assumption is that individual accident years are independent, with the same underlying development factors f_j. One reason why this assumption may be invalid is if there is a calendar year effect in the data. This could arise, for example, as a result in changes in claims underwriting processes, that could speed up settlements, or if inflation rates are very different over the calendar years represented in $\mathcal{D}_I$. The calendar years are represented in the run-off triangle by the anti-diagonals. Mack (1994) suggests a test based on classifying the development factors in each development year as high or low, relative to the other values in the same DY. A value of $f_{i,j}$ that is higher than the median value for DY j is labelled L (for large); a value below the median is labelled S (small). Values on the median are disregarded. We then consider the frequency of L and S development factors across each anti-diagonal. If there is no calendar year effect, roughly half of the development factors in each calendar year should be large, and half small, and the numbers should be binomially distributed, with probability parameter 0.5. Let S_k denote the number of 'S' labels in calendar year k, and let L_k denote the number of 'L' labels in calendar year k. We exclude calendar year 0 which has only one value. We test the null hypothesis, that calendar years are independent, by considering the distribution of the smaller of S_k and L_k each year. This is not binomially distributed under the

[1] Mack (1994) suggests an aggregate test of the rank correlations that assumes that the correlations, $r_j \sim N(0, (n_j - 1)^{-1})$. This test is used in the R ChainLadder package.

null hypothesis, but its mean and variance can be derived from the binomial distribution. The test procedure is as follows.

1. For $k = 1, 2, ..., I-1$, find S_k, L_k, $n_k = S_k + L_k$, and $m_k = \lfloor (n_k - 1)/2 \rfloor$. Under the null hypothesis of no calendar year effect, both L_k and S_k have a bin(n_k, 0.5) distribution.

2. Let $Z_k = \min(S_k, L_k)$ for $k = 1, 2, ..., I-1$. Using the binomial distribution, it can be shown that

$$\mathrm{E}[Z_k] = \frac{n_k}{2} - \binom{n_k - 1}{m_k} \frac{n_k}{2^{n_k}} \qquad \text{and}$$
$$\mathrm{Var}[Z_k] = \frac{n_k(n_k-1)}{4} - \binom{n_k - 1}{m_k} \frac{n_k(n_k-1)}{2^{n_k}} + \mathrm{E}[Z_k] - \mathrm{E}[Z_k]^2$$

3. The test statistic is $Z = \sum\limits_{k=1}^{I-1} Z_k$. We assume that, approximately,

$$Z \sim N(\mathrm{E}[Z],\ \mathrm{Var}[Z]) \quad \text{where} \quad \mathrm{E}[Z] = \sum_{k=1}^{I-1} \mathrm{E}[Z_k] \quad \text{and} \quad \mathrm{Var}[Z] = \sum_{k=1}^{I-1} \mathrm{Var}[Z_k]$$

(using the null hypothesis that calendar years are independent).

Then the p - value for the 2-sided test is

$$p = 2\left(1 - \Phi\left(\frac{|Z - \mathrm{E}[Z]|}{\sqrt{\mathrm{Var}[Z]}}\right)\right).$$

Example 3.2. Test the Table 2.1 data for calendar year effects.

Solution The L-S triangle for the Table 2.1 data is shown in Table 3.1.

We review each anti-diagonal, starting from the second, to assess the S-L distribution by calendar year. A summary of the results is given in Table 3.2. The p–value for the test is $p = 2\left(1 - \Phi\left(\frac{|14 - 12.6875|}{\sqrt{3.6621}}\right)\right) = 0.4928$, indicating that there is no significant evidence of a calendar year effect in this data. □

We can also explore the chain ladder assumptions graphically. For example, we have assumed that the development factors for each accident year follow a similar pattern. A graph of the

	Development Year j								
AY i	0	1	2	3	4	5	6	7	8
0	L	L	S	L	S	L	*	L	*
1	L	L	S	S	S	S	L	S	
2	L	L	*	L	*	L	S		
3	S	L	L	L	L	S			
4	S	S	L	S	L				
5	S	S	S	S					
6	L	S	L						
7	*	S							
8	S								

Table 3.1: Large-Small categorisation of run-off triangle development factors for Table 2.1 data; median results are indicated by * .

CY, k	S_k	L_k	Z_k	n_k	m_k	$\mathrm{E}[Z_k]$	$\mathrm{Var}[Z_k]$
1	0	2	0	2	0	0.5000	0.2500
2	1	2	1	3	1	0.7500	0.1875
3	2	2	2	4	1	1.2500	0.4375
4	3	1	1	4	1	1.2500	0.4375
5	3	3	3	6	2	2.0625	0.6211
6	2	3	2	5	2	1.5625	0.3711
7	3	4	3	7	3	2.4063	0.5537
8	6	2	2	8	3	2.9063	0.8037
Total	20	19	**14**	39		12.6875	3.6621

Table 3.2: Summary of calendar year effect test of Table 2.1 data.

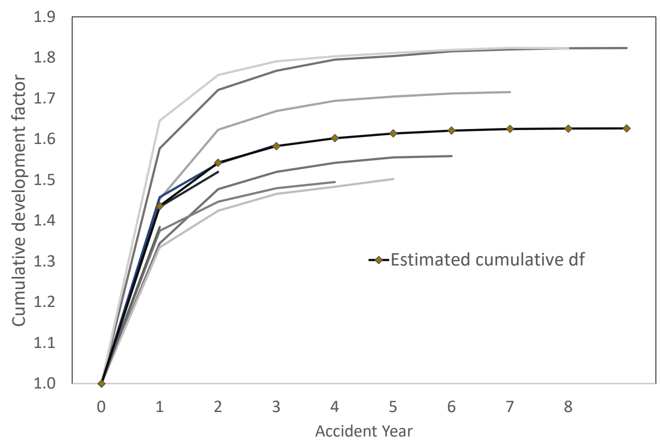

Figure 3.1: Cumulative development factors for data in Table 2.1. The marked line is the estimated cumulative development factor curve. Longer lines are for earlier AYs, shorter lines are for later AYs.

development factors will indicate whether there are any anomalies. For example, a jumbo claim settled early, or reported late, can distort the development factors. Typically, jumbo claims would be excluded from the chain ladder estimation, and managed through case estimates. In practice, the development factors are less informative than the cumulative development factors, that is, for each AY i, plot $f_{i,0} \times f_{i,1} \times \ldots \times f_{i,j}$ for $j = 0, 1, \cdots, I-i$. In Figure 3.1 we show the cumulative development factors for the data from Table 2.1. The longer curves correspond to the earlier AY, where we have more development year data. The thicker line indicates the cumulative development factors using the estimated $\hat{f}_j$ values. There is some evidence of a trend downwards in the cumulative developments, shown by the fact that the shorter curves (later accident years) lie below the longer curves (earlier accident years). This did not show up in the statistical tests, and would need to be investigated further.

3.3 The Bornhuetter-Ferguson (BF) method

One critique of the chain ladder method is that it can be very reliant on a small amount of data. In the run-off triangle in Table 2.1, we only have one observation for AY 2020, but that accident year is responsible for over 60% of the chain ladder estimate of outstanding claims. The BF method reduces the influence of the limited data available from the latest accident years, by using an estimate that can be viewed as a weighted average of the Chain Ladder estimate of the ultimate cost, and of a (possibly exogenous) prior estimate of the ultimate cost. For earlier accident years, where there are several years of development available, the weight is higher for the chain ladder estimate. For later accident years, where there are only one or two years of development data available, the weight is higher for the prior estimate.

Let μ_i represent the prior estimate of losses arising from AY i. This is often based on the expected loss ratio for the policies covered in that year. Let $\widetilde{C}_{i,J}$ denote the BF estimate of ultimate cumulative claims incurred in AY i. As above, we assume the data in $\mathcal{D}_I$ is available, and we let $\widehat{C}_{i,J}$ denote the estimated ultimate claims using the chain ladder method, so that $\widehat{C}_{i,J} = C_{i,I-i}\,\hat{\lambda}_{I-i}$.

The underlying assumptions of the BF method are given by Wüthrich and Merz (2008) as

BF Assumption (1) Given parameters β_j, $j=0,1,\cdots,J$, with $\beta_J=1$, and μ_i, $i=0,1,\cdots,I$, we have:

$$\mathrm{E}\big[C_{i,0}\big] = \beta_0\,\mu_i \tag{3.6}$$

$$\mathrm{E}\big[C_{i,j+1}\big|C_{i,0},\,C_{i,1},\,\cdots,C_{i,j}\big] = C_{i,j} + (\beta_{j+1}-\beta_j)\,\mu_i \tag{3.7}$$

BF Assumption (2) $C_{i,j}$ and $C_{l,k}$ are independent for $i \neq l$, and for all j, k.

From these assumptions, iterating equation (3.7), from $j = I-i$ to $j = J-1$, and given estimates $\hat{\beta}_j$ of the β_j parameters, we have the Bornhuetter-Ferguson estimate of cumulative claims,

$$\widetilde{C}_{i,J} = C_{i,I-i} + \Big(1-\hat{\beta}_{I-i}\Big)\,\mu_i \tag{3.8}$$

Example 3.3. Show that under the BF assumptions $\mathrm{E}[C_{i,j}] = \beta_j\,\mu_i$.

Solution

Again, we can prove this by induction. For $j = 0$ the equation is directly stated in Assumption (1) above. Assume the result is true for some $j \leq J-1$, i.e., assume that $\mathrm{E}[C_{i,j}] = \beta_j\,\mu_i$. Now consider

$$\begin{aligned}
\mathrm{E}\left[C_{i,j+1}\middle|C_{i,j}\right] &= C_{i,j} + (\beta_{j+1} - \beta_j)\,\mu_i \\
\mathrm{E}\left[C_{i,j+1}\right] = \mathrm{E}\big[\mathrm{E}\left[C_{i,j+1}\middle|C_{i,j}\right]\big] &= \mathrm{E}\left[C_{i,j} + (\beta_{j+1} - \beta_j)\,\mu_i\right] \\
&= \beta_j\mu_i + (\beta_{j+1} - \beta_j)\,\mu_i \quad \text{(from the induction assumption)} \\
&= \beta_{j+1}\,\mu_i, \qquad \text{as required.} \qquad \Box
\end{aligned}$$

To estimate the β_j parameters, we note that

$$\frac{\mathrm{E}[C_{i,j+1}]}{\mathrm{E}[C_{i,j}]} = \frac{\beta_{j+1}}{\beta_j}$$

This can be compared with the similar result for the Chain Ladder method, under which

$$\frac{\mathrm{E}[C_{i,j+1}]}{\mathrm{E}[C_{i,j}]} = \frac{\mathrm{E}[\mathrm{E}[C_{i,j+1}|C_{ij}]]}{\mathrm{E}[\mathrm{E}[C_{i,j}|C_{ij}]]} = \frac{f_j\,C_{i,j}}{C_{i,j}} = f_j$$

This indicates the connection between the β_j of the BF method, and the f_j of the CL method; given that $\beta_J = 1$ (from assumption (1)), we have

$$\beta_j = \frac{1}{f_j\,f_{j+1}\,\cdots\,f_{J-1}} = \frac{1}{\lambda_j}$$

which leads to estimates of the β_j parameters,

$$\hat{\beta}_j = \frac{1}{\hat{\lambda}_j} = \frac{1}{\hat{f}_j\,\hat{f}_{j+1}\,\cdots\,\hat{f}_{J-1}}.$$

Now, noting that the chain ladder estimate of the cumulative claims, given C_{I-i}, is $\widehat{C}_{i,J} = C_{i,I-i}\,\hat{\lambda}_{I-i}$ we can rewrite the BF cumulative claims estimate as

$$\widetilde{C}_{i,J} = \widehat{C}_{i,J}\,\frac{1}{\hat{\lambda}_{I-i}} + \left(1 - \hat{\beta}_{I-i}\right)\mu_i$$

$$\text{that is,}\quad \widetilde{C}_{i,J} = \hat{\beta}_{I-i}\,\widehat{C}_{i,J} + \left(1 - \hat{\beta}_{I-i}\right)\mu_i \tag{3.9}$$

So we see that the BF estimate can be thought of as a credibility estimate of ultimate claims, where the prior estimate for AY i is μ_i, and the data-based estimate is $\widehat{C}_{i,J}$. The credibility factor associated with the data-based estimate is $\hat{\beta}_{I-i}$, which is lower for higher values of i, when there is less data available to support the chain ladder estimate, and increases as i gets closer to 0, when the chain ladder estimate, $\widehat{C}_{i,J}$, is based on many years of development, indicating that it has more credibility than the loss ratio estimate.

Example 3.4.

You are given the following premium information for the data in Table 2.1.

Year	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Earned Premium	8825	8859	8850	8920	9120	9515	9010	11512	12240	14810

The expected loss ratio in each year is 0.85.

(a) Calculate the BF estimate of the outstanding claims reserve for the Table 2.1 data, assuming (i) no discounting, and (ii) payments are discounted at 5% per year effective. Assume payments are made 1/2-way through the calendar year.

(b) Compare the results with the chain ladder estimate, for the undiscounted reserves.

Solution

(a) (i) The prior estimates of claims costs for each AY are $\mu_i = 0.85 \times$ (Earned Premium). The estimated β parameters are found by inverting the $\hat{\lambda}_j$ from Example 2.2. We could use the $\widehat{C}_{i,J}$ values from Example 2.2, to calculate the $\widetilde{C}_{i,J}$ values directly, using equation (3.9), but since we are going to need the incremental payments for (ii), we will use equation (3.7) to determine the estimated cumulative claims for each development year.

The results, with data values in the top left triangle, and projected cumulative claims in the bottom right, are given in Table 3.3, along with the estimated $\hat{\beta}_j$.

The total estimated outstanding claims reserve, without discounting, is $\sum_{i=0}^{9} \widetilde{R}_i = 7026$.

(ii) Proceeding exactly as in Example 2.2, we find the discounted outstanding claims reserve is 6637.

(b) A summary of the chain ladder and Bornhuetter-Ferguson results is given in Table 3.4. P_i are the earned premiums in AY i; $\widehat{C}_{i,J}$ and $\widetilde{C}_{i,J}$ are, respectively, the chain ladder and Bornhuetter-Ferguson estimated ultimate claims, and $\widehat{R}_i$ and $\widetilde{R}_i$ are, respectively, the chain ladder and Bornhuetter-Ferguson estimated outstanding claims, by accident year. The final two columns show the estimated loss ratios by accident year for the chain ladder and Bornhuetter-Ferguson methods, respectively.

We note that the difference between the chain ladder and BF estimates is slight for earlier accident years, where the credibility factor for the chain ladder estimate is high, but there is more divergence in the estimates for later accident years, where there is less data

AY i	DY, j 0	1	2	3	4	5	6	7	8	9	$\widetilde{R}_i$
0	4360	6876	7501	7708	7826	7865	7916	7936	7947	7950	0
1	3996	6574	7023	7157	7206	7237	7268	7288	7292	7295	3
2	3840	5578	6233	6408	6504	6544	6573	6587	6594	6597	10
3	5108	6865	7545	7761	7875	7944	7960	7979	7986	7989	29
4	4585	6117	6531	6720	6800	6887	6920	6939	6947	6950	63
5	5767	7931	8341	8534	8620	8679	8713	8733	8741	8744	124
6	5550	8090	8548	8800	8891	8947	8979	8998	9006	9009	209
7	6525	9353	9915	10160	10277	10348	10389	10413	10423	10427	512
8	6620	9164	9844	10105	10229	10304	10348	10374	10384	10388	1224
9	7014	10385	11208	11523	11673	11764	11818	11849	11862	11866	4852
$\hat{\beta}_j$	0.6145	0.8823	0.9477	0.9727	0.9847	0.9919	0.9962	0.9986	0.9996	1.0	

Table 3.3: Bornhuetter-Ferguson cumulative claims projections for data from Table 2.1

supporting the chain ladder estimate. Note also that the projected loss ratio from the BF method will always lie between the projected loss ratio for the chain ladder method, and the prior loss ratio, 0.85.

In this example, the expected proportion of ultimate loss settled by the end of DY 0 is relatively high, at over 60%. Also the average projected loss ratio using the chain ladder is 85.8%, close to the prior assumption of 85% in the BF estimates. It is therefore not surprising that the two methods give similar results, although the projections for the final accident year are somewhat larger for the BF method. The two methods will diverge more significantly if the prior estimated loss ratio is very inaccurate, or when applied to longer tailed business, especially where there is a long period of IBNR claims. □

Exercise 3.2. You are given premium and expected loss ratio information for the two run-off triangles in Table 2.4.

(a) Calculate the outstanding claims reserve using the Bornhuetter-Ferguson method.

(b) Comment on the advantages and disadvantages of the Bornhuetter-Ferguson method, using your results as illustrations.

AY, i	**Short-tail** 0	1	2	3	4
Premiums	52600	54000	55520	57500	59850
Exp Loss Ratio	0.84	0.82	0.80	0.80	0.80

i	P_i	$\widehat{C}_{i,J}$	$\widetilde{C}_{i,J}$	$\widehat{R}_i$	$\widetilde{R}_i$	$\widehat{C}_{i,J}/P_i$	$\widetilde{C}_{i,J}/P_i$
0	8825	7950	7950	0	0	0.901	0.901
1	8859	7295	7295	3	3	0.823	0.823
2	8850	6596	6597	9	10	0.745	0.745
3	8920	7991	7989	31	29	0.896	0.896
4	9120	6943	6950	56	63	0.761	0.762
5	9515	8754	8744	134	124	0.920	0.919
6	9010	9047	9009	247	209	1.004	1.000
7	11512	10462	10427	547	512	0.909	0.906
8	12240	10386	10388	1222	1224	0.849	0.849
9	14810	11413	11866	4399	4852	0.771	0.801

Table 3.4: Summary of chain ladder and BF outstanding claims calculations and projected loss ratios, for the Table 2.1 data.

	Long-tail										
AY, i	0	1	2	3	4	5	6	7	8	9	10
Premiums	5641	5980	6339	6719	7122	7549	9002	9482	9491	9531	10103
Exp loss ratio	0.95	0.95	0.95	0.95	0.95	0.95	0.90	0.90	0.90	0.90	0.90

Solution

(a) The BF estimate of outstanding claims for the short-tail example is 62,870. The BF estimate of outstanding claims for the long-tail example is 34,570.

(b) The projected chain ladder loss rations for the short tail business are consistently lower than the estimated loss ratios, so the BF method gives significantly higher reserves in this case. The long-tail business is the other way around – the chain ladder projected loss ratios average 98%, so the BF reserve is substantially lower than the chain ladder. This indicates the importance of the expected loss ratio element of the BF calculation.

An advantage of the BF approach is that the importance of the most recent accident years' experience is reduced. The AY in the long tail data accounts for 27.5% of the total reserve, but the estimation of the ultimate cumulative claims is based on data for only 1 year, representing only about 1% of the total. Under the BF approach, this figure is given very little weight in the final calculation with a credibility factor of 1%.

3.4 The Bühlmann-Straub credibility model

We have seen that the Bornhuetter-Ferguson method can be viewed as a credibility estimate, where the chain ladder estimation of cumulative claims, $\widehat{C}_{i,J}$ is the estimate based on the accident year i experience up to calendar year I, and μ_i is the prior or exogenous estimate of the cumulative claims. The credibility factor is $Z_i = \hat{\beta}_{I-i} = \dfrac{1}{\hat{\lambda}_{I-i}}$.

As $\hat{\lambda}_j$ is an increasing function of j (assuming $\hat{f}_k \geq 1$), the credibility factor gives more weight to the data based estimate when there are more years of development data available, which makes sense. However, the approach assumes that we have a reliable estimate of μ_i available, and does not take into consideration the variance of the claims process, or the uncertainty associated with the μ_i estimates.

These are implicitly allowed for if we use the Bühlmann-Straub credibility model. The Bühlmann-Straub model constructs a linear approximation to a Bayesian estimate of the mean of a random variable, whose mean and variance are treated as functions of an unknown random parameter vector θ_i.

In applying the Bühlmann-Straub method to the outstanding claims reserve problem, we will use an iterated credibility approach. This will ensure that our reserve estimate for each AY is non-negative; we effectively use the Bühlmann-Straub estimate of ultimate cumulative claims as the μ_i factor in the BF formula. That means that the estimated ultimate cumulative claims will be

$$\widehat{C}_{i,J}^{BS2} = C_{i,I-i} + (1-\hat{\beta}_{I-i})\widehat{C}_{i,J}^{BS} = \hat{\beta}_{I-i}\widehat{C}_{i,J} + (1-\hat{\beta}_{I-i})\widehat{C}_{i,J}^{BS}$$

where $\widehat{C}_{i,J}^{BS}$ is the Bühlmann-Straub credibility estimate of ultimate claims, described below, and $\widehat{C}_{i,J}^{BS2}$ is the iterated credibility estimate, using $\hat{\beta}_{I-i}$ as the credibility factor. As above, $\widehat{C}_{i,J}$ is the chain ladder estimate. This estimate can be thought of as regular credibility estimate, with credibility factor Z_i^* for AY i, where

$$Z_i^* = 1 - \left(1-\hat{\beta}_{I-i}\right)(1-Z_i) \qquad \text{and} \qquad \widehat{C}_{i,J}^{BS2} = Z_i^*\,\widehat{C}_{i,J} + (1-Z_i^*)\,\mu$$

where μ is the prior mean value of ultimate cumulative claims in the Bühlmann-Straub credibility estimate.

For developing $\widehat{C}_{i,j}^{BS}$, we use the parameter $\gamma_j = \beta_j - \beta_{j-1}, j = 1, 2, \cdots, J$, with $\gamma_0 = \beta_0$. The estimated parameters, $\hat{\gamma}_j$ are evaluated from the $\hat{\beta}_j$ estimates. Loosely, the γ_j parameters represent the expected proportion of ultimate claims that are paid in DY j, while β_j represents the expected proportion of ultimate claims paid by the end of DY j; f_j represents the ratio of expected cumulative claims in year $j+1$ to year j, and λ_j represents the ratio of the expected ultimate cumulative claims to the cumulative claims in year j, and each of these is estimated using the chain ladder development factors. Note that

$$\sum_{j=0}^{k} \gamma_j = \beta_k \qquad \text{and} \quad \sum_{j=0}^{J} \gamma_j = 1$$

We let γ_j play the role of the volume measure in the Bühlmann-Straub framework, leading to the following assumptions.

Bühlmann-Straub Assumption (1) Given θ_i, the $X_{i,j}$ are conditionally independent random variables, for $j = 0, 1, \cdots, J$.

Bühlmann-Straub Assumption (2) We assume that for i, $j \in \{0, 1, \cdots, J\}$,

$$\mathrm{E}\left[X_{i,j}|\theta_i\right] = \gamma_j\, \mu(\theta_i) \qquad \text{and} \qquad \mathrm{Var}\left[X_{i,j}|\theta_i\right] = \gamma_j\, v(\theta_i)$$

Bühlmann-Straub Assumption (3) θ_i are assumed to be independent and identically distributed for $i = 0, 1, \cdots, I$, and the pairs $(X_{i,j}, \theta_i)$ and $(X_{l,j}, \theta_l)$ are jointly independent and identically distributed.

Let $\mu = \mathrm{E}[\mu(\theta_i)]$; note that this does not depend on i, as the θ_i are i.i.d. We also let $v = \mathrm{E}[v(\theta_i)]$ and $a = \mathrm{Var}[\mu(\theta_i)]$. Then the Bühlmann-Straub estimate of $\mathrm{E}[C_{i,J}|\mathcal{D}_I]$ is

$$\widehat{C}_{i,J}^{BS} = Z_i\, \widehat{C}_{i,J} + (1 - Z_i)\hat{\mu} \quad \text{where} \quad Z_i = \frac{\hat{\beta}_{I-i}}{\hat{\beta}_{I-i} + \hat{v}/\hat{a}} \quad \text{and} \quad \hat{\mu} = \frac{\sum_{i=0}^{I} Z_i\, \widehat{C}_{i,J}}{\sum_{i=0}^{I} Z_i}$$

The estimated parameters $\hat{\mu}$, $\hat{v}$, and $\hat{a}$ are the same for all $i = 1, 2, \cdots, J$. The credibility factors, Z_i, are different for each accident year, because of the $\hat{\beta}_{I-i}$ parameters.

The estimators for v and a, can be calculated using the empirical Bayes formulas (see, for example, Klugman *et al.* (2019)) as follows.

$$\text{Let } m_i = \hat{\beta}_{I-i}; \quad m = \sum_{i=0}^{I} m_i; \quad \overline{C} = \frac{\sum\limits_{i=0}^{I} C_{i,I-i}}{m}$$

$$\text{For } \ i = 0, 1, ..., I-1, \quad s_i^2 = \frac{1}{I-i} \sum_{j=0}^{I-i} \hat{\gamma}_j \left(\frac{X_{i,j}}{\hat{\gamma}_j} - \widehat{C}_{i,J} \right)^2$$

$$\text{Then} \quad \hat{v} = \frac{1}{I} \sum_{i=0}^{I-1} s_i^2 \quad \text{and} \quad \hat{a} = \frac{\sum\limits_{i=0}^{I} m_i \left(\widehat{C}_{i,J} - \overline{C} \right)^2 - I\hat{v}}{m - \frac{1}{m} \sum\limits_{i=0}^{I} m_i^2}$$

This method requires few assumptions, and provides an estimate for the μ_i in the BF equation that is estimated from the data, and is therefore less subjective, and is also automatically updated as new data arrives. Also, the credibility factors in the $\widehat{C}_{i,J}^{BS}$ calculation reflect not only the volume of data behind the $\widehat{C}_{i,J}$ estimate, but also the volatility of the claims process within each accident year (through $\hat{v}$), and the uncertainty in the estimate of μ (through $\hat{a}$). A major disadvantage over the BF approach is that we have assumed a common μ for all accident years. This assumption can be relaxed by working with per-premium claims data, rather than raw claims, and multiplying the resulting estimate by the premiums. That means in the equations above, $X_{i,j}$ and $C_{i,j}$ are replaced throughout with $X_{i,j}/P_i$ and $C_{i,j}/P_i$, where P_i is the earned premium in AY i.

Exercise 3.3.
Apply the Bühlmann-Straub model to the data in Table 2.1. Comment on the difference between the chain ladder, BF, and BS estimates of outstanding claims for this data. Use (a) the raw data and (b) the premium adjusted data.

Solution The estimated reserves, by accident years, are shown in the following table. We also show the BF and CL estimates for comparison, as well as the BF and BS credibility factors. In the column headers, BS2 is the Bühlmann-Straub estimate without premium adjustment, and BS2PA is the estimate with premium adjustment. The credibility factors are the iterated (Z_i^*) factors.

	$\widehat{R}_i$	$\widetilde{R}^{\mathrm{BF}}$	$\widehat{R}^{\mathrm{BS2}}$	$\widehat{R}^{\mathrm{BS2PA}}$	Z_i^{BF}	Z_i^{BS2}	Z_i^{BS2PA}	μ_i^{BF}	$\mu_i^{\mathrm{BS-raw}}$	$\mu_i^{\mathrm{BS-pa}}$
1	3	3	3	3	0.9996	1.0000	0.9999	7530	8669	7613
2	9	10	9	9	0.9986	0.9999	0.9996	7523	8669	7606
3	31	29	31	30	0.9962	0.9997	0.9988	7582	8669	7666
4	56	63	57	58	0.9919	0.9994	0.9975	7752	8669	7838
5	134	124	134	131	0.9847	0.9988	0.9953	8088	8669	8177
6	247	209	246	236	0.9727	0.9979	0.9916	7659	8669	7743
7	547	512	540	538	0.9477	0.9958	0.9836	9785	8669	9893
8	1222	1224	1205	1235	0.8823	0.9900	0.9610	10404	8669	10519
9	4399	4852	4275	4640	0.6145	0.9547	0.8396	12589	8669	12728
Total	6648	7026	6499	6881						

Note that the estimates are all fairly similar. The BF values are higher, as the weights for the later AYs are quite low, pushing the estimated cumulative claims towards the (higher) subjective μ_i values. The Bühlmann-Straub μ_i values for the raw data are all the same, based on the common $\hat{\mu}$ value estimated from the run-off data. This is too low for the later accident years, which are the most critical, so the Bühlmann-Straub method (without premium adjustment) tends to underestimate the reserve, Using the premium adjusted method improves the estimate; in this calculation, we assume a common mean loss ratio for each accident year, and do the Bühlmann-Straub calculations on the triangle of claims divided by earned premiums. This gives a common estimate of the mean loss ratio for each accident year. The values of $\mu_i^{\mathrm{BS-pa}}$ recorded in the table are the earned premiums in each year, multiplied by the estimated loss ratio for the table as a whole.

3.5 The Poisson model

In the Poisson model, we assume that incremental values in the run-off triangle are independent Poisson random variables. This model is more suited to a run-off triangle of claim frequency than claim amounts, as the claim numbers are more likely to be approximately Poisson distributed, and the incremental claim amounts may be negative, which is not possible under the Poisson assumption. However, the model provides an interesting perspective on the chain ladder method and the Bornhuetter-Ferguson methods, and also provides a possible starting point for measuring uncertainty associated with outstanding claims estimates.

The Poisson assumptions are

Poisson Assumption (1) $X_{i,j}$ are independent random variables, for all $i,\, j \in \{0, 1, \cdots, I\}$.

Poisson Assumption (2) There exist parameters $\mu_i,\ i = 0, 1, .., I$, and $\gamma_j,\ j = 0, 1, \cdots, J$ such that

$$X_{i,j} \sim \mathrm{Poi}\left(\mu_i \gamma_j\right), \text{ where } \mu_i,\, \gamma_j > 0, \text{ and } \sum_{j=0}^{J} \gamma_j = 1$$

The μ_i parameters here have a similar interpretation as in the BF method, as the expected ultimate claims cost (or frequency) for AY i. As before, the γ_j parameters indicate the proportion of ultimate claims paid in DY j.

From these assumptions, the following results immediately follow:

$$C_{i,j} \text{ and } C_{l,k} \text{ are independent for } i \neq l \tag{3.10}$$

$$\mathrm{E}\left[C_{i,j}\right] = \mu_i \sum_{k=0}^{j} \gamma_k \tag{3.11}$$

$$\mathrm{E}\left[C_{i,J}\right] = \mu_i \tag{3.12}$$

$$\mathrm{E}\left[C_{i,j+1} \middle| C_{i,0},\, C_{i,1},\, \cdots,\, C_{i,j}\right] = C_{i,j} + \mu_i\, \gamma_{j+1} \tag{3.13}$$

$$\mathrm{E}\left[C_{i,J} \middle| C_{i,0},\, C_{i,1},\, \cdots,\, C_{i,j}\right] = C_{i,j} + \mu_i \sum_{k=j+1}^{J} \gamma_k \tag{3.14}$$

Noting that $\beta_j = \sum\limits_{k=0}^{j} \gamma_k$, equations (3.13) and (3.14) can be written as

$$\mathrm{E}\left[C_{i,j+1} \middle| C_{i,0},\, C_{i,1},\, \cdots,\, C_{i,j}\right] = C_{i,j} + \left(\beta_{j+1} - \beta_j\right) \mu_i$$
$$\mathrm{E}\left[C_{i,J} \middle| C_{i,0},\, C_{i,1},\, \cdots,\, C_{i,I-i}\right] = C_{i,I-i} + \left(1 - \beta_{I-i}\right) \mu_i$$

The Poisson model estimate of ultimate cumulative claims for AY i is therefore

$$\widehat{C}_{i,J}^{P} = C_{i,I-i} + \hat{\mu}_i \sum_{k=j+1}^{J} \hat{\gamma}_k = C_{i,I-i} + \hat{\mu}_i \left(1 - \hat{\beta}_{I-i}\right) \tag{3.15}$$

where $\hat{\mu}_i$ and $\hat{\gamma}_j$ are estimated parameter values. We note that the BF assumptions are met by the Poisson model, and we see from equation (3.15) that the Poisson estimate of $C_{i,J}$ can be written in an identical form to the BF estimator. If $\hat{\mu}_i$ is treated in this model as a prior estimate of outstanding claims, then the estimated ultimate claims will follow the BF method. However, if we estimate the μ_i and the γ_j parameters from the data, $\mathcal{D}_I$, using maximum likelihood estimation, then we can express the Poisson estimate similarly to the chain ladder formula, as follows.

Because the $X_{i,j}$ are assumed to be independent, the likelihood function is just the product of the probability functions for each $X_{i,j}$ for $i+j \leq I$, based on the appropriate Poisson distributions, so we have

$$L = \prod_{i+j \leq I} \frac{e^{-\mu_i \gamma_j} \, (\mu_i \gamma_j)^{X_{i,j}}}{X_{i,j}\,!}$$

$$\implies l = \log L = -\sum_{i+j \leq I} \mu_i \gamma_j + \sum_{i+j \leq I} X_{i,j} \left(\log \mu_i + \log \gamma_j\right) - \sum_{i+j \leq I} \log X_{i,j}!$$

$$\implies \frac{\partial l}{\partial \mu_i} = -\sum_{j=0}^{I-i} \gamma_j + \frac{1}{\mu_i} \sum_{j=0}^{I-i} X_{i,j}$$

$$\text{set this equal to } 0 \implies \hat{\mu}_i \sum_{j=0}^{I-i} \hat{\gamma}_j = \sum_{j=0}^{I-i} X_{i,j} = C_{i,I-i} \tag{3.16}$$

So we have

$$\widehat{C}^P_{i,J} = C_{i,I-i} + \hat{\mu}_i \sum_{j=I-i+1}^{J} \hat{\gamma}_j \qquad \text{(from (3.15))}$$

$$C_{i,I-i} = \hat{\mu}_i \sum_{j=0}^{I-i} \hat{\gamma}_j \qquad \text{(from (3.16))}$$

$$\begin{aligned}
\implies \widehat{C}^P_{i,J} &= \hat{\mu}_i \sum_{j=0}^{I-i} \hat{\gamma}_j + \hat{\mu}_i \sum_{j=I-i+1}^{J} \hat{\gamma}_k = \hat{\mu}_i \sum_{j=0}^{J} \hat{\gamma}_j \\
&= \left(\hat{\mu}_i \sum_{j=0}^{I-i} \hat{\gamma}_j \right) \times \frac{\sum_{j=0}^{I-i+1} \hat{\gamma}_j}{\sum_{j=0}^{I-i} \hat{\gamma}_j} \times \frac{\sum_{j=0}^{I-i+2} \hat{\gamma}_j}{\sum_{j=0}^{I-i+1} \hat{\gamma}_j} \cdots \times \frac{\sum_{j=0}^{J} \hat{\gamma}_j}{\sum_{j=0}^{J-1} \hat{\gamma}_j} \\
&= C_{i,I-i} \times \hat{f}_{I-i} \times \hat{f}_{I-i+1} \times \cdots \times \hat{f}_{J-1}
\end{aligned}$$

$$\text{where } \hat{f}_k = \frac{\sum_{j=0}^{k+1} \hat{\gamma}_j}{\sum_{j=0}^{k} \hat{\gamma}_j}.$$

This demonstrates that the Poisson model estimator, with MLE parameters, can be written in the same form as the chain ladder estimator. Furthermore, the MLE formulae for the parameters generate development factor estimates that are identical to the chain ladder formulas, so the outstanding claims reserve calculated using the Poisson MLE model is identical to the reserve using the chain ladder model (see Wüthrich and Merz (2008) for a proof of this result).

The fact that the Poisson MLE estimate of outstanding claims is identical to the chain ladder estimate does not mean the two models are the same. The chain ladder method is non-parametric, and without additional assumptions offers no information about the uncertainty associated with the estimated outstanding claims. The Poisson model assumptions are much stronger than the chain ladder assumptions, and in fact will be too restrictive in most cases involving claim amounts. This does not affect the estimated outstanding claims, but it does mean that any inference with respect to uncertainty derived using the Poisson model assumptions will be very questionable for most run-off triangle data, where the data is unlikely to be consistent with the Poisson assumptions. The main problem is that outstanding claims data, whether by frequency or amount, is typically significantly overdispersed, compared with the Poisson distribution - that is, the variance is greater than the mean. That leads us to a more suitable choice, the **overdispersed Poisson (ODP) model**, which has the same mean structure as the Poisson model, – that is, $\mathrm{E}[X_{i,j}] = \mu_i \gamma_j$, but with variance incorporating an overdispersion parameter, such that $\mathrm{Var}[X_{i,j}] = \phi \mu_i \gamma_j$. Because the estimated cumulative claims depend only on the mean values of the $X_{i,j}$, not on the variances, the outstanding claims reserve calculated using the ODP model, with maximum likelihood estimated parameters, is identical to the Poisson estimate, which is, as we have noted, identical to the chain ladder estimate. However, allowing for overdispersion in the random variables will allow us to estimate the standard errors of the estimates more accurately, compared wth the Poisson distribution. We consider the ODP model in more detail in Chapter 5.

Chapter 4

Mack's model

4.1 Introduction and assumptions

The methods presented in previous sections all focus on a single point estimate for outstanding claims, with the chain ladder providing the most recognized approach. However, it is common in reserving to add a margin for adverse experience or for uncertainty in the parameter estimates. Mack (1993) proposed an extension of the chain ladder method that reproduces the chain ladder estimate of outstanding claims, but adds an assumption about the variance of the cumulative claims that allows users to estimate the uncertainty associated with the chain ladder estimates. This can be used in setting prudent reserves and solvency margins, and also acts as a benchmark to assess whether the reserve assumptions continue to be appropriate. The Mack model assumptions are

Mack Assumption (1) $C_{i,j}$ and $C_{l,k}$ are independent for $i \neq l$, and for all j, k.

Mack Assumption (2) For a given accident year i, $\{C_{i,j}\}_{j=0,1,2,\cdots}$ is a Markov Chain.

Mack Assumption (3) There exist f_j such that $\mathrm{E}\left[C_{i,j+1} \middle| C_{i,j}\right] = f_j\, C_{i,j}$.

Mack Assumption (4) There exist σ_j^2 such that $\mathrm{Var}\left[C_{i,j+1} \middle| C_{i,j}\right] = \sigma_j^2\, C_{i,j}$.

Assumptions (1) and (3) are exactly the chain ladder assumptions. Assumptions (2) and (4) are added to allow us to explore prediction uncertainty, without having to make any stronger assumptions about specific distributions for incremental or cumulative claims.

The independence of accident years assumption, together with the Markov assumption, allow us to replace $C_{i,J}|\mathcal{D}_I$ with $C_{i,J}|C_{i,I-i}$, as values in $\mathcal{D}_I$ relating to other accident years are irrelevant

(Assumption (1)), and values relating to earlier development years for AY i are superfluous (Assumption (2)).

The estimator for f_j in Mack's model is identical to the chain ladder estimate:

$$\hat{f}_j = \frac{\sum_{i=0}^{I-1-j} C_{i,j+1}}{\sum_{i=0}^{I-1-j} C_{i,j}} = \frac{\sum_{i=0}^{I-1-j} C_{i,j}\, f_{i,j}}{\sum_{i=0}^{I-1-j} C_{i,j}} \qquad j \in \{0, 1, \cdots, J-1\}$$

In Section 3.1 we demonstrated that $\hat{f}_j$ is an unbiased estimator of f_j.

The estimator for σ_j^2 in Mack's model is

$$\hat{\sigma}_j^2 = \frac{1}{I-1-j} \sum_{i=0}^{I-1-j} C_{i,j} \left(f_{i,j} - \hat{f}_j\right)^2 \qquad \text{for } j \leq I-2$$

We state without proof that $\hat{\sigma}_j^2$ is an unbiased estimator of σ_j^2. We see that $\hat{f}_j$ is a weighted mean of the sample development factors, and $\hat{\sigma}_j^2$ is a weighted variance.

If $J = I$ (which is often the case) then we do not have enough information to estimate σ_{J-1}^2, as we only have a single observation of $f_{i,j}$. Mack (1994) suggests using

$$\hat{\sigma}_{J-1}^2 = \min\left(\hat{\sigma}_{J-2}^2,\, \hat{\sigma}_{J-3}^2,\, \hat{\sigma}_{J-2}^4/\hat{\sigma}_{J-3}^2\right)$$

Note that every individual development factor $f_{i,j}$, $i+j \leq I$, is an unbiased estimator of f_j, and so any weighted average of the $f_{i,j}$'s, across accident years, is also an unbiased estimate of f_j.

Exercise 4.1. Calculate $\hat{\sigma}_j$ for the data in Table 2.1. Use Mack's method for $\hat{\sigma}_8$.

Solution

DY	0	1	2	3	4	5	6	7	8
$\hat{\sigma}_j$	7.028	1.907	0.330	0.288	0.290	0.162	0.026	0.052	0.026

□

Theorem 2.
Under the Mack model assumptions, $\hat{f}_j$ is the minimum variance estimator of f_j from all linear combinations of the individual development factors, $f_{i,j}$, conditional on $\mathcal{B}_j$.

Proof

Let $\tilde{f}_j$ denote a general weighted average of $f_{i,j}$, for $i = 0, 1, \cdots, I-1-j$, i.e.

$$\tilde{f}_j = \sum_{i=0}^{I-1-j} w_i \, f_{i,j} \quad \text{where} \quad \sum_{i=0}^{I-1-j} w_i = 1$$

Let $\sigma_{i,j}^2 = \text{Var}\left[f_{i,j} \middle| C_{i,j}\right]$. Because claims in different accident years are assumed to be independent, we have

$$\text{Var}\left[\tilde{f}_j\right] = \sum_{i=0}^{I-1-j} w_i^2 \, \sigma_{i,j}^2$$

We want to find the w_i to minimize this variance, subject to the constraint that $\sum_{i=0}^{I-1-j} w_i = 1$. Letting λ denote the Lagrange multiplier, we minimize

$$L = \sum_{i=0}^{I-1-j} w_i^2 \, \sigma_{i,j}^2 - \lambda \left(\sum_{i=0}^{I-1-j} w_i - 1 \right)$$

Differentiating with respect to w_k, and setting equal to 0, will give the minimum variance weights, denoted w_k^*, for $k \in \{0, 1, \cdots, I-1-j\}$:

$$\frac{\partial L}{\partial w_k} = 2 w_k \sigma_{k,j}^2 - \lambda \implies w_k^* = \frac{\lambda}{2\sigma_{k,j}^2}$$

Now we use the constraint to solve and substitute for λ:

$$1 = \sum_{i=0}^{I-1-j} w_i^* = \frac{\lambda}{2} \sum_{i=0}^{I-1-j} \frac{1}{\sigma_{i,j}^2} \implies \lambda = \frac{2}{\sum_{i=0}^{I-1-j} 1/\sigma_{i,j}^2} \implies w_k^* = \frac{1/\sigma_{k,j}^2}{\sum_{i=0}^{I-1-j} 1/\sigma_{i,j}^2}$$

This is a fairly general result, that is, for any estimator which is a weighted average of independent unbiased estimators, the minimum variance weights are inversely proportional to the variance of the individual estimators. To apply it in the Mack model case, we condition on $C_{0,j}, \cdots, C_{I-1-j,j}$, and use Assumption (4):

$$\begin{aligned} \sigma_{i,j}^2 &= \text{Var}\left[f_{i,j} | C_{i,j}\right] = \text{Var}\left[C_{i,j+1}/C_{i,j} \middle| C_{i,j}\right] \\ &= \frac{1}{C_{i,j}^2} \text{Var}\left[C_{i,j+1} \middle| C_{i,j}\right] = \frac{1}{C_{i,j}^2} \left(\sigma_j^2 \, C_{i,j}\right) = \frac{\sigma_j^2}{C_{i,j}} \end{aligned}$$

$$\implies w_k^* = \frac{C_{k,j}/\sigma_j^2}{\sum_{i=0}^{I-1-j} C_{i,j}/\sigma_j^2} = \frac{C_{k,j}}{\sum_{i=0}^{I-1-j} C_{i,j}} \qquad \text{as required.}$$

□

In Section 2.2 we stated, rather vaguely, that using a weighted average of the $f_{i,j}$'s with $C_{i,j}$ weights, to estimate f_j, was more efficient than taking a straight unweighted average. This theorem provides more rigorous support for that statement.

4.2 Uncertainty measures in Mack's model

We consider the prediction uncertainty through the **mean square error of prediction** (MSEP) of of $\widehat{C}_{i,J}$ given $\mathcal{D}_I$, where $\widehat{C}_{i,J}$ is the chain ladder estimate of $C_{i,J}$, based on the run-off triangle $\mathcal{D}_I$. The (conditional) MSEP is defined as

$$\text{MSEP}\left(\widehat{C}_{i,J}|\mathcal{D}_I\right) = \text{E}\left[\left(C_{i,J} - \widehat{C}_{i,J}\right)^2 |C_{i,I-i}\right]$$

Now, $\widehat{C}_{i,J}|\mathcal{D}_I$ is not a random variable, as $\widehat{C}_{i,J}$ is a function of the $C_{i,j}$ values contained in the set $\mathcal{D}_I$. The MSEP equation can therefore be rearranged as

$$\begin{aligned}\text{MSEP}\left(\widehat{C}_{i,J}|\mathcal{D}_I\right) &= \text{E}\left[C_{i,J}^2\middle| C_{i,I-i}\right] - 2\widehat{C}_{i,J}\,\text{E}\left[C_{i,J}\middle| C_{i,I-i}\right] + \widehat{C}_{i,J}^2\\ &= \text{Var}\left[C_{i,J}\middle| C_{i,I-i}\right] + \left(\widehat{C}_{i,J} - \text{E}\left[C_{i,J}\middle| C_{i,I-i}\right]\right)^2 \qquad (4.1)\end{aligned}$$

The first term on the right side of equation (4.1) is not related to the estimator. This is the **process variance**; it is innate to the claims settlement process. The second term is the **estimation error**, which arises from the discrepancy between the estimated and actual values of the conditional expected ultimate claims. Both the process and estimation error are important in the assessment of reserve uncertainty.

4.2.1 Process variance

The model assumptions specify the relationship between cumulative claims in successive development years, so we derive an equation for $\text{Var}[C_{i,J}| C_{i,I-i}]$ by iteratively developing formulas for $\text{Var}[C_{i,j+1}| C_{i,j}]$, as follows.

Assume that we know $C_{i,j}$. Then from the model assumptions, for one-step ahead we have

$$\text{Var}\left[C_{i,j+1}\middle| C_{i,j}\right] = \sigma_j^2\, C_{i,j} \qquad (4.2)$$

For two steps ahead, we condition first on $C_{i,j+1}$, and then on $C_{i,j}$:

$$\mathrm{Var}\left[C_{i,j+2}\middle|\, C_{i,j}\right] = \mathrm{E}\left[\mathrm{Var}\left[C_{i,j+2}\middle|\, C_{i,j+1}\right] \middle|\, C_{i,j}\right] + \mathrm{Var}\left[\mathrm{E}\left[C_{i,j+2}\middle|\, C_{i,j+1}\right] \middle|\, C_{i,j}\right]$$

From the model assumptions, we know that

$$\mathrm{Var}\left[C_{i,j+2}\middle|\, C_{i,j+1}\right] = C_{i,j+1}\,\sigma^2_{j+1} \quad \text{and} \quad \mathrm{E}\left[C_{i,j+2}\middle|\, C_{i,j+1}\right] = C_{i,j+1}\, f_{j+1}$$

so

$$\begin{aligned}
\mathrm{Var}\left[C_{i,j+2}\middle|\, C_{i,j}\right] &= \mathrm{E}\left[C_{i,j+1}\,\sigma^2_{j+1}\middle|\, C_{i,j}\right] + \mathrm{Var}\left[C_{i,j+1}\, f_{j+1}\middle|\, C_{i,j}\right] && (4.3)\\
&= \sigma^2_{j+1}\, C_{i,j}\, f_j + f^2_{j+1}\,\sigma^2_j\, C_{i,j} && (4.4)
\end{aligned}$$

Now we move another step ahead, conditioning first on $C_{i,j+2}$ and then on $C_{i,j}$:

$$\begin{aligned}
\mathrm{Var}\left[C_{i,j+3}\middle|\, C_{i,j}\right] &= \mathrm{E}\left[\mathrm{Var}\left[C_{i,j+3}\middle|\, C_{i,j+2}\right] \middle|\, C_{i,j}\right] + \mathrm{Var}\left[\mathrm{E}\left[C_{i,j+3}\middle|\, C_{i,j+2}\right] \middle|\, C_{i,j}\right]\\
&= \mathrm{E}\left[\sigma^2_{j+2}\, C_{i,j+2}\middle|\, C_{i,j}\right] + \mathrm{Var}\left[f_{j+2} C_{i,j+2}\middle|\, C_{i,j}\right]\\
&= \sigma^2_{j+2}\mathrm{E}\left[C_{i,j+2}\middle|\, C_{i,j}\right] + f^2_{j+2}\,\mathrm{Var}\left[C_{i,j+2}|C_{i,j}\right] && (4.5)\\
&= \sigma^2_{j+2}\, C_{i,j}\, f_j\, f_{j+1} + f^2_{j+2}\left(\sigma^2_{j+1}\, C_{i,j}\, f_j + f^2_{j+1}\,\sigma^2_j\, C_{i,j}\right) \qquad \text{(from (4.4))} && (4.6)\\
&= \sigma^2_{j+2}\, f_j\, f_{j+1}\, C_{i,j} + \sigma^2_{j+1} f^2_{j+2}\, f_j\, C_{i,j} + \sigma^2_j\, f^2_{j+1}\, f^2_{j+2}\, C_{i,j} && (4.7)
\end{aligned}$$

We see the iterative nature of the calculation in (4.5) and (4.6), where we slot the result of the previous iteration into the formula.

Exercise 4.2. Show that

$$\begin{aligned}
\mathrm{Var}\left[C_{i,j+4}\middle|\, C_{i,j}\right] =& \sigma^2_{j+3}\, f_j\, f_{j+1}\, f_{j+2}\, C_{i,j} + \sigma^2_{j+2}\, f_j\, f_{j+1}\, f^2_{j+3}\, C_{i,j}\\
&+ \sigma^2_{j+1}\, f_j\, f^2_{j+2}\, f^2_{j+3}\, C_{i,j} + \sigma^2_j\, f^2_{j+1}\, f^2_{j+2}\, f^2_{j+3}\, C_{i,j}
\end{aligned}$$

□

We can use the iteration to derive the following general equation for $\mathrm{Var}[C_{i,j+K}|C_{i,j}]$, which we rearrange in order to make the calculation easier, as follows.

$$
\begin{aligned}
\mathrm{Var}\left[C_{i,j+K}|C_{i,j}\right] &= \sigma_j^2\, C_{i,j}\, f_{j+1}^2\, \cdots f_{j+K-1}^2 + \sum_{k=1}^{K-1} \sigma_{j+k}^2\, C_{i,j}\, \left(f_j \cdots f_{j+k-1}\right) \left(f_{j+k+1}^2 \cdots f_{j+K-1}^2\right) \\
&= \sum_{k=0}^{K-1} \frac{\sigma_{j+k}^2}{f_{j+k}} \left(C_{i,j}\, f_j\, f_{j+1}\, \cdots f_{j+K-1}\right) \left(f_{j+k+1}\, f_{j+k+2} \cdots f_{j+K-1}\right) \\
&= \sum_{k=0}^{K-1} \frac{\sigma_{j+k}^2}{f_{j+k}^2} \frac{\left(C_{i,j}\, f_j\, f_{j+1}\, \cdots f_{j+K-1}\right)^2}{\left(C_{i,j}\, f_j\, f_{j+1}\, \cdots f_{j+k-1}\right)}
\end{aligned}
$$

The reason for this rearrangement is that $\left(C_{i,j}\, f_j\, f_{j+1}, \cdots f_{j+k-1}\right) = \mathrm{E}[C_{i,j+k}|C_{i,j}]$, which means that the process variance can be written as

$$
\mathrm{Var}\left[C_{i,j+K}|C_{i,j}\right] = \mathrm{E}\left[C_{i,j+K}\middle|\, C_{i,j}\right]^2 \sum_{k=0}^{K-1} \frac{\sigma_{j+k}^2}{f_{j+k}^2\, \mathrm{E}\left[C_{i,j+k}|C_{i,j}\right]}.
$$

In the context of the run off triangle, we are estimating the variance of $C_{i,J}$, given the latest development year claims data, $C_{i,I-i}$, so we have

$$
\mathrm{Var}\left[C_{i,J}|C_{i,I-i}\right] = \mathrm{E}\left[C_{i,J}\middle|\, C_{i,I-i}\right]^2 \sum_{j=I-i}^{J-1} \frac{\sigma_j^2}{f_j^2\, \mathrm{E}\left[C_{i,j}|C_{i,I-i}\right]}. \tag{4.8}
$$

To estimate the process variance, we approximate $\mathrm{E}[C_{i,J}|C_{I-i}]$ with $\widehat{C}_{i,J}$, f_j with $\hat{f}_j$, and σ_j^2 with $\hat{\sigma}_j^2$, giving

$$
\mathrm{Var}\left[C_{i,J}|C_{i,I-i}\right] \approx \widehat{C}_{i,J}^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2\, \widehat{C}_{i,j}}. \tag{4.9}
$$

The process variance for the estimated outstanding claims from AY i is therefore estimated as

$$
\mathrm{Var}\left[C_{i,J} - C_{i,I-i}|C_{i,I-i}\right] \;=\; \mathrm{Var}\left[C_{i,J}|C_{i,I-i}\right] \;\approx\; \widehat{C}_{i,J}^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2\, \widehat{C}_{i,j}}. \tag{4.10}
$$

The total process variance for the outstanding claims is just the sum of the variances for individual AY's, as the accident years are assumed to be independent.

Exercise 4.3. Calculate the estimated process standard deviation for the outstanding claims from each AY in Table 2.1.

Solution

AY	1	2	3	4	5	6	7	8	9
$\sqrt{\text{Process Variance}}$	2.23	4.69	5.67	14.53	31.70	42.36	56.72	200.72	699.44

Note that in this case, most of the uncertainty in the outstanding claims estimate is generated by the latest accident year. □

4.2.2 Estimation error, individual AY

The estimation error is measured in the second term in the MSEP, $\left(\widehat{C}_{i,J} - \mathrm{E}\left[C_{i,J}|\mathcal{D}_I\right]\right)^2$. This term is more complicated to work with, as most of what we know about $\mathrm{E}\left[C_{i,J}|\mathcal{D}_I\right]$ is already contained in $\widehat{C}_{i,j}$. Several methods have been proposed to to estimate this term, all involving either additional assumptions, or some substantial approximation. In this note, we follow the original approach from Mack (1993), which analyzes the impact of successive discrepancies between estimated and underlying development factors, and estimates that impact by successively conditioning on all the information available up to the latest relevant development year.

Theorem 3.

$$\left(\widehat{C}_{i,J} - \mathrm{E}\left[C_{i,J}|\mathcal{D}_I\right]\right)^2 \approx \widehat{C}_{i,J}^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2\, S_j} \qquad \text{where } S_j = \sum_{i=0}^{I-j-1} C_{i,j}. \tag{4.11}$$

See the Appendix for a proof of this result.

The function S_j is the sum of the column j cumulative claims from the run-off triangle, excluding the most recent value. It is the denominator in the calculation of $\hat{f}_j$.

Exercise 4.4. Calculate the square root of the estimation error for each AY for the data in Table 2.1.

Solution

AY	1	2	3	4	5	6	7	8	9
$\sqrt{\text{Estimation error}}$	2.13	3.36	4.31	7.53	16.31	20.84	28.28	81.93	274.66

□

4.2.3 MSEP for aggregate outstanding claims

MSEP by accident year

We can combine the process variance and the estimation error into a single term for the conditional MSEP for accident year i,

$$\begin{aligned}\text{MSEP}\left(\widehat{C}_{i,J}|\mathcal{D}_I\right) = MSEP(\widehat{R}_i|\mathcal{D}_I) &\approx \underbrace{\widehat{C}_{i,J}^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2\ \widehat{C}_{i,j}}}_{\text{Process variance}} + \underbrace{\widehat{C}_{i,J}^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2\ S_j}}_{\text{Estimation error}} \\ &\approx \widehat{C}_{i,J}^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2}\left(\frac{1}{\widehat{C}_{i,j}} + \frac{1}{S_j}\right) \end{aligned} \tag{4.12}$$

MSEP aggregated over accident years

Because the individual accident years are independent, we can add the process variance terms from the individual AYs for the aggregate process variance:

$$\begin{aligned}\text{Var}\left[R|\mathcal{D}_I\right] &= \text{Var}\left[C_{1,J} + C_{2,J} + \ldots + C_{I,J}\middle|\mathcal{D}_I\right] \\ &\approx \sum_{i=1}^{I} \widehat{C}_{i,J}^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2\,\widehat{C}_{i,j}} \qquad \text{(from equation (4.10))}\end{aligned}$$

However, we cannot add the MSEPs for the individual accident years together for an aggregate MSEP, because the estimated $\widehat{C}_{i,J}$ values are not independent – they are all functions of overlapping subsets of $\{\hat{f}_0, \hat{f}_1, \cdots, \hat{f}_{J-1}\}$. We therefore have some covariance terms to consider, as follows.

$$
\begin{aligned}
\mathrm{MSEP}\left(\widehat{R}\,\middle|\mathcal{D}_I\right) =& \mathrm{Var}\left[R\,\middle|\mathcal{D}_I\right] + \left(\hat{R} - \mathrm{E}\left[R\,\middle|\mathcal{D}_I\right]\right)^2 \\
\left(\widehat{R} - \mathrm{E}\left[R\,\middle|\mathcal{D}_I\right]\right)^2 =& \left(\widehat{C}_{1,J} + \widehat{C}_{2,J} + \ldots + \widehat{C}_{I,J} - \mathrm{E}\left[C_{1,J} + C_{2,J} + \ldots + C_{I,J}\,\middle|\mathcal{D}_I\right]\right)^2 \\
=& \sum_{i=1}^{I}\left(\widehat{C}_{i,J} - \mathrm{E}\left[C_{i,J}\,\middle|\mathcal{D}_I\right]\right)^2 \\
&+ 2\sum_{i=1}^{I-1}\sum_{l=i+1}^{I}\left(\widehat{C}_{i,J} - \mathrm{E}\left[C_{i,J}\,\middle|\mathcal{D}_I\right]\right)\left(\widehat{C}_{l,J} - \mathrm{E}\left[C_{l,J}\,\middle|\mathcal{D}_I\right]\right)
\end{aligned}
$$

The terms in the first sum on the right side are the individual accident year estimation errors, which we estimate using equation (4.11). The second term is a covariance term, which Mack approximates using the same approach as the estimation errors, giving the approximation for any i, l such that $1 \leq i < l \leq I$

$$
\left(\widehat{C}_{i,J} - \mathrm{E}\left[C_{i,J}\,\middle|\mathcal{D}_I\right]\right)\left(\widehat{C}_{l,J} - \mathrm{E}\left[C_{l,J}\,\middle|\mathcal{D}_I\right]\right) \approx \widehat{C}_{i,J}\,\widehat{C}_{l,J}\sum_{j=I-i}^{J-1}\frac{\hat{\sigma}_j^2}{\hat{f}_j^2\,S_j}
$$

Putting all the pieces together, we have an approximation for the aggregate MSEP for the outstanding claims, given $\mathcal{D}_I$:

$$
\begin{aligned}
\mathrm{MSEP}\left(\widehat{R}\,\middle|\mathcal{D}_I\right) =& \sum_{i=1}^{I}\left(\mathrm{Var}[C_{i,J}\,|\mathcal{D}_I] + \left(\widehat{C}_{i,J} - \mathrm{E}\left[C_{i,J}\,\middle|\mathcal{D}_I\right]\right)^2\right) \\
&+ 2\sum_{i=1}^{I-1}\sum_{l=i+1}^{I}\left(\widehat{C}_{i,J} - \mathrm{E}\left[C_{i,J}\,\middle|\mathcal{D}_I\right]\right)\left(\widehat{C}_{l,J} - \mathrm{E}\left[C_{l,J}\,\middle|\mathcal{D}_I\right]\right) \\
\approx& \sum_{i=1}^{I}\widehat{C}_{i,J}^2\sum_{j=I-i}^{J-1}\frac{\hat{\sigma}_j^2}{\hat{f}_j^2}\left(\frac{1}{\widehat{C}_{i,j}} + \frac{1}{S_j}\right) + 2\sum_{i=1}^{I-1}\widehat{C}_{i,J}\left(\sum_{j=I-i}^{J-1}\frac{\hat{\sigma}_j^2}{\hat{f}_j^2\,S_j}\right)\left(\sum_{l=i+1}^{I}\widehat{C}_{l,J}\right)
\end{aligned}
$$

The square root of the MSEP is the standard error of the estimated outstanding claims reserve.

Example 4.1.

(a) Estimate the standard error of $\widehat{R}\,\middle|\mathcal{D}_I$, for the data in Table 2.1.

(b) Identify the contribution to the standard deviation of the estimation error.

(c) Comment on the impact of the covariance term.

Solution

(a) We have calculated the individual AY process variance and estimation error in previous exercises. The total of the process variance terms for the individual accident years is 535, 794. The total of the individual AY estimation errors is 83, 739.

Let $\xi_j = \hat{\sigma}_j^2/(\hat{f}_j^2\, S_j)$. Then the contribution to the covariance terms for $i = 1$ to $i = 8$ is

$\boldsymbol{i = 1}, \quad \boldsymbol{I - i = 8}:$
$2\,\widehat{C}_{1,9}\ \ \xi_8\ \left(\widehat{C}_{2,9} + \widehat{C}_{3,9} + \ldots + \widehat{C}_{9,9}\right) = 89.27$

$\boldsymbol{i = 2}, \quad \boldsymbol{I - i = 7}:$
$2\,\widehat{C}_{2,9}\ \ (\xi_7 + \xi_8)\ \left(\widehat{C}_{3,9} + \widehat{C}_{4,9} + \ldots + \widehat{C}_{9,9}\right) = 222.97$

$\boldsymbol{i = 3}, \quad \boldsymbol{I - i = 6}:$
$2\,\widehat{C}_{3,9}\ \ (\xi_6 + \xi_7 + \xi_8)\ \left(\widehat{C}_{4,9} + \widehat{C}_{4,9} + \ldots + \widehat{C}_{9,9}\right) = 265.93$

$\vdots$

$\boldsymbol{i = 8}, \quad \boldsymbol{I - i = 1}:$
$2\,\widehat{C}_{8,9}\ \ (\xi_1 + \xi_2 + \ldots + \xi_8)\ \left(\widehat{C}_{9,9}\right) = 14,751.01$

The total of the covariance terms is 25, 083.82.

Summing the three parts of the MSEP gives an aggregate estimated MSEP of 644, 617, so the standard error is 802.9.

(b) The estimated aggregate process standard deviation is $\sqrt{535,794} = 732.0$. The difference between the aggregate MSEP and the aggregate SD is the contribution from estimation error, which in this case is 70.9.

(c) If we had ignored the covariance term, we would have estimated the aggregate standard error as $\sqrt{535,794 + 83,739} = 787.1$, so the covariance term increased the standard error by 15.8 which is around 2.0% of the total.

□

Exercise 4.5.
Calculate the estimated process variance and standard error for the outstanding claims reserves from the run-off triangles in Table 2.5. Comment on the differences between the short tail and long tail results.

4.3 Mack's Model in R

The ChainLadder package in R (Gesmann *et al.*, 2022) can be used to fit Mack's model to run-off data. First use the `triangle` command to create an object recognized as a triangle from the vectors of cumulative claims paid. A summary of the Mack model results described in this chapter, for a run-off triangle called `ROT` is obtained using the command

```
MackChainLadder(ROT,est.sigma="Mack")
```

Using the data from Table 2.1 generates the output shown in Table 4.1. Details of the output are as follows.

- The first column is the AY, but note that the earliest AY is labelled 1 in the R output, where we have used AY 0 for the first accident year.
- The second column shows the cumulative claims from the run-off data, up to the latest CY, for each AY.
- The third column gives $\hat{\beta}_{I-i}$, which is the estimated proportion of total claims paid to date for AY i.
- The fourth column gives $\widehat{C}_{i,J}$, the estimated total claims paid by the end of DY J in respect of AY i.
- The fifth column gives the outstanding claims for AY i, which is calculated by subtracting the second column from the fourth column. Note the use of 'IBNR' for the full outstanding claims reserve. In this note we have reserved IBNR for unreported claims, but it is also used by many practitioners for the full outstanding claims estimate.
- The sixth column shows the AY standard errors. This is is the square root of the MSEP value for each AY, using equation (4.12).
- The seventh column shows the coefficient of variation for each AY, calculated as the ratio of the column 4 and column 5 values.
- The 'Latest' value under the totals header is the total cumulative claims to date for the given accident years, which is the sum of the column 2 values.
- The 'Ultimate' value is estimated total cumulative claims from the given accident years. It is the sum of the column 4 value.

```
MackChainLadder(Triangle = ROT, est.sigma = "Mack")
     Latest  Dev.To.Date  Ultimate      IBNR  Mack.S.E  CV(IBNR)
  1   7,950        1.000     7,950      0.00      0.00       NaN
  2   7,292        1.000     7,295      2.75      3.08     1.120
  3   6,587        0.999     6,596      8.98      5.78     0.643
  4   7,960        0.996     7,991     30.63      7.12     0.232
  5   6,887        0.992     6,943     56.18     16.36     0.291
  6   8,620        0.985     8,754    134.15     35.65     0.266
  7   8,800        0.973     9,047    246.53     47.20     0.191
  8   9,915        0.948    10,462    546.96     63.37     0.116
  9   9,164        0.882    10,386  1,222.18    216.79     0.177
 10   7,014        0.615    11,413  4,399.33    751.44     0.171

     Totals
 Latest:      80,189.00
 Dev:              0.92
 Ultimate:    86,836.69
 IBNR:         6,647.69
 Mack.S.E        802.88
 CV(IBNR):         0.12
```

Table 4.1: R output using `MackChainLadder(ROT,est.sigma=‘‘Mack’’)` `ROT` is the Table 2.1 data triangle.

- The ‘Dev’ value is the ratio of the 'Latest' total to the 'Ultimate' total, representing the estimated proportion of losses already paid.
- the ‘IBNR’ value is the total estimated outstanding claims. this is the difference between the 'Ultimate' and 'Latest' totals, and is also the sum of the column 5 values.
- The ‘Mack S.E.’ is the estimated aggregate standard error of the outstanding claims estimate, including the covariance term.
- The ‘CV(IBNR)’ value is the ratio of the estimated outstanding claims to the estimated standard error of the outstanding claims.

The ChainLadder package will also generate diagnostic plots from fitting data to Mack's model. Figure 4.1 was produced using the command:

```
plot(MackChainLadder(ROT,est.sigma="Mack"))
```

- The top left plot of Figure 4.1 shows the cumulative paid and forecast outstanding claims by AY (where AY 0 is labelled as Origin Period 1), with box and whiskers plots for the outstanding claims. We see that the last two accident years are responsible for almost all of the outstanding claims liability MSEP.

- The top right plot shows the actual and projected cumulative claims paid by development period. In this plot we are looking for changes in the shape of the settlement pattern. As mentioned earlier, there is some evidence of an AY effect, in that the earlier accident years are largely in the lower part of the graph, and the later accident years in the upper part.

The next four plots are all standardized residuals. R calculates the standardized residuals taking into consideration the uncertainty in the parameter estimates.

The fitted cumulative claim values used to calculate the residuals are $\widehat{C}_{i,j+1} = C_{i,j} \times \hat{f}_j$ for $i \in \{0, 1, ..., I-1\}$, $j = 0, ..., I-i-1$ (that is, for the cumulative claims in $\mathcal{D}_I$).

The residuals are $e_{i,j} = C_{i,j+1} - \widehat{C}_{i,j+1}$ (we are implicitly conditioning on $C_{i,j}$).

The variance of $e_{i,j}$ is $\sigma_j^2 \times C_{i,j} \times \left(1 - C_{i,j}/S_j\right)$. The first two terms represent the variance of $C_{i,j+1}|C_{i,j}$, but the residual variance is smaller, as $C_{i,j}$ is used in $\widehat{C}_{i,j+1}$, through the $\hat{f}_j$ term.[1]

The standardized residuals plotted by R are calculated as

$$r_{i,j} = \frac{C_{i,j} - \widehat{C}_{i,j}}{\sqrt{\sigma_{j-1}^2 \, C_{i,j-1} \left(1 - C_{i,j-1}/S_{j-1}\right)}} = \frac{\left(f_{i,j-1} - \hat{f}_{j-1}\right) C_{i,j-1}^{0.5}}{\sigma_{j-1}\sqrt{1 - C_{i,j-1}/S_j}}$$

If the data are consistent with the model the standardized residuals should be independent and approximately N(0,1) distributed.

- The middle left plot shows the residuals against the $\widehat{C}_{i,j}$ values, along with a fitted trend line that indicates higher residuals for smaller values of $\widehat{C}_{i,j}$.

- The middle right plot shows the residuals by AY. The trend is similar to the middle left plot, because the higher $\widehat{C}_{i,j}$ are associated with the later accident years.

- The bottom right plot shows residuals by calendar year. Note that for this plot, we have more data on the right side, representing the most recent calendar years. Again, we see a slight decreasing trend.

[1]The $C_{i,j}/S_j$ term is the **leverage** associated with the $C_{i,j}$ value in the Mack model regression.

- The bottom left plot shows residuals by development year. There is no indication from this plot that settlement patterns have significantly changed over the period of these data.

The decreasing trends of residuals by fitted values, AY and CY suggests that some further investigation of the data might be reasonable.

Exercise 4.6. Figure 4.2 shows the R diagnostic graphs resulting from fitting Mack's model to the long-tail data from Table 2.5. Comment on the major differences between the results for this data, compared with the results for the Table 2.1 data, shown in Figure 4.1.

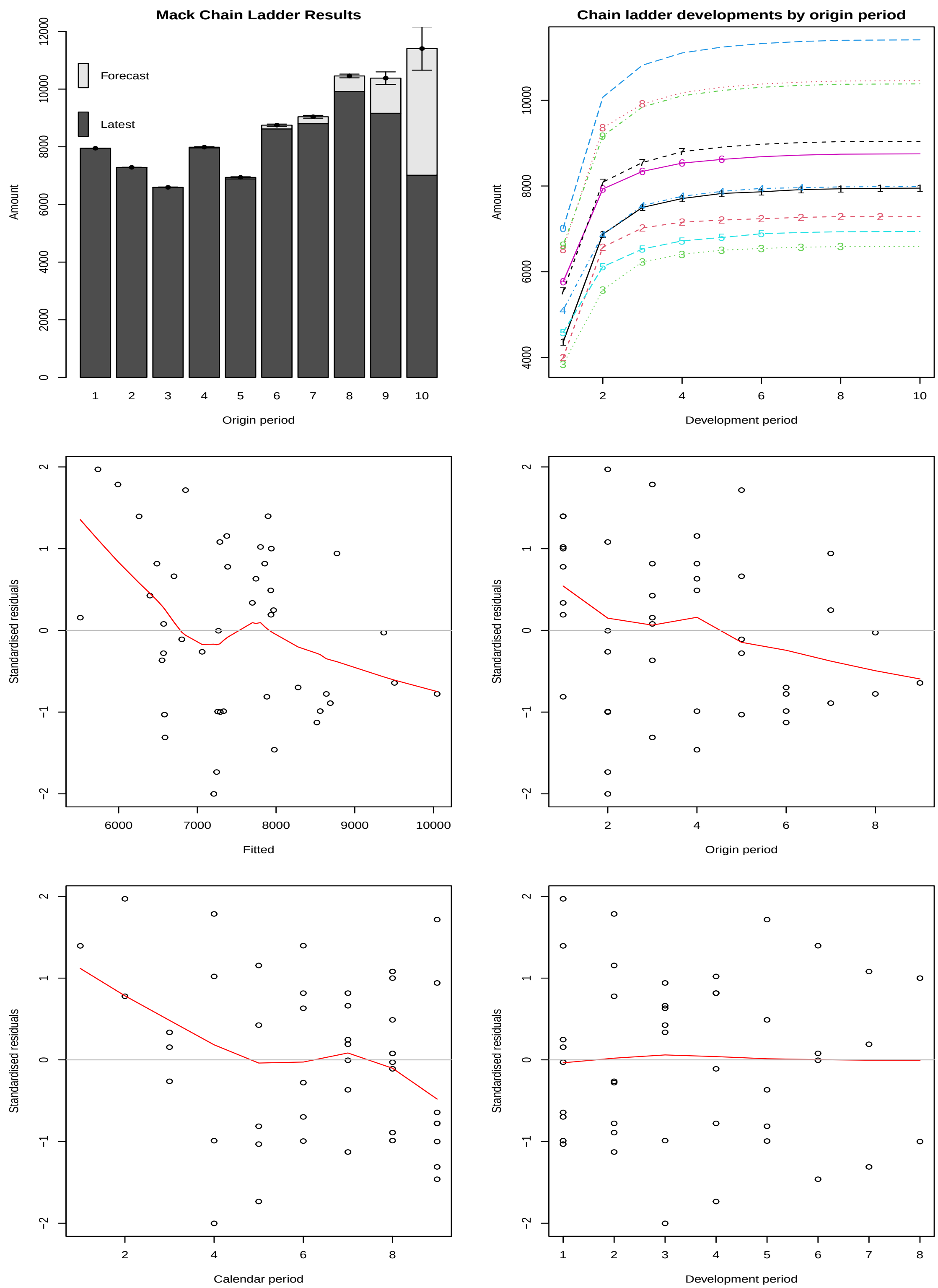

Figure 4.1: Results and diagnostic graphs for the Mack model, applied to the Table 2.1 data.

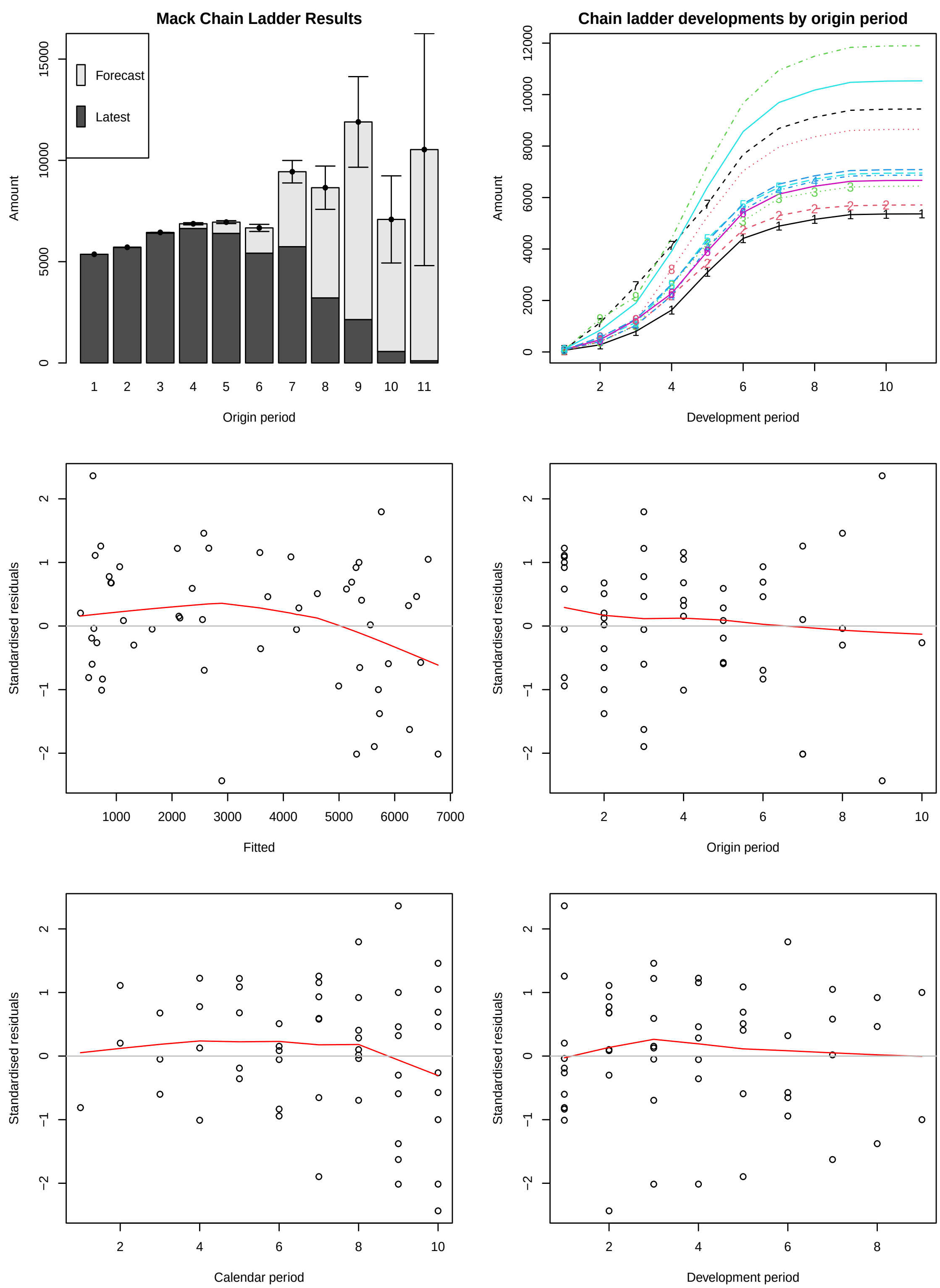

Figure 4.2: Results and diagnostic graphs for the long-tailed data in Table 2.5 data, Mack's model.

Chapter 5

The overdispersed Poisson model

The Mack model is non-parametric, in the sense that we do not assume any distributional models for incremental or cumulative claims. In Chapter 3 we mentioned that both the Poisson and the overdispersed Poisson (ODP) models, with maximum likelihood estimated parameters, are parametric models which give identical values for estimated outstanding claims as the chain ladder method. Given a parametric model that is an adequate fit to the data, we can use the distributional information to estimate standard errors of the outstanding claims projections. Typically, the Poisson model has too small a variance compared with the data, even for claim number triangles, but the overdispersed Poisson is a more viable model that provides us with an alternative to Mack's formulas for estimating standard errors.

The ODP assumptions are

ODP Assumption (1) $X_{i,j}$ are independent random variables for all i, j.

ODP Assumption (2) There exist parameters μ_i, $i = 0, 1, .., I$; γ_j, $j = 0, 1, \cdots, J$, and ϕ such that $X_{i,j} \sim$ ODP, with $\mathrm{E}[X_{i,j}] = \mu_i \, \gamma_j$ and $\mathrm{Var}[X_{i,j}] = \phi \mu_i \gamma_j$.

The μ_i and γ_j parameters here play the same role as in the Poisson model assumptions in Chapter 3. However, it is more convenient to formulate the ODP as a generalized linear model (GLM), with a log-link function, which means that we model the mean as a linear function of the log of the random variable, rather than a multiplicative function of the random variable itself. For the GLM formulation then, we assume there exist parameters:

$$c > 0; \quad \phi > 0; \quad \alpha_i, \, i = 0, 1, \cdots, I, \text{ with } \alpha_0 = 0; \quad \beta_j, \, j = 0, 1, \cdots, J; \text{ with } \beta_0 = 0, \text{ such that}$$
$$\mathrm{E}\left[X_{i,j}\right] = e^{c+\alpha_i+\beta_j}; \qquad \mathrm{Var}\left[X_{i,j}\right] = \phi \, e^{c+\alpha_i+\beta_j}.$$

	Mack Model			OPD		
AY i	OCR	Standard error	CV	OCR	Standard error	CV
1	3	3.08	1.12	3	12.34	4.11
2	9	5.78	1.04	9	19.99	2.22
3	31	7.12	0.33	31	36.08	1.16
4	56	16.36	0.32	56	46.33	0.83
5	134	35.65	0.27	134	72.03	0.54
6	247	47.20	0.24	247	96.42	0.39
7	4547	63.40	0.14	547	144.45	0.26
8	1222	216.79	0.22	1222	218.70	0.18
9	4399	751.44	0.22	4399	490.34	0.11
Totals	6648	802.88	0.12	6648	637.44	0.10

Table 5.1: Mack's model and ODP model compared, for Table 2.1 data. Output from R Chain-Ladder package (Gesmann *et al.*, 2022)

.

The α_0 and β_0 parameters are set to 0 to avoid over-parametrisation.

The technical details of GLM estimation and model fitting are beyond the scope of this note, but we can utilize the ChainLadder package in R (Gesmann *et al.*, 2022) to see the results and diagnostics from fitting the ODP model with maximum likelihood parameter estimates, using the `glmReserve` function.

In Table 5.1 we present the results for estimated outstanding claims (OCR), standard errors by AY, and coefficient of variation (standard error/estimated outstanding claims) using the OPD and Mack model.

As expected, the OCR estimates are identical. The expected value under the ODP model is the same as the expected value under a regular Poisson model, and we noted in Chapter 3 that this perfectly replicated the chain ladder model, given that we have used MLE parameters. The Mack model standard errors are higher overall, but much lower for the early accident years.

One advantage of the parametric approach is that the assumptions and model are well defined, and can be tested. The model fitting process generates estimates for c, α_i and β_j, along with standard errors of the estimates. The results for the adjusted Table 2.1 data are given in Table 5.2. The table makes it clearer that this is a highly parametrized model. We also note that, there is a significant AY effect, particularly for the most recent years. The algorithm also generates the estimated dispersion parameter, $\hat{\phi} = 28.8$, and the null and residual deviance. In our case, the null deviance is 138,709 on 54 degrees of freedom, and the residual deviance is 1,017 on 36

Parameter	Estimate	SE	t-value	p-value	significant?
c	8.4941	0.06235	136.243	<2e-16	***
$\hat{\alpha}_1$	-0.0860	0.08705	-0.988	0.330	
$\hat{\alpha}_2$	-0.1867	0.08946	-2.087	0.044	*
$\hat{\alpha}_3$	0.0051	0.08516	0.060	0.953	
$\hat{\alpha}_4$	-0.1354	0.08843	-1.531	0.134	
$\hat{\alpha}_5$	0.0964	0.08357	1.153	0.257	
$\hat{\alpha}_6$	0.1292	0.08321	1.553	0.129	
$\hat{\alpha}_7$	0.2746	0.08106	3.387	0.002	**
$\hat{\alpha}_8$	0.2673	0.08272	3.231	0.003	**
$\hat{\alpha}_9$	0.3616	0.08942	4.044	3e-4	***
$\hat{\beta}_1$	-0.8307	0.04526	-18.354	<2e-16	***
$\hat{\beta}_2$	-2.2405	0.08641	-25.928	<2e-16	***
$\hat{\beta}_3$	-3.2008	0.14791	-21.641	<2e-16	***
$\hat{\beta}_4$	-3.9421	0.23229	-16.971	<2e-16	***
$\hat{\beta}_5$	-4.4422	0.33072	-13.432	1e-15	***
$\hat{\beta}_6$	-4.9723	0.47763	-10.41	2e-12	***
$\hat{\beta}_7$	-5.5157	0.73160	-7.539	6e-9	***
$\hat{\beta}_8$	-6.4371	1.38683	-4.642	5e-5	***
$\hat{\beta}_9$	-7.3954	3.09995	-2.386	0.022	*

Table 5.2: Parameter estimates for ODP fit of Table 2.1 incremental data. Significance codes: < 0.001, *** ; $(0.001, 0.01]$, **; $(0.01, 0.05]$, *; $(0.5, 0.1]$, $\circ$.

degrees of freedom, which means that the model has accounted for over 99% of the deviance. In Figure 5.1 the QQ plot of the residuals indicates that they are somewhat more disperse than they would be if the model completely explained the systematic variance, especially in the right tail. The lower plot shows the residuals against the fitted claim values – but note that the fitted values are incremental, not cumulative, so this plot cannot be directly compared with the fitted value - residual plot on the previous section, which fitted cumulative, not incremental values. The plot does not indicate any systematic correlation or non-normality in the residuals.

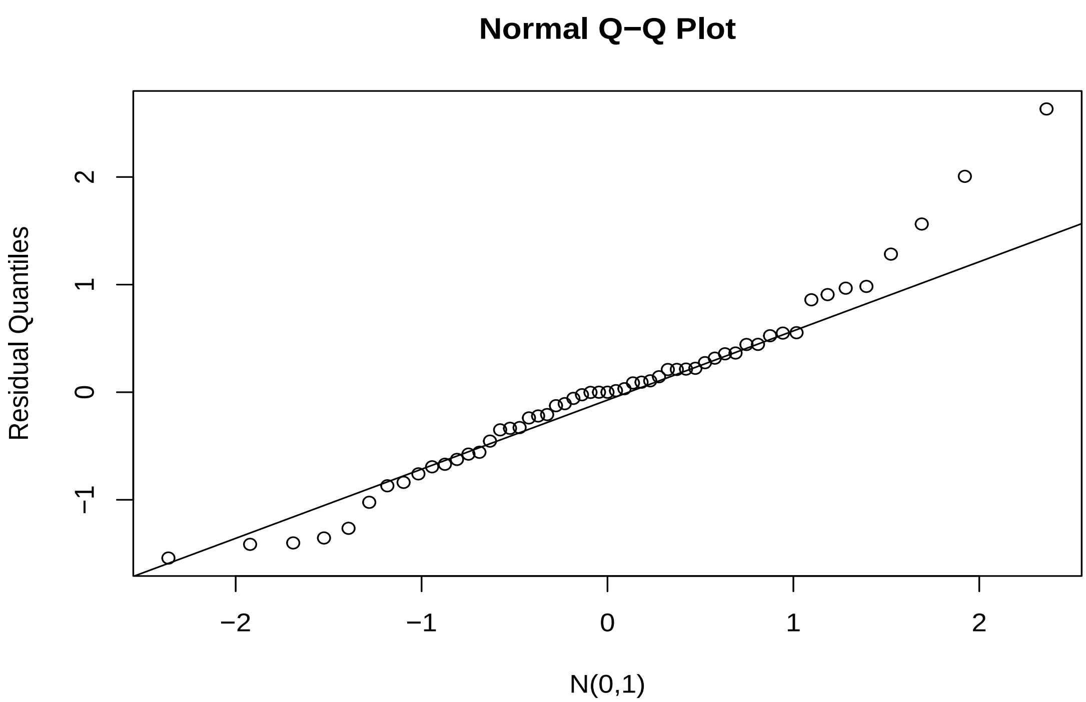

Figure 5.1: QQ plot of residuals vs N(0,1) quantiles, Table 2.1 data, ODP model.

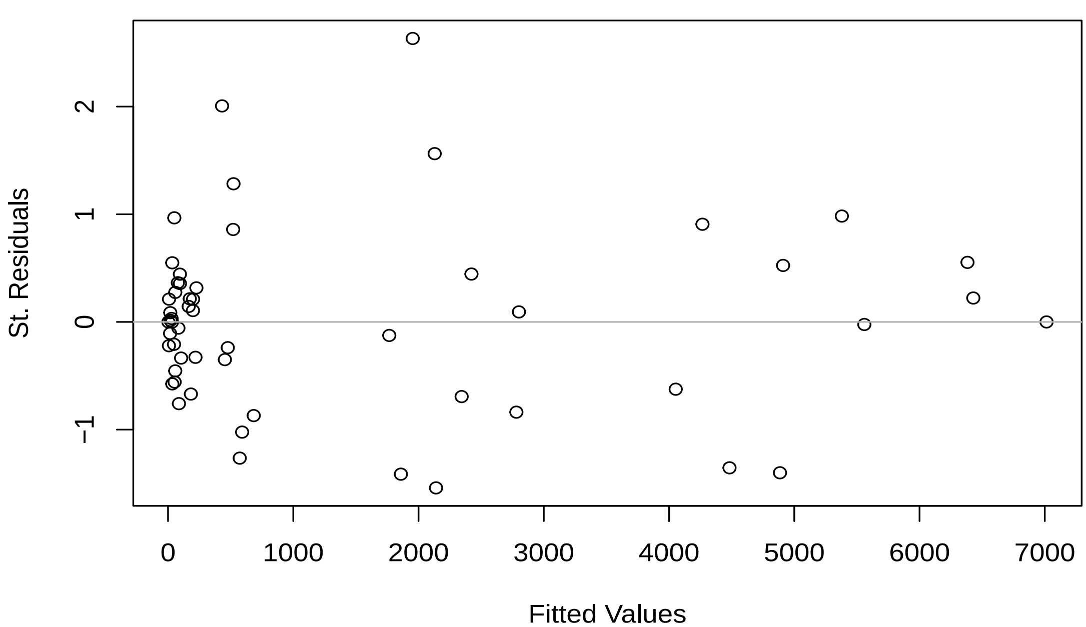

Figure 5.2: Plot of residuals vs fitted values (incremental), Table 2.1 data, ODP model.

Chapter 6

Frequency-severity models

In the previous section, we have used the run-off triangle of aggregate claims to project the ultimate cumulative claims. It is sometimes useful to separate the analysis of outstanding claim counts, from the outstanding claim severity. The separate analyses can reveal trends or outliers that are not apparent from the aggregate table; it may be useful to track the relationship between the settlement delay, and the ultimate cost of a claim. It may also be more straightforward to accommodate changes in claims processing, or in the nature of the underlying risks, using this approach.

The principle of the frequency-severity approach is straightforward, though the severity estimation is often rather ad hoc. While the chain ladder approach is generally quite suited to claim number data, it does not work well when applied to average costs per claim. These numbers are not cumulative, and the chain ladder works best when estimating a cumulative process, with, typically, decreasing increments after some initial period.

We illustrate the methods described in this section with the example short-tailed run-off triangle of cumulative claims shown in Table 6.1. In Table 6.2 we show the run-off triangles for the cumulative reported claim numbers, and settled claim numbers underlying the data in Table 6.1. A claim is deemed to be reported as soon as the insurer is informed of a loss event; it is not settled until the claim is fully paid and closed.

6.1 Claim frequency

The projection of claim numbers usually follows the chain ladder method. If we work with reported claims, then the tail may be quite short, even for some long-tailed business, if the

	DY				
AY	0	1	2	3	4
0	168,830	442,760	1,062,807	1,311,257	1,333,517
1	177,540	436,618	873,088	1,013,083	
2	203,860	499,301	1,027,061		
3	215,988	405,472			
4	191,753				

Table 6.1: Example cumulative claims run-off triangle for frequency-severity section.

	DY, Reported claims					DY, Settled claims				
AY	0	1	2	3	4	0	1	2	3	4
0	315	44	0	0	0	95	150	90	23	1
1	325	30	0	0		97	148	84	25	
2	341	30	1			99	163	76		
3	330	32				97	137			
4	315					83				

Table 6.2: Incremental numbers of claims reported and settled by AY and DY, for aggregate loss data from Table 6.1.

delays are mostly in the processing, rather than the reporting stage. If we work with settled claims, that is, we use the chain ladder to project the number of claims that are fully paid and closed in each development year, then we will have a long tail, and we may have very little data in the early years if the settlement process is prolonged, even if claims are reported promptly.

Projecting both triangles from Table 6.2, using the chain ladder method, gives the estimates in Table 6.3.

AY	DY, Reported claims				
i	0	1	2	3	4
0	315	359	359	359	359
1	325	355	355	355	355
2	341	371	372	372	372
3	330	362	362	362	362
4	315	348	348	348	348

	DY, Settled claims				
AY	0	1	2	3	4
0	95	245	335	358	359
1	97	245	329	354	355
2	99	262	338	362	363
3	97	234	312	334	335
4	83	211	281	301	302

Table 6.3: Cumulative projected numbers of claims reported and settled by AY and DY. Table 6.2 data.

We see that applying the chain ladder to the settled claims data gives projections that are inconsistent with the reported claims projections. For example, for AY 4, the projected number of settled claims by DY 4 is less than the number of reported claims in DY 0, even though we are assuming that all claims are settled by the end of DY 4.

A better approach to projecting the settled claims numbers is to use the settled claim incremental triangle to estimate the proportion of claims settled in each DY, and apply these proportions to the projected total cumulative numbers from the reported claim chain ladder values. The estimated proportion of claims settled in year j is $\hat{\gamma}_j$; we can calculate the $\hat{\gamma}_j$ from the settled claims triangle, as

$$\hat{\gamma}_0^s = \hat{\beta}_0^s; \quad \hat{\gamma}_j^s = \hat{\beta}_j^s - \hat{\beta}_{j-1}^s \quad j = 1, 2, 3, ..., J$$

Let $\widehat{C}_{i,J}^r$ denote the cumulative ultimate reported claims for AY i, and let $\widetilde{X}_{i,j}^s$ denote the estimated incremental settled claims for $i + j > I$. Then we can estimate $X_{i,j}^s$ by combining the γ_j^s with the $\widehat{C}_{i,J}^r$, as

$$\widetilde{X}_{i,j}^s = \hat{\gamma}_j^s \, \widehat{C}_{i,j}^r$$

The resulting cumulative settled claims triangle is given in Table 6.4.

AY	DY, Settled claims				
i	0	1	2	3	4
0	95	245	335	358	359
1	97	245	329	354	355
2	99	262	338	363	364
3	97	234	346	371	372
4	83	243	324	347	348

Table 6.4: Cumulative projected numbers of claims settled, using reported ultimate claims numbers, and $\hat{\gamma}_j$ from the settled claim run-off triangle; Table 6.2 data.

6.2 Claim severity

The simplest approach to estimating claim severity is to calculate cumulative average severity for each accident year by dividing the cumulative claims by the projected cumulative claims settled. In Table 6.5 we show the cumulative averages, corresponding to the claim frequency tables above, using the Table 6.1 data above.

This does not provide much information. The average severity triangle does not generate well-behaved development factors, and there is no reason to suppose that the chain ladder assumptions apply. However, the use of the chain ladder as a heuristic is still quite common. The resulting projected average severity by accident year, and estimated OCR, are given in Table 6.6.

	Average cost per settled claim				
AY	0	1	2	3	4
0	1777.2	1807.2	3172.6	3662.7	3714.5
1	1830.3	1782.1	2653.8	2861.8	
2	2059.2	1905.7	3038.6		
3	2226.7	1732.8			
4	2310.3				

Table 6.5: Average severity per cumulative claim settled, using Table 6.1 claims data and Table 6.2 claim number data.

A more appropriate approach is to use the average severity by development year of settlement, because in many lines of business, longer delays signal higher expected severity. Long delays are associated with more complex claim underwriting and processing, caused, for example, by

AY	Average Severity	Projected Claim numbers	Freq × Severity	$\hat{R}_i$
0	3714.5	359	1,333,517	0
1	2902.3	355	1,030,313	17,231
2	3450.9	372	1,283,742	256,681
3	3174.7	362	1,150,314	744,841
4	3875.9	348	1,348,801	1,157,049
Total				2,175,801

Table 6.6: Estimated OCR; frequency-severity approach; chain ladder projected average severity; Table 6.1 and 6.5 data.

uncertain medical outcomes in cases of severe injury, or by protracted legal proceedings, each of which signals a higher expected claim severity.

We can explore the claim severity by settlement year using the incremental settled claim numbers from the claim frequency analysis. In Table 6.7 the ratio of the incremental payments in DY j to the number of claims settled in DY j. the incremental claim amounts are determined by subtracting successive cumulative claim values in Table 6.1, and the number of claims settled in each year is obtained by subtracting successive settled claim numbers from Table 6.3.

	Ave incremental cost per settled claim				
AY	0	1	2	3	4
0	1,777	1,826	6,889	10,802	22,260
1	1,830	1,751	5,196	5,600	
2	2,059	1,813	6,944		
3	2,227	1,383			
4	2,310				
Average	2,041	1,693	6,343	8,201	22,260

Table 6.7: Average severity per incremental claim settled; Table 6.1 and 6.3 data.

Table 6.7 shows much more clearly the impact of settlement delay on the average claims, compared with Table 6.5. We can now combine the average severity for each DY with the projected settled claim counts by DY, from the claim frequency analysis, to estimate the outstanding claims reserve. Some intermediate calculations are given in Table 6.8. We see that, although this method has captured the increasing claim size by development year of settlement, it relies

DY	Average Severity	Outstanding Claim numbers	$\hat{R}_i$
0	2,041	0	0
1	1,693	128	216,584
2	6,343	148	938,798
3	8,201	67	549,466
4	22,260	4	89,040
Total			1,793,888

Table 6.8: Average severity and incremental frequency of settled claims by DY. Table 6.2 and 6.5 data.

quite heavily on the single entry for claim numbers in AY 4, and the single entry for claim size in DY 4.

In Table 6.9 we compare results from the two approaches described here, and from the standard chain ladder method. Column (1) uses the chain ladder estimates of cumulative average severity, and chain ladder estimates of claim numbers, using the reported claims triangle. Column (2) uses the incremental average severity by development year of settlement, together with the same projected claim numbers as column (1), and column (3) uses the aggregate chain ladder approach. The incremental average severity approach can be useful, particularly if average severities are changing, for example, through changing claims inflation expectations.

AY	Chain ladder of aggregate severity (1)	Average severity by settlement DY (2)	Chain ladder, aggregate claims (3)
0	0	0	0
1	17,231	22,260	17,198
2	256,681	219,084	227,018
3	744,841	705,655	658,333
4	1,157,049	846,890	979,933
Total	2,175,801	1,793,888	1,882,701

Table 6.9: Estimated OCR, using two different frequency-severity methods, and using the aggregate chain ladder method . Table 6.2 and 6.5 data.

Bibliography

England PD, Verrall RJ (2002). "Stochastic Claims Reserving in General Insurance." *British Actuarial Journal*, **8**(3), 443–518.

Enz R, Holzheu T (2004). "The economics of liability losses – insuring the moving target." *Sigma*, **6**.

Gesmann M, Murphy D, Zhang W, Carrato A, Wüthrich M, Concina F, Dal Moro E (2022). "Claims reserving with R: Chainladder-0.2.16 package vignette." *The Comprehensive R Archive Network*.

Hindley D (2017). *Claims Reserving in General Insurance*. Cambridge University Press.

Klugman SA, Panjer HH, Willmot GE (2019). *Loss Models: from Data to Decisions*. 5th edition. John Wiley & Sons.

Mack T (1993). "Distribution-free calculation of the standard error of chain ladder reserve estimates." *ASTIN Bulletin: The Journal of the IAA*, **23**(2), 213–225.

Mack T (1994). "Measuring the variability of chain ladder reserve estimates." In *Casualty Actuarial Society Forum*.

Wüthrich MV, Merz M (2008). *Stochastic Claims Reserving Methods in Insurance*. John Wiley & Sons.

Appendix A

Proof of Theorem 3

$$\begin{aligned}
\text{Let } T_j &= \hat{f}_{I-i}\,\hat{f}_{I-i+1}...\hat{f}_{j-1}\left(f_j - \hat{f}_j\right) f_{j+1}\,f_{j+2}...f_{J-1} \\
&= \hat{f}_{I-i}\,\hat{f}_{I-i+1}...\hat{f}_{j-1}\,f_j\,f_{j+1}...\,f_{J-1} - \hat{f}_{I-i}\,\hat{f}_{I-i+1}...\hat{f}_{j-1}\,\hat{f}_j\,f_{j+1}...\,f_{J-1}
\end{aligned}$$

If we sum the T_j, we find successive terms cancel out, leaving the first part of T_{I-i} and the last part of T_{J-1}, i.e.

$$\begin{aligned}
&\sum_{j=I-i}^{J-1} T_j = f_{I-i}\,f_{I-i+1}\,...\,f_{J-1} - \hat{f}_{I-i}\,\hat{f}_{I-i+1}\,...\,\hat{f}_{J-1} \\
&\implies C_{i,I-i}\left(\sum_{j=I-i}^{J-1} T_j\right) = C_{i,I-i}\,f_{I-i}\,f_{I-i+1}\,...\,f_{J-1} - C_{i,I-i}\,\hat{f}_{I-i}\,\hat{f}_{I-i+1}\,...\,\hat{f}_{J-1} \\
&\qquad\qquad\qquad\qquad\quad = \widehat{C}_{i,J} - \mathrm{E}\left[C_{i,J}|\mathcal{D}_I\right] \\
&\implies C_{i,I-i}^2\left(\sum_{j=I-i}^{J-1} T_j\right)^2 = \left(\widehat{C}_{i,J} - \mathrm{E}\left[C_{i,J}|\mathcal{D}_I\right]\right)^2.
\end{aligned}$$

$$\begin{aligned}
\text{Now}\quad &\left(\sum_{j=I-i}^{J-1} T_j\right)^2 = \sum_{j=I-i}^{J-1} T_j^2 + 2\sum_{j=I-i}^{J-2}\;\sum_{k=j+1}^{J-1-j} T_j\,T_k \\
&\implies \left(\widehat{C}_{i,J} - \mathrm{E}\left[C_{i,J}|\mathcal{D}_I\right]\right)^2 = C_{i,I-i}^2\left(\sum_{j=I-i}^{J-1} T_j^2 + 2\sum_{j=I-i}^{J-2}\;\sum_{k=j+1}^{J-1-j} T_j\,T_k\right)
\end{aligned}$$

At this point, we need to approximate T_j^2 for $j = I-i, I-i+1, ..., J-1$, and $T_j\,T_k$, for j, k such that $I - i \leq j < k$ and $j + k \leq J - 1$. We do this in each case by taking conditional expected values.

- For T_j, T_k, $k \geq j+1$, we condition on $\mathcal{B}_k$. Note that because $j < k$, $\mathcal{B}_k$ contains all the $C_{i,r}$ used in T_j, so $\mathrm{E}[T_j|\mathcal{B}_k] = T_j$. It also contains all the $C_{i,r}$ used in T_k, except for the terms in $C_{i,k+1}$ used in the numerator of $\hat{f}_k$. This means that $\hat{f}_k$ is the only random variable in the conditional expectation $T_j\, T_k \big| \mathcal{B}_k$, so that

$$\mathrm{E}\left[T_j\, T_k \big| \mathcal{B}_k\right] = T_j \mathrm{E}\left[T_k \big| \mathcal{B}_k\right] \propto \mathrm{E}\left[(f_k - \hat{f}_k)|\mathcal{B}_k\right]$$

From Theorem 1 we know that $\hat{f}_k \big| \mathcal{B}_k$ is an unbiased estimator of f_k, so

$$\mathrm{E}\left[(f_k - \hat{f}_k)|\mathcal{B}_k\right] = 0 \implies \mathrm{E}\left[T_j\, T_k \big| \mathcal{B}_k\right] = 0$$

- For T_j^2 we condition on $\mathcal{B}_j$, which means that $\hat{f}_j$ is the only random variable in the conditional expectation $\mathrm{E}[T_j^2 \big| \mathcal{B}_j]$, i.e.

$$\begin{aligned}
\mathrm{E}\left[T_j^2 \big| \mathcal{B}_j\right] &= \left(\hat{f}_{I-i}\, \hat{f}_{I-i+1} \ldots \hat{f}_{j-1}\right)^2 \mathrm{E}\left[(f_j - \hat{f}_j)^2 \big| \mathcal{B}_j\right] \left(f_{j+1}\, f_{j+2} \ldots f_{J-1}\right)^2 \\
&= \left(\hat{f}_{I-i}\, \hat{f}_{I-i+1} \ldots \hat{f}_{j-1}\right)^2 \mathrm{Var}\left[\hat{f}_j \big| \mathcal{B}_j\right] \left(f_{j+1}\, f_{j+2} \ldots f_{J-1}\right)^2
\end{aligned}$$

because $\hat{f}_1, \ldots, \hat{f}_{j-1}$ are all known when conditioning on $\mathcal{B}_j$, and $f_{j+1}\, f_{j+2} \ldots f_{J-1}$ are all unknown parameters, not random variables. Note also that $S_j|\mathcal{B}_j$ is not a random variable, so

$$\begin{aligned}
\mathrm{Var}\left[\hat{f}_j \big| \mathcal{B}_j\right] &= \mathrm{Var}\left[\left.\frac{\sum_{l=0}^{I-1-j} C_{l,j+1}}{\sum_{l=0}^{I-1-j} C_{l,j}}\right| \mathcal{B}_j\right] = \mathrm{Var}\left[\left.\frac{\sum_{l=0}^{I-1-j} C_{l,j+1}}{S_j}\right| \mathcal{B}_j\right] \\
&= \frac{1}{S_j^2} \sum_{l=0}^{I-1-j} C_{l,j}\, \sigma_j^2 \quad = \quad \frac{\sigma_j^2}{S_j} \\
\implies \mathrm{E}\left[T_j^2 \big| \mathcal{B}_j\right] &= \frac{\sigma_j^2 \left(\hat{f}_{I-i}\, \hat{f}_{I-i+1} \ldots \hat{f}_{j-1}\, f_{j+1} \ldots f_{J-1}\right)^2}{S_j} \\
&= \frac{\sigma_j^2 \left(\hat{f}_{I-i}\, \hat{f}_{I-i+1} \ldots \hat{f}_{j-1}\, f_j\, f_{j+1} \ldots f_{J-1}\right)^2}{f_j^2\, S_j}
\end{aligned}$$

If we approximate f_k by $\hat{f}_k$, and σ_j^2 by $\hat{\sigma}_j^2$, we have

$$
\begin{aligned}
C_{i,I-i}^2 \, T_j^2 &\approx C_{i,I-i}^2 \left(\hat{f}_{I-i} \, \hat{f}_{I-i+1} \dots \hat{f}_{J-1} \right)^2 \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \frac{1}{S_j} \\
&\approx \frac{\widehat{C}_{i,J}^2 \, \hat{\sigma}_j^2}{\hat{f}_j^2 \, S_j}
\end{aligned}
$$

Summing over the development years from $j = I-i$ to $j = J-1$ gives the approximate estimation error for AY i as

$$
\left(\widehat{C}_{i,J} - \mathrm{E}\left[C_{i,J} | \mathcal{D}_I \right] \right)^2 \approx \widehat{C}_{i,J}^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2 \, S_j} \tag{A.1}
$$