DRAFT
09.04.14

Outstanding Claims in General Insurance

Walther Neuhaus

April 2014

ABSTRACT Estimation of outstanding claims is an essential part of actuarial work in general insurance. This book is an introduction to delays involved in claim settlement, mathematical models of the claim settlement process, and methods that actuaries use to estimate outstanding claims. It provides a coherent modelling framework that is illustrated by worked examples and exercises.

Contents

I have crossed out all sections that have not been taught in class and that will not be examined.

1 Introduction **1**
1.1 Estimation of outstanding claims 1
1.2 Notes 6

2 Claim development in general insurance **7**
2.1 Introduction 7
2.2 General insurance contracts 7
2.3 Claim attachment principles 9
2.4 Stages in the life of a claim 10
2.5 The length of the tail 12
2.6 Valuation of outstanding claims 13
2.6.1 Purpose of valuation 13
2.6.2 Accounting 13
2.6.3 Pricing 14
2.6.4 Statutory reporting 14
2.6.5 Sale or purchase of a portfolio 15
2.6.6 Commutation 15
2.7 Case estimates 15
2.8 Nominal and discounted values 16
2.9 Time dimensions 17
2.9.1 Underwriting date 18
2.9.2 Accident date 18
2.9.3 Reporting date 18
2.9.4 Valuation date 18

2.9.5 Development date 18
2.9.6 Acronyms 19

3 Framework model **21**
3.1 Introduction 21
~~3.2 Continuous time model . . . . 21~~
3.3 Discrete time model with three dimensions 24
3.4 Discrete time model with two dimensions 26

4 The number of unreported claims **29**
4.1 Introduction 29
4.2 Modelling the number of claims IBNR 29
4.3 Constant claim frequency: The Bornhuetter-Ferguson method 30
4.4 Varying claim frequency: The chain-ladder method 32
~~4.5 Bootstrapping in fixed-parameter models . . . . 35~~
~~4.6 Varying claim frequency: A Bayesian model . . . . 36~~
~~4.6.1 The model . . . . 36~~
~~4.6.2 Parameter estimation . . . . 38~~
4.7 Varying claim frequency: The Bühlmann-Straub model . . . 39
4.7.1 The model 39
4.7.2 Parameter estimation 41
~~4.8 Varying claim frequency: A random walk model . . . . 43~~
~~4.9 Varying claim frequency: A general credibility model . . . . 46~~
4.10 The number of claims CBNI 48
4.11 Measures of risk exposure 49
4.12 Tail development factors 49

5 The cost of unreported claims **51**
5.1 Introduction 51
5.2 The cost of claims IBNR 51
5.3 The cost of claims CBNI 52
~~5.4 Estimating the severity distribution . . . . 53~~

~~**6 The cost of reported claims** **55**~~
~~6.1 Introduction . . . . 55~~
~~6.2 Estimating claims RBNS by payment data . . . . 55~~
~~6.2.1 Background . . . . 55~~
~~6.2.2 The Dirichlet distribution . . . . 56~~
~~6.2.3 Model of payment pattern . . . . 57~~
~~6.2.4 Prediction of ultimate claims . . . . 57~~
~~6.2.5 Estimation of parameters . . . . 60~~
~~6.3 Estimating claims RBNS by incurred data . . . . 61~~
~~6.3.1 Background . . . . 61~~
~~6.3.2 Compound poisson model of reporting events . . . . 62~~

6.3.3 Prediction of ultimate claims 63
6.3.4 Estimation of parameters 65
6.4 Actuarial case estimates 66
6.5 Markov chain models 67

7 Two-dimensional models 69
7.1 Introduction 69
7.2 Generic notation 70
7.3 The Bornhuetter-Ferguson method 70
7.4 The Chain-ladder method 71
7.5 Bühlmann-Straub's model 71
7.6 A model with random delay probabilities 74
7.6.1 The Hesselager-Witting model 74
7.6.2 Parameter estimation 77
7.7 Benktander´s method 79
7.8 Mack's model 80
7.9 Hertig's model 85

8 Log-linear models 89
8.1 Introduction 89
8.2 Log-linear models 89
8.3 Bootstrapping 96

9 General linear models 101
9.1 Introduction 101
9.2 Exponential family of distributions 102
9.3 Applications 104
9.4 Bootstrapping the GLM 104
9.5 Notes 105

10 Dynamic linear models 107

11 Miscellaneous topics 109
11.1 Introduction 109
11.2 Simple model diagnostics 109
11.3 Analysis of development 110
11.4 The NP approximation 110
11.5 The mean squared error of the one-period run-off result . . 112
11.6 When is a credibility estimator worse than the apriori mean? 115

12 Inflation and discounting 117
12.1 Introduction 117
12.2 Inflation and discounting 117
12.3 Estimating inflation by the separation method 118

13 Reinsurance recoveries **121**

14 Data requirements **123**
14.1 Introduction 123

15 Appendix. Credibility estimation in the regression case **127**

16 Literature **131**

Preface

This text is based on lecture notes for a course on Loss Reserving, that the author has held at the Universidade Técnica de Lisboa (UTL) since 2002. It is the second text on Loss Reserving to emanate from the Master in Actuarial Science Program at UTL, the first being the monograph by Taylor (2000).

In its present form the text is still incomplete inasfar as numerical examples and exercises are lacking. Examples and exercises will be given separately during the course.

1 Introduction

1.1 Estimation of outstanding claims

Estimation of outstanding claims, also known as loss reserving, is an essential part of actuarial work in general insurance. In almost every actuarial task the actuary needs to address the question: have outstanding claims been taken into account? Neither the underwriting nor the accounting function of an insurance company can be performed reliably, unless the company has a good idea about the likely cost of its outstanding claims at any point in time.

Estimation of outstanding claims is a challenging and an interesting task. It involves the analysis of past data that is often sparse and usually heterogeneous, prediction of future outcomes over many years, and an assessment of the effect of changes that have occurred in the past or may occur in the future.

Actuaries have devised several heuristic methods to assist them in estimating outstanding claims. The most commonly-used methods are known as the *Chain-ladder method* and the *Bornhuetter-Ferguson method*. Both methods can be derived using rudimentary model assumptions which, in a certain sense, are diametrical opposites of each other. This does not diminish the practical usefulness and popularity of those methods.

Actuaries are facing increasing demands in the modern financial environment, including

1. to be able to estimate outstanding claims reliably and consistently over time,

2. to be able to explain, discuss and modify model assumptions, and
3. to be able to illustrate predictive uncertainty with a quantitative statement.

Several lines of research are currently *en vogue* in the actuarial profession. For want of better terms, I will refer to these lines of research as "fitting to method", "fitting to data" and "stochastic micro-modelling". Before explaining what I mean, I will very briefly define a few terms.

- Mechanism: the totality of internal and external processes that generate claim development.
- Model: a simplified mathematical description of the claim development mechanism.
- Method: an algorithm for turning observed data into projections of future data.

Fitting to method

A number of academics and actuaries are attempting to analyse the statistical properties of the heuristic methods, most often the Chain-ladder method. The seminal paper is Mack (1993), see also England & Verrall (2002) and Wüthrich & Merz (2008) for comprehensive descriptions. This line of research involves finding a model within which a given method is optimal or at least justifiable, for example, because its predictions coincide with maximum likelihood estimates. Thereafter the statistical properties of the method are computed within constraints of that model. As a result, the actuary will be able to produce an estimate of predictive uncertainty.

Fitting to data

Other authors fit models not to methods, but to data. An extensive treatment can be found in Taylor (2000). This approach typically involves definition of a stochastic model with a finite number of fixed, unknown model parameters. Statistical software is used to calculate parameter estimates and their standard errors. To calculate predictive uncertainty, bootstrap methodology is often proposed.

The main difference between fitting to method and fitting to data is that in the former approach the method being studied puts à priori constraints on the admissible models, while in the latter approach the model is built with the objective of capturing important aspects of the mechanism that underlies claim development.

Stochastic micro-modelling

Several authors, most prominently Arjas (1989) and Norberg (1986, 1993, 1999), are advocating a stochastic micro-modelling approach. According to their prescription, the actuary should start with an *à priori* description of the stochastic nature of the mechanisms that generate the claim process. Having modelled the important features of the claim process, calibrated its parameters to the available data, and decided on an optimality criterion, the actuary can then derive the estimation method that is optimal to his criterion. He or she will also be able to assess the performance of methods that, even if not optimal with respect to the chosen criterion, may be legitimate alternatives.

The paper of Arjas (1989) does not provide operational models, but states as a general principle that the estimation of outstanding claims should utilise the conditional distribution of future claim development, given observed claim development. Norberg (1986, 1993, 1999) is more specific and suggests models that allow estimation by linear Bayes methods, known to actuaries as Credibility methods.

In the humble opinion of this author, fitting to method is putting the cart before the horse. The real-world meaning of the uncertainty remains hazy because normally no attempt is made to verify that the model that produces a method, also reflects the mechanism that generates the data. The assumptions needed to produce (in particular) the Chain-ladder method and its variations, seem to provide a poor description of the claim development mechanismn. In any case, the need to make certain model assumptions for the sake of the method is not conducive to thinking about the properties of the claim development mechanism.

Properly calibrated, a statistical model can provide a meaningful quantification of predictive uncertainty. However, it also suffers drawbacks. The ability to fit a statistical model that produces consistent and reliable results over time requires a good volume of reasonably well-behaved data. To many actuaries that situation is an exception rather than the rule. Communication of model assumptions may also be difficult if modelling requires transformation of the data or the parameters. An old adage says that "insurance executives cannot think in log space".

In the view of this author, estimation of outstanding claims fits squarely into the decision-theoretic framwork formulated by Wald (1950), see also Ferguson (1967). The actuary is confronted with an uncertain "state of nature" - being the claim development mechanism that is only imperfectly understood - and must make a decision, i.e. provide an estimate. The precision of the estimate will depend on both the estimation method used and the claim development mechanism. Bayesian and linear Bayesian estimation procedures are a natural part of the decision-theoretic toolbox.

Coming back to the main theme, actuarial students face a multifarious array of approaches when they study estimation of outstanding claims. The academic proclivity to publish papers not because something essential has been recognised, but because publication is essential to being recognised, only adds to the cacophony. To theoretically-minded students the lack of a coherent and credible model framework can be a turn-off from the subject.

Less theoretically-minded students may never get to see the forest because logs are blocking their view. It is symptomatic of the state of the art that the General Insurance Reserving Issues Task Force of the Institute of Actuaries in a recent report (GRIT, 2006) both recommends greater transparency of reserving methods, and rejects calls for more sophisticated methods. In the opinion of this author, while transparency undoubtedly is a desirable property, to achieve transparency at the expense of methodological development would lead to a dead-end.

This book is an introduction to mechanisms involved in claim settlement, modelling of the claim settlement process, and to methods that actuaries can use to estimate outstanding claims. Its main focus is on stochastic micro-modelling.

Estimation of outstanding claims is not exclusively a statistical discipline. In most practical situations, the actuary's estimates are by necessity based on a combination of statistical estimation and actuarial judgement. There is nothing wrong with such a compromise. Quite on the contrary, the necessity of applying judgement makes estimation of outstanding claims a more interesting exercise, than it would be if it amounted to pure number-crunching. The mathematical framework can be likened to the white cane used by a blind person: the cane is an indispensible aid while the person is finding his orientation in an unfamiliar environment, but once the environment has become familiar, the person can move around more freely.

Chapter 2 describes the main aspects of claim development in general insurance, and introduces the reader to some actuarial terminology in speaking of outstanding claims.Chapter 3 outlines the model framework of Arjas (1989) and Norberg (1986) in preparation for the more detailed modelling in Chapters 4 (estimating the number of unreported claims), 5 (estimating the cost of unreported claims) and 6 (estimating the cost of reported claims). The common theme of chapters 4 to 6 is what Norberg (1986) calls a stochastic micro-theory.

Chapter 7 presents some of the more summary models of claim development that are commonly used, most of which can be viewed as applications of models presented in chapter 4.

Chapters 8 and 9 are an excursion into the statistical model-fitting approach. Chapter 8 treats a log-linear model, while Chapter 9 gives an introduction to generalised linear models (GLM). Chapter 10 gives a brief introduction to dynamic linear models (DLM).

Chapter 11 takes up some practical issues, including model diagnostics, the comparison between actual and expected claim development, and the philosophical question of what estimation methods one is justified in using in an imperfect world.

Chapter 12 deals briefly with the technicalities of inflation adjustment and discounting. The more practical question of what inflation and discount rates one should use, is not addressed in this book.

Chapter 13 addresses quite briefly the conversion of estimated gross liabilities (i.e., before taking reinsurance into account) to estimated net liabilities (after reinsurance). No attempt at an exhaustive treatment is made in this book, as most reinsurance contracts have very specific clauses, frustrating attempts at a unified theoretical treatment.

Chapter 14 is devoted to the important issue of data. Many actuaries have to make do with data that is far from perfect for the purpose of estimating outstanding claims. In the opinion of this author, well-organised and disciplined data collection is an indispensable prerequisite for the estimation of outstanding claims.

A few words about the order of appearance in chapters 4-7.Chapters 4-6 present an outline of a stochastic micro-model in the spirit of Arjas (1989) and Norberg (1986, 1993, 1999). It is only an outline because, as will become apparent, the details of the model can be varied. Chapter 7 represents a few methods that are tradionally being taught and used - their main characteristic being that the two main sources of delay (reporting delay and settlement delay) are merged into just one aggregate delay pattern. One could ask, would it not be better to start teaching the traditional methods before advancing to more sophisticated ones? I have chosen the opposite approach for two reasons:

- From my own time as a student, I remember vividly that what I remember best of most courses, was what I learnt first. It provided the scaffolding for what came after, so to speak.
- The chosen approach makes it possible to present a sound theoretical derivation of the two classical methods (Chain-ladder method and Bornhuetter-Ferguson method). Both methods provide maximum likelihood estimates in a very intuitive Poisson model of claim counts. There is no need to appeal to heuristics or strange conditional distributions to motivate those methods, if they are used in the appropriate context.

For the teacher wanting to teach the traditional approaches first, I have kept Chapter 7 free of references to chapters 4-6. For the sceptic I would like to mention that the model framework propounded in chapters 4-6 has been used continuously for more than 15 years by a major Nordic insurer.

It is common for papers on loss reserving to be full of ritual apologies for the inadequacy of models, as opposed to so-called expert knowledge. Let me therefore say this just once:

- All models are wrong, but some are useful.
- Actuaries use models because they are not clairvoyants.
- Most of our predictions will miss the mark, but hopefully by less than if we hadn't built a model.

As will be see, moreover, stochastic micro-models allow the incorporation of expert opinion in a natural way.

1.2 Notes

There exists an extensive literature on estimation of outstanding claims, or loss reserving. A good source of references is Schmidt (2011). The interested reader will also find a great number or references Wüthrich & Merz (2008). Another standard textbook is Taylor (2000).

2
Claim development in general insurance

2.1 Introduction

This chapter describes the main aspects of delayed claim development in general insurance, and introduces the reader to the common actuarial terminology in speaking of outstanding claims.

2.2 General insurance contracts

A general insurance contract is normally valid for a period of one year. The contract stipulates that the insured may claim compensation from the insurer, for financial losses incurred during the contract period. The right to compensation depends on the fulfilment of contractual conditions. Among such conditions are:

- The cause of the loss. This could be fire, burglary, accident, disease, an act of negligence, and so on.
- The type of the loss. This could be fire damage, loss of property, loss of income, medical expenses, liability for the loss incurred by a third party, and so on.

If a right to compensation has been established, the compensation will normally be limited by a deductible and an upper limit. Only in rare cases do insurance contracts provide for full compensation.

Property insurance compensates losses incurred by the insured himself. Liability insurance compensates the insured for a financial liability that the insured has incurred towards a third party. In both cases the contractual relation is limited to the two parties, the insurer and the insured.

Statutory liability insurances, like Motor Vehicle Third Party insurance and Workers' Compensation insurance, cover losses incurred by a third party. In those insurances, the third party can lodge a direct claim against the insurer, thereby bypassing the insured.

General insurance is customarily subdivided into lines of business. Each of these lines of business has its own characteristics, which may differ between countries. A list of the most common lines is given below, together with an indication of what each type of insurance typically would cover.

- Domestic property insurance compensates a private home owner for losses due to fire, water damage and burglary.
- Commercial property insurance compensates a commercial property owner for losses due to fire, water damage and burglary, as well as the consequential loss of rent income.
- Commercial business insurance compensates a business for losses due to fire, water damage, burglary and fraud by employees, as well as consequential loss of profits and recovery of business archives.
- Motor vehicle insurance compensates a car owner for losses from collision damage, burglary/theft and fire.
- Motor vehicle third party insurance compensates a third party for personal injuries or property damage that have been caused by the insured's use of a motor vehicle.
- Travel insurance covers travellers for loss of property, hospital expenses while abroad, repatriation, accidental injury or death, cancellation.
- Liability insurance compensates the insured for a financial liability that the insured has incurred towards a third party.
- Accident insurance pays compensation on the insured's accidental injury or death.
- Health insurance pays doctors' fees, hospital fees and medical expenses.
- Disability insurance compensates loss of income arising out of a disability that can be the result of an accident or illness.
- Workers' compensation insurance compensates employed workers for accidents or diseases contracted at the workplace.

- Marine insurance covers damage to the vessel (hull insurance) and/or the liability of the owner or operator (P&I insurance).
- Aviation insurance covers damage to the aircraft (hull insurance) and/or the liability of the owner or operator.
- Transport insurance covers damage to or loss of goods in transport.
- Reinsurance covers an insurer against exposures that it cannot bear in its own balance sheet.

One should not attach too much importance to the label that is given to a product of insurance. The content of the product can vary significantly between countries, and even between insurers within a country. In particular, insurance in personal lines like health insurance and workers' compensation insurance, will depend on the legislation and the social security framework of the country. Therefor, lumping together products of insurance on the basis of only their name can be fraught with danger.

2.3 Claim attachment principles

When a claim is received, the insurer must check whether that claim can rightfully be attached to an insurance policy (contract) that the insurer has issued to the insured. The time of the event leading to the loss, as well as the time when the claim is reported, may determine whether the claim can be attached to a policy.

Claims incurred principle

Under the *Claims incurred principle*, a loss is covered if the event leading to the loss happened during the policy period. The claim itself may be lodged even when the policy period has expired. The Claims incurred principle is common in lines of business where the event leading to the loss can be identified (property insurance, business insurance, motor vehicle insurance, accident insurance, marine insurance).

Claims made principle

Under the *Claims made principle*, a loss is covered only if the claim is received while the policy is in force. There may be a limited time extension after the policy has expired. The Claims made principle is sometimes used in liability insurance, and is common in health insurance. The reason is that in those lines, it is often impossible to pinpoint the exact time of the event leading to the loss.

Claims manifestation principle

An intermediate form between Claims incurred and Claims made is the *Claim manifestation principle* that is sometimes used for Workers' Compensation insurance. Under the Claim manifestation principle, the loss is covered by the policy that the insured had at the time when the right to claim for the first time became manifest - for example, through diagnosis of an occupational disease.

Under all attachment principles, there will normally be time limits on the notification of claims. For the actuary estimating outstanding claims, the attachment principle is important as it determines the potential length of the reporting period.

2.4 Stages in the life of a claim

Reporting (Notification)

A claim is reported when the insured contacts the insurance company, and the insurance company opens a file on the claim. There will almost invariably be a delay between the event leading to an insured loss and the notification of a claim. The delay can vary between a few days and several years. Common reasons for delayed notification of claims are:

- Gradual deterioration of the insured's health as the result of an accident or disease,
- The insured not being aware of his or her right to claim,
- Initial reliance on social security institutions for financial support,
- Public holidays just prior to the end of year, for example Christmas,
- Periodic bulk reporting of claims by intermediaries,
- The insurer's administrative routines.

Assessment

Once a claim has been reported, if it cannot be settled immediately, an assessment of its likely cost should be made. Depending on the nature of the claim, the assessment could involve

- On-site inspections,
- A technical report by a builder, plumber etc.,
- A quote by a motor vehicle repair workshop,

- Medical examinations,
- Legal opinions.

These processes take time. In particular, medical examinations will often have to be conducted over an extended period of time in order to assess the degree of permanent disability. Normally, no payments other than fees are made during the assessment period.

Handling

After the initial assessment the claim will be handled, which may involve payments for

- Repairs,
- Reconstruction,
- Medical treatment,
- Rehabilitation,
- Legal assistance,
- Loss of income.

During the time that the claim is actively being handled it is often called an "open claim". The ultimate cost of the claim will not be known with certainty.

Depending on the nature of the claim, the claimant and the insurer, the claim handling phase can take any length of time. Claim handling may involve litigation, possibly passing through several law courts before the claim can be finally settled.

Payment

Partial payments may be made throughout the handling phase of the claim, to cover losses that the insured has had (such as medical expenses and lost income) or the cost of ongoing repairs. A final payment is made when the claim is settled, unless the entire compensable loss already has been covered by partial payments.

In personal lines, the final payment will often be for future loss of income and future cost of care. Future benefits can be paid in a lump sum (one payment) or an annuity (a series of payments). Once the final lump sum has been paid, or the amount and the duration of the annuity has been determined, the insurance company normally considers the claim to be settled.

Settlement

A claim is considered to be settled when there is no reasonable expectation that it will give rise to any more payments or adjustments by the insurance company. Most claims, once they are settled, are never looked at again by claim handling staff. IT staff tend to consider closed claims as space-consuming rubbish, to be removed from the data base as quickly as possible. For the actuarial analysis, the information provided by closed claims is indispensable.

Reopening

Some claims are reopened. This occurs when the claimant decides that the original compensation was inadequate for some reason, or when the health of the claimant has deteriorated beyond the stage that was the basis of the original compensation. A reopened claim enters again into the stages of assessment, claim handling and finally settlement.

The processes of claim settlement and claim reopening are strongly influenced by the administrative routines of the insurer, and by the incentives that claim handlers have to settle claims. As a general rule, when claim handlers have a strong incentive to settle claims, one must expect a significant proportion of settled claims to be reopened later.

Recoveries

The insurer will sometimes be able to recover some of its outlay from other parties. Common forms of recovery are

- Recovery from reinsurers,
- Recovery from the liability insurer of the negligent party,
- Subrogation (the insurer takes possession of the damaged goods and sells them for a residual value).

2.5 The length of the tail

In actuarial and insurance parlance, it is normal to distinguish between short-tailed lines and long-tailed lines. Short-tailed lines are those where the majority of claims incurred during a year will be settled within, say two years. Long-tailed lines are those where there remains a significant proportion of open claims after, say 3-5 years.

The table below is only meant to give an indication of the placement of different lines of business between the two categories. In practice, the development time of a line of business depends on local legislation, policy terms and time limits.

Line of insurance	Reporting	Handling	Reopening	Overall
Domestic property	short	medium	rare	short
Commercial property	short	short	rare	short
Commercial business	short	medium	rare	short
Motor vehicle	short	short	rare	short
Third party	medium	long	frequent	long
Travel	short	short	rare	short
Liability	long	long	frequent	long
Accident	short/med.	long	frequent	long
Health	short	short	N/A	short
Disability	short/med.	long	frequent	long
Worker´s compensation	med./long	long	frequent	long
Marine	short/med.	long	rare	long
Aviation	med./long	long	rare	long
Transport	short	short	rare	short
Reinsurance	med./long	long	N/A	long

I have put "N/A" on reopening for health insurance and reinsurance, because in those lines it is difficult to identify a single claim and to decide whether it is open or closed. Again, this is only a rule of thumb which may not hold in all situations.

In all lines of business there will be some claims that are notified and settled almost immediately, and others that drag on forever. The characterisation above is only meant to show the average delays in notification and assessment/handling that one should expect to see in the different lines.

2.6 Valuation of outstanding claims

Valuation of outstanding claims involves establishing a value of future payments that will be generated by claims that are currently outstanding. I am deliberately saying *a* value, not *the* value, as valuations made for different purposes and with different assumptions, can produce different values.

2.6.1 Purpose of valuation

There are five main purposes for which a valuation of outstanding claims is required: Accounting, pricing, statutory reporting, portfolio transfer, and commutation. They will be briefly outlined in the following sections.

2.6.2 Accounting

An general insurer's balance sheet has the following main entries:

Assets
- Financial assets
- Reinsured share of premium provision
- Reinsured share of outstanding claim provision
- Receivables (from reinsurers and policyholders)
- Other assets

Liabilities
- Gross premium provision
- Gross outstanding claim provision
- Prudential margin or risk margin
- Payables (to reinsurers and policyholders)
- Other liabilities

Equity = Assets - Liabilities

Insurance provisions (gross premium provision and gross outstanding claim provision) normally form the largest part of the liabilities in the balance sheet of a general insurer. Insurance provisions should represent a fair and prudent estimate of the value of outstanding claim payments. This means that the provisions should give a realistic view of the liabilities. They should not be excessive but, in case of doubt, they should rather be a little too high than too low. The difference between the actual provision and a strictly unbiased estimate (or "central estimate") is called a prudential margin or risk margin.

2.6.3 Pricing

Pricing in general insurance is an ongoing process. Pricing calculations are often complicated by the fact that the latest years' claims have not been finally settled, maybe not even fully notified, so that one must estimate the outstanding claim cost of the latest years in order to use those in pricing calculations.

As a general rule, outstanding claim estimates used in pricing should be strictly central estimates with no risk margins. Expense loadings and profit margins should be entered into the pricing formula separately from the pure claim cost estimate.

2.6.4 Statutory reporting

Some countries have statutory requirements for the calculation of technical provisions in the balance sheet. These requirements are normally designed to protect the policyholders by ensuring adequate provisions, and the resulting provisions are likely to be higher than a strictly central estimate.

The opposite may also occur, however, if the statutory rules are not sufficiently adaptive to actual claim developments.

This book does not contain a description of statutory requirements. The only point I should like to make is that the actuary should not rely on the statutory provisions being adequate, without having formed an independent opinion.

2.6.5 Sale or purchase of a portfolio

When a block of business is transferred from one insurance company to another, or from an insurance company to a reinsurer, the buyer and the seller must form an opinion on the value of the outstanding claim payments that follow with the business.

It is equally important for both parties to have a view on the fair value of the outstanding claim payments. The parties will normally keep their respective fair value estimates to themselves and negotiate on the basis of estimates that serve their commercial self-interest. Unless the portfolio is strictly to be run off, the price that is ultimately paid for the transfer will reflect a number of other consideration besides the value of outstanding claim payments - such as expected future profits, synergy benefits, cost savings and so on.

2.6.6 Commutation

Commutation means the settlement of uncertain future claim payment liabilities for a fixed price. Commutation is similar to the sale of a portfolio for run-off, and a valuation of the outstanding claim payments is indispensable.

2.7 Case estimates

When the insurance company has received notification of a claim, and while the claim goes through the different stages, the claim handler will normally have an opinion of what the ultimate cost of the claim will be. That opinion may be based on past experience and an educated guess, or it may be based on more systematic analysis using estimation tools.

Actuaries usually refer to the claim handler's estimate of the outstanding cost (i.e., future payments) on a claim as its "case estimate". The case estimate of a claim should be adjusted whenever the claim handler has reason to believe that the existing case estimate does not reflect the expected outstanding claim cost. The change in opinion could stem from new information on the individual claim (e.g., a new medical report) or from information on another, similar claim (e.g., a landmark court decision).

This author likes to distinguish between two types of case estimate,

- The *outstanding case estimate*, being an estimate of future payments;
- The *ultimate case estimate*, being an estimate of total payments on the claim.

A claim's ultimate case estimate is the sum of past payments on the claim and its outstanding case estimate. One can extend the terminology to closed claims, where the ultimate case estimate is the sum of payments, and the outstanding case estimate is zero. After all, a number that one knows with certainty is just a very reliable estimate.

While the outstanding case estimate normally should decrease towards zero as a claim is handled and approaches final settlement, the ultimate case estimate should ideally be stable. An upward trend in ultimate case estimates of a given body of reported claims indicates under-estimation by the claim handler.A persistent downward trend indicates that early case estimates overstate the ultimate cost.

The sum of ultimate case estimates of a portfolio of reported claims is often referred to as the *reported incurred claims* or simply *incurred claims* in the actuarial literature. The second term, while widely used, is unsatisfactory because in other contexts, incurred claims also comprises unreported claims. A short and unambiguous synonym for *ultimate case estimates* is the term *reported claim cost*, and it will be used in what follows.

2.8 Nominal and discounted values

The estimated value of outstanding claims should always include an allowance for future inflation. That allowance should comprise the expected normal inflation that occurs over time in most countries. In addition, an allowance should be made for so-called superimposed inflation, if any. Superimposed inflation is inflationary growth in claim cost over and above normal inflation. Superimposed inflation may be the result of specific influences, for example:

- Higher than normal inflation for specific products or services,
- An increasing propensity to claim,
- Gradual change in legal practice.

An argument can be made for not separating normal inflation from superimposed inflation at all, because the inflationary forces that drive the cost of insurance claims may be quite different from those that drive normal inflation.

The effect that inflation can have on the cost of outstanding claims depends on policy terms and the length of time before the claims are settled.

The policy terms are important inasfar as they regulate if claim payments are made in the currency of the policy period (for example, a fixed nominal sum insured) or in the currency of the payment period (for example, unspecified medical cost); in the former case there will be no future inflation, while in the latter case there may be. Generally, the longer the time before payments are made, the greater will be the potential inflationary effect. As a result, inflation must always be considered in long-tailed lines of business, while it often can be ignored in short-tailed lines.

If the value of outstanding claim payments includes an allowance for inflation, we say that it is expressed in *nominal values*. The nominal value of outstanding claims represents the amount of Euro notes that the insurance company would need to stash away in its vault, if it wanted to make expected claim payments as they fall due, dispensing Euro notes from its vault.

In reality, the insurance company would have a bank account at the very least. More sensibly, it would invest some of its money in interest-bearing securities (notes and bonds). The money thus invested is expected to generate interest that can also be used to finance future claim payments.

The cash outlay required to buy notional notes and bonds that with interest will generate sufficient cash inflow to finance the expected future cash outflow, is less than the value of the cash needed in the vault. Therefore it is cheaper for the company to express its liability in terms of the cash value of notes and bonds; in doing so, however, the company allocates future investment income to payment of outstanding claims.

If the value of outstanding claim payments includes an allowance for inflation and makes allowance for expected future investment income, we say that it is expressed in *discounted value* or *present value*.

Whether outstanding claim provisions in the balance sheet are expressed in nominal value or in present value, depends on local accounting rules. Considerations of financial prudence are often quoted in favour of holding provisions in nominal values, while considerations of economic realism work in favour of expressing provisions in discounted values.

If outstanding claims are valued for the purpose of pricing insurances in a long-tailed line of business, the estimates should always be expressed in discounted values, using a conservative discount rate.

Outstanding claims valued for the purchase or sale of a portfolio, or for a commutation, must by necessity be expressed in discounted values in order to make economic sense to buyer and seller.

2.9 Time dimensions

Depending on the context, premiums and claims in insurance are analysed along several different time dimensions, which will be defined now.

2.9.1 Underwriting date

The underwriting date of an insurance policy is the date of its inception or its most recent renewal. For a claim, the underwriting date is the underwriting date of the policy to which the claim is attached.

From a risk perspective it is often argued that a claim's underwriting date is is irrelevant. Notwithstanding, the actuary sometimes needs to analyse claim cost by underwriting period. One reason for this is that reinsurance is often written by underwriting year, i.e., covering claims from policies that are underwritten during a the reinsurance contract's period. Another reason is that claim cost allocated to underwriting periods can be readily compared with premium income during the corresponding periods, thus allowing profitability analyses.

2.9.2 Accident date

Accident date is the date of the loss event leading to the claim. The term accident date is used generically, also when the cause of loss is not an accident in the narrow sense of the word.

2.9.3 Reporting date

Reporting date is the date when the claim it reported or, more precisely, recognised as a claim in the official files of the insurer. There is some discussion as to whether reopened claims should be considered to be new claims with a new reporting date, or simply as a continuation of known claims. Both approaches are possible and valid, as long as the chosen approach is used consistently.

2.9.4 Valuation date

This is the date as at which claims are valued.

2.9.5 Development date

Development date (normally: development year) is a term used generically in the actuarial literature to describe delay after a certain event. Development date normally starts with zero (0), which is the date of the anchor event. Thus development date could be delay after the underwriting date, the accident date, or the reporting date. The actuary should always take care to define precisely what he or she means by development date and, consequently, what it is that constitutes claim development.

2.9.6 Acronyms

Several acronyms are common in the actuarial literature:

- RBNS – reported but not settled;
- IBNR – incurred but not reported;
- IBNS – incurred but not settled;
- CBNI – covered but not incurred;
- CBNR – covered but not reported.

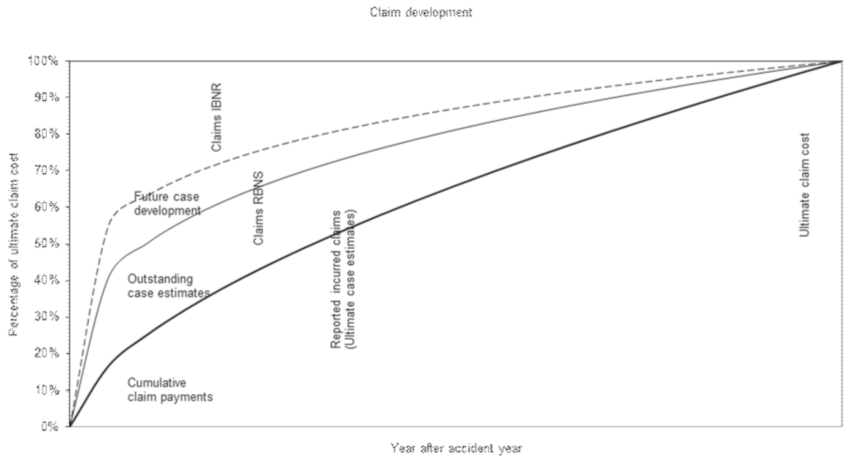
RBNS claims are claims that have been notified to the insurer. This book uses the term RBNS for the totality of reported claims, including settled claims, because even settled claims may be reopened and generate more payments. Of RBNS claims one definitely knows their number, but the ultimate claim amount can be uncertain. Some actuaries use the acronym IBNER (incurred but not enough reported) for future development of reported claims. I dislike the acronym IBNER because it uses the term "incurred" in an ambiguous way.

IBNR claims are claims where the event that will give rise to a claim (the fire, accident, burglary or whatever) has occurred, but where the claim has not been notified to the insurer. One knows from experience that there will be unreported claims around at any point in time, but neither their number nor their ultimate claim amount is known.

The term IBNS is sometimes used to denote the sum of claims IBNR and claims RBNS. Claims IBNS is the totality of claims that have been incurred at a given point in time.

CBNI claims are claims that are linked to events that will occur in future, i.e., after the balance date, and that are covered by policies in force on the balance date. For example, if a policy was renewed for one year on 1st October and the balance date is 31st December, then any insured events between ist January and 30th September of the following year will be covered by the policy and constitute a CBNI liability for the insurer.

Claims CBNI behave very similarly to claims IBNR. Both types of claim are unknown in number and amount, and both are reported after the balance date. Thus one can introduce a category of claims covered but not reported (CBNR) that comprises claims CBNI and claims IBNR. One does not need to split claims CBNR into separate components if one is only interested in the liability for unreported claims, not whether the loss events have occurred before, or will occur after after, the balance date. For example, claim cost development in marine insurance is customarily analysed by underwriting year. The accident date of claims may be unavailable, which makes it impossible to distinguish between claims IBNR and claims CBNI. In such a setting, the most fruitful division is between claims RBNS and claims CBNR.



The most common way of examining claims, however, is by accident period. In that case one can distinguish between claims reported (RBNS), claims incurred in the past but not reported (IBNR), and claims that will be incurred in the future and obviously have not been reported (CBNI).

It is important to distinguish between future claims that are covered by existing contracts (CBNI), and claims that will arise from new contracts and renewals written after the balance date. A general insurer normally has no legal obligation to write new contracts or to renew existing ones. As a consequence, expected premiums and claim liabilities related to future business are normally not recorded in the insurer's accounts. For the purpose of budgeting and forecasting, however, the premiums and claim cost generated by future business is very interesting.

The graph below shows the typical development of the claim cost of an accident year. As time goes by after the accident year, more and more claims will be paid (cumulative claim payments) or reported (reported claims). More often then not, the reported claims are under-estimated, leading to positive future case development. Negative future case development is also known to occur. Also as time goes by, fewer and fewer claims will be unreported (IBNR).

Loosely speaking, the actuary can observe the development of the two solid lines - cumulative claim payments and reported claims - at any point in time, and his task is to estimate the magnitude of future case development and claims IBNR.

3

Framework model

3.1 Introduction

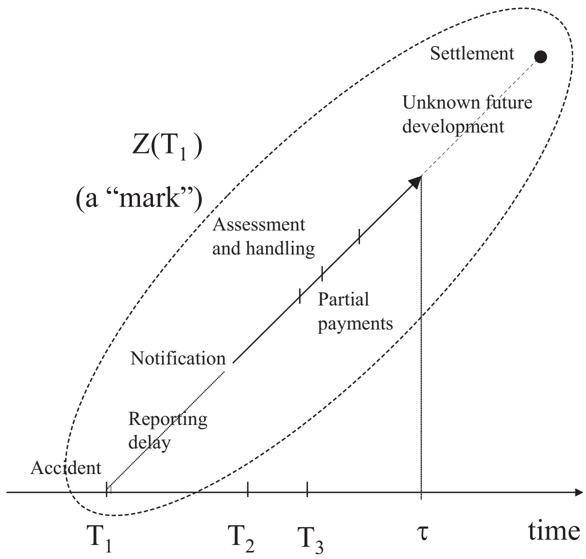
This chapter starts with a brief outline of a framework model in continuous time. The main consequence of the framework model is that the cost of reported and unreported claims ought to be estimated separately, whenever possible.Two discrete time models are then presented. The first model has three time dimensions along which claims evolve, while the second model only has two time dimensions.

3.2 Continuous time model

A general framework model for the estimation of outstanding claims in general insurance has been developed in papers by Arjas (1989), Norberg (1993, 1999), and Jewell (1989, 1990).

The main features of the framework model of Norberg are the following:

- Prior to discretisation of the data, claims develop in continuous time.
- Claims are incurred in accordance with a point process that generates random times $T_1, T_2, \cdots$. More specifically, Norberg assumes the existence of an exposure function $\{p(t) : t \geq 0\}$ and a frequency function $\{\theta(t) : t \geq 0\}$, so that claims are incurred in accordance with a Poisson process with an infinitesimal rate of claim incurrence of $p(t)\theta(t)$. Given $\{p(t) : 0 \leq t \leq \tau\}$ and $\{\theta(t) : 0 \leq t \leq \tau\}$, the number of claims



incurred up to time τ has a Poisson distribution with expected value $\int_0^\tau p(t)\theta(t)dt$.

- A claim incurred at time t has an evolution $Z(t)$ that encapsulates the development of all the relevant aspects of the claim, including time to notification, partial payments, development of case estimates, time to settlement, and whatever else could be of interest. Statisticians refer to $Z(t)$ as a mark, and to the whole process model as a marked point process. A mark is a generalisation of the random claim amount that actuaries commonly use in the compound Poisson process.
- The marks $\{Z(t) : t \geq 0\}$ are stochastically independent of each other and of the claim incurrence process. The marks are not necessarily identically distributed, although this is a common assumption. Internally in each mark, there may be a number of stochastic dependencies, such as, for example, dependency between the ultimate claim amount and the time it takes to settle the claim.

The diagram below illustrates the concept of a marked point process.

I will not elaborate on Norberg´s framework model any further in this book, just state some of its main implications for our work.

- Given complete information about the claim process up to time τ, the future development of reported claims is stochastically independent of the future development of unreported claims. That is so because the future development of reported claims occurs entirely within the marks that are assumed to be stochastically independent of each other and of the claim incurrence process.

- In order to estimate the future development of all claims (reported and unreported), one should separate the estimation of reported claims from the estimation of unreported claims. Being conditionally independent, the two processes do not carry any information about each other. By mixing them one loses information.

- In the estimation of unreported claims one should separate the estimation of claim counts from the estimation of claim amounts. Claim amounts, being part of the marks, do not carry information about claim counts.

- The estimated amount of unreported claims becomes the estimated number of unreported claims, multiplied with an average amount. The average amount may depend on the reporting delay, because both the claim amount and the reporting delay are part of the mark.

The humble actuary will normally only have access to discretised data, so that work in continuous time is not a practical option. However, in recognition of the conclusions above, the actuary should organise his or her work in the following five-step procedure:

1. Estimate the *amount* of reported claims (RBNS).

2. Estimate the *number* of incurred, unreported claims (IBNR).

3. Estimate the *amount* of incurred, unreported claims (IBNR).

4. Estimate the *number* of covered, future claims (CBNI).

5. Estimate the *amount* of covered, future claims (CBNI).

Each of these steps is not a trivial task, and the actuary has to fill the generalities with specific assumptions in order to arrive at a number. A short comment on the sequence of the analysis is in order.

Reported claims (RBNS) come with a great deal of information that is potentially valuable: number of claims, claim characteristics, settlement status, settled amounts, partial payments, case estimates etc.

Of unreported claims we know neither their number nor their cost, in other words: nothing. We only know that they are there and will be reported later. To establish the expected cost of unreported claims, one normally has no other source of information than already reported claims. That is why an analysis of reported claims usually must come first.

A final, important observation is that the available factual information for claims CBNI and claims IBNR is identical: nothing. From a methodologial standpoint, therefore, it makes more sense to treat claims CBNI similarly to claims IBNR, than to confound claims RBNS with claims IBNR.

It may come as a surprise that it is often easier to estimate the cost of unreported claims (IBNR and CBNI), than it is to estimate the ultimate cost of reported claims (RBNS).

Notwithstanding the above, the presentation that follows starts with estimating the number of incurred, unreported claims in Chapter 4, followed by estimating the amount of incurred, unreported claims in Chapter 5, and then estimating the amount of reported claims in 6. The reason for the inverted ordering is that the two most basic methods - the Chain-ladder method and the Bornhuetter-Ferguson method - are most convincingly motivated and explained in the context of estimating the number of unreported claims.

3.3 Discrete time model with three dimensions

Let us now consider a discrete time model with three time dimensions: accident period, reporting period and valuation period.

Conventional actuarial terminology usually speaks of 'years'. In practice it is entirely possible and usually advisable to build the model with shorter time periods (quarters or months). The initial investment in doing so is more than compensated by the facility with which one can calculate updated estimates at shorter time intervals, using a consistent set of assumptions. We will call the smallest discrete time units 'periods'.

We denote accident periods by j. For an accident period j, we denote the amount of risk exposed by p_j. The number of claims reported with delay d is denoted by N_{jd}. The individual severities of those claims we denote by $\{Y_{jd}^{(k)} : k = 1, \cdots, N_{jd}\}$ and their sum as Y_{jd}.

For a given claim, its ultimate severity $Y_{jd}^{(k)}$ is made up of a series of partial payments $U_{jdt}^{(k)}$ that occur at delay $t = 0, \cdots, \infty$ after the reporting date:

$$Y_{jd}^{(k)} = \sum_{t=0}^{\infty} U_{jdt}^{(k)} \tag{3.1}$$

The use of the symbol ∞ only means that we do not impose a fixed upper limit on the possible number of terms; there will of course be a finite number of terms. In addition to partial payments we may observe outstanding case estimates. Denote by $V_{jdt}^{(k)}$ the change in the outstanding case estimate at delay t after the reporting date. Finally, let $W_{jdt}^{(k)} = U_{jdt}^{(k)} + V_{jdt}^{(k)}$ denote the change in the reported claim cost. Note that

$$Y_{jd}^{(k)} = \sum_{t=0}^{\infty} U_{jdt}^{(k)} = \sum_{t=0}^{\infty} W_{jdt}^{(k)}, \tag{3.2}$$

which states the obvious fact that the total change in the outstanding case estimate from the time when the claim is reported to the time when it is settled, is zero. The outstanding case estimate starts at zero and ends at zero.

Now assume that the last calendar period and the current valuation period is J.

At time J we will have recorded the reported number of claims $\{N_{jd} : j = 1, \cdots, J, d = 0, \cdots, J - j\}$, while $\{N_{jd} : j = 1, \cdots, J, d > J - j\}$ will still be unreported. The only partial payments that we have had the chance to observe are those for which $j + d + t \leq J$. The accumulated payments to the end of period J are

$$U_{jd,\leq J-(j+d)}^{(k)} = \sum_{t=0}^{J-(j+d)} U_{jdt}^{(k)}, \tag{3.3}$$

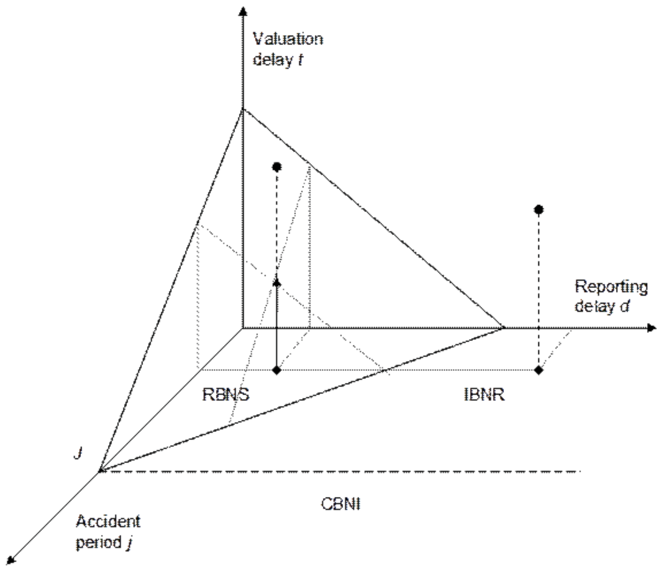
with corresponding formulas for the current outstanding case estimate and ultimate case estimate. The outstanding payments in respect of reported claims are

$$\mathrm{RBNS}_J = \underbrace{\sum_{j=1}^{J} \underbrace{\sum_{d=0}^{J-j} \underbrace{\sum_{t=J-(j+d)+1}^{\infty} U_{jdt}}_{\text{Future payments}}}_{\text{Reported claims}}}_{\text{Incurred claims}}, \tag{3.4}$$

and the future cost of claims IBNR is

$$\mathrm{IBNR}_J = \underbrace{\sum_{j=1}^{J} \underbrace{\sum_{d=J-j+1}^{\infty} \underbrace{\sum_{t=0}^{\infty} U_{jdt}}_{\text{All payments}}}_{\text{Unreported claims}}}_{\text{Incurred claims}} . \tag{3.5}$$

The development tetrahedron below illustrates the three dimensions of claim development. Claims that are RBNS at time J have been reported inside the horizontal triangle given by $j+d \leq J$, as indicated by a diamond. The observed development of a reported claim is indicated by a solid vertical line lying inside the tetrahedron which is delimited by $j + d + t \leq J$, and its future development is indicated by the dotted extension of that line. The development of a claim ends at settlement, indicated by a bullet. A



claim that is IBNR starts its observed development outside the horizontal triangle and its development lifeline is dotted all the way to settlement. The current status of reported claims can be 'read off' on the simplex given by $j + d + t = J$.

The following abbreviation will be used in the rest of this paper: a variable with a subscript omitted denotes the sum of the underlying variables across all values of the subscript that has been omitted. A variable with a subscript replaced by an inequality (for example, $U_{jd,\leq J-(j+d)}$) denotes the sum of the underlying variables that satisfy the inequality. A variable with a subscript replaced $\leq$ by is usually the sum of the underlying variables that lie inside the tetrahedron, while a variable with a subscript replaced by $>$ is the sum of the underlying variables that lie outside the tetrahedron.

3.4 Discrete time model with two dimensions

Most loss-reserving studies analyse only two-dimensional development models, where the two dimensions are accident period (j) and accident-to-valuation delay ($e = d + t$). The delay dimension in those triangles is normally referred to as development period. In those models, no attempt is made to separate the development of reported claims from that of unreported claims.

The most common data to analyse in the two-dimensional setting is paid claims. One starts with a development triangle containing the accumulated claim payments per accident period and valuation delay:

$$\begin{array}{llll} \tilde{U}_{10} & \tilde{U}_{11} & \cdots & \tilde{U}_{1,J-1} \\ \tilde{U}_{20} & \tilde{U}_{21} & \iddots & \\ \vdots & \iddots & & \\ \tilde{U}_{J0} & & & \end{array} \tag{3.6}$$

Here we have defined the accumulated claim payments $\tilde{U}_{je}$ by

$$\tilde{U}_{je} = \sum_{d=0}^{e-j} \sum_{t=0}^{e-(j+d)} U_{jdt}, \tag{3.7}$$

i.e., $\tilde{U}_{je}$ is the sum of all payments recorded until valuation period $j+e$, comprising payments from all reporting periods as they emerge.

Development in the accumulated claim payments between two different valuation dates is driven by two different processes:

- Partial payments on claims that had been reported already on the first valuation date; and
- Partial payments on new claims that were reported between the valuation dates.

  The task then becomes to predict the entries in the south-east corner of the development square.

One may triangulate reported claims in the same way:

$$\begin{array}{llll} \tilde{W}_{10} & \tilde{W}_{11} & \cdots & \tilde{W}_{1,J-1} \\ \tilde{W}_{20} & \tilde{W}_{21} & \iddots & \\ \vdots & \iddots & & \\ \tilde{W}_{J0} & & & \end{array} \tag{3.8}$$

Here we have defined the accumulated reported claims $\tilde{W}_{je}$ by

$$\tilde{W}_{je} = \sum_{d=0}^{e-j} \sum_{t=0}^{e-(j+d)} W_{jdt} \tag{3.9}$$

Development in the amount of reported claims between two valuation dates comprises the effect of four processes:

- Partial payments on claims that had been reported already on the first valuation date;

- Partial payments on new claims that were reported between the valuation dates;
- Revaluation of case estimates on claims that had been reported on the first date;
- New case estimates for claims that were reported between the valuation dates.

We will study some of these methods in Chapter 7.

4

The number of unreported claims

4.1 Introduction

This chapter considers specific models and estimation methods for the claim reporting process. In this chapter we only consider the number of claims reported, not their cost.

4.2 Modelling the number of claims IBNR

Let us assume that we have observed claim notifications for the accident periods $j = 1, \cdots J$, where J denotes the current period. Thus the set of observations is $\{N_{jd} : j + d \leq J\}$. For a specific accident period j we have observed $\{N_{jd} : d = 0, \cdots J - j\}$. The data can be arranged in the well-known triangular array.

$$\begin{array}{cccc} N_{10} & N_{11} & \cdots & N_{1,J-1} \\ N_{20} & N_{21} & \iddots & \\ \vdots & \ddots & & \\ N_{J0} & & & \end{array} \tag{4.1}$$

The task of the actuary is to fill the lower part of the triangle with predicted values that denote by $\{\overline{N}_{jd} : j + d > J\}$. Let us denote the highest *observed* reporting delay by $D = J - 1$; it is not necessarily the highest *possible* reporting delay.

Now, let us assume that the amount of risk exposed in accident period j has been p_j. The expected claim frequency relative to the risk measure we denote by θ_j, and the probability of a notification delay of d periods, we denote by π_d. We now make the assumption that N_{jd} has a Poisson distribution with expected value $p_j\theta_j\pi_d$:

$$N_{jd} \sim \text{Poisson}(p_j\theta_j\pi_d) \tag{4.2}$$

Before continuing, a few words on the assumptions are in order. The assumption that claim numbers have a Poisson distribution is standard in insurance modelling. The assumption that the expected number of claims is proportional to the amount of risk exposed p_j is very reasonable, provided of course that it is possible to measure the amount of risk exposed reliably. The assumption that every accident period has a certain ultimate claim frequency θ_j, possibly different from period to period, is almost axiomatic. Finally, assuming that the reporting of claims follows a fixed statistical pattern π_d, is quite reasonable as a first approach. What has not been modelled so far, is possible interactions, i.e. models when the expected number of claims has the form $p_j\theta_{jd}$, where θ_{jd} cannot be written as a product of two marginal factors. We defer the more complex models until later.

If the claim frequencies $\theta_1, \cdots, \theta_J$ and the delay probabilities $\pi_0, \pi_1, \cdots$ were known, an obvious predictor of N_{jd} for $j + d > J$ would be

$$\overline{N}_{jd} = p_j\theta_j\pi_d \tag{4.3}$$

In practice, of course, both the claim frequencies and the delay probabilities are unknown and must be estimated first. We now consider different models for estimating claim frequencies and delay probabilities. For the time being, let us assume that $\sum_{d=0}^{D} \pi_d = 1$, which means that the claims of at least one accident period are fully reported.

4.3 Constant claim frequency: The Bornhuetter-Ferguson method

For some lines of insurance a quite reasonable assumption is that claim frequencies are constant over time, i.e., $\theta_1 = \theta_2 = \cdots = \theta_J = \theta$. For instance, claim frequencies in personal accident insurance normally do not vary greatly from period to period, provided that the insurance cover is unchanged, and barring catastrophic events.

A convenient re-parametrisation is $\theta_d = \theta\pi_d$ for $d = 0, \cdots, D$.The parameters θ_d can be interpreted as delay-specific claim frequencies. To estimate the parameters one can maximise the likelihood of the observations:

$$L = \prod_{j=1}^{J} \prod_{d=0}^{J-j} \frac{(p_j \theta_d)^{N_{jd}}}{N_{jd}!} e^{-p_j \theta_d} \tag{4.4}$$

The log-likelihood function is then:

$$\ln(L) = \sum_{d=0}^{D} N_{\leq J-d,d} \ln(\theta_d) - \sum_{d=0}^{D} p_{\leq J-d} \theta_d + \text{terms not involving } \theta_d \tag{4.5}$$

Differentiating the log-likelihood function, one obtains for $d = 0, \cdots, D$,

$$\frac{\partial \ln(L)}{\partial \theta_d} = \frac{N_{\leq J-d,d}}{\theta_d} - p_{\leq J-d} \tag{4.6}$$

Equating these expression to zero one obtains the maximum likelihood estimates of the delay-specific claim frequencies,

$$\theta_d^* = \frac{N_{\leq J-d,d}}{p_{\leq J-d}} = \frac{\text{Number of claims reported with delay } d \text{ (column } d)}{\text{Sum of exposure for which delay } d \text{ has been observed}} \tag{4.7}$$

for $d = 0, \cdots, D$. From this and the constraint $\sum_{d=0}^{D} \pi_d = 1$ one can derive the maximum likelihood estimates of the original parameters:

$$\theta^* = \sum_{d=0}^{D} \theta_d^* \tag{4.8a}$$

and, for $d = 0, \cdots, D$,

$$\pi_d^* = \frac{\theta_d^*}{\theta^*} \tag{4.9}$$

The maximum likelihood predictions of claims IBNR for $j + d > J$ then become

$$\overline{N}_{jd} = p_j \theta_d^* = p_j \theta^* \pi_d^* = p_j \left(\frac{N_{\leq J-d,d}}{p_{\leq J-d}} \right) \tag{4.10}$$

This method, where the same delay-specific claim frequencies are assumed for all accident periods, is often referred to as the Bornhuetter-Ferguson method. There exist several interpretations as to what constitutes the essence of the Bornhuetter-Ferguson method. One aspect that is frequently cited is that the claim frequencies θ should be chosen *à priori*, i.e. using expert knowledge. I do not see any conflict with the current approach, because estimation is a way of acquiring knowledge. In my view, the essential characteristic of the Bornhuetter-Ferguson method is that predictions of outstanding claims of a specific accident year only depend on the risk exposed and a delay pattern, but not on the already observed claim history of that accident year.

An advantage of the Bornhuetter-Ferguson method is that is provides stable estimates for the newer, undeveloped accident periods, based on experience from older periods.

A disadvantage of the Bornhuetter-Ferguson method is that its projections are relatively unresponsive to changes in observed claim frequencies – but then, if one à priori believed that changes in claim frequency are a possibility, one should not have made the assumption $\theta_1 = \theta_2 = \cdots = \theta_J = \theta$ in the first place. Let us therefore consider ways in which the assumption of constant claim frequencies can be relaxed.

4.4 Varying claim frequency: The chain-ladder method

An extreme alternative to the assumption underlying the Bornhuetter-Ferguson method, is to postulate that all claim frequencies are potentially different and need to be estimated in their own right. Instead of estimating θ and $\pi_0, \cdots, \pi_D$, we must then estimate the parameters $\theta_1, \cdots, \theta_J$ and $\pi_0, \cdots, \pi_D$.In that case, the likelihood function is

$$L = \prod_{j=1}^{J} \prod_{d=0}^{J-j} \frac{(p_j \theta_j \pi_d)^{N_{jd}}}{N_{jd}!} e^{-p_j \theta_j \pi_d} \tag{4.11}$$

The log-likelihood function is then:

$$\begin{aligned} \ln(L) &= \sum_{j=1}^{J} N_{j,\leq J-j} \ln(\theta_j) + \sum_{d=0}^{D} N_{\leq J-d,d} \ln(\pi_d) \\ &- \sum_{j=1}^{J} \sum_{d=0}^{J-j} p_{jd} \theta_j \pi_d + \text{terms not involving } \theta_j \text{ or } \pi_d \end{aligned} \tag{4.12}$$

Differentiating the log-likelihood and equating the derivatives to zero, we find the defining equations of the maximum likelihood estimates:

$$\underbrace{N_{j,\leq J-j}}_{\text{Row sum}} = p_j \theta_j^* \pi_{\leq J-j}^* \text{ for } j = 1, \cdots, J \tag{4.13}$$

and

$$\underbrace{N_{\le J-d,d}}_{\text{Column sum}} = \left(\sum_{j=1}^{J-d} p_j\theta_j^*\right)\pi_d^* \text{ for } d = 0,\cdots,D \tag{4.14}$$

Using the constraint $\sum_{d=0}^{D}\pi_d = 1$, the sets of equations (4.13) and (4.14) can be solved by backward recursion, starting in the far north-east corner of the triangle with θ_1^*, then finding π_{J-1}^*, then carrying on with θ_2^*, then π_{J-2}^*, and so on.

Due to (4.13), the maximum likelihood prediction of claims IBNR for $j+d > J$ becomes

$$\overline{N}_{jd} = p_j\theta_j^*\pi_d^* = p_j\left(\frac{N_{j,\le J-j}}{p_j\pi^*_{\le J-j}}\right)\pi_d^* = N_{j,\le J-j}\left(\frac{\pi_d^*}{\pi^*_{\le J-j}}\right) \tag{4.15}$$

In the literature is is called the Chain-ladder method. The essence of this method is that it extrapolates the observed claim number of accident period j into the future, using "grossing up" of the observed claim number in proportion with the estimated delay probabilities.

An advantage of the Chain-ladder method is that its predicted claim numbers are highly responsive to changes in the observed claim numbers. Any change in reported claim numbers of a specific accident period is extrapolated in the same proportion to the estimated, unreported claim numbers of that accident period.

A disadvantage of the Chain-ladder method is its sensitivity. In long-tail lines of business, where $\pi^*_{\le J-j}$ will be small for the newer accident periods, a small and possibly random fluctuation in the reported claim number can lead to a much larger change in the Chain-ladder prediction of the unreported claim number. In essence, a small number will be wagging a heavy tail.

The backward recursion implied by (4.13) and (4.14) is very awkward. One can avoid it using an equivalence that is proved by Taylor (2000, pp. 34). To this end we define development factors.

Definition 1 *The development factors $\delta_d, d = 1, 2, \cdots$ are defined as the relative increase in the reported proportion from development period $d-1$ to development period d:*

$$\delta_d = \frac{\pi_{\le d}}{\pi_{\le d-1}} \tag{4.16}$$

Empirical development factors are easy to calculate as ratios between numbers of claims reported in successive development periods:

$$\delta_d^* = \frac{\sum_{j=1}^{J-d} N_{j,\le d}}{\sum_{j=1}^{J-d} N_{j,\le d-1}} \tag{4.17}$$

Note that the summation in the numerator and the denominator extends over the accident periods for which both numbers are available. Now we can prove the equivalence that makes life a lot easier.

Lemma 2 *The ratios between successive estimates of the cumulative delay probabilities* $\pi^*_{\leq d}$ *are equal to the empirical development factors* δ^*_d*. In formulas, this means that*

$$\frac{\pi^*_{\leq d}}{\pi^*_{\leq d-1}} = \frac{\pi^*_0 + \pi^*_1 + \cdots + \pi^*_{d-1} + \pi^*_d}{\pi^*_0 + \pi^*_1 + \cdots + \pi^*_{d-1}} = \delta^*_d \tag{4.18}$$

Proof. We prove equation (4.18) by backwards induction. The induction step begins with assuming that for an arbitrary $d \leq J-1$ the equality

$$\sum_{j=1}^{J-d} N_{j,\leq d} = \sum_{j=1}^{J-d} p_j \theta^*_j \pi^*_{\leq d} \tag{4.19}$$

holds. Use (4.13) to verify that $N_{1,\leq J-1} = p_1\theta^*_1\pi^*_{\leq J-1}$, which means that (4.19) holds for $d = D = J-1$.From (4.19) and using (4.14) we then derive that

$$\begin{aligned} \sum_{j=1}^{J-d} N_{j,\leq d-1} &= \sum_{j=1}^{J-d} (N_{j,\leq d} - N_{jd}) \\ &= \sum_{j=1}^{J-d} p_j\theta^*_j \left(\pi^*_{\leq d} - \pi^*_d\right) \\ &= \sum_{j=1}^{J-d} p_j\theta^*_j \pi^*_{\leq d-1} \end{aligned} \tag{4.20}$$

Dividing both sides of (4.19) by (4.20) we find

$$\delta^*_d = \frac{\sum_{j=1}^{J-d} N_{j,\leq d}}{\sum_{j=1}^{J-d} N_{j,\leq d-1}} = \frac{\pi^*_{\leq d}}{\pi^*_{\leq d-1}} \tag{4.21}$$

Thus the sought-after equation (4.18) holds for this particular value of d. Now it only remains to show that (4.19) carries over to $d-1$. Using (4.13) again we find

$$\begin{aligned} \sum_{j=1}^{J-(d-1)} N_{j,\leq d-1} &= \sum_{j=1}^{J-d} N_{j,\leq d-1} + N_{J-(d-1),\leq d-1} \\ &= \sum_{j=1}^{J-d} p_j\theta^*_j\pi^*_{\leq d-1} + p_{J-(d-1)}\theta^*_{J-(d-1)}\pi^*_{\leq d-1} \\ &= \sum_{j=1}^{J-(d-1)} p_j\theta^*_j\pi^*_{\leq d-1} \end{aligned} \tag{4.22}$$

So (4.19) indeed holds for $d-1$. This concludes the induction. ■

Thus an easy way to calculate the maximum likelihood estimates of the delay probabilities and claim frequencies is via the empirical development factors. The algorithm goes as follows:

1. Calculate empirical development factors $\delta_d^* = \frac{\sum_{j=1}^{J-d} N_{j,\le d}}{\sum_{j=1}^{J-d} N_{j,\le d-1}}$ for $d = 1, \cdots, D$. Set $\delta_0^* = 1$.
2. Calculate cumulative development factors $\Delta_d^* = \prod_{d'=0}^{d} \delta_{d'}^*$ for $d = 0, \cdots, D$.
3. Calculate cumulative delay probabilities $\pi_{\le d}^* = \Delta_d^* / \Delta_D^*$ for $d = 0, \cdots, D$.
4. Calculate incremental delay probabilities $\pi_d^* = \pi_{\le d}^* - \pi_{\le d-1}^*$ for $d = 1, \cdots, D$, and $\pi_0^* = \pi_{\le 0}^*$.
5. Calculate claim frequencies $\theta_j^* = N_{j,\le J-j} / p_j \pi_{\le J-j}^*$ for $j = 1, \cdots, J$.

This algorithm is straightforward to implement in a spreadsheet program.

Actuaries love their development factors, and major parts of the theory for loss reserving are formulated exclusively in terms of development factors. In the opinion of this author, development proportions (here: delay probabilities) provide a far more understandable representation of claim development than the ubiquitous development factors.

Since the prediction of future claim numbers (4.15) does not depend on the exposures p_j, one can use any exposure measure and get the same predictions (but not the same claim frequencies). The Chain-ladder method is often referred to as being exposure-independent. This author prefers to view the Chain-ladder method as a specific way of estimating the claim frequencies.

4.5 Bootstrapping in fixed-parameter models

Having derived estimates of the claim frequencies $\left\{\theta_j^* : j = 1, \cdots, J\right\}$ and delay distributions $\{\pi_d^* : d = 0, \cdots, D\}$, the next question is about the uncertainty of the predictions. Under the Poisson assumption (4.2), one can apply a parametric bootstrap procedure (**reference needed**). The procedure amounts to simulating random outcomes of past and future claims, then applying the prediction method, and finally comparing the simulated outcome of future claims with the simulated predictions.

The algorithm goes as follows:

For $i = 1, \cdots, M$:

1. Generate $\{N_{jd}^{(i)} : j = 1, \cdots, J, d = 0, \cdots, D\}$ in such a way that $N_{jd}^{(i)}$ ~Poisson$(p_j \theta_j^* \pi_d^*)$ and independent.
2. Based on pseudo-observations $\{N_{jd}^{(i)} : j + d \leq J\}$, calculate pseudo-predictions $\{\overline{N}_{jd}^{(i)} : j + d > J\}$.
3. Calculate pseudo-errors of any desired form, for example $E^{(i)} = \sum_{j+d>J} \left(N_{jd}^{(i)} - \overline{N}_{jd}^{(i)}\right)$.

For each of the M simulations, store the measures that are of interest, for example, $\left\{E^{(i)} : i = 1, \cdots, M\right\}$. At the end, one can analyse the simulated probability distribution of the measures under investigation.

4.6 Varying claim frequency: A Bayesian model

4.6.1 The model

In the previous two sections we have seen two models that both are based on extreme assumptions. In section 4.3 we assumed that the claim frequencies pertaining to different accident period were identical and derived the Bornhuetter-Ferguson method. In section 4.4 we assumed that every accident period's claim frequency needed to be established in isolation and found the Chain-ladder method. Both Bornhuetter-Ferguson's method and the Chain-ladder method are in widespread use. Still, the extreme assumptions that underly both methods are somewhat unsatisfactory. Claim frequencies do differ from period period, but in many lines of insurance we have an idea of the claim frequency we can expect.

In trying to strike a balance between the assumption of constant claim frequency on the one hand, and no assumption about claim frequencies on the other hand, one could consider the following model.

Assume that instead of being fixed model parameters, the claim frequencies $\theta_1, \cdots, \theta_J$ are the realisations of random variables $\Theta_1, \cdots, \Theta_J$, which follow some probability distribution. Then we assume that the *conditional* probability distribution of N_{jd}, given $\Theta_j = \theta_j$, is a Poisson with expected value $p_j \theta_j \pi_d$. The N_{jd} are still assumed to be independent, but now in the conditional distribution given $\Theta_1, \cdots, \Theta_J$. To start with, let us assume that the claim frequencies $\Theta_1, \cdots, \Theta_J$ are independent and identically distributed with a continuous distribution U, the density of which we denote by u.

From Bayesian estimation theory it is well known that the best predictor of N_{jd} $(j+d > J)$ with respect to expected squared error is $\overline{N}_{jd}$ =E$(N_{jd}|\mathbf{D})$, where $\mathbf{D}$ represents the observed data. Using the assumption that claims develop independently in different accident periods, we can further write

$\overline{N}_{jd}$ =E($N_{jd}|\mathbf{D}_j$), where $\mathbf{D}_j = (N_{j0}, \cdots, N_{j,J-j})$ represents the observed data from accident period j alone.

Now using the law of iterated expectations and the assumed independence between the N_{jd} in the conditional distribution given $\Theta_j = \theta_j$, one can find the conditional expectation of N_{jd} $(j+d>J)$, given the vector of observations $\mathbf{D}_j$ to be:

$$\begin{aligned} \mathrm{E}(N_{jd}|\mathbf{D}_j) &= \mathrm{E}(\mathrm{E}(N_{jd}|\Theta_j, \mathbf{D}_j)|\mathbf{D}_j) \\ &= \mathrm{E}(\mathrm{E}(N_{jd}|\Theta_j)|\mathbf{D}_j) \\ &= \mathrm{E}(p_j\Theta_j\pi_d|\mathbf{D}_j) \\ &= p_j\mathrm{E}(\Theta_j|\mathbf{D}_j)\pi_d \end{aligned} \tag{4.23}$$

Thus we need to find the conditional distribution of Θ_j, given the observations $\mathbf{D}_j$. In general terms, the Bayes inversion rule states that

$$u(\theta_j|\mathbf{D}_j) = \frac{p(\mathbf{D}_j|\theta_j)u(\theta_j)}{\displaystyle\int p(\mathbf{D}_j|\theta)u(\theta)d\theta} \tag{4.24}$$

To compute this expression, a distributional assumption is necessary. Thus let us assume that each Θ_j follows a $\Gamma(\alpha, \beta)$ distribution. The probability density of the gamma distribution with parameters (α, β) is

$$u(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}I(\theta > 0) \tag{4.25}$$

The mean and variance of $\Theta_1, \cdots, \Theta_J$ are then $\mathrm{E}(\Theta_j) = \alpha/\beta$ and $\mathrm{Var}(\Theta_j) = \alpha/\beta^2$.

The family of gamma distributions form a family of conjugate priors to the Poisson distributions. By the Bayes inversion rule (4.24), the conditional density of Θ_j, given $\mathbf{D}_j$, is proportional to the conditional density of $\mathbf{D}_j$, given Θ_j, and the unconditional density of Θ_j :

$$\begin{aligned} u(\theta_j|\mathbf{D}_j) &\propto p(\mathbf{D}_j|\theta_j)u(\theta_j) \\ &= \prod_{d=0}^{J-j} \frac{(p_j\theta_j\pi_d)^{N_{jd}}}{N_{jd}!}e^{-p_j\theta_j\pi_d} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)}\theta_j^{\alpha-1}e^{-\beta_j\theta_j} \\ &\propto \theta_j^{\alpha+N_{j,\leq J-j}-1}e^{-(\beta+p_j\pi_{\leq J-j})\theta_j} \\ &\quad \text{(ignoring terms that do not involve } \theta_j) \end{aligned} \tag{4.26}$$

Therefore the conditional distribution of Θ_j, given the observations D_j, must again be a gamma with updated parameters:

$$\begin{aligned} \overline{\alpha} &= \alpha + N_{j,\leq J-j} \\ \overline{\beta} &= \beta + p_j\pi_{\leq J-j} \end{aligned} \tag{4.27}$$

Thence we derive conditional expected value of Θ_j, given the observations $\mathbf{D}_j$:

$$\overline{\Theta}_j = \mathrm{E}\left(\Theta_j|\mathbf{D}_j\right) = \frac{\alpha + N_{j,\leq J-j}}{\beta + p_j\pi_{\leq J-j}} \tag{4.28}$$

Note that the formula may also be written as

$$\overline{\Theta}_j = \zeta_j\left(\frac{N_{j,\leq J-j}}{p_j\pi_{\leq J-j}}\right) + \left(1-\zeta_j\right)\left(\frac{\alpha}{\beta}\right) \tag{4.29}$$

where he have defined so-called credibility factors,

$$\zeta_j = \frac{p_j\pi_{\leq J-j}}{p_j\pi_{\leq J-j} + \beta} \tag{4.30}$$

From (4.29) one sees that the estimate $\overline{\Theta}_j$ is a weighted average of a Chain-ladder type estimate of θ_j and the prior mean. The resulting credibility predictor of outstanding claims,

$$\overline{N}_{jd} = p_j\overline{\Theta}_j\pi_d \tag{4.31}$$

is a blend of a Bornhuetter-Ferguson type estimate and a Chain-ladder type estimate. The Bornhuetter-Ferguson method is a limiting case for $\beta \to \infty$ (no variation in claim frequencies), while the Chain Ladder method is a limiting case for $\beta \to 0$ (no statistical likeness between claim frequencies).

The mean squared error of prediction of the predictor $\overline{N}_{jd}$ for $j+d > J$ is

$$\begin{aligned}
\mathrm{E}\left(\left(N_{jd} - \overline{N}_{jd}\right)^2|\mathbf{D}_j\right) &= \mathrm{E}\left(\left(N_{jd} - \overline{N}_{jd}\right)^2|\mathbf{D}_j\right) \\
&= \mathrm{E}\left(\left(N_{jd} - \mathrm{E}(N_{jd}|\mathbf{D}_j)\right)^2|\mathbf{D}_j\right) \\
&\quad \mathrm{Var}\left(N_{jd}|\mathbf{D}_j\right) \\
&\quad \mathrm{E}\left(\mathrm{Var}\left(N_{jd}|\Theta_j\right)|\mathbf{D}_j\right) + \mathrm{Var}\left(\mathrm{E}\left(N_{jd}|\Theta_j\right)|\mathbf{D}_j\right) \\
&\quad \mathrm{E}\left(p_j\Theta_j\pi_d|\mathbf{D}_j\right) + \mathrm{Var}\left(p_j\Theta_j\pi_d|\mathbf{D}_j\right) \\
&\quad p_j\pi_d\tfrac{\alpha+N_{j,\leq J-j}}{\beta+p_j\pi_{\leq J-j}} + \left(p_j\pi_d\right)^2\tfrac{\alpha+N_{j,\leq J-j}}{\left(\beta+p_j\pi_{\leq J-j}\right)^2}
\end{aligned} \tag{4.32}$$

4.6.2 Parameter estimation

In a genuine Bayesian framework, the gamma prior and posterior distributions represent uncertainty about the true values of the claim frequencies $\Theta_1, \cdots, \Theta_J$.

In an empirical Bayesian framework, the gamma distributions have a frequentist interpretation and, as a consequence, one may attempt to estimate the parameters (α, β). This can be done in a two-step procedure:
a) First, estimate the delay probabilities $\pi_0, \cdots, \pi_D$;
b) Then estimate the parameters (α, β) in the gamma distribution, treating the estimated delay probabilities $\pi_0^*, \cdots, \pi_D^*$ as fixed parameters.

In step (b) one can estimate the parameters (α, β) by maximising the unconditional likelihood function of the data, which is a product of negative binomial densities:

$$\begin{aligned} L &= \prod_{j=1}^{J} \int_0^\infty \prod_{d=0}^{J-j} \frac{(p_j\theta_j\pi_d)^{N_{jd}}}{N_{jd}!} e^{-p_j\theta_j\pi_d} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_j^{\alpha-1} e^{-\beta_j\theta_j} d\theta_j \\ &\propto \prod_{j=1}^{J} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma\left(\alpha+N_{j,\le J-j}\right)}{\left(\beta+p_j\pi_{\le J-j}\right)^{\alpha+N_{j,\le J-j}}} \end{aligned} \tag{4.33}$$

If the distribution of $\Theta_1, \cdots, \Theta_J$ is not assumed to be a gamma distribution, the derivation of the conditional distribution of Θ_j, given the observations, becomes much harder.

4.7 Varying claim frequency: The Bühlmann-Straub model

4.7.1 The model

In previous section we assumed that the claim frequencies $\Theta_1, \cdots, \Theta_J$ have a gamma distribution, which allowed us to use a Bayesian argument to arrive at the formula (4.29). The credibility factors could also be expressed in terms of the mean $\tau = \alpha/\beta$ and variance $\lambda = \alpha/\beta^2$ in the gamma distribution, in the following way:

$$\zeta_j = \frac{p_j\pi_{\le J-j}\lambda}{p_j\pi_{\le J-j}\lambda + \tau} \tag{4.34}$$

There is another theoretical avenue to the formulas (4.29) and (4.31), using linear least squares credibility theory. Assume that

1. The claim frequencies $\Theta_1, \cdots, \Theta_J$ are independent and identically distributed with mean $\tau = \mathrm{E}(\Theta_j)$ and variance $\lambda = \mathrm{Var}(\Theta_j)$. Assume for the present that the mean $\tau = \mathrm{E}(\Theta_j)$ and the variance $\lambda = \mathrm{Var}(\Theta_j)$ are known.
2. Conditional on Θ_j, the reported claim numbers N_{jd} are independent, and Poisson distributed, $N_{jd} \mid \Theta_j = \theta_j \sim \mathrm{Poisson}(p_j\theta_j\pi_d)$.
3. Variables belonging to different accident periods are stochastically independent.

4. The estimator of Θ_j is restricted to be linear and in the form: $\overline{\Theta}_j = z_j\left(\frac{N_{j,\leq J-j}}{p_j\pi_{\leq J-j}}\right) + (1-z_j)\,\tau$.
5. The predictor of N_{jd} for $j+d>J$ is $\overline{N}_{jd} = p_j\overline{\Theta}_j\pi_d$.

Lemma 3 *For an* arbitrary, non-random *value of z_j, the estimator $\overline{\Theta}_j$ has the following mean squared error,*

$$Q(z_j) = E\left(\overline{\Theta}_j - \Theta_j\right)^2 = z_j^2\frac{\tau}{p_j\pi_{\leq J-j}} + (1-z_j)^2\lambda \tag{4.35}$$

and the mean squared error of prediction of the predictor $\overline{N}_{jd}$ is

$$E\left(\overline{N}_{jd} - N_{jd}\right)^2 = (p_j\pi_d)^2\,Q(z_j) + p_j\pi_d\tau \tag{4.36}$$

and

$$E\left(\overline{N}_{j,>J-j} - N_{j,>J-j}\right)^2 = (p_j\pi_{>J-j})^2\,Q(z_j) + p_j\pi_{>J-j}\tau \tag{4.37}$$

and

$$E\left(\sum_{j=1}^{J}\left(\overline{N}_{j,>J-j} - N_{j,>J-j}\right)\right)^2 = \sum_{j=1}^{J}\left((p_j\pi_{>J-j})^2\,Q(z_j) + p_j\pi_{>J-j}\tau\right) \tag{4.38}$$

Proof. We start by proving (4.35).

$$\begin{aligned} Q(z_j) &= \mathrm{E}\left(z_j\frac{N_{j,\leq J-j}}{p_j\pi_{\leq J-j}} + (1-z_j)\,\tau - \Theta_j\right)^2 \\ &= \mathrm{E}\left(z_j\left(\frac{N_{j,\leq J-j}}{p_j\pi_{\leq J-j}} - \Theta_j\right) + (1-z_j)\left(\tau-\Theta_j\right)\right)^2 \\ &= z_j^2\mathrm{E}\left(\frac{N_{j,\leq J-j}}{p_j\pi_{\leq J-j}} - \Theta_j\right)^2 + (1-z_j)^2\,\mathrm{E}\left(\tau-\Theta_j\right)^2 \\ &+ 2z_j(1-z_j)\,\mathrm{E}\left(\frac{N_{j,\leq J-j}}{p_j\pi_{\leq J-j}} - \Theta_j\right)\left(\tau-\Theta_j\right) \end{aligned} \tag{4.39}$$

Now we calculate

$$\begin{aligned} \mathrm{E}\left(\frac{N_{j,\leq J-j}}{p_j\pi_{\leq J-j}} - \Theta_j\right)^2 &= \mathrm{EE}\left(\left(\frac{N_{j,\leq J-j}}{p_j\pi_{\leq J-j}} - \Theta_j\right)^2|\Theta_j\right) \\ &= \mathrm{EVar}\left(\frac{N_{j,\leq J-j}}{p_j\pi_{\leq J-j}}|\Theta_j\right) \\ &= \frac{\tau}{p_j\pi_{\leq J-j}} \end{aligned} \tag{4.40}$$

and

$$\mathrm{E}\left(\tau-\Theta_j\right)^2 = \lambda \tag{4.41}$$

and

$$\mathrm{E}\left(\frac{N_{j,\leq J-j}}{p_j\pi_{\leq J-j}}-\Theta_j\right)(\tau-\Theta_j)=\mathrm{EE}\left(\left(\frac{N_{j,\leq J-j}}{p_j\pi_{\leq J-j}}-\Theta_j\right)(\tau-\Theta_j)\,|\Theta_j\right)=0 \tag{4.42}$$

This gives us (4.35). To prove (4.36) we can write

$$\begin{aligned}\mathrm{E}\left(\overline{N}_{jd}-N_{jd}\right)^2 &= \mathrm{E}\left(p_j\overline{\Theta}_j\pi_d-N_{jd}\right)^2\\ &= \mathrm{E}\left(p_j\overline{\Theta}_j\pi_d-p_j\Theta_j\pi_d+p_j\Theta_j\pi_d-N_{jd}\right)^2\\ &= \mathrm{E}\left(p_j\overline{\Theta}_j\pi_d-p_j\Theta_j\pi_d\right)^2+\mathrm{E}\left(p_j\Theta_j\pi_d-N_{jd}\right)^2\\ &+ 2\mathrm{E}\left(p_j\overline{\Theta}_j\pi_d-p_j\Theta_j\pi_d\right)\left(p_j\Theta_j\pi_d-N_{jd}\right)\end{aligned} \tag{4.43}$$

Now we calculate

$$\mathrm{E}\left(p_j\overline{\Theta}_j\pi_d-p_j\Theta_j\pi_d\right)^2=(p_j\pi_d)^2\,\mathrm{E}\left(\overline{\Theta}_j-\Theta_j\right)^2=(p_j\pi_d)^2\,Q(z_j) \tag{4.44}$$

and

$$\begin{aligned}\mathrm{E}\left(p_j\Theta_j\pi_d-N_{jd}\right)^2 &= \mathrm{EE}\left(\left(p_j\Theta_j\pi_d-N_{jd}\right)^2|\Theta_j\right)\\ &= \mathrm{EVar}\left(N_{jd}|\Theta_j\right)\\ &= \mathrm{E}\left(p_j\Theta_j\pi_d\right)\\ &= p_j\pi_d\tau\end{aligned} \tag{4.45}$$

and

$$\mathrm{E}\left(p_j\overline{\Theta}_j\pi_d-p_j\Theta_j\pi_d\right)\left(p_j\Theta_j\pi_d-N_{jd}\right)=\mathrm{EE}\left(\left(p_j\overline{\Theta}_j\pi_d-p_j\Theta_j\pi_d\right)\left(p_j\Theta_j\pi_d-N_{jd}\right)|\Theta_j\right)=0 \tag{4.46}$$

The last equation holds because $\overline{\Theta}_j$ and N_{jd} are conditionally independent, given Θ_j, for $j+d>J$.

The equation (4.37) follows similarly, and (4.38) follows from the assumed independence of different accident years' claim development. ■

An arbitrary value of z_j could be, for example, $z_j=1$ (Chain-ladder) or $z_j=0$ (Bornhuetter-Ferguson) or another value between zero and one. The question naturally arises what the best choice of z_j would be within the constraints of the model. Minimising (4.35) it is easy to see that the *optimal* choice of z_j is exactly the one defined in (4.34).

4.7.2 Parameter estimation

We start with the following

Lemma 4 *Define chain ladder estimates* $\hat{\theta}_j=N_{j,\leq J-j}/p_j\pi^*_{\leq J-j}$ *and*

$$\tau^*=\left(\sum_{j=1}^{J}\zeta_j\right)^{-1}\sum_{j=1}^{J}\zeta_j\hat{\theta}_j \tag{4.47}$$

and

$$\lambda^* = (J-1)^{-1} \sum_{j=1}^{J} \zeta_j \left(\hat{\theta}_j - \tau^*\right)^2 \tag{4.48}$$

Under the assumptions of the Bühlmann-Straub model, the following equations hold

$$E(\tau^*) = \tau \tag{4.49}$$

and

$$E(\lambda^*) = \lambda \tag{4.50}$$

Proof. *The first equation (4.49) follows directly from*

$$\begin{aligned} E\left(\hat{\theta}_j\right) &= EE\left(N_{j,\leq J-j}|\Theta_j\right)/p_j\pi^*_{\leq J-j} \\ &= E\left(p_j\Theta_j\pi_{\leq J-j}\right)/p_j\pi^*_{\leq J-j} \\ &= \tau \end{aligned} \tag{4.51}$$

To prove the second equation (4.50), we first verify that

$$\begin{aligned} Var(\hat{\theta}_j) &= EVar\left(\hat{\theta}_j|\Theta_j\right) + VarE\left(\hat{\theta}_j|\Theta_j\right) \\ &= E\left(\frac{\Theta_j}{p_j\pi_{\leq J-j}}\right) + Var(\Theta_j) \\ &= \frac{\tau}{p_j\pi_{\leq J-j}} + \lambda \\ &= \lambda/\zeta_j \end{aligned} \tag{4.52}$$

Defining $\zeta = \sum_{j=1}^{J} \zeta_j$*, we find that*

$$\begin{aligned} Var(\tau^*) &= \zeta^{-2} \sum_{j=1}^{J} \zeta_j^2\, Var\left(\hat{\theta}_j\right) \\ &= \zeta^{-2} \sum_{j=1}^{J} \zeta_j^2 \lambda/\zeta_j \\ &= \lambda/\zeta \end{aligned} \tag{4.53}$$

and

$$\begin{aligned} Cov(\hat{\theta}_j, \tau^*) &= \zeta^{-1}\zeta_j\, Var\left(\hat{\theta}_j\right) \\ &= \zeta^{-1}\zeta_j\lambda/\zeta_j \\ &= \lambda/\zeta \end{aligned} \tag{4.54}$$

Thence we can verify that

$$
\begin{aligned}
E(\lambda^*) &= (J-1)^{-1}\sum_{j=1}^{J}\zeta_j E\left(\hat{\theta}_j-\tau^*\right)^2\\
&= (J-1)^{-1}\sum_{j=1}^{J}\zeta_j\, Var\left(\hat{\theta}_j-\tau^*\right)\\
&= (J-1)^{-1}\sum_{j=1}^{J}\zeta_j\left(Var\left(\hat{\theta}_j\right)+Var(\tau^*)-2Cov\left(\hat{\theta}_j,\tau^*\right)\right)\\
&= (J-1)^{-1}\sum_{j=1}^{J}\zeta_j\left(\lambda/\zeta_j+\lambda/\zeta-2\lambda/\zeta\right)\\
&= (J-1)^{-1}\lambda(J+1-2)\\
&= \lambda
\end{aligned}
\tag{4.55}
$$

■

The estimators (4.47) and (4.48) cannot be applied directly because the weights ζ_j involve the estimands τ and λ. They are therefore called pseudo-estimators. To estimate the parameters, one can proceed as follows: First, estimate the delay probabilities $\pi_0, \cdots, \pi_D$. Then apply the iteration method of De Vylder to estimate τ and λ, treating the delay probabilities as fixed. The iteration method goes a follows:

1. Pick starting values $\tau^*_{(0)}$ and $\lambda^*_{(0)}$. Set the iteration number to $i = 0$.
2. For $j = 1, \cdots, J$, calculate the Chain Ladder estimates $\hat{\theta}_j = N_{j,\le J-j}/p_j\pi^*_{\le J-j}$.
3. For $j = 1, \cdots, J$, calculate credibility factors $z_j^{(i)} = \frac{p_j\pi^*_{\le J-j}\lambda^*_{(i)}}{p_j\pi^*_{\le J-j}\lambda^*_{(i)}+\tau^*_{(i)}}$.
4. Calculate a new estimate of the mean $\tau^*_{(i+1)} = \sum_{j=1}^{J} z_j^{(i)}\hat{\theta}_j / \sum_{j=1}^{J} z_j^{(i)}$.
5. Calculate a new estimate of the variance $\lambda^*_{(i+1)} = (J-1)^{-1}\sum_{j=1}^{J} z_j^{(i)}\left(\hat{\theta}_j - \tau^*_{(i+1)}\right)^2$.

Repeat (3)-(5) until convergence is reached.

4.8 Varying claim frequency: A random walk model

The Bühlmann-Straub model represents a relaxation of the strong assumption underlying Bornhuetter-Ferguson's method (identical claim frequencies) and a tightening of the slack assumption underlying the chain ladder method (unrelated claim frequencies). Its underlying asumption is that claim frequencies are independent and identically distributed replicates from an underlying probability distribution. In particular, this means that for every new accident year the *à priori* claim frequency is τ.

In real-life situations, claim frequencies are neither constant over time nor independent, but behave like an auto-correlated time series. Claim frequencies are determined by a number of factors that change gradually over time, including underlying risk conditions, policy terms and deductibles, and the insureds´ overall propensity to make a claim. Therefore claim frequencies develop in a way that suggests that there should be some benefit in including data from previous accident periods, when one is trying to estimate the claim frequency in a given accident period.

A simple assumption is that the claim frequencies follow a random walk process:

$$\Theta_j = \Theta_{j-1} + \varepsilon_j \tag{4.56}$$

where $\varepsilon_1, \cdots, \varepsilon_J$ are independent and identically distributed error terms with mean zero and variance σ^2. Assume also, *pro forma*, that there exists an initial random variable Θ_0 that has mean $\tau = \mathrm{E}(\Theta_0)$ and variance $\sigma_0^2 = \mathrm{Var}(\Theta_0)$. Then it is easy to verify that the random vector of claim frequencies $\boldsymbol{\Theta}_J = (\Theta_1, \cdots, \Theta_J)'$ has mean

$$\boldsymbol{\tau}_J = \mathrm{E}\left(\boldsymbol{\Theta}_J\right) = \tau\,[1, \cdots, 1]' = \tau\mathbf{1} \tag{4.57}$$

and a covariance matrix

$$\boldsymbol{\Lambda}_J = \mathrm{Cov}\left(\boldsymbol{\Theta}_J\right) = [\lambda_{jk}]_{j,k=1,\cdots,J} \tag{4.58}$$

with elements $\lambda_{jk} = \sigma_0^2 + \min(j,k)\sigma^2$.

As before, let us assume that

$$N_{jd}|\Theta_j \sim \mathrm{Poisson}(p_j \Theta_j \pi_d) \tag{4.59}$$

and that the N_{jd} are stochastically independent, given $\boldsymbol{\Theta}_J$. Write

$$\hat{\Theta}_j = \frac{N_{j,\leq J-j}}{p_j \pi_{\leq J-j}} \text{ for } j = 1, \cdots, J \tag{4.60}$$

and verify that $\mathrm{E}(\hat{\boldsymbol{\Theta}}_J|\boldsymbol{\Theta}_J) = \boldsymbol{\Theta}_J$ and $\mathrm{Var}(\hat{\boldsymbol{\Theta}}_J|\boldsymbol{\Theta}_J) = \mathrm{diag}(\boldsymbol{\Theta}_j/p_j\pi_{\leq J-j})$. A linear estimator of $\boldsymbol{\Theta}_J$ is then

$$\overline{\boldsymbol{\Theta}}_J = \mathbf{Z}_J\hat{\boldsymbol{\Theta}}_J + (\mathbf{I} - \mathbf{Z}_J)\,\boldsymbol{\tau}_J \tag{4.61}$$

and, as we will see in the next section, the best linear estimator of $\boldsymbol{\Theta}_J$ is given when one uses the credibility matrix

$$\mathbf{Z}_J = \mathbf{\Lambda}_J \left(\mathbf{\Lambda}_J + \begin{bmatrix} \tau_1/p_1\pi_{\leq J-1} & 0 & \dots & 0 \\ 0 & \tau_2/p_2\pi_{\leq J-2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \tau_J/p_J\pi_{\leq 0} \end{bmatrix} \right)^{-1} \tag{4.62}$$

As in the the Bühlmann-Straub model, one has the choice between a subjectivist approach, where the mean τ and the variances σ_0^2 and σ^2 represent a prior guess and its uncertainty; and a frequentist approach, where τ and σ^2 may be estimated from the data (the prior variance σ_0^2 cannot be estimated on the basis of just one realisation of the process).

Assume that $\pi_0, \cdots, \pi_D$ have already been estimated. Empirical estimation of τ and σ^2 can then start with the relations

$$\mathrm{E}\left(\hat{\mathbf{\Theta}}_J\right) = \tau \cdot \mathbf{1} \tag{4.63}$$

and

$$\mathrm{E}\left(\hat{\Theta}_j - \hat{\Theta}_{j-1}\right)^2 = \sigma^2 + \tau\left(\frac{1}{p_j\pi_{\leq J-j}} + \frac{1}{p_{j-1}\pi_{\leq J-(j-1)}}\right) \tag{4.64}$$

for $j = 1, \cdots, J$. Given an estimate τ^* one could attempt to estimate σ^2 by

$$\sigma^{*2} = \left(\sum_{j=2}^{J} w_j\right)^{-1} \sum_{j=2}^{J} w_j \left(\left(\hat{\Theta}_j - \hat{\Theta}_{j-1}\right)^2 - \tau^*\left(\frac{1}{p_j\pi_{\leq J-j}} + \frac{1}{p_{j-1}\pi_{\leq J-(j-1)}}\right)\right) \tag{4.65}$$

using suitable weights w_j.

One could argue that strictly positive claim frequencies cannot be modelled as a random walk, as claim frequencies cannot become negative. Like all statistical models, the random walk model is only an approximation to reality. The purpose of the model is to allow for a pattern of varying claim frequencies, while at the same time retaining some transfer of information between consecutive accident periods.

One can develop more sophisticated models for the time series of claim frequencies. For example, if the basic time period is shorter than a period, it may be necessary to model seasonal variation. This can be done within the framework of section 4.9, at the expense of having to specify a larger number of model parameters.

4.9 Varying claim frequency: A general credibility model

We turn to a general credibility model. Let us therefore assume that the evolution of claim frequencies is governed by a stochastic process of some form. Denote the mean of the vector $\boldsymbol{\Theta}_J = (\Theta_1, \cdots, \Theta_J)'$ by $\boldsymbol{\tau}_J$ and its covariance matrix by $\boldsymbol{\Lambda}_J$. Assume for the present that $\boldsymbol{\tau}_J$ and $\boldsymbol{\Lambda}_J$ are known quantities.

We assume as before that conditional on unknown claim frequencies $(\Theta_1, \cdots, \Theta_J)$, the claim numbers N_{jd} are independent random variables, each with a Poisson distribution,

$$N_{jd} \mid \Theta_j = \theta_j \sim \text{Poisson}(p_j \theta_j \pi_d) \tag{4.66}$$

with fixed, non-negative delay probabilities $\{\pi_d : d = 0, 1, \cdots\}$ that add to one.

Define the diagonal matrix

$$\mathbf{V}_J = \begin{bmatrix} p_1 \pi_{\leq J-1} & 0 & \cdots & 0 \\ 0 & p_2 \pi_{\leq J-2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & p_J \pi_{\leq 0} \end{bmatrix} \tag{4.67}$$

At any time J, the vector of reported claim counts

$$\mathbf{N}_J = \begin{bmatrix} N_{1,\leq J-1} \\ N_{2,\leq J-2} \\ \vdots \\ N_{J,\leq 0} \end{bmatrix} \tag{4.68}$$

is linearly regressed on the vector of claim frequencies $\boldsymbol{\Theta}_J = (\Theta_1, \cdots, \Theta_J)'$ through the equation

$$\mathrm{E}\left(\mathbf{N}_J \mid \boldsymbol{\Theta}_J\right) = \mathbf{V}_J \boldsymbol{\Theta}_J \tag{4.69}$$

and has a covariance matrix given by

$$\mathrm{Cov}\left(\mathbf{N}_J \mid \boldsymbol{\Theta}_J\right) = \mathbf{V}_J \mathrm{diag}\left(\boldsymbol{\Theta}_J\right) \tag{4.70}$$

'Chain-ladder estimates' of the unknown claim frequencies $(\Theta_1, \cdots, \Theta_J)$ are given by

$$\hat{\boldsymbol{\Theta}}_J = \mathbf{V}_J^{-1} \mathbf{N}_J = \begin{bmatrix} N_{1,\leq J-1}/p_1 \pi_{\leq J-1} \\ N_{2,\leq J-2}/p_2 \pi_{\leq J-2} \\ \vdots \\ N_{J,\leq 0}/p_J \pi_{\leq 0} \end{bmatrix} \tag{4.71}$$

'Bornhuetter-Ferguson estimates' of the unknown claim frequencies $(\Theta_1, \cdots, \Theta_J)$ are given by the prior mean vector $\boldsymbol{\tau}_J$. A general linear mixture of the two types of estimate is

$$\overline{\boldsymbol{\Theta}}_J = \mathbf{Z}_J \hat{\boldsymbol{\Theta}}_J + (\mathbf{I} - \mathbf{Z}_J)\, \boldsymbol{\tau}_J \tag{4.72}$$

The mean squared error matrix of the estimator $\overline{\boldsymbol{\Theta}}_J$ is

$$\begin{aligned} \mathbf{Q}(\mathbf{Z}_J) &= \mathrm{E}\left(\overline{\boldsymbol{\Theta}}_J - \boldsymbol{\Theta}_J\right)\left(\overline{\boldsymbol{\Theta}}_J - \boldsymbol{\Theta}_J\right)' \\ &= \mathbf{Z}_J \mathbf{V}_J^{-1} \mathrm{diag}\,(\boldsymbol{\tau}_J)\, \mathbf{Z}_J' + (\mathbf{I} - \mathbf{Z}_J)\, \boldsymbol{\Lambda}_J\, (\mathbf{I} - \mathbf{Z}_J)' \end{aligned} \tag{4.73}$$

Using the apparatus of credibility theory, we know that the best linear estimator of $\boldsymbol{\Theta}_J$ based on the vector of observations $\mathbf{N}_J$ is given by the following choice of credibility matrix:

$$\mathbf{Z}_J = \boldsymbol{\Lambda}_J \left(\boldsymbol{\Lambda}_J + \mathbf{V}_J^{-1} \mathrm{diag}\,(\boldsymbol{\tau}_J)\right)^{-1} \tag{4.74}$$

Having thus estimated $(\Theta_1, \cdots, \Theta_J)$, the credibility predictor of the number of claims IBNR in respect of accident period j, is

$$\overline{N}_{j,>J-j} = p_j \overline{\Theta}_j \pi_{>J-j} \tag{4.75}$$

and its mean squared error is

$$\mathrm{E}\left(\overline{N}_{j,>J-j} - N_{j,>J-j}\right)^2 = (p_j \pi_{>J-j})^2 \left[\mathbf{Q}(\mathbf{Z}_J)\right]_{jj} + p_j \pi_{>J-j} \tau_j \tag{4.76}$$

The credibility predictor of the total number of claims IBNR is

$$\overline{N}_> = \sum_{j=1}^{J} p_j \overline{\Theta}_j \pi_{>J-j} \tag{4.77}$$

with mean squared error

$$\mathrm{E}\left(\overline{N}_> - N_>\right)^2 = \sum_{j=1}^{J} \sum_{j'=1}^{J} (p_j \pi_{>J-j}) \left[\mathbf{Q}(\mathbf{Z}_J)\right]_{jj'} (p_{j'} \pi_{>J-j'}) + \sum_{j=1}^{J} p_j \pi_{>J-j} \tau_j$$

Many different time series models can be formulated with the proper specification of the matrices $\boldsymbol{\tau}_J$ and $\boldsymbol{\Lambda}_J$. In the previous section we saw one possible specification, where claim frequencies perform a random walk. Rather than to derive an "explicit" form of the best linear estimator for every specification of the mean-covariance structure, it is much easier to treat (4.74) as the explicit form that covers all possible specifications.

4.10 The number of claims CBNI

Claims CBNI are claims that will occur under policies which are in force on the balance date and which will remain in force for some time after the balance date. If no policy has a term of longer than one period, all claims CBNI at the end of period J will occur during period $J+1$. Denote the number of claims CBNI at the end of period J by $N_{J+1|J}$ and the unexpired risk exposure by $p_{J+1|J}$.

Traditionally, the unearned premium was seen to be a sufficient provision for unexpired risk. Under newer accounting rules, if the actuary´s estimate of the unexpired risk exceeds the unearned premium, an additional provision must be made. That provision is alternatively called an "unexpired risk provision", or a "premium deficiency provision". I do not know of any accounting standard that allows a provision for claims CBNI that is less than the unearned premium. Anyway, the actuary needs an estimate of the number and the amount of claims CBNI.

In the current model framework, the best estimate of the number of claims CBNI becomes

$$\overline{N}_{J+1|J} = p_{J+1|J}\overline{\Theta}_{J+1|J}, \tag{4.78}$$

where $\overline{\Theta}_{J+1|J}$ is a predictor of Θ_{J+1} that is based on data up to time J.

Within the Bornhuetter-Ferguson model with constant claim frequencies, a natural course of action is to equate $\overline{\Theta}_{J+1|J}$ to the previously estimated $\overline{\Theta}_1 = \cdots = \overline{\Theta}_J$.

Within the chain ladder model, one has no guidance as to what one should think of the accident period $j+1$. Being a pragmatist, one could of course set $\overline{\Theta}_{J+1|J} = \overline{\Theta}_J$, believing that risk conditions next period should not be too different from risk conditions this period. But in doing so, one would base one´s expectation for the period $J+1$ on the estimate that has the least amount of data to substantiate it. Alternatively, one could set $\overline{\Theta}_{J+1|J}$ =average$\left(\overline{\Theta}_1, \cdots, \overline{\Theta}_J\right)$ or some moving average of the latest periods. None of these fixes, however practical, has any theoretical foundation in the assumptions underlying of the chain ladder model.

Within the Bühlmann-Straub model, the theoretically right estimate would be $\overline{\Theta}_{J+1|J} = \tau$. The mean squared error of prediction would be

$$E\left(\overline{N}_{J+1|J} - N_{J+1|J}\right)^2 = p_{J+1|J}^2\lambda + p_{J+1|J}\tau. \tag{4.79}$$

Within the random walk model, the theoretically right estimate is $\overline{\Theta}_{J+1|J} = \overline{\Theta}_J$. However, unlike in the chain ladder model, the estimate will be based on the entire claim history, not just the claims of the last period. The mean squared error of prediction would be

$$E\left(\overline{N}_{J+1|J} - N_{J+1|J}\right)^2 = p_{J+1|J}^2\left(\left[\mathbf{Q}\left(\mathbf{Z}_J\right)\right]_{JJ} + \sigma^2\right) + p_{J+1|J}\tau. \tag{4.80}$$

See chapter 10 for more examples of modelling the claims frequency process.

4.11 Measures of risk exposure

We have defined p_j as a measure of risk exposure. Let us briefly consider what that could be.

In lines of business with fairly homogeneous risks, the number of risks or number of policies will normally be adequate. For example domestic property insurance, motor vehicle insurance, travel insurance, accident insurance, health insurance.

In lines of business where individual policies can cover large or small collectives, one should use the number of risks covered, like insured persons or labour-years. Example: Workers Compensation insurance, motor vehicle fleet insurance, or any other collective form of insurance where an insurance can cover a variable number of otherwise homogenous risks.

In some lines of business it is difficult to quantify the amount of risk. Examples are liability insurance and most business insurances. There one can use the premium as a proxy, but only if the average premium rate has been reasonably steady over time.

If the line of business generates a substantial number of claims that are notified in the accident period ($d = 0$), then $p_j = N_{j0}$ can be used as a proxy measure of risk exposed. This method is very useful in practice when other exposure measures are unavailable or unreliable.

4.12 Tail development factors

We assumed above that the claims belonging to the first accident period were fully notified, i.e. $\sum_{d=0}^{D} \pi_d = 1$. In practice it is very common to encounter lines of business where one must expect further claim notifications even for the oldest accident periods. In those cases one must extend the sequence of delay probabilities beyond $D = J - 1$. As there will be no data to base estimates on, the common procedure is to extrapolate the “tail” by some type of exponential curve.

If the delay probabilities are estimated as in the Chain-ladder method, using empirical development factors δ_d^*, a typical formula would be

$$\delta_d^* = 1 + \gamma \left(\delta_{d-1}^* - 1\right) \tag{4.81}$$

for $d > D$, with a suitably chosen value of γ. If the delay probabilities are estimated as in the Bornhuetter-Ferguson method, using delay-specific claim frequencies θ_d^*, a typical formula would be

$$\theta^*_d = \gamma \theta^*_{d-1}. \tag{4.82}$$

In both cases on would choose $\gamma \in [0, 1)$.

If the empirical development factors or claim frequencies exhibit erratic behaviour already at earlier delays, it may be appropriate to start the smoothing formula at an earlier delay than D.If necessary, different values of γ can be applied to different sections of the tail. Note that the outstanding claim estimates are very sensitive to the choice of tail factors, because whatever additional development one is assuming beyond the observed time horizon D, will affect every single accident year.

5

The cost of unreported claims

5.1 Introduction

This chapter treats the estimation of the cost of unreported claims, in a situation where it is possible to estimate separately the number of unreported claims.

5.2 The cost of claims IBNR

We now turn to estimating the cost of claims IBNR.

The severities of individual claims reported in period $j + d$ in respect of accidents incurred in period j we denote by $\left\{Y_{jd}^{(k)} : k = 1, \cdots, N_{jd}\right\}$. We assume that $Y_{jd}^{(k)}$ are independent random variables with a distribution G_d that may depend on the reporting delay d. We also assume that the severities are independent of the claim counts. Denote the mean and variance of $Y_{jd}^{(k)}$ by ξ_d and σ_d^2 , and let $\rho_d = \sigma_d^2 + \xi_d^2$ denote the non-central second order moment.

In the conditional distribution given $\mathbf{\Theta}_J = (\Theta_1, \cdots, \Theta_J)'$, the amounts $\{Y_{j,>J-j} : j = 1, \cdots, J\}$ of claims IBNR are independent random variables, and $Y_{j,>J-j}$ has a compound Poisson distribution with frequency parameter $p_j \theta_j \pi_{>J-j}$ and a mixed severity distribution (the tail severity distribution)

$$G_{\overline{>J-j}} = \pi_{>J-j}^{-1} \sum_{d=J-j+1}^{\infty} \pi_d G_d \tag{5.1}$$

We let the inequality subscript in conjunction with an overbar denote a π-weighted average. The non-central first and second order moments of the tail severity distribution are then

$$\xi_{\overline{>J-j}} = \pi_{>J-j}^{-1} \sum_{d=J-j+1}^{\infty} \pi_d \xi_d \tag{5.2}$$

and

$$\rho_{\overline{>J-j}} = \pi_{>J-j}^{-1} \sum_{d=J-j+1}^{\infty} \pi_d \rho_d \tag{5.3}$$

The credibility predictor of the amount of claims IBNR in respect of accidents incurred in period j, is then

$$\overline{Y}_{j,>J-j} = p_j \overline{\Theta}_j \pi_{>J-j} \xi_{\overline{>J-j}} \tag{5.4}$$

and its mean squared error is

$$\mathrm{E}\left(\overline{Y}_{j,>J-j} - Y_{j,>J-j}\right)^2 = \left(p_j \pi_{>J-j} \xi_{\overline{>J-j}}\right)^2 \left[\mathbf{Q}\left(\mathbf{Z}_J\right)\right]_{jj} + p_j \pi_{>J-j} \tau_j \rho_{\overline{>J-j}} \tag{5.5}$$

The credibility predictor of the total amount of claims IBNR in respect of all accident periods is

$$\overline{Y}_{>} = \sum_{j=1}^{J} p_j \overline{\Theta}_j \pi_{>J-j} \xi_{\overline{>J-j}} \tag{5.6}$$

with mean squared error

$$\begin{aligned} \mathrm{E}\left(\overline{N}_{>} - N_{>}\right)^2 &= \textstyle\sum_{j=1}^{J} \sum_{j'=1}^{J} \left(p_j \pi_{>J-j} \xi_{\overline{>J-j}}\right) \left[\mathbf{Q}\left(\mathbf{Z}_J\right)\right]_{jj'} \left(p_{j'} \pi_{>J-j'} \xi_{\overline{>J-j'}}\right) \\ &+ \textstyle\sum_{j=1}^{J} p_j \pi_{>J-j} \tau_j \rho_{\overline{>J-j}} \end{aligned} \tag{5.7}$$

5.3 The cost of claims CBNI

Let $G = \sum_{d=0}^{\infty} \pi_d G_d$ denote the probability distribution of all claims incurred in an accident period and denote by $\xi = \sum_{d=0}^{\infty} \pi_d \xi_d$ and $\rho = \sum_{d=0}^{\infty} \pi_d \rho_d$ its first and non-central second order moment.

An estimator of the amount of claims CBNI at time J is

$$\overline{Y}_{J+1|J} = p_{J+1|J}\overline{\Theta}_{J+1|J}\xi \tag{5.8}$$

and its mean squared error is

$$\mathrm{E}\left(\overline{Y}_{J+1|J} - Y_{J+1|J}\right)^2 = \left(p_{J+1|J}\xi\right)^2 \mathrm{E}\left(\overline{\Theta}_{J+1|J} - \Theta_{J+1}\right)^2 + p_{J+1|J}\tau_{J+1}\rho \tag{5.9}$$

The exact form of the mean squared error $\mathrm{E}\left(\overline{\Theta}_{J+1|J} - \Theta_{J+1}\right)^2$ and the expected claim frequency τ_{J+1} depends on the model that is used for the evolution of the claim frequencies.

5.4 Estimating the severity distribution

Until now we have assumed that the severity distributions G_d are known. They must be estimated too, of course, or at least their first and second order moments, $\xi_d = \mathrm{E}\left(Y_{jd}^{(k)}\right)$ and $\rho_d = \mathrm{E}\left(Y_{jd}^{(k)}\right)^2$. In this section we just briefly touch some of the practical issues.

Given the ultimate claim amounts $Y_{jd}^{(k)}$ for sufficiently many claims, the severity distribution or its moments can be estimated by a range of statistical techniques, ranging from the non-parametric to the fully parametric. Of parametric distributions, the lognormal or pareto are usually good candidates for severity distributions in insurance, with the caveat that the pareto distribution may not possess all the required moments.

As we noted in the previous sections, however, the ultimate claim amounts will normally not be known for all claims. One will then have to use some approximation, adjusting the results for any biases that they may contain.

- If reasonably reliable case estimates are available, one can base the estimation on the total case estimates, treating them as if they were the ultimate cost of the claims. Whether the average claim amount will need to be adjusted, and in what direction, depends on the quality of the case estimates.

- In an attempt to alleviate a problem with unreliable case estimates, one could omit, say, claims reported in the last period $(j + d = J)$ or the last two periods from the estimation. This procedure works well for short-tailed lines of insurance where claims are settled rapidly. For long-tailed lines, if one wanted to be 100% that sure that the case estimates were reliable or the claims settled, one would need to omit so many reporting periods that there was no data left.

- One could base the estimation on settled claims only. Unless one has a significant volume of claim data, this method is likely to underestimate the mean severity and the variability of claims, as small claims tend to be over-represented among settled claims.

- In order to avoid the bias inherent in using only settled claims, one could apply a procedure that consists of estimating separately the settlement pattern, and the severity distributions as a function of the time to settlement. The overall severity distribution can then be estimated as the weighted average of those distributions, with weights provided by the settlement pattern.

- Having to specify (potentially) a different severity distribution for each delay can require a great number of parameters, unless one is able to find a suitable parametric representation of the dependency between severity and reporting delay. Norberg (1999) proposes a joint probability distribution for severity and reporting delay in continuous time, that requires only three parameters (α, β, μ). In Norberg's model, $Y \sim \Gamma(\alpha, \beta)$ and $D|Y = y \sim \Gamma(1, \mu y)$, where Y denotes the claim amount and D denotes reporting delay in continuous time. An implication of Norberg's model is that large claims tend to be reported more promptly than small claims. This makes the model suitable for property insurance, less so for casualty insurance. An advantage of Norberg's model is that its discretised version is a mixture of gamma distributions - employing only three parameters (α, β, μ) - that is mathematically tractable.

Generally, estimating the severity distribution and its dependency on the reporting delay requires ad-hoc adjustments and a good deal of judgement.

6
The cost of reported claims

6.1 Introduction

Let us now turn to the problem of estimating the ultimate cost of a cohort of claims that has been reported in calendar period $j + d$ and was incurred in accident period j, where of course $j + d \leq J$. We know with certainty the number of claims that have been reported (N_{jd}) and any activity that has already been recorded on the claims. We pretend to know the ultimate cost of claims that are closed, but they could be reopened.

In this section two models will be proposed to estimate the ultimate cost of reported claims. One model is based on payments and the other model is based on reported claim cost, i.e. payments plus case estimates. Which basis one should choose, depends on the situation. This author normally prefers using reported claims, but in high-volume and reasonably short-tailed lines of business, using paid claims may work just as well. An interesting area for further research would be to formulate an elegant and tractable model of the joint development of payments and case estimates.

6.2 Estimating claims RBNS by payment data

6.2.1 Background

For the cohort of claims that has been reported in calendar period $j+d$ and was incurred in accident period j, we denote the payments at delay t after the reporting period by U_{jdt}. The unknown ultimate claim cost we denote by U_{jd} and the unknown average severity by Ξ_{jd}. Thus $U_{jd} = N_{jd}\Xi_{jd}$ is the product of an observed number of claims and an unobserved average severity,

Under the (reasonable) assumption that claim payments after the reporting date follow a certain pattern $\{\upsilon_t : t = 0, 1, \cdots\}$ with $\sum_{t=0}^{\infty} \upsilon_t = 1$, one can readily propose estimators of the outstanding cost of reported claims in the cell (j, d):

- Alluding to a Bornhuetter-Ferguson estimator, one could estimate the outstanding cost by $\overline{U}^{\text{"BF"}}_{jd,>J-(j+d)} = N_{jd}\xi_d\upsilon_{>J-(j+d)}$.
- Alluding to a Chain-ladder estimator, one could estimate the outstanding cost by $\overline{U}^{\text{"CL"}}_{jd,>J-(j+d)} = U_{jd,\leq J-(j+d)}\left(\frac{\upsilon_{>J-(j+d)}}{\upsilon_{\leq J-(j+d)}}\right)$.
- Alluding to credibility estimators, one could use a convex combination of the two above.

In order to evaluate the mean squared error and to find an optimal credibility estimator, further assumption are needed. Let us consider one possible set of assumptions about the evolution of claim payments after the reporting date.

6.2.2 The Dirichlet distribution

The Dirichlet distribution is a generalised Beta distribution on the $(n-1)$-dimensional simplex $\{x_1, \cdots, x_n \geq 0 : x_1 + \cdots + x_n = 1\}$, with density function

$$f(x_1, \cdots, x_n) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_n)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_n)} x_1^{\alpha_1 - 1} \cdots x_n^{\alpha_n - 1} \tag{6.1}$$

for $x_1, \cdots, x_n \geq 0$ and $x_1 + \cdots + x_n = 1$. Its parameters $\alpha_1, \alpha_2, \cdots$ are non-negative and fixed and sum to $\alpha > 0$. Let $\upsilon_t = \alpha_t/\alpha$.. Let us quickly recapitulate the first and second order moment structure of the Dirichlet distribution.

Lemma 5 *The first and second order moments of the Dirichlet distribution are*

$$E(X_t) = \frac{\alpha_t}{\alpha_t + (\alpha - \alpha_t)} = \upsilon_t \tag{6.2}$$

and

$$Var(X_t) = \frac{\alpha_t(\alpha - \alpha_t)}{\alpha^2(\alpha + 1)} = \frac{\upsilon_t(1 - \upsilon_t)}{\alpha + 1} \tag{6.3}$$

and

$$Cov(X_t, X_{t'}) = \frac{-\upsilon_t\upsilon_{t'}}{\alpha + 1}. \tag{6.4}$$

Proof. Observe that each component X_t in a Dirichlet distribution has a marginal Beta distribution with parameters $(\alpha_t, \alpha - \alpha_t)$, giving us 6.2 and 6.3. To establish the covariance $\text{Cov}(X_t, X_{t'})$ for $t \neq t'$, calculate

$$\begin{aligned}
\text{E}(X_t X_{t'}) &= \textstyle\int_{x_i \geq 0\,,\, x_1 + \cdots + x_n = 1} (x_t x_{t'}) f\left(x_1, \cdots, x_n\right) d(x_1, \cdots, x_{n-1}) \\
&= \tfrac{\Gamma(\alpha)}{\Gamma(\alpha_t)\Gamma(\alpha_{t'})} \cdot \tfrac{\Gamma(\alpha_t+1)\Gamma(\alpha_{t'}+1)}{\Gamma(\alpha+2)} \\
&= \tfrac{\alpha_t \alpha_{t'}}{\alpha(\alpha+1)} \\
&= \upsilon_t \upsilon_{t'} \tfrac{\alpha}{(\alpha+1)}
\end{aligned} \tag{6.5}$$

Subtracting $\text{E}(X_t)\text{E}(X_{t'})$ from (6.5) we find the expression 6.4. ■

6.2.3 *Model of payment pattern*

To model the development payment ratios $\left\{ \frac{U_{jdt}}{U_{jd}} : t = 0, 1, \cdots \right\}$, the Dirichlet distribution is an obvious candidate. Let us assume that the process of payment ratios after the reporting period has a Dirichlet distribution:

$$\left(\frac{U_{jd0}}{U_{jd}}, \frac{U_{jd1}}{U_{jd}}, \cdots \right) \mid N_{jd}, \Xi_{jd} \sim \text{Dirichlet}(\alpha_0, \alpha_1, \cdots) \tag{6.6}$$

with non-negative fixed parameters $\alpha_0, \alpha_1, \cdots$ summing to $\alpha > 0$. Please note that $U_{jd} = N_{jd}\Xi_{jd}$ is the unknown ultimate cost. Let $\upsilon_t = \alpha_t/\alpha$. The conditional moments of the partial payments are then

$$\text{E}\left(U_{jdt} \mid N_{jd}, \Xi_{jd}\right) = \upsilon_t N_{jd} \Xi_{jd} \tag{6.7}$$

and

$$\text{Cov}\left(U_{jdt}, U_{jdt'} \mid N_{jd}, \Xi_{jd}\right) = \left(\frac{\delta_{tt'}\upsilon_t - \upsilon_t \upsilon_{t'}}{\alpha + 1} \right) (N_{jd}\Xi_{jd})^2 \tag{6.8}$$

6.2.4 *Prediction of ultimate claims*

Conditional on only N_{jd} and before any payments have been recorded, the average severity Ξ_{jd} has a 'prior mean' of ξ_d and a variance of σ_d^2/N_{jd}. We now use the apparatus of credibility theory to find the best linear predictor of Ξ_{jd} in the conditional model, given N_{jd}.

Proposition 6 *Define a general linear predictor of the average severity by*

$$\overline{\Xi}_{jd} = z_{jd}\hat{\Xi}_{jd} + (1 - z_{jd})\,\xi_d \tag{6.9}$$

with 'Chain-ladder estimate'

$$\hat{\Xi}_{jd} = \frac{U_{jd,\leq J-(j+d)}}{N_{jd}\upsilon_{\leq J-(j+d)}} \tag{6.10}$$

The conditional mean squared error of the predictor (6.9), given the number of claims N_{jd}, is

$$\begin{aligned} q_d\left(z_{jd} \mid N_{jd}\right) &= E\left(\left(\overline{\Xi}_{jd} - \Xi_{jd}\right)^2 \mid N_{jd}\right) \\ &= N_{jd}^{-1}\left(z_{jd}^2 \tfrac{\left(\sigma_d^2 + N_{jd}\xi_d^2\right)\upsilon_{>J-(j+d)}}{(\alpha+1)\upsilon_{\leq J-(j+d)}} + \left(1 - z_{jd}\right)^2 \sigma_d^2\right) \end{aligned} \tag{6.11}$$

A linear predictor of the outstanding payments is then

$$\overline{U}_{jd,>J-(j+d)} = N_{jd}\overline{\Xi}_{jd} - U_{jd,\leq J-(j+d)} \tag{6.12}$$

with conditional mean squared error

$$E\left(\left(\overline{U}_{jd,>J-(j+d)} - U_{jd,>J-(j+d)}\right)^2 \mid N_{jd}\right) = N_{jd}^2 \cdot q_d\left(z_{jd} \mid N_{jd}\right) \tag{6.13}$$

Due to the independence between the different cohorts, the mean squared error of the overall amount of outstanding payments for reported claims is additive. The conditional mean squared error is minimised if one chooses the credibility factors in the following way:

$$z_{jd} = \frac{\sigma_d^2\left(\alpha+1\right)\upsilon_{\leq J-(j+d)}}{\sigma_d^2\left(\alpha+1\right)\upsilon_{\leq J-(j+d)} + \left(\sigma_d^2 + N_{jd}\xi_d^2\right)\upsilon_{>J-(j+d)}} \tag{6.14}$$

Proof. In order to prove 6.9 to 6.13, let us simplify the notation a bit.

We omit the indexes j and d, as every combination of accident period and reporting delay is being considered separately. Thus the reported number of claims is now denoted by N, the unknown ultimate cost is denoted by U, and the unknown average severity is denoted by Ξ . Let us now write the observed payments as a vector $\mathbf{U} = \left(U_0, U_1, \cdots, U_{J-(j+d)}\right)'$ and the corresponding payment ratios as $\boldsymbol{\upsilon} = \left(\upsilon_0, \upsilon_1, \cdots, \upsilon_{J-(j+d)}\right)'$. Further, let $\mathbf{1}^{(J-(j+d)+1)\times 1} = (1, \cdots, 1)'$ and verify that $\mathbf{1}'\boldsymbol{\upsilon} = \upsilon_{\leq J-(j+d)}$, the accumulated payment proportion which we simply denote by $\upsilon_{\leq}$. Its complement we denote by $\upsilon_{>} = 1 - \upsilon_{\leq}$. Now please verify that, given the number of reported claims and the unknown average severity, the vector of observed payments has conditional expectation

$$\mathrm{E}\left(\mathbf{U} \mid N, \Xi\right) = N\Xi\boldsymbol{\upsilon} \tag{6.15}$$

and conditional variance

$$\mathrm{Var}\left(\mathbf{U} \mid N, \Xi\right) = (N\Xi)^2 \frac{1}{\alpha+1}\left(\mathrm{diag}(\boldsymbol{\upsilon}) - \boldsymbol{\upsilon}\boldsymbol{\upsilon}'\right) = (N\Xi)^2 \frac{1}{\alpha+1}\left(\mathbf{I} - \boldsymbol{\upsilon}\mathbf{1}'\right)\mathrm{diag}(\boldsymbol{\upsilon}) \tag{6.16}$$

To allude to the notation used in Appendix A, we write

$$\mathrm{E}\left((\mathbf{U} \mid N, \Xi) \mid N\right) = N\boldsymbol{\upsilon}\Xi = \mathbf{Y}^{(J-(j+d)+1)\times 1}\Xi \tag{6.17}$$

$$\boldsymbol{\Phi} = \mathrm{E}\left(\mathrm{Var}\left(\mathbf{U} \mid N, \Xi\right) \mid N\right) = (N\sigma^2 + N^2\xi^2)\frac{1}{\alpha+1}\left(\mathbf{I} - \boldsymbol{\upsilon}\mathbf{1}'\right)\mathrm{diag}(\boldsymbol{\upsilon}) \tag{6.18}$$

and note that the conditional mean and variance of Ξ, given N, are

$$\mathrm{E}\left(\Xi \mid N\right) = \xi \tag{6.19}$$

(remember we dropped the subscript d),

$$\mathrm{Var}\left(\Xi \mid N\right) = \sigma^2/N \tag{6.20}$$

Thus the conditional model (given N) fits into the mould of the credibility regression model of Appendix A. To find the credibility estimator of Ξ, we need only to go through the motions of Appendix A.

Using (15.15) it is easy to prove that

$$\left(\mathbf{I} - \boldsymbol{\upsilon}\mathbf{1}'\right)^{-1} = \mathbf{I} + \frac{1}{\upsilon_>}\boldsymbol{\upsilon}\mathbf{1}' \tag{6.21}$$

Therefore,

$$\boldsymbol{\Phi}^{-1} = \frac{\alpha+1}{N\sigma^2 + N^2\xi^2}\left(\mathrm{diag}^{-1}(\boldsymbol{\upsilon}) + \frac{1}{\upsilon_>}\mathbf{1}\mathbf{1}'\right) \tag{6.22}$$

$$\mathbf{Y}'^{-1}\boldsymbol{\Phi}^{-1}\mathbf{U} = \frac{\alpha+1}{\sigma^2 + N\xi^2} \cdot \frac{U_{\leq}}{\upsilon_>} \tag{6.23}$$

$$\mathbf{Y}'^{-1}\boldsymbol{\Phi}^{-1}\mathbf{Y} = \frac{\alpha+1}{\sigma^2 + N\xi^2} \cdot \frac{N\upsilon_{\leq}}{\upsilon_>} \tag{6.24}$$

$$\hat{\Xi} = \frac{U_{\leq}}{N\upsilon_{\leq}} \tag{6.25}$$

and

$$z = \frac{(\sigma^2/N)\mathbf{Y}'^{-1}\boldsymbol{\Phi}^{-1}\mathbf{Y}}{1 + (\sigma^2/N)\mathbf{Y}'^{-1}\boldsymbol{\Phi}^{-1}\mathbf{Y}} = \frac{\sigma^2(\alpha+1)\upsilon_{\leq}}{\sigma^2(\alpha+1)\upsilon_{\leq} + \left(\sigma^2 + N\xi^2\right)\,\upsilon_>} \tag{6.26}$$

Equation 6.11 is (15.14) with the appropriate substitutions. This concludes the proof. ■

The assumption of the payment pattern being the same for claims at all notification delays, is not necessarily realistic. To see why this need not be the case, contrast claims notified in the accident period ($d = 0$) with claims notified in the subsequent period ($d = 1$). If accidents are spread evenly over the accident period, claim notifications in the accident period will be skewed towards the end of the period because of the notification delay. On the other hand, unless the reporting pattern is very flat-tailed, claim notifications in the subsequent period will occur mostly at the start of the period before they start tailing off. Thus on average, claims that are reported in the accident period will have less time for the first batch of payments ($t = 0$) to be processed, than claims reported in the subsequent period. Therefore one should expect that υ_0 is smaller for $d = 0$ than for $d = 1$. The formulas above extend readily to a model with payment patterns that depend on d. However, this comes at the expense of having to set more parameters.

6.2.5 Estimation of parameters

Let us briefly consider the estimation of the parameters $\alpha_0, \alpha_1, \cdots$ by a maximum likelihood method. For each combination of $j = 1, \cdots, J$ and $d = 0, \cdots, J - j$, define the payment-ratios-to-date

$$A_{jdt} = U_{jdt}/U_{jd,\leq J-(j+d)} \text{ for } t = 0, \cdots, J-(j+d) \tag{6.27}$$

and note that the vectors $\mathbf{A}_{jd} = \left(A_{jd0}, \cdots, A_{jd,J-(j+d)}\right)$ are stochastically independent and that

$$\mathbf{A}_{jd} \sim \text{Dirichlet}(\alpha_0, \cdots, \alpha_{J-(j+d)}) \tag{6.28}$$

Thus the likelihood function of the set of payment-ratios-to-date is

$$L = \prod_{j=1}^{J} \prod_{d=0}^{J-j} \frac{\Gamma\left(\alpha_{\leq J-(j+d)}\right)}{\Gamma\left(\alpha_0\right) \cdots \Gamma\left(\alpha_{J-(j+d)}\right)} \prod_{t=0}^{J-(j+d)} A_{jdt}^{\alpha_t - 1} \tag{6.29}$$

Then we can calculate

$$\frac{\partial \ln(L)}{\partial \alpha_s} = \sum_{j=1}^{J-s} \sum_{d=0}^{J-s-j} \psi\left(\alpha_{\leq J-(j+d)}\right) - n_s \psi\left(\alpha_s\right) + \sum_{j=1}^{J-s} \sum_{d=0}^{J-s-j} Z_{jds} \tag{6.30}$$

Here we have used the digamma function $\psi(x) = \Gamma'(x)/\Gamma(x)$ and defined

$$n_s = \sum_{j=1}^{J-s} \sum_{d=0}^{J-s-j} 1 = \frac{(J-s)(J-s+1)}{2} \tag{6.31}$$

as well as

DRAFT
09.04.14

$$Z_{jds} = \ln(A_{jds}) \tag{6.32}$$

The solution to $\frac{\partial \ln(L)}{\partial \alpha_s} = 0$ for $s = 0, \cdots, J-1$, if it exists, is given by the equations

$$\psi\left(\alpha_s^*\right) = n_s^{-1} \sum_{j=1}^{J-s} \sum_{d=0}^{J-s-j} \psi\left(\alpha^*_{\leq J-(j+d)}\right) + \overline{Z}_{\cdot\cdot s} \text{ for } s = 0, \cdots, J-1 \tag{6.33}$$

Here $\overline{Z}_{\cdot\cdot s} = n_s^{-1} \sum_{j=1}^{J-s} \sum_{d=0}^{J-s-j} Z_{jds}$ denotes the average of $\ln\left(A_{jds}\right)$, for the cells where payment delay s has been observed.

The Hessian matrix $\mathbf{H}(\boldsymbol{\alpha})$ of $\ln(L)$ is given by its elements

$$\begin{aligned} h_{st}(\boldsymbol{\alpha}) &= \frac{\partial^2 \ln(L)}{\partial \alpha_s \partial \alpha_t} \\ &= \sum_{j=1}^{J-\max(s,t)} \left(\sum_{d=0}^{J-\max(s,t)-j} \psi'\left(\alpha_{\leq J-(j+d)}\right)\right) - \delta_{st} n_s \psi'\left(\alpha_s\right) \end{aligned} \tag{6.34}$$

Since ψ' is a non-negative, decreasing and convex function, this author suspects that the matrix $\mathbf{H}(\boldsymbol{\alpha}^*)$ is negative definite, which would prove that (6.33) defines a maximum.

Providing a solution to (6.33) exists, it should be possible to evaluate numerically. One could also try to determine $\boldsymbol{\alpha}^*$ by a Newton-Raphson iteration of the form

$$\boldsymbol{\alpha}^*_{i+1} = \boldsymbol{\alpha}^*_i - \mathbf{H}^{-1}(\boldsymbol{\alpha}^*_i)\left(\frac{\partial \ln(L)}{\partial \boldsymbol{\alpha}}(\boldsymbol{\alpha}^*_i)\right) \tag{6.35}$$

The author has not tried to do this in practice yet.

6.3 Estimating claims RBNS by incurred data

6.3.1 Background

For the cohort of claims that has been reported in calendar period $j+d$ and was incurred in accident period j, we denote the change in the reported incurred claim amount at delay t after the reporting period by W_{jdt}. We denote the unknown ultimate claim cost by W_{jd} and the unknown average severity by Ξ_{jd}.

Under the (reasonable) assumption that claim reports after the reporting date follow a certain pattern $\{\omega_t : t = 0, 1, \cdots\}$ with $\sum_{t=0}^{\infty} \omega_t = 1$, one can readily propose estimators of the outstanding cost of reported claims in the cell (j, d):

- Alluding to a Bornhuetter-Ferguson estimator, one could estimate the outstanding cost by $\overline{W}^{\text{"BF"}}_{jd,>J-(j+d)} = N_{jd}\xi_d\omega_{>J-(j+d)}$.
- Alluding to a Chain-ladder estimator, one could estimate the outstanding cost by $\overline{W}^{\text{"CL"}}_{jd,>J-(j+d)} = W_{jd,\leq J-(j+d)}\left(\frac{\omega_{>J-(j+d)}}{\omega_{\leq J-(j+d)}}\right)$.
- Alluding to credibility estimators, one could use a convex combination of the two above.

In order to evaluate the mean squared error and to find an optimal credibility estimator, further assumption are needed. Let us consider one possible set of assumptions about the evolution of claim payments after the reporting date.

To model the development of

$$\{W_{jdt} : t = 0, 1, \cdots\} \tag{6.36}$$

conditionally on the number of claims and the unknown average severity, one needs a distribution that allows negative as well as positive increments. That requirement excludes the Dirichlet model that we used to model payment ratios. An alternative model is proposed in what follows.

6.3.2 Compound poisson model of reporting events

Consider the following model: given the number of reported claims N_{jd} and the average claim amount Ξ_{jd}, we assume that the W_{jdt} at different delays t are conditionally independent, and that W_{jdt} is a compound Poisson random variable with a frequency parameter that is proportional to $N_{jd}\Xi_{jd}$ and a jump size distribution H_t that allows negative jumps:

$$W_{jdt} \mid N_{jd}, \Xi_{jd} \sim \text{Compound Poisson } (N_{jd}\Xi_{jd}, H_t) \tag{6.37}$$

The assumption (6.37) implies that the expected number of claim reassessments at delay t (a claim reassessment being a partial payment and/or a change to the outstanding case estimate) is proportional to the unknown overall claim amount $W_{jd} = N_{jd}\Xi_{jd}$, and that the individual reassessments have a size distribution H_t. Let us briefly discuss this assumption.

To assume that the expected number of claim reassessments is proportional to the number of claims reported, is quite reasonable. To assume that it is actually proportional not to the number of claims, but to the amount of claims, stretches the imagination a bit more. That assumption could be wrong, but it could also be approximately right. It will be postulated here that it is approximately right, because this assumption makes for tractable mathematics. That does not imply that the expected number of claim reassessments must be equal to the aggregate claim amount (expressed in some currency or other); it is only the proportionality that counts. The distribution function H_t will have a high point mass at zero, so that the number of actual claim reassessments with a jump, will be much smaller. One could generate the same compound Poisson distribution using a different model formulation with an explicit proportionality factor in the claim frequency parameter and a distribution function H_t that is strictly non-zero.

Also note that we are not constraining the aggregate claim development to equal the aggregate severity, i.e. we are not demanding that $\sum_{t=0}^{\infty} W_{jdt} = N_{jd}\Xi_{jd}$, as we did in the payment model. Thus the aggregate severity takes on the role of the expected level of ultimate payments, given the (abstract) severities of claims reported, rather than the definitive level of ultimate payments.

Denote the non-central first and second order moments of the distribution H_t by

$$\omega_t = \int_{-\infty}^{\infty} u \, dH_t(u) \tag{6.38}$$

and

$$\eta_t = \int_{-\infty}^{\infty} u^2 \, dH_t(u) \tag{6.39}$$

This formalistic definition (using integrals from $-\infty$ to ∞) is made only to emphasise that the distributions H_t must allow negative jumps.

Then we can easily establish the following conditional moments:

$$\mathrm{E}\left(W_{jdt} \mid N_{jd}, \Xi_{jd}\right) = N_{jd}\Xi_{jd}\omega_t \tag{6.40}$$

and

$$\mathrm{Var}\left(W_{jdt} \mid N_{jd}, \Xi_{jd}\right) = N_{jd}\Xi_{jd}\eta_t \tag{6.41}$$

We are assuming that $\sum_{t=0}^{\infty} \omega_t = 1$ and $\sum_{t=0}^{\infty} \eta_t < \infty$, but not all ω_t need to be non-negative.

6.3.3 Prediction of ultimate claims

Proposition 7 *Conditional on only N_{jd} and before any payments have been recorded, the average severity Ξ_{jd} has a 'prior mean' of ξ_d and a variance of σ_d^2/N_{jd}. This leads to the following proposition.Define a general linear predictor of the average severity by*

$$\overline{\Xi}_{jd} = z_{jd}\hat{\Xi}_{jd} + \left(1 - z_{jd}\right)\xi_d \tag{6.42}$$

with

$$\hat{\Xi}_{jd} = \left(\sum_{t=0}^{J-(j+d)} \frac{\omega_t^2}{\eta_t}\right)^{-1} \sum_{t=0}^{J-(j+d)} \frac{\omega_t}{\eta_t} \cdot \frac{W_{jdt}}{N_{jd}} \tag{6.43}$$

The conditional mean squared error of $\bar{\bar{\Xi}}_{jd}$ is

$$\begin{aligned} r_d\left(z_{jd} \mid N_{jd}\right) &= E\left(\left(\bar{\bar{\Xi}}_{jd} - \Xi_{jd}\right)^2 \mid N_{jd}\right) \\ &= N_{jd}^{-1}\left(z_{jd}^2\left(\sum_{t=0}^{J-(j+d)} \frac{\omega_t^2}{\eta_t}\right)^{-1} \xi_d + (1-z_{jd})^2\sigma_d^2\right) \end{aligned}$$

Having estimated the average severity by the credibility formula (6.42), an estimator of outstanding claim development becomes

$$\bar{W}_{jd,>J-(j+d)} = N_{jd}\bar{\bar{\Xi}}_{jd}\omega_{>J-(j+d)} \tag{6.44}$$

with mean squared error

$$E\left(\left(\bar{W}_{jd,>J-(j+d)} - W_{jd,>J-(j+d)}\right)^2 \mid N_{jd}\right) = N_{jd}\xi_d\eta_{>J-(j+d)} + \left(N_{jd}\omega_{>J-(j+d)}\right)^2 r_d\left(z_{jd} \mid N_{jd}\right) \tag{6.45}$$

The conditional mean squared error is minimised if one chooses the credibility factors in the following way:

$$z_{jd} = \sigma_d^2 \sum_{t=0}^{J-(j+d)} \frac{\omega_t^2}{\eta_t} \cdot \left(\xi_d + \sigma_d^2 \sum_{t=0}^{J-(j+d)} \frac{\omega_t^2}{\eta_t}\right)^{-1} \tag{6.46}$$

Remark 8 *It is interesting to note that the number of claims N_{jd} does not enter into the credibility factor z_{jd}. The reason for this lies in the assumption that the 'prior' variance of the unknown Ξ_{jd} is inversely proportional to N_{jd} in the conditional model.*

Proof. This proof follows the same lines as the one in the previous section, but it is easier. Without going through all the definitions again, let us simply write a regression model in the following form:

$$\mathrm{E}\left(\left(\mathbf{W} \mid N, \Xi\right) \mid N\right) = N\boldsymbol{\omega}\Xi = \mathbf{Y}\Xi \tag{6.47}$$

$$\mathbf{\Phi} = \mathrm{E}\left(\mathrm{Var}\left(\mathbf{W} \mid N, \Xi\right) \mid N\right) = N\xi \cdot \mathrm{diag}(\boldsymbol{\omega}) \tag{6.48}$$

$$\mathbf{Y}'^{-1}\mathbf{\Phi}^{-1}\mathbf{W} = \frac{1}{\xi}\sum_{t=0}^{J-(j+d)} \frac{\omega_t}{\eta_t} W_t \tag{6.49}$$

$$\mathbf{Y}'^{-1}\mathbf{\Phi}^{-1}\mathbf{Y} = \frac{N}{\xi}\sum_{t=0}^{J-(j+d)} \frac{\omega_t^2}{\eta_t} \tag{6.50}$$

This gives the required expressions. ■

6.3.4 Estimation of parameters

Let us briefly consider the estimation of the parameters ω_t and η_t under the simplifying assumption that

$$\eta_t = \eta\omega_t \tag{6.51}$$

It is easy to see that, in this case,

$$\hat{\Xi}_{jd} = \frac{W_{jd,\leq J-(j+d)}}{N_{jd}\omega_{\leq J-(j+d)}} \tag{6.52}$$

i.e. the reported claim cost to date, $W_{jd,\leq J-(j+d)}$, is a linear sufficient statistic. Conditional, given N_{jd} and Ξ_{jd}, we have that the increments W_{jdt} are independent and

$$\mathrm{E}\left(W_{jdt} \mid N_{jd}, \Xi_{jd}\right) = N_{jd}\Xi_{jd}\omega_t \tag{6.53}$$

and, by virtue of (6.51),

$$\mathrm{Var}\left(W_{jdt} \mid N_{jd}, \Xi_{jd}\right) = N_{jd}\Xi_{jd}\eta\omega_t \tag{6.54}$$

Assume that the parameters ω_t are known or have been estimated. Estimating those parameters is straightforward and entirely analogeous to estimating a claim reporting pattern or a payment pattern. Then consider the statistics

$$V_{jd} = \frac{1}{J-(j+d)} \sum_{t=0}^{J-(j+d)} N_{jd}\omega_t \left(\frac{W_{jdt}}{N_{jd}\omega_t} - \hat{\Xi}_{jd}\right)^2 \tag{6.55}$$

One can verify that

$$\mathrm{E}\left(V_{jd} \mid N_{jd}, \Xi_{jd}\right) = \Xi_{jd}\eta \tag{6.56}$$

and, consequently,

$$\mathrm{E}\left(V_{jd} \mid N_{jd}\right) = \xi_d\eta \tag{6.57}$$

Assuming that the ξ_d are known or have been estimated, one could then estimate η by a suitably weighted average of V_{jd}/ξ_d. One could suggest, for example, the estimator

$$\eta^* = \sum_{d=0}^{J-1} w_d \frac{1}{\xi_d} \left(\sum_{j=1}^{J-d} (J - (j+d)) \right)^{-1} \sum_{j=1}^{J-d} (J - (j+d)) V_{jd} \qquad (6.58)$$

with suitable weights w_d that add to 1.

The variance estimates V_{jd} will be unstable if ω_t is small or negative. If that is the case, one could consider using the regression equation

$$\mathrm{E}\left(\hat{\Xi}_{jd}^2 \mid N_{jd}\right) = \frac{\xi_d}{N_{jd}\omega_{\le J-(j+d)}} \cdot \eta + \left(\frac{\sigma^2}{N_{jd}} + \xi_d^2\right) \qquad (6.59)$$

to estimate η - again, assuming the other quantities on the right hand side of (6.59) have been estimated aforehand.

6.4 Actuarial case estimates

In a real life situation, the actuary should consider whether he or she can do better than just to extrapolate the evolution of claim payments or of reported claim cost. He or she could attempt to predict the ultimate outcome of reported claims by a statistical model that gives regard to individual claim characteristics.

Individual claim characteristics that can be utilised depend on the type of insurance and on the available data. Examples of individual claim characteristics would be:

- In Workers Compensation insurance: the type of injury sustained, the current degree of disability (which may not be the same as the degree of permanent disability), the claimant's age, his salary before the accident, and so on.
- In Disability insurance: the time the claimant has been off work (the longer the claimant has been off work, the less the chance of a recovery will be), the type of illness, age, the claimant's employability etc.
- In Motor Vehicle insurance: the type of vehicle, type of damage etc.

Case estimates by claim handlers, if available, should of course also be taken into account in the estimation, but not necessarily at their face value.

The theoretical actuarial literature offers little general guidance on that kind of modelling, probably because such models will be dependent on the individual circumstances and can seldom be generalised. There is still a lot of work to be done.

6.5 Markov chain models

In some lines of insurance one can model the development inside a "mark" as a Markov chain in continuous time, with a finite number of states. For example, in disability insurance, a claimant could be in one of several states:

- Off work for the first time.
- On work again.
- Off work for a subsequent period.
- In rehabilitation.
- Permanently disabled.
- and so on.

Assume that transitions between states are governed by transition intensities $\lambda_{ij}(t)$, where $t = 0$ refers to the start of the first period off work. Assume further that a pension benefit of $b_i(t)dt$ is payable if the claimant is in state i at time $(t, t + dt)$. If the claimant transits from state i to state j at time t, a lump sum benefit of $A_{ij}(t)$ is payable. Of course, both $b_i(t)$ and $A_{ij}(t)$ will be zero for all but a few defined states and transitions.

Hesselager (1994) shows how one can calculate the expected future claim payments, given the state of the claimant at a specified time. His model has been applied in the calculation of actuarial case estimates in Danish Workers' Compensation claims.

7
Two-dimensional models

7.1 Introduction

Most loss-reserving studies analyse only two-dimensional development models, where the two dimensions are accident period (j) and accident-to-valuation delay ($e = d + t$). The delay dimension in those triangles is normally referred to as development period. In those models, no attempt is made to separate the development of reported claims from that of unreported claims.

The most common data to analyse in the two-dimensional setting is paid claims. One starts with a development triangle containing the accumulated claim payments per accident period and valuation delay:

$$\begin{array}{llll} \tilde{U}_{10} & \tilde{U}_{11} & \cdots & \tilde{U}_{1,J-1} \\ \tilde{U}_{20} & \tilde{U}_{21} & \ddots & \\ \vdots & \ddots & & \\ \tilde{U}_{J0} & & & \end{array} \tag{7.1}$$

The task then becomes to predict the entries in the south-east corner of the development square.

One may triangulate reported claims in the same way:

$$\begin{array}{llll} \tilde{W}_{10} & \tilde{W}_{11} & \cdots & \tilde{W}_{1,J-1} \\ \tilde{W}_{20} & \tilde{W}_{21} & \ddots & \\ \vdots & \ddots & & \\ \tilde{W}_{J0} & & & \end{array} \tag{7.2}$$

The methods derived for claim numbers, in particular the Chain ladder method and the Bornhuetter-Ferguson method, can also be applied to claim amounts. We will study some of these methods in this chapter.

7.2 Generic notation

Let us use X as a generic notation for the quantity that is being analysed, whether that is the number of claims (N), claim payments (U) or reported claim cost (W). More specifically, let X_{je} denote the incremental change in development period e and $\tilde{X}_{je}$ the accumulated quantity at the end of development period e, so that

$$X_{je} = \tilde{X}_{je} - \tilde{X}_{j,e-1} \tag{7.3}$$

and

$$\tilde{X}_{je} = X_{j,\leq e} \tag{7.4}$$

7.3 The Bornhuetter-Ferguson method

In the opinion of this author, the essential characteristic of the Bornhuetter-Ferguson method is the following prediction formula:

$$\overline{X}_{je} = p_j \theta_j^* \pi_e^* \text{ for } j + e > J, \tag{7.5}$$

where p_j is a measure of risk exposure, θ_j^* is an expected (or à priori) claim rate per unit of risk exposed in year j that does not directly depend on the observed experience from year j, and π_e^* is the expected proportion of claims that will materialise at delay e.

Estimating future claim by the Bornhuetter-Ferguson method is similar to "budgeting" a certain amount of claims to arrive in future periods, in proportion with an estimated delay pattern.

The expected claim rate, as well as the delay pattern, may come from different sources:

- Estimation from the observed data,

- Expert knowledge, or a budgeted claims cost per unit of exposure,
- Industry statistics.

Especially for the delay pattern, industry statistics can be very helpful if the data at hand is not sufficiently developed or organised to allow an estimation.

7.4 The Chain-ladder method

In the opinion of this author, the essential characteristic of the Chain ladder method is the following prediction formula:

$$\overline{X}_{je} = X_{j,\leq J-j}\left(\frac{\pi_e^*}{\pi_{j,\leq J-j}^*}\right) \text{ for } j+e>J, \tag{7.6}$$

where π_e^* is the expected proportion of claims that will materialise at delay e.

Estimating future claim by the Chain ladder method amounts to "grossing up" the observed claim experience of the actual year, in proportion with an estimated delay pattern.

The delay pattern may come from different sources:

- Estimation from the observed data,
- Expert assessment,
- Industry statistics.

Industry statistics can be very helpful if the data at hand is not sufficiently developed or organised to allow an estimation.

7.5 Bühlmann-Straub's model

The Bühlmann-Straub model is very similar to the model we have seen for claim frequencies. To motivate the model, imagine that conditional on a risk parameter Θ_j, the claim statistic X_{je} has a compound Poisson distribution with frequency parameter $p_j\lambda(\Theta_j)\pi_e$ and a severity distribution $F(y|\Theta_j)$. In this case, however, we only observe the compound Poisson random variable, not the number of jumps. Then

$$\mathrm{E}(X_{je}|\Theta_j) = p_j\pi_e\lambda(\Theta_j)\int ydF(y|\Theta_j) =: p_j\pi_e b(\Theta_j) \tag{7.7}$$

$$\mathrm{Var}\left(X_{je}|\Theta_j\right) = p_j \pi_e \lambda\left(\Theta_j\right) \int y^2 dF\left(y|\Theta_j\right) =: p_j \pi_e v\left(\Theta_j\right) \tag{7.8}$$

That means that there are two functions $b\left(\Theta_j\right) = \lambda\left(\Theta_j\right) \int y dF\left(y|\Theta_j\right)$ and $v\left(\Theta_j\right) = \lambda\left(\Theta_j\right) \int y^2 dF\left(y|\Theta_j\right)$ of the risk parameter Θ_j. For the Poisson model (claim counts), the two functions coincide.

The assumptions of Bühlmann-Straub's model are:

- Conditional on an unobserved risk parameter that we denote by Θ_j, the increments $X_{j0}, X_{j1}, \cdots$ are stochastically independent with conditional mean $\mathrm{E}(X_{je}|\Theta_j) = p_j b\left(\Theta_j\right) \pi_e$ and variance $\mathrm{Var}(X_{je}|\Theta_j) = p_j v\left(\Theta_j\right) \pi_e$. The quantity p_j denotes an observed measure of risk exposure, while the quantity π_e denotes the expected amount of increment in development period e. We assume that $\sum_{e=0}^{\infty} \pi_e = 1$.

- The unobserved risk parameter Θ_j is assumed to be the outcome of a random variable.

- The risk parameters $\Theta_1, \cdots, \Theta_J$ are assumed to be stochastically independent and identically distributed. We denote the mean and variance of the function $b\left(\Theta_j\right)$ by $\beta = \mathrm{E}(b\left(\Theta_j\right))$ and $\lambda = \mathrm{Var}(b\left(\Theta_j\right))$. Let use further denote the mean of the function $v\left(\Theta_j\right)$ by $\varphi = \mathrm{E}(v\left(\Theta_j\right))$.

- The sets $\{\Theta_j, X_{j0}, X_{j1}, \cdots\}$ and $\{\Theta_k, X_{k0}, X_{k1}, \cdots\}$ are assumed to be independent for different accident periods (i.e., $j \neq k$).

We assume for the present that the delay probabilities π_e as well as the distribution moments $(\beta, \lambda, \varphi)$ are known. We restrict the estimator of $b\left(\Theta_j\right)$ to be a linear combination of a Chain-ladder estimate and the à prior mean, i.e.

$$\bar{b}_j = z_j \left(\frac{X_{j,\leq J-j}}{p_j \pi_{\leq J-j}}\right) + (1 - z_j)\beta \tag{7.9}$$

Then it is easy to verify that for an arbitrary choice of z_j, the mean squared error of the estimator $\bar{b}_j$ is

$$Q\left(z_j\right) = \mathrm{E}\left(\bar{b}_j - b\left(\Theta_j\right)\right)^2 \quad = \quad z_j^2 \frac{\varphi}{p_j \pi_{\leq J-j}} + \left(1 - z_j\right)^2 \lambda \tag{7.10}$$

The mean squared error of the predictor

$$\bar{X}_{je} = p_j \bar{b}_j \pi_e \tag{7.11}$$

for $j + e > J$ is

$$E\left(\bar{X}_{je} - X_{je}\right)^2 = \left(p_j \pi_e\right)^2 Q\left(z_j\right) + p_j \pi_e \varphi \tag{7.12}$$

An estimator of the total outstanding development is

$$\bar{X}_{j,>J-j} = p_j \bar{b}_j \pi_{>J-j} \tag{7.13}$$

with mean squared error

$$\mathrm{E}\left(\bar{X}_{j,>J-j} - X_{j,>J-j}\right)^2 = \left(p_j \pi_{>J-j}\right)^2 Q\left(z_j\right) + p_j \pi_{>J-j} \varphi \tag{7.14}$$

Minimising (7.10) it is easy to see that the optimal choice of z_j is

$$\zeta_j = \frac{\lambda p_j \pi_{\leq J-j}}{\lambda p_j \pi_{\leq J-j} + \varphi} \tag{7.15}$$

Note that $\zeta_j \to 1$ when $\lambda \to \infty$; thus [a version of] the Chain-ladder method is the appropriate choice if there is much heterogeneity between accident periods. On the other hand, $\zeta_j \to 0$ when $\lambda \to 0$, which means that [a version of] Bornhuetter-Ferguson's method is the appropriate choice if there is no, or very little, heterogeneity between accident periods. These conditions correspond exactly to the models which were used to derive those methods for claim frequencies.

In practice, of course, the delay probabilities π_e and the moments $(\beta, \lambda, \varphi)$ will not be known, unless they reflect a subjective, apriori assessment. They can be estimated by De Vylder's iteration method in essentially the same way as in section 4.7.

First, estimate the delay probabilities $\pi_0, \pi_1, \cdots$. Then treat the delay probabilities as known, fixed parameters and estimate the remaining parameters in the following way:

1. Pick starting values $\beta^*_{(0)}$and $\lambda^*_{(0)}$. Set the iteration number $i = 0$.

2. Calculate the Chain Ladder estimates $\hat{b}_j = X_{j,\leq J-j}/p_j\pi_{\leq J-j}$ for $j = 1, \ldots, J$.

3. Calculate the empirical variances $\hat{v}_j = \frac{1}{J-j}\sum_{e=0}^{J-j} p_j\pi_e \left(\frac{X_{je}}{p_j\pi_e} - \hat{b}_j\right)^2$ for $j = 1, \ldots, J-1$, and then $\varphi^* = \frac{2}{(J-1)J}\sum_{j=1}^{J-1}(J-j)\hat{v}_j$. This is an unbiased estimator of φ.

4. Calculate credibility factors $z_j^{(i)} = \frac{\lambda^*_{(i)} p_j \pi_{\leq J-j}}{\lambda^*_{(i)} p_j \pi_{\leq J-j} + \varphi^*}$ for $j = 1, \ldots, J$.

5. Calculate an estimate of the mean $\beta^*_{(i+1)} = \sum_{j=1}^{J} z_j^{(i)}\hat{b}_j / \sum_{j=1}^{J} z_j^{(i)}$.

6. Calculate the estimated variance $\lambda^*_{(i+1)} = \frac{1}{J-1}\sum_{j=1}^{J} z_j^{(i)}\left(\hat{b}_j - \beta^*_{(i+1)}\right)^2$. Repeat (4)-(6) until convergence is reached.

7.6 A model with random delay probabilities

7.6.1 The Hesselager-Witting model

In the Bühlmann-Straub model of section 6.3, the delay probabilities $\{\pi_0, \cdots, \pi_D\}$ were interpreted as fixed (non-random) parameters. As a result of this assumption, the optimal credibility factor ζ_j defined in (7.15) will always be non-negative, and there will always be a non-negative linkage between the (observed) past claim development and the (predicted) future claim development. In the words of Taylor (2000), the stochastic risk parameter $b(\Theta_j)$ can do little more than act as a scaling constant between different periods of origin.

Sometimes the actuary will meet the argument that there should be a negative linkage, in the sense that above-average claim development in the past should be seen as an indication of below-average development in the future. The argument is intuitively very plausible.

Hesselager & Witting (1988) have developed a model with random delay probabilities to explain negative linkages, or at least linkages that are weaker than in the Bühlmann-Straub model. In our context and notation, a version of their model looks as follows:

1. Given an unobserved risk parameter Θ_j and a vector of unknown delay probabilities $\Pi_j = (\Pi_{j0}, \cdots, \Pi_{jD})'$, the increments $X_{j0}, X_{j1}, \cdots$ are conditionally independent with conditional mean $\mathrm{E}(X_{je} \mid \Theta_j, \Pi_j) = p_j b(\Theta_j)\Pi_{je}$ and variance $\mathrm{Var}(X_{je} \mid \Theta_j, \Pi_j) = p_j v(\Theta_j)\Pi_{je}$. The quantity p_j denotes an observed measure of risk exposure, while the quantity Π_{je} denotes the expected amount of increment in development

period e.

2. The unobserved risk pair (Θ_j, Π_j) is the outcome of a random variable. For each j, Θ_j is assumed to be stochastically independent of Π_j, i.e., there is no stochastic linkage between the level of claims and their development pattern. The risk pairs $(\Theta_1, \Pi_1), \cdots, (\Theta_J, \Pi_J)$ are stochastically independent and identically distributed.

3. We define $\beta = \mathrm{E}(b(\Theta_j))$, $\lambda = \mathrm{Var}(b(\Theta_j))$and $\varphi = \mathrm{E}(v(\Theta_j))$ with no further specification of the distribution of Θ_j.

4. We assume that each Π_j has a Dirichlet distribution with parameters $\alpha = (\alpha_0, \cdots, \alpha_D)'$. Let $\alpha = \sum_{e=0}^{D} \alpha_e$, and define the expected delay proportions $\pi_e = \alpha_e/\alpha$.

5. The sets $\{(\Theta_j, \Pi_j), X_{j0}, X_{j1}, \cdots\}$ and $\{(\Theta_k, \Pi_k), X_{k0}, X_{k1}, \cdots\}$ are independent for different accident periods (i.e., $j \neq k$).

Using the assumed independence between Θ_j and Π_j and the moments of the Dirichlet distribution, it is straightforward to derive that

$$\mathrm{E}(X_{je}) = p_j \beta \pi_e \tag{7.16}$$

and

$$\begin{aligned}
\mathrm{Var}(X_{je}) &= \mathrm{E}\mathrm{Var}(X_{je} \mid \Theta_j, \Pi_j) + \mathrm{Var}\mathrm{E}(X_{je} \mid \Theta_j, \Pi_j) \\
&= \mathrm{E}(p_j v(\Theta_j)\Pi_{je}) + \mathrm{Var}(p_j b(\Theta_j)\Pi_{je}) \\
&= p_j \varphi \pi_e + p_j^2(\lambda + \beta^2)\left(\pi_e^2 \tfrac{\alpha}{\alpha+1} + \pi_e \tfrac{1}{\alpha+1}\right) - p_j^2 \beta^2 \pi_e^2 \\
&= p_j \pi_e \left(\varphi + \tfrac{p_j(\lambda+\beta^2)}{\alpha+1}\right) + (p_j \pi_e)^2 \left(\tfrac{\lambda\alpha - \beta^2}{\alpha+1}\right) \\
&= p_j \pi_e \varphi_j(\alpha) \quad + \quad (p_j \pi_e)^2 \lambda(\alpha)
\end{aligned} \tag{7.17}$$

and, similarly, for $e \neq e'$:

$$\mathrm{Cov}(X_{je}, X_{je'}) = p_j^2 \pi_e \pi_{e'} \lambda(\alpha) \tag{7.18}$$

Here we have defined the quantities

$$\varphi_j(\alpha) = \varphi + \frac{p_j(\lambda + \beta^2)}{\alpha + 1} \tag{7.19}$$

and

$$\lambda(\alpha) = \frac{\lambda\alpha - \beta^2}{\alpha + 1} \tag{7.20}$$

These quantities are not variances, as one can easily see by checking that $\lambda(\alpha)$ becomes negative for small values of α. They are important because they allow us write the mean-covariance structure of the observed and unobserved claims in the same formal way as in the Bühlmann-Straub model. Please note that $\varphi_j(\infty) = \varphi$ and $\lambda(\infty) = \lambda$.

Let us assume that the Dirichlet parameters $\alpha = (\alpha_0, \cdots, \alpha_D)'$ as well as the distribution moments $(\beta, \lambda, \varphi)$ are known.

A linear estimator of the outstanding development is

$$\bar{X}_{j,>J-j} = p_j \pi_{>J-j} \left(z_j \left(\frac{X_{j,\leq J-j}}{p_j \pi_{\leq J-j}} \right) + (1 - z_j)\beta \right) \tag{7.21}$$

Then it is easy to verify that the mean squared error of the estimator $\bar{X}_{j,>J-j}$ is

$$\begin{aligned} \mathrm{E}(\bar{X}_{j,>J-j} - X_{j,>J-j})^2 &= \mathrm{Var}\left(z_j \tfrac{\pi_{>J-j}}{\pi_{\leq J-j}} X_{j,\leq J-j} - X_{j,>J-j} \right) \\ &= (p_j \pi_{>J-j})^2 \left(z_j^2 \tfrac{\varphi_j(\alpha)}{p_j \pi_{\leq J-j}} + (1 - z_j)^2 \lambda(\alpha) \right) \\ &+ p_j \pi_{>J-j} \varphi_j(\alpha) \end{aligned} \tag{7.22}$$

Minimising (7.22) it is easy to see that the optimal choice of the credibility factor z_j is

$$\zeta_j(\alpha) = \frac{\lambda(\alpha) p_j \pi_{\leq J-j}}{\lambda(\alpha) p_j \pi_{\leq J-j} + \varphi_j(\alpha)} \tag{7.23}$$

Note that this is the same form as in the Bühlmann-Straub model, with only the parameter α added. Thus we have an optimal credibility factor that is always smaller than the credibility factor in the Bühlmann-Straub model, with equality only in the limiting case of $\alpha \to \infty$ while $\alpha_e/\alpha \to \pi_e$. In that case, the Dirichlet distribution approaches a degenerate distribution.

On the other hand, for $\alpha \to 0$ while $\alpha_e/\alpha \to \pi_e$, the Dirichlet distribution attains its maximum dispersion. The credibility factor then approaches

$$\zeta_j(0) = -\frac{p_j \beta^2 \pi_{\leq J-j}}{p_j \beta^2 \pi_{>J-j} + \varphi + p_j \lambda} < 0 \tag{7.24}$$

The model of Hesselager & Witting can thus be used to justify lower or even negative credibility factors than those that derive from the Bühlmann-Straub model. Even if one is not using the credibility factors (7.23), the expression for the mean squared error (7.22) allows one to pose the question: Given that we are using credibility factors z_j, what would the mean squared error of our estimates if the delay probabilities were random, and how would that mean squared error compare to the optimum that we could achieve if we knew α?

The notion of negative credibility factors is not as far-fetched as it may appear at first glance. In many situations, the actuary would have a prior opinion on the ultimate claim cost of an accident period, and would not change that estimate even if early claim development is different from what he had expected - typically arguing that early claim development is subject to so much random fluctuation that it does not warrant a reassessment of the ultimate outcome. That actuary is implicitly using negative credibility factors in the sense that more (less) development in the past implies less (more) development in the future. The model of Hesselager & Witting is an attempt to provide a theoretical grounding to such an approach.

The alert reader may object that before using an estimator of the form (7.21), we should have checked that the sum of observed claim development to date (i.e., $X_{j,\leq J-j}$) is linear sufficient. Otherwise, a better linear estimator could be designed using some other linear combination of the $\{X_{j,e} : e \leq J-j\}$. Let it suffice to say that the linear sufficiency of $X_{j,\leq J-j}$ can be proved, see Hesselager & Witting (1988).

7.6.2 Parameter estimation

The model is not operational before one has established values for the parameters $(\beta, \varphi, \lambda)$, $\{\pi_0, \cdots, \pi_D\}$and α, whether that be by subjective judgement or by estimation or, most likely, a combination of both. Here an estimation procedure will be outlined that is inspired by the suggestions in Hesselager & Witting's paper.

We begin by noting the regression equations

$$\mathrm{E}(X_{je}) = p_j \beta \pi_e \tag{7.25}$$

$$E(X_{je} X_{je'}) = \delta_{ee'} p_j \pi_e \left(\varphi + p_j \frac{\lambda + \beta^2}{\alpha + 1} \right) + p_j^2 \pi_e \pi_{e'} \left(\frac{(\lambda + \beta^2)\alpha}{\alpha + 1} \right) \tag{7.26}$$

Now defining

$$Z_{j,ee'} = X_{je} X_{je'} \tag{7.27}$$

$$\mathbf{a}_{j,ee'} = \left(\delta_{ee'} p_j \pi_e, \delta_{ee'} p_j^2 \pi_e, p_j^2 \pi_e \pi_{e'}\right) \tag{7.28}$$

$$\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3)' = \left(\varphi, \frac{\lambda + \beta^2}{\alpha + 1}, \frac{(\lambda + \beta^2)\alpha}{\alpha + 1}\right)' \tag{7.29}$$

we write (7.26) as a linear regression

$$\mathrm{E}(\mathbf{Z}) = \mathbf{A}\boldsymbol{\tau} \tag{7.30}$$

where the vector $\mathbf{Z}$ is a collection of all available cross-products of the form (7.27) and the matrix $\mathbf{A}$ consists of the corresponding row vectors (7.28). Note that $\tau_2 + \tau_3 = \lambda + \beta^2$ and $\alpha = \tau_2/\tau_3$. Let us assume that $\mathbf{A}$ is of full column rank.

To begin the estimation, the parameters β and $\pi_0, \cdots, \pi_D$ must be estimated. The estimates β^* and $\pi_0^*, \cdots, \pi_D^*$ are thereafter treated as fixed parameters.

In order to estimate the remaining parameters, we use the regression (7.30). For any fixed weighting matrix $\mathbf{W}$ of full rank, the estimator $\boldsymbol{\tau}_{\mathbf{W}}^* = (\mathbf{A}'\mathbf{W}\mathbf{A})^{-1}\mathbf{A}'\mathbf{W}\mathbf{Z}$ is unbiased. A simple solution would be to use the identity matrix so that $\boldsymbol{\tau}^* = \boldsymbol{\tau}_{\mathbf{I}}^* = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Z}$. That estimator can be calculated as soon as one has estimated the $\pi_0, \cdots, \pi_D$ that enter into the matrix $\mathbf{A}$.

Having estimated $\boldsymbol{\tau}^*$, we can estimate λ by

$$\lambda^* = \tau_2^* + \tau_3^* - (\beta^*)^2 \tag{7.31}$$

(crossing our fingers and praying that it will be non-negative!) and α by

$$\alpha^* = \tau_2^*/\tau_3^* \tag{7.32}$$

(say another prayer...).

By the Gauss-Markov theorem, however, the optimal choice of $\mathbf{W}$ would be

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{Z}) \tag{7.33}$$

. Obviously, that matrix depends on the unknown parameters, a fact we can indicate by writing $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\tau})$. Thus one could be tempted to iterate the procedure and use $\mathbf{W} = \boldsymbol{\Sigma}(\boldsymbol{\tau}_{\mathbf{I}}^*)$, hoping that this will produce a more reliable estimate. A complication is that the covariance matrix cannot be calculated without a further specification of the underlying probability distribution of both Θ_j and $(X_{j0}, \cdots, X_{jD}) \mid \Theta_j, \Pi_j$.

To enable the calculation of an approximate $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\tau})$, one could assume that Θ_j and Π_j have a degenerate (one-point) distribution. In that case, the X_{je} will be mutually independent with mean $\mathrm{E}(X_{je}) = p_j \beta \pi_e$ and variance $\mathrm{Var}(X_{je}) = p_j \varphi \pi_e$, and the covariances

$$\mathrm{Cov}(Z_{j,ee'}, Z_{j,ff'}) = \mathrm{E}(X_{je}X_{je'}X_{jf}X_{jf'}) - \mathrm{E}(X_{je}X_{je'})\mathrm{E}(X_{jf}X_{jf'}) \quad (7.34)$$

involve only expressions of the form $\mathrm{E}(X_{je}^k)$ for k=1,2,3,4. Calculating the $\mathrm{E}(X_{je}^k)$ is normally straightforward if the X_{je} have a two-parametric probability distribution that is completely characterised by its mean and variance. Candidates for the parametric distribution of the X_{je} are the Poisson distribution (if X_{je} represents claim counts), the lognormal or gamma distribution (if X_{je} represents claim amounts) or even the normal distribution.

To exemplify, let us assume that the observations are gamma distributed, $X_{je} \sim \Gamma(\gamma_{je}, \delta_{je})$. Equating the mean and the variance of the gamma distribution to the mean and variance of X_{je}, one easily finds that $\gamma_{je} = p_j \pi_e \beta^2/\varphi$ and $\delta_{je} = \beta/\varphi$. The k'th order moment of X_{je} is then $\mathrm{E}(X_{je}^k) = \frac{\Gamma(\gamma_{je}+k)}{\Gamma(\gamma_{je})\delta_{je}^k}$, and this can be used to assemble estimates of the covariances (7.34), using previously calculated estimates β^*, $\pi_0^*, \cdots, \pi_D^*$ and φ^*.

7.7 Benktander´s method

As we have seen, every credibility method forms a compromise between the Bornhuetter-Ferguson method and the Chain Ladder method. Within a Bayesian framework the credibility estimates are optimal. However, assessing or estimating the parameters requires hard work. Especially when data is sparse or of poor quality - as is the case most of the time - estimating the variances φ and λ degrades to pure guesswork.

In Neuhaus (1992), a pragmatic variation of the formula (7.9) is explored, where the credibility factors are set in the following way:

$$z_j = \pi_{\leq j}$$

The advantage of this choice is that it does not require estimation of the variances φ and λ.

Mack (2000) has traced the method back to Benktander and proposed to call it Benktander's method. In Sweden, Benktander's home country, the method is known as Hovinen's method, after a Finnish actuary. Personally, I like the name “Golden Stairs method”. After all, Golden Stairs are a lot more comfortable to climb than a chain ladder (as well as a lot more scenic)!

Benktander's method is close to Bornhuetter-Ferguson's method as long as $\pi_{\leq j}$ is small, and approaches the chain ladder method as $\pi_{\leq j}$ approaches

1. In that way it avoids both the sensitivity of the chain ladder methods for immature accident periods, and the inflexibility of Bornhuetter-Ferguson's method for mature periods. Within the model of Bühlmann and Straub, the mean squared error of its predictions can of course be calculated by (7.10)-(7.12) with $z_j = \pi_{\leq j}$.

It can be shown that Benktander's method is superior to Bornhuetter-Ferguson's method most of the time, when judged by mean squared prediction error within the Bühlmann-Straub model. By the same criterion it is also often, but not always, better than the Chain Ladder method. Expressed somewhat informally, the method is “not too far from optimal, most of the time”.

An interesting feature of Benktander's method appears when one looks at the predicted outstanding claim development:

$$\begin{aligned} \bar{X}_{je} &= p_j \left(\pi_{\leq J-j} \left(\frac{X_{j,\leq J-j}}{p_j \pi_{\leq J-j}} \right) + (1 - \pi_{\leq J-j})\, \beta \right) \pi_e \\ &= \left(X_{j,\leq J-j} + p_j\, (1 - \pi_{\leq J-j})\, \beta \right) \pi_e \\ &= p_j \tilde{b}_j \pi_e \end{aligned} \tag{7.35}$$

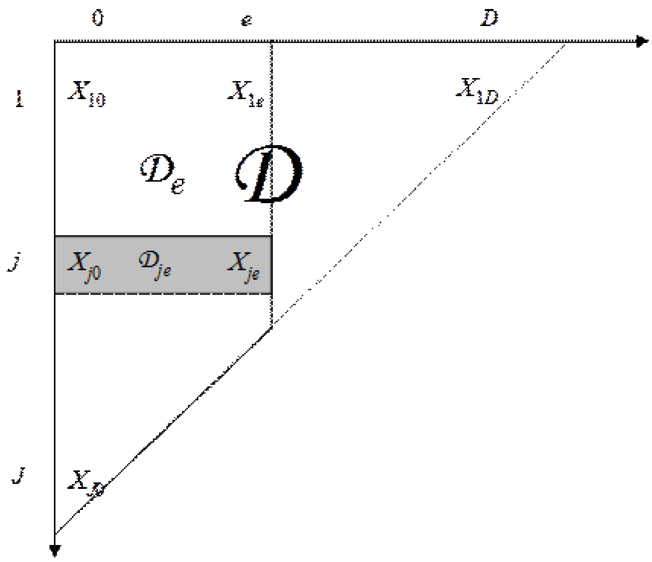
The quantity $\tilde{b}_j$ is just the ultimate claim level of period j, when estimated by the Bornhuetter-Ferguson method. Thus in Benktander's method, we are not predicting the outstanding claim development by the Bornhuetter-Ferguson method, but rather the unobserved claim level in the tail.

7.8 Mack's model

The chain ladder method has already been presented in chapter 4.4, where it was derived as a maximum likelihood method for estimating claim frequencies and predicting the number of unreported claims within the framework of a Poisson model. Due to its simplicity and intuitive appeal, the chain ladder method is being used in many other contexts too, for example to predict the amount of future payments. As Taylor (2000) correctly states, the chain ladder method has a pre-eminent role in actuarial practice.

Let us recapitulate by writing out the chain ladder method in the notation of this chapter. For the quantity that is being analysed, we denote by X_{je} its incremental change during development period e and by $\tilde{X}_{je}$ its cumulative status at the end of development period e. Let us further denote the maximum observed development period by $D = J - 1$. Although it is usually not the case in practice, it is quite common to assume that at least one period (period one) is fully developed, so that there will be no development beyond D. As an abbreviation, let us denote the unknown ultimate claims of accident period j by $\tilde{X}_j = \tilde{X}_{jD}$.

The chain ladder method can be succinctly formulated in the following two instructions:



1. For delay $e = 2, \cdots, D$, calculate empirical development factors $\delta^*_e = \sum_{j=1}^{J-e} \tilde{X}_{je} / \sum_{j=1}^{J-e} \tilde{X}_{j,e-1}$.

2. For accident period $j = 2, \cdots, J$, predict its ultimate claims by $\bar{X}_j = \tilde{X}_{j,J-j} \cdot \prod_{e=J-j+1}^{D} \delta^*_e$.

Thus for every accident period, its current cumulative level is extrapolated in a multiplicative fashion, using the average development that has been observed for earlier accident periods. It will also be useful to have defined the chain ladder predictions at delays before the ultimate delay D, by $\bar{X}_{je} = \tilde{X}_{j,J-j} \cdot \prod_{e'=J-j+1}^{e} \delta^*_{e'}$ for $e > J - j$.

Mack (1993) has proposed a distribution-free model to calculate the mean squared error of chain ladder predictions, which will be presented briefly here. As will be explained later, this author has certain reservations about Mack's approach. As you will almost certainly be confronted with Mack's model in actuarial practice, you need to know about it.

Let $D = \left\{ \tilde{X}_{je} : j = 1, \cdots, J, e = 0, \cdots, J - j \right\}$ denote the observed data, and let us write $D_{je} = \left\{ \tilde{X}_{je'} : e' = 0, \cdots, e \right\}$ for the development of accident period j up to development period e. Finally, let $D_e = \bigcup_{j=1}^{J} D_{j,e\wedge(J-j)}$denote the observed development of all accident periods up to development period e. The drawing below indicates the demarkations.

Conditional on the observed data, the mean squared error of the predictor $\bar{X}_j$ is

$$E\left(\left(\tilde{X}_j - \bar{X}_j\right)^2 \mid D\right) = Var\left(\tilde{X}_j \mid D\right) + \left(\mathrm{E}\left(\tilde{X}_j \mid D\right) - \bar{X}_j\right)^2 \tag{7.36}$$

This formula can easily be verified by noting that $\tilde{X}_j$ is a random variable, while $\bar{X}_j$ is fully determined in the conditional distribution, given D.

To be able to give substance to the conditional expression in (7.36), Mack makes three model assumptions:

1. There exist constants $\delta_1, \delta_2, \cdots, \delta_D$ such that $\mathrm{E}\Big(\tilde{X}_{je} \mid D_{j,e-1}\Big) = \delta_e \tilde{X}_{j,e-1}$ for $e = 1, \cdots, D$.

2. There exist constants $\gamma_1, \gamma_2, \cdots, \gamma_D$ such that $\mathrm{Var}\Big(\tilde{X}_{je} \mid D_{j,e-1}\Big) = \gamma_e \tilde{X}_{j,e-1}$ for $e = 1, \cdots, D$.

3. The ensembles $D_{j,D}$ and $D_{k,D}$ are stochastically independent for $j \neq k$.

Under these assumptions, Mack proves that the estimated development factors $\delta_1^*, \delta_2^*, \cdots, \delta_D^*$ are unbiased and uncorrelated in the unconditional distribution, i.e. $\mathrm{E}(\delta_e^*) = \delta_e$ and $\mathrm{E}(\delta_e^* \delta_{e'}^*) = \delta_e \delta_{e'}$ for $e \neq e'$. By repeated conditioning it is easy to verify that $\mathrm{E}\Big(\tilde{X}_{je} \mid D_{j,J-j}\Big) = \tilde{X}_{j,J-j} \prod_{e'=J-j+1}^{e} \delta_{e'}$ for $e = J-j+1, \cdots, D$. It is also easy to verify that the estimators $\gamma_e^* = \frac{1}{J-e-1} \sum_{j=1}^{J-e} \tilde{X}_{j,e-1} \left(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}} - \delta_e^*\right)^2$ are unbiased for γ_e.

Using assumptions (1)-(3) we shall now calculate (7.36). We begin with

$$\begin{aligned}
\mathrm{Var}\left(\tilde{X}_j \mid D\right) &= \mathrm{Var}\left(\tilde{X}_{jD} \mid D_{j,J-j}\right) \\
&= \mathrm{E}\left(\mathrm{Var}\left(\tilde{X}_{jD} \mid D_{j,D-1}\right) \mid D_{j,J-j}\right) \\
&+ \mathrm{Var}\left(\mathrm{E}\left(\tilde{X}_{jD} \mid D_{j,D-1}\right) \mid D_{j,J-j}\right) \\
&= \gamma_D \mathrm{E}\left(\tilde{X}_{j,D-1} \mid D_{j,J-j}\right) \;+\; \delta_D^2 \mathrm{Var}\left(\tilde{X}_{j,D-1} \mid D_{j,J-j}\right) \\
&= \gamma_D \prod_{e=J-j+1}^{D-1} \delta_e \tilde{X}_{j,J-j} \;+\; \delta_D^2 \mathrm{Var}\left(\tilde{X}_{j,D-1} \mid D_{j,J-j}\right) \\
&= \cdots \\
&= \tilde{X}_{j,J-j} \sum_{e=J-j+1}^{D} \left(\prod_{e'=J-j+1}^{e-1} \delta_{e'}\right) \gamma_e \left(\prod_{e'=e+1}^{D} \delta_{e'}^2\right)
\end{aligned} \tag{7.37}$$

This expression can be estimated if one substitutes the unknown quantities by their empirical counterparts to obtain

$$\begin{aligned} \mathrm{Var}^{*}\left(\tilde{X}_j \mid D\right) &= \tilde{X}_{j,J-j} \sum_{e=J-j+1}^{D} \left(\prod_{e'=J-j+1}^{e-1} \delta^*_{e'}\right) \gamma^*_e \left(\prod_{e'=e+1}^{D} \delta^{*\,2}_{e'}\right) \\ &= \bar{X}_j^2 \sum_{e=J-j+1}^{D} \frac{\gamma^*_e}{\delta^{*2}_e} \left(\frac{1}{\bar{X}_{j,e-1}}\right) \end{aligned} \tag{7.38}$$

Next, we find

$$\begin{aligned} \left(\mathrm{E}\left(\tilde{X}_j \mid D\right) - \bar{X}_j\right)^2 &= \tilde{X}^2_{j,J-j} \left(\prod_{e=J-j+1}^{D} \delta_e - \prod_{e=J-j+1}^{D} \delta^*_e\right)^2 \\ &= \tilde{X}^2_{j,J-j} \left(\sum_{e=J-j+1}^{D} \left(\prod_{e'=J-j+1}^{e-1} \delta^*_{e'}\right) (\delta_e - \delta^*_e) \left(\prod_{e'=e+1}^{D} \delta_{e'}\right)\right)^2 \\ &= \tilde{X}^2_{j,J-j} \left(\sum_{e=J-j+1}^{D} \;\; S_e\right)^2 \end{aligned} \tag{7.39}$$

If one were to simply substitute unknown quantities by their their empirical counterparts in this expression one would obtain zero, which admittedly is a bit optimistic. Mack therefore proposes to expand the expression to

$$\left(\mathrm{E}\left(\tilde{X}_j \mid D\right) - \bar{X}_j\right)^2 = \tilde{X}^2_{j,J-j} \left(\sum_{e=J-j+1}^{D} S_e^2 + 2 \sum_{e=J-j+1}^{D} \sum_{e'=e+1}^{D} S_e S_{e'} \right) \tag{7.40}$$

He then approximates S_e^2 by $\mathrm{E}\big(S_e^2 \mid D_{e-1}\big)$ and $S_e S_{e'}$ by $\mathrm{E}(S_e S_{e'} \mid D_{e'-1})$, i.e. the distribution where only one term $(\delta_e - \delta^*_e)$is random while everything else is fixed.

Using conditioning it is not difficult to verify that

$$\mathrm{E}\left(S_e^2 \mid D_{e-1}\right) = \left(\prod_{e'=J-j+1}^{e-1} \delta^*_{e'}\right)^2 \frac{\gamma_e}{\sum_{j=1}^{J-e} \tilde{X}_{j,e-1}} \left(\prod_{e'=e+1}^{D} \delta_{e'}\right)^2 \tag{7.41}$$

and that $\mathrm{E}(S_e S_{e'} \mid D_{e'-1}) = 0$ for $e < e'$. Expression (7.41) may be estimated by substituting unknown quantities with their empirical counterparts, and Mack obtains the estimator

$$\mathrm{E}^{*}\left(S_e^2 \mid D_{e-1}\right) = \left(\prod_{e'=J-j+1}^{D} \delta^*_{e'}\right)^2 \frac{\gamma^*_e}{\delta^{*2}_e} \frac{1}{\sum_{j=1}^{J-e} \tilde{X}_{j,e-1}} \tag{7.42}$$

Assembling all the terms, one obtains the following estimator of the mean squared error:

$$\mathrm{E}^*\left(\left(\tilde{X}_j - \bar{X}_j\right)^2 \mid D\right) = \bar{X}_j^2 \sum_{e=J-j+1}^{D} \frac{\gamma_e^*}{\delta_e^{*2}} \left(\frac{1}{\tilde{X}_{j,e-1}} + \frac{1}{\sum_{j=1}^{J-e} \tilde{X}_{j,e-1}}\right) \tag{7.43}$$

The mean squared error of the overall estimate of outstanding claim development is

$$\begin{aligned} \mathrm{E}\left(\left(\sum_{j=1}^{J}(\tilde{X}_j - \bar{X}_j)\right)^2 \mid D\right) &= \sum_{j=1}^{J} \mathrm{E}\left(\left(\tilde{X}_j - \bar{X}_j\right)^2 \mid D\right) \\ &+ 2\sum_{j=1}^{J}\sum_{k=j+1}^{J} \mathrm{E}\left(\left(\tilde{X}_j - \bar{X}_j\right)\left(\tilde{X}_k - \bar{X}_k\right) \mid D\right) \end{aligned}$$

The first sum on the right hand side can be taken directly from (7.43). For the second sum, the assumed independence between accident periods implies

$$\begin{aligned} \mathrm{E}\left(\left(\tilde{X}_j - \bar{X}_j\right)\left(\tilde{X}_k - \bar{X}_k\right) \mid D\right) &= \left(\mathrm{E}\left(\tilde{X}_j \mid D_{j,J-j}\right) - \bar{X}_j\right)\left(\mathrm{E}\left(\tilde{X}_k \mid D_{k,J-k}\right) - \bar{X}_k\right) \\ &= \tilde{X}_{j,J-j}\tilde{X}_{k,J-k}\left(\prod_{e=J-j+1}^{D} \delta_e - \prod_{e=J-j+1}^{D} \delta_e^*\right) \\ &\times \left(\prod_{e=J-k+1}^{D} \delta_e - \prod_{e=J-k+1}^{D} \delta_e^*\right) \end{aligned} \tag{7.44}$$

Using a similar approximation as that leading to (7.42), Mack obtains

$$\begin{aligned} \mathrm{E}^*\left(\left(\sum_{j=1}^{J}\left(\tilde{X}_j - \bar{X}_j\right)\right)^2 \mid D\right) &= \sum_{j=1}^{J} \mathrm{E}^*\left(\left(\tilde{X}_j - \bar{X}_j\right)^2 \mid D\right) \\ &+ 2\sum_{j=1}^{J} \bar{X}_j \left(\sum_{k=j+1}^{J} \bar{X}_j\right) \sum_{e=J-j+1}^{D} \frac{\gamma_e^*}{\delta_e^{*2}} \cdot \frac{1}{\sum_{n=1}^{J-e} \tilde{X}_{n,e-1}} \end{aligned} \tag{7.45}$$

Mack (1993) claims that his model is distribution-free. However, one should note that in order to derive expressions for the conditional mean squared error given the data, he is making very specific assumptions (1)-(2), without checking whether there exist sensible models where those assumptions are satisfied. Mack seems to imply that in using the chain ladder method, one is accepting a model of the form (1)-(3). This author disagrees with such an implication, while reserving the right to use the chain ladder method.

One model that would fit into Mack's mould, is the following: Conditional on the development of an accident period's claims up to development period e-1, the incremental development in development period e is generated by a compound Poisson random variable with a frequency parameter that is [proportional to] $\tilde{X}_{j,e-1}$ and a severity distribution H_e that is independent of $\tilde{X}_{j,e-1}$ but depends on the development period e. In formulas

$$X_{je} \mid D_{j,e-1} \sim \text{Compound Poisson } (\tilde{X}_{j,e-1}, H_e) \tag{7.46}$$

$$\delta_e - 1 = \int_{-\infty}^{\infty} u \, dH_e(u) \text{ and } \gamma_e = \int_{-\infty}^{\infty} u^2 \, dH_e(u) \tag{7.47}$$

This author believes that these assumptions exaggerate the dependency between what happened before and what happens next. One (the only?) situation in which they are asymptotically satisfied, is in the Bayesian Poisson/Gamma model for claim counts, when the variance of the Gamma prior tends to infinity.

7.9 Hertig's model

Hertig has proposed a parametric model for development factors.

Before we get stuck into Hertig's model, let us quickly recapitulate some properties of the lognormal distribution. A random variable Z is said to be lognormally distributed with parameters (ξ, σ^2) if $L = \ln(Z) \sim$Normal(ξ, σ^2). Using the moment generating function of the normal distribution of L, it is easy to verify that $\text{E}(Z) = \exp\left(\xi + \sigma^2/2\right)$ and Var(Z) =$\text{E}^2(Z)\left(e^{\sigma^2} - 1\right)$. Finally, if

$$\begin{pmatrix} L_1 \\ L_2 \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right) \tag{7.48}$$

and $Z_i = e^{L_i}$ for $i = 1, 2$, then Cov(Z_1, Z_2) =E(Z_1)E$(Z_2)\left(e^{\rho\sigma_1\sigma_2} - 1\right)$.

Let $D = \left\{\tilde{X}_{je} : j = 1, \cdots, J, e = 0, \cdots, J - j\right\}$ denote the observed (cumulative) data , and let us write $D_{je} = \left\{\tilde{X}_{je'} : e' = 0, \cdots, e\right\}$ for the development of accident period j up to development period e. Denote the development ratios of accident period j by $Z_{je} = \tilde{X}_{je}/\tilde{X}_{j,e-1}$ for $e = 1, 2, \cdots$. As before, let the maximum observed delay be denoted by $D = J - 1$.

Assume that the random variables $\{Z_{je} : j = 1, \cdots, J, e = 1, \cdots, D\}$ are independent and that there exist constants $\left\{(\xi_e, \sigma_e^2) : e = 1, \cdots, D\right\}$ so that $Z_{je} \sim$lognormal(ξ_e, σ_e^2). That is Hertig's model in a nutshell.

Under those assumptions, it is straightforward to derive the conditional distribution of the ultimate claims $\tilde{X}_j$ of accident period j, given its observed history:

$$\tilde{X}_j \mid D_{j,J-j} \sim \text{lognormal}\left(\ln \tilde{X}_{j,J-j} + \sum_{e=J-j+1}^{D} \xi_e, \sum_{e=J-j+1}^{D} \sigma_e^2\right) \quad (7.49)$$

Here we have used that $\tilde{X}_j = \tilde{X}_{j,J-j} \prod_{e=J-j+1}^{D} Z_{je}$.

It is also straightforward to find estimators of the parameters $\{(\xi_e, \sigma_e^2) : e = 1, \cdots, D\}$. They are

$$\xi_e^* = \frac{1}{n_e} \sum_{j=1}^{n_e} \ln(Z_{je}) \ (e = 1, 2, \cdots, D) \quad (7.50)$$

and

$$\sigma_e^{*2} = \frac{1}{n_e - 1} \sum_{j=1}^{n_e} (\ln(Z_{je}) - \xi_e^*)^2 \ \ (e = 1, 2, \cdots, D-1) \quad (7.51)$$

where we have denoted the number of observed development factors in development period e by $n_e = J - e$. Using Student-Fisher's theorem and the assumed independence of development factors pertaining to different development periods, we conclude that all ξ_e^* and σ_e^{*2} are independent and that $\xi_e^* \sim$Normal$(\xi_e, \sigma_e^2/n_e)$ and $\sigma_e^{*2} \sim \frac{\sigma_e^2}{n_e-1}\chi^2_{n_e-1}$.

Hertig now proposes the following predictor for the ultimate claims of accident period j:

$$\bar{X}_j = \tilde{X}_{j,J-j}\exp\left(\sum_{e=J-j+1}^{D}\left(\xi_e^* + \frac{1}{2}\sigma_e^2\left(1 + \frac{1}{n_e}\right)\right)\right) \quad (7.52)$$

Let us for the moment ignore the fact that σ_e^2 are unknown. The rationale for Hertig's predictor becomes apparent when one considers the probability distribution of the ratio

$$\frac{\tilde{X}_j}{\bar{X}_j} = \exp\left(\sum_{e=J-j+1}^{D}\left(\ln(Z_{je}) - \xi_e^* - \frac{1}{2}\sigma_e^2\left(1 + \frac{1}{n_e}\right)\right)\right) \quad (7.53)$$

Using that the estimators ξ_e^* (being based on past data for accident periods other than j) are independent of the future development of accident period j, the conditional distribution of $\frac{\tilde{X}_j}{\bar{X}_j}$ is found to be

$$\frac{\tilde{X}_j}{\bar{X}_j} \mid D_{j,J-j} \sim \text{lognormal}\left(-\frac{1}{2}\sum_{e=J-j+1}^{D}\sigma_e^2\left(1+\frac{1}{n_e}\right), \sum_{e=J-j+1}^{D}\sigma_e^2\left(1+\frac{1}{n_e}\right)\right) \tag{7.54}$$

so that $\mathrm{E}\Big(\frac{\tilde{X}_j}{\bar{X}_j} \mid D_{j,J-j}\Big) = 1$. Thus the predictors (7.52) are "relatively unbiased".

Using the lognormal distribution, one can immediately conclude that

$$\mathrm{Var}\left(\frac{\tilde{X}_j}{\bar{X}_j} \mid D_{j,J-j}\right) = \exp\left(\sum_{e=J-j+1}^{D}\sigma_e^2\left(1+\frac{1}{n_e}\right)\right) - 1 \tag{7.55}$$

Thus one can calculate the conditional mean squared error of the predictor $\bar{X}_j$ as

$$\begin{aligned}
\mathrm{E}\left(\left(\tilde{X}_j - \bar{X}_j\right)^2 \mid D_{j,J-j}\right) &= \mathrm{E}\left(\bar{X}_j^2\left(\tfrac{\tilde{X}_j}{\bar{X}_j}-1\right)^2 \mid D_{j,J-j}\right) \\
&= \bar{X}_j^2 \mathrm{Var}\left(\tfrac{\tilde{X}_j}{\bar{X}_j} \mid D_{j,J-j}\right) \\
&= \bar{X}_j^2\left(\exp\left(\textstyle\sum_{e=J-j+1}^{D}\sigma_e^2\left(1+\frac{1}{n_e}\right)\right) - 1\right)
\end{aligned} \tag{7.56}$$

This expression may be estimated if one replaces the unknown σ_e^2 by the estimators σ_e^{*2}.

Similarly one can calculate an expression for the mean squared error of the overall predictor:

$$\begin{aligned}
\mathrm{E}\left(\textstyle\sum_{j=1}^{J}\left(\tilde{X}_j - \bar{X}_j\right)\right)^2 &= \textstyle\sum_{j=1}^{J}\sum_{k=1}^{J}\mathrm{E}\left(\tilde{X}_j - \bar{X}_j\right)\left(\tilde{X}_k - \bar{X}_k\right) \\
&= \textstyle\sum_{j=1}^{J}\sum_{k=1}^{J}\mathrm{E}\left(\bar{X}_j\bar{X}_k\left(\frac{\tilde{X}_j}{\bar{X}_j}-1\right)\left(\frac{\tilde{X}_k}{\bar{X}_k}-1\right)\right) \\
&= \textstyle\sum_{j=1}^{J}\sum_{k=1}^{J}\mathrm{EE}\left(\bar{X}_j\bar{X}_k\left(\frac{\tilde{X}_j}{\bar{X}_j}-1\right)\left(\frac{\tilde{X}_k}{\bar{X}_k}-1\right) \mid D_{j,J-j}, D_{k,J-k}\right) \\
&= \textstyle\sum_{j=1}^{J}\sum_{k=1}^{J}\bar{X}_j\bar{X}_k\mathrm{ECov}\left(\frac{\tilde{X}_j}{\bar{X}_j}, \frac{\tilde{X}_k}{\bar{X}_k} \mid D_{j,J-j}, D_{k,J-k}\right)
\end{aligned} \tag{7.57}$$

Using the same representation as in (7.53) for both j and k, one can derive the covariance

$$\mathrm{Cov}\left(\frac{\tilde{X}_j}{\bar{X}_j}, \frac{\tilde{X}_k}{\bar{X}_k} \mid D_{j,J-j}, D_{k,J-k}\right) = \exp\left(\sum_{e=J-j+1}^{D}\frac{\sigma_e^2}{n_e}\right) - 1 \tag{7.58}$$

for $j<k$. Here we have used the fact that the only dependency comes through the common summands $\xi^*_{J-j+1}, \cdots, \xi^*_D$.

Assembling (7.57) up we find

$$\mathrm{E}\left(\sum_{j=1}^{J}\left(\tilde{X}_j - \bar{X}_j\right)\right)^2 = \sum_{j=1}^{J}\sum_{k=1}^{J}\bar{X}_j\bar{X}_k\left(\exp\left(\sum_{e=J-(j\wedge k)+1}^{D}\left(\delta_{jk} + \frac{\sigma_e^2}{n_e}\right)\right) - 1\right) \tag{7.59}$$

Again, to calculate this expression one must replace the unknown σ_e^2 by the estimators σ_e^{*2}.

This derivation of the mean squared error of Hertig's estimator is due to Taylor (2000). As we noted above, Hertig assumed in his derivation that the variances were known, and only were replaced by estimates at the last stage. Taylor is more circumspect and defines

$$\bar{X}_j^* = \tilde{X}_{j,J-j}\exp\left(\sum_{e=J-j+1}^{D}\xi_e^* + \frac{1}{2}\sum_{e=J-j+1}^{D}\sigma_e^{*2}\left(1 + \frac{1}{n_e}\right)\right) \tag{7.60}$$

from the outset. Using Student-Fisher's theorem and the moment generating function of the χ^2-distribution, one can derive closed expressions for the variances and covariances of $\tilde{X}_j/\bar{X}_j^*$. Not surprisingly, those expressions are very complicated and still involve the σ_e^2.

It is interesting to compare the conditional mean-variance structure of Hertig's model with that of Mack's model. In Hertig's model we have

$$\mathrm{E}\left(\tilde{X}_{je}|\tilde{X}_{j,e-1}\right) = \tilde{X}_{j,e-1}\mathrm{E}\left(Z_{je}\right) = \tilde{X}_{j,e-1}\exp\left(\xi_e + \frac{1}{2}\sigma_e^2\right) \tag{7.61}$$

$$\mathrm{Var}\left(\tilde{X}_{je}|\tilde{X}_{j,e-1}\right) = \tilde{X}_{j,e-1}^2\mathrm{Var}\left(Z_{je}\right) = \tilde{X}_{j,e-1}^2\exp^2\left(\xi_e + \frac{1}{2}\sigma_e^2\right)\left(\exp\left(\sigma_e^2\right) - 1\right) \tag{7.62}$$

8
Log-linear models

8.1 Introduction

The following chapters will present a number of statistical model-fitting approaches rather briefly, starting with log-linear models. The reader can consult Taylor (2000) for an extensive treatment of statistical model-fitting techniques.

An apology concerns the use of notation. Throughout the previous chapters I have attempted to use a standard notation for data and model parameters. In this chapter it is necessary to deviate from the standard, which includes recycling some previously defined symbols with new meanings attached to them. The reason is that the statistical theories that underly the different models have their own notations, which utilise many of the same symbols as we have used before. Rather than defining a bevy of new symbols, I have chosen to tolerate some duplication and trust that the reader will understand the meaning of symbols from their context.

A second apology is for the extensive use of matrix notation. In the opinion of this author that without resorting to matrix notation, one does not stand a chance of seeing through the forest of detail.

8.2 Log-linear models

Let us denote the quantity that is subject to analysis by X_{je} as before. The common definition of X_{je} is $X_{je} = U_{je}$, i.e. X_{je} means incremental claim payments. A common and useful distribution in modelling insurance claims is the lognormal, i.e. assuming that the logarithms of X_{je} follow a normal distribution. Since logarithms are required, the X_{je} should be strictly positive with high probability (a few isolated exceptions we can always deal with). Thus it would not be a good idea to define $X_{je} = W_{je}$ (reported claim cost), because incremental reported claim cost often will be negative. As before, p_j will denote a non-random measure of risk exposed in accident period j.

If one studies the two-dimensional models of the previous chapter, one finds that they all comprise an accident period effect and a development period effect. Sometimes one would like to model additional effects, for example: a calendar period effect; a trend in the accident period effect; a parametric function for the development period effect; and so on. All this can be modelled within the framework of log-linear models.

Let us therefore assume that the random variables X_{je}/p_j are stochastically independent and lognormally distributed, i.e., $Z_{je} = \ln(X_{je}/p_j) \sim \mathrm{N}\big(\mu_{je}, \sigma_{je}^2\big)$. This means inter alia that

$$\mathrm{E}\left(X_{je}/p_j\right) = \exp\left(\mu_{je} + \sigma_{je}^2/2\right) \tag{8.1}$$

$$\mathrm{Var}\left(X_{je}/p_j\right) = \mathrm{E}^2\left(X_{je}/p_j\right)\left(\exp\left(\sigma_{je}^2\right) - 1\right) \tag{8.2}$$

The question is, how could the parameters μ_{je}, σ_{je}^2 be modelled? Let us consider some possibilities.

1. In 4.4 we saw that the chain ladder method could be derived as a maximum likelihood procedure in a Poisson model with multiplicative means. A model with multiplicative means could be obtained by defining $\mu_{je} = \alpha_j + \beta_e$.

2. In 4.3 we saw that the Bornhuetter-Ferguson method could be derived as a maximum likelihood procedure in a Poisson model with multiplicative means and a fixed accident period parameter. Analogeously, we could define $\mu_{je} = \alpha + \beta_e$ here.

3. Maybe we are quite happy with the assumption of a fixed accident period parameter, for all but one exceptional accident period j'. We could model this by setting $\mu_{je} = \alpha \cdot I(j \neq j') + \alpha' \cdot I(j = j') + \beta_e$.

4. Or maybe we would like to capture a calendar period effect. This we can achieve by modelling $\mu_{je} = \alpha_j + \beta_e + \lambda_{j+e}$. If the calendar period effect is caused by inflation at a constant rate λ, we could model $\mu_{je} = \alpha_j + \beta_e + \lambda \cdot (j + e - 1)$.

5. We could attempt to model the development period effect β_e as a parametric function in order to reduce the number of parameters. For

example, $\beta_e = \delta \cdot \ln(k+e) + \gamma \cdot (k+e)$ with a known location parameter $k > 0$ results in a gamma-shaped function that only involves two parameters δ and γ. It is also known as a Hoerl curve.

6. If we are happy to use gamma-shaped function in the tail ($e > e'$) but not at the early development periods, then we could model $\mu_{je} = \alpha_j + \beta_e \cdot I(e \leq e') + (\delta \cdot \ln(k+e) + \gamma \cdot (k+e)) \cdot I(e > e')$.

7. ... and so forth.

I think you have got the picture by now. In every model specification, the mean μ_{je} could be written in the form

$$\mu_{je} = \mathbf{m}_{je}^{'} \boldsymbol{\beta} \tag{8.3}$$

with $\mathbf{m}_{je}^{'}$ denotes a row vector of knowable covariates associated with the distribution of Z_{je} and $\boldsymbol{\beta}$ denotes a column vector of unknown regression parameters. Let us denote the number of observations by n, the number of unknown parameters by p, and the degrees of freedom by $f = n - p$. Let us further assume that $\sigma_{je}^2 = \sigma^2 / w_{je}$, where the weights w_{je} are yet to be determined.

We now stack the observed variables $\{Z_{je} : j + e \leq J\}$ into a column vector that we denote by $\mathbf{Z}^{n \times 1}$, and the corresponding covariate vectors $\{\mathbf{m}_{je}^{'} : j + e \leq J\}$ into a matrix that we denote by $\mathbf{M}^{n \times p}$. Finally, we arrange the weights

$$\{w_{je} : j + e \leq J\} \tag{8.4}$$

in a diagonal matrix $\mathbf{W}^{n \times n}$.

It is easy to see that under the model assumptions, the random vector $\mathbf{Z}$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu} = \mathbf{M}\boldsymbol{\beta}$ and covariance matrix

$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{W}^{-1} \tag{8.5}$$

. We assume that $\mathbf{M}$ is of full column rank p. Let $\mathbf{Q}^{p \times p} = \mathbf{M}'\mathbf{W}\mathbf{M}$.

Using the standard theory of normal linear models, we can immediately write down maximum likelihood estimators of the model parameters:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{M}'\mathbf{W}\mathbf{M}\right)^{-1} \mathbf{M}'\mathbf{W}\mathbf{Z} \tag{8.6}$$

and

$$\hat{\sigma}^2 = n^{-1} \left(\mathbf{Z} - \mathbf{M}\hat{\boldsymbol{\beta}}\right)^{'} \mathbf{W} \left(\mathbf{Z} - \mathbf{M}\hat{\boldsymbol{\beta}}\right) \tag{8.7}$$

The Student-Fisher Theorem tells us that $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are stochastically independent with

$$\hat{\boldsymbol{\beta}} \sim \mathrm{N}_p\left(\boldsymbol{\beta}\,,\, \sigma^2\left(\mathbf{M}'\mathbf{W}\mathbf{M}\right)^{-1}\right) \sim \mathrm{N}_p\left(\boldsymbol{\beta}\,,\, \sigma^2\mathbf{Q}^{-1}\right) \tag{8.8}$$

and

$$\hat{\sigma}^2 \sim \left(\frac{\sigma^2}{n}\right)\chi^2_{n-p} \sim \left(\frac{\sigma^2}{n}\right)\Gamma\left(\frac{f}{2}, \frac{1}{2}\right) \tag{8.9}$$

We are not finished yet, because insurance liabilities are not settled in log(money). What we actually need, is a predictor of the future payments $X_> = \{\, X_{je} = p_j \exp\left(Z_{je}\right) : j+e > J \,\}$ and an estimate of its mean squared error.

The expected value of X_{je} is $p_j\xi_{je}$, where ξ_{je} is defined by

$$\xi_{je} = \exp\left(\mathbf{m}^{'}_{je}\boldsymbol{\beta} + \frac{1}{2}\frac{\sigma^2}{w_{je}}\right) \tag{8.10}$$

The maximum likelihood estimator of ξ_{je} is

$$\hat{\xi}_{je} = \exp\left(\mathbf{m}^{'}_{je}\hat{\boldsymbol{\beta}} + \frac{1}{2}\frac{\hat{\sigma}^2}{w_{je}}\right) \tag{8.11}$$

Using the independence between $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$, the multivariate normal distribution of $\hat{\boldsymbol{\beta}}$, and the chi-square distribution of $\hat{\sigma}^2$ and, we find that

$$\begin{aligned}
\mathrm{E}\left(\hat{\xi}_{je}\right) &= \mathrm{E}\exp\left(\mathbf{m}^{'}_{je}\hat{\boldsymbol{\beta}}\right) \times \mathrm{E}\exp\left(\tfrac{1}{2}\tfrac{\hat{\sigma}^2}{w_{je}}\right) \\
&= \exp\left(\mathbf{m}^{'}_{je}\boldsymbol{\beta} + \tfrac{1}{2}\sigma^2\mathbf{m}^{'}_{je}\mathbf{Q}^{-1}\mathbf{m}_{je}\right) \times M_f\left(\tfrac{1}{2}\tfrac{\sigma^2}{nw_{je}}\right) \\
&= \exp\left(\mathbf{m}^{'}_{je}\boldsymbol{\beta} + \tfrac{1}{2}\sigma^2\mathbf{m}^{'}_{je}\mathbf{Q}^{-1}\mathbf{m}_{je}\right) \times \left(1 - \tfrac{\sigma^2}{nw_{je}}\right)^{-\frac{f}{2}}
\end{aligned} \tag{8.12}$$

In this expression, $M_f(t) = (1-2t)^{-f/2}$ denotes the moment generating function of the χ^2_f distribution (remember that $f = n - p$ denotes the degrees of freedom). We assume that $\sigma^2 < nw_{je}$ so that the m.g.f. is defined. The estimators $\hat{\xi}_{je}$ are biased in finite samples.

A first-order bias corrected estimator of ξ_{je} is

$$\tilde{\xi}_{je} = \exp\left(\mathbf{m}^{'}_{je}\hat{\boldsymbol{\beta}} + \frac{1}{2}\hat{\sigma}^2\left(w_{je}^{-1} - \mathbf{m}^{'}_{je}\mathbf{Q}^{-1}\mathbf{m}_{je}\right)\right) \tag{8.13}$$

Using the same reasoning as in (8.12), we see that this estimator has mean

$$\begin{aligned} \mathrm{E}\left(\tilde{\xi}_{je}\right) &= \exp\left(\mathbf{m}_{je}^{'}\boldsymbol{\beta} + \tfrac{1}{2}\sigma^2\mathbf{m}_{je}^{'}\mathbf{Q}^{-1}\mathbf{m}_{je}\right) \\ &\times \left(1 - \frac{\sigma^2\left(w_{je}^{-1} - \mathbf{m}_{je}^{'}\mathbf{Q}^{-1}\mathbf{m}_{je}\right)}{n}\right)^{-\frac{f}{2}} \end{aligned}$$

These estimators are still biased in finite samples.

In order to find an unbiased estimator of ξ_{je}, we compute the moments of all orders of $\hat{\sigma}^2$. Using (8.9) and the moments of a gamma distribution, we find that

$$\mathrm{E}\left(\hat{\sigma}^{2k}\right) = \sigma^{2k} \cdot \left(\frac{f}{n}\right)^k \cdot \gamma_k\left(\frac{f}{2}, \frac{f}{2}\right) \quad \text{for } k = 0, 1, 2, \cdots \tag{8.14}$$

where $\gamma_k(\alpha, \beta) = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\beta^k}$ denotes the k'th order moment of a $\Gamma(\alpha, \beta)$distribution. Now let us denote the well-known, unbiased estimator of σ^2 by

$$\bar{\sigma}^2 = \hat{\sigma}^2\left(\frac{n}{f}\right) = (n-p)^{-1}\left(\mathbf{Z} - \mathbf{M}\hat{\boldsymbol{\beta}}\right)^{'}\mathbf{W}\left(\mathbf{Z} - \mathbf{M}\hat{\boldsymbol{\beta}}\right) \tag{8.15}$$

Using (8.14) and (8.15) and the Taylor expansion of e^x, is is now easy to verify that

$$\mathrm{E}\left(\sum_{k=0}^{\infty}\gamma_k^{-1}\left(\frac{f}{2}, \frac{f}{2}\right)\frac{\bar{\sigma}^{2k}}{k!}\right) = \sum_{k=0}^{\infty}\frac{\sigma^{2k}}{k!} = \exp\left(\sigma^2\right) \tag{8.16}$$

If we define the function

$$G_f(t) = \sum_{k=0}^{\infty}\gamma_k^{-1}\left(\frac{f}{2}, \frac{f}{2}\right)\frac{t^k}{k!} \tag{8.17}$$

then it is easy to verify that

$$\mathrm{E}\, G_f\left(\bar{\sigma}^2 b\right) \quad = \quad \exp\left(\sigma^2 b\right) \tag{8.18}$$

for any constant b.

In particular, an unbiased estimator of ξ_{je} is

$$\bar{\xi}_{je} = \exp\left(\mathbf{m}_{je}^{'}\hat{\boldsymbol{\beta}}\right) \cdot G_{n-p}\left(\frac{1}{2}\bar{\sigma}^2\left(w_{je}^{-1} - \mathbf{m}_{je}^{'}\mathbf{Q}^{-1}\mathbf{m}_{je}\right)\right) \tag{8.19}$$

Next, let us derive a formula for the MSEP of an overall estimate based on $\bar{\xi}_{je}$.

The overall outstanding claim liability is $X_> = \sum_{j+e>J} X_{je}$. Let us write $\bar{X}_{je} = p_j \bar{\xi}_{je}$ for $j+e>J$, and $\bar{X}_> = \sum_{j+e>J} \bar{X}_{je}$. Using the unbiasedness of $\bar{X}_{je}$ and the independence between past and future observations, we can write the MSE as

$$\begin{aligned} \mathrm{E}\left(X_> - \bar{X}_>\right)^2 &= \sum_{j+e>J}\sum_{j'+e'>J} \mathrm{E}\left(X_{je} - \bar{X}_{je}\right)\left(X_{j'e'} - \bar{X}_{j'e'}\right) \\ &= \sum_{j+e>J}\sum_{j'+e'>J} \left(\mathrm{Cov}\left(X_{je}, X_{j'e'}\right) + \mathrm{Cov}\left(\bar{X}_{je}, \bar{X}_{j'e'}\right)\right) \end{aligned} \tag{8.20}$$

Now

$$\mathrm{Cov}\left(X_{je}, X_{j'e'}\right) \quad = \quad \delta_{jj'}\,\delta_{ee'}\,p_j^2\,\xi_{je}^2\,\left(\exp\left(\sigma^2/w_{je}\right) - 1\right) \tag{8.21}$$

and

$$\mathrm{Cov}\left(\bar{X}_{je}, \bar{X}_{j'e'}\right) = p_j\,p_{j'}\,\left(\mathrm{E}\left(\bar{\xi}_{je}\bar{\xi}_{j'e'}\right) - \xi_{je}\xi_{j'e'}\right) \tag{8.22}$$

Using the abbreviation $v_{je} = \left(w_{je}^{-1} - \mathbf{m}_{je}^{'}\mathbf{Q}^{-1}\mathbf{m}_{je}\right)/2$, we find the expression

$$\begin{aligned} \mathrm{E}\left(\bar{\xi}_{je}\bar{\xi}_{j'e'}\right) &= \mathrm{E}\left(\exp\left(\mathbf{m}_{je}^{'}\hat{\boldsymbol{\beta}}\right) G_f\left(\bar{\sigma}^2 v_{je}\right) \exp\left(\mathbf{m}_{j'e'}^{'}\hat{\boldsymbol{\beta}}\right) G_f\left(\bar{\sigma}^2 v_{j'e'}\right)\right) \\ &= \mathrm{E}\left(\exp\left(\left(\mathbf{m}_{je}^{'} + \mathbf{m}_{j'e'}^{'}\right)^{'}\hat{\boldsymbol{\beta}}\right)\right) \times \mathrm{E}\left(G_f\left(\bar{\sigma}^2 v_{je}\right) G_f\left(\bar{\sigma}^2 v_{j'e'}\right)\right) \\ &= [1] \;\times\; [2] \end{aligned} \tag{8.23}$$

with

$$
\begin{aligned}
[1] &= \mathrm{E}\left(\exp\left(\left(\mathbf{m}_{je}^{'}+\mathbf{m}_{j'e'}^{'}\right)^{'}\hat{\boldsymbol{\beta}}\right)\right) \\
&= \exp\left(\left(\mathbf{m}_{je}^{'}+\mathbf{m}_{j'e'}^{'}\right)^{'}\boldsymbol{\beta}\right) \\
&\times \exp\left(\tfrac{1}{2}\sigma^2\left(\mathbf{m}_{je}^{'}+\mathbf{m}_{j'e'}^{'}\right)^{'}Q^{-1}\left(\mathbf{m}_{je}^{'}+\mathbf{m}_{j'e'}^{'}\right)\right)
\end{aligned}
\tag{8.24}
$$

and, using the abbreviation $\gamma_k = \gamma_k\left(\frac{f}{2}, \frac{f}{2}\right)$,

$$
\begin{aligned}
[2] &= \mathrm{E}\left(G_f\left(\bar{\sigma}^2 v_{je}\right)\cdot G_f\left(\bar{\sigma}^2 v_{j'e'}\right)\right) \\
&= \mathrm{E}\left(\sum_{k=0}^{\infty}\sum_{k'=0}^{\infty}\gamma_k^{-1}\frac{\left(\bar{\sigma}^2 v_{je}\right)^k}{k\,!}\cdot\gamma_{k'}^{-1}\frac{\left(\bar{\sigma}^2 v_{j'e'}\right)^{k'}}{k'\,!}\right) \\
&= \sum_{k=0}^{\infty}\sum_{k'=0}^{\infty}\left(\frac{\gamma_{k+k'}}{\gamma_k\,\gamma_{k'}}\right)\cdot\left(\frac{v_{je}^k\,v_{j'e'}^{k'}}{k!\,k'!}\right)\cdot\sigma^{2(k+k')}
\end{aligned}
\tag{8.25}
$$

Now one can collect terms from (8.21)-(8.25) to assemble the MSEP in (8.20). A consistent estimator of the MSEP, albeit not an unbiased one, can be obtained by replacing the parameters ξ, β and σ^2 with estimates in (8.21)-(8.25).

The function G_f in the estimator $\bar{\xi}_{je}$ may appear to be a bit unsatisfactory at first glance, as it involves an infinite series. To that objection, let it suffice to respond that $G_f(t)$ and (8.25) can be evaluated to any desired degree of precision, using any modern programming language.

The meaning of the weights w_{je} has not been made clear yet. Since they affect the variance of the logarithmic variables $Z_{je} = \ln\left(X_{je}/p_j\right)$, the w_{je} are not necessarily proportional to p_j, as one would intuitively assume if the variance of X_{je}/p_j itself was affected. Taylor (2000) determines the weights by visual inspection of the residuals. Let us briefly summarise his procedure.

One starts with observing that $\frac{Z_{je}-\mathbf{m}_{je}^{'}\hat{\boldsymbol{\beta}}}{\sigma/\sqrt{w_{je}}} \sim \mathrm{N}(0,1)$ if the model is correctly specified; moreover, in large samples, the Pearson standardised residuals R_{je} will have a mean of approximately zero and a variance of approximately one:

$$
R_{je} \;=\; \frac{Z_{je}-\mathbf{m}_{je}^{'}\hat{\boldsymbol{\beta}}}{\hat{\sigma}/\sqrt{w_{je}}} \;\overset{\text{approx.}}{\sim}\; [0,1] \tag{8.26}
$$

In order to fit a model using the standardised residuals and the requirement (8.26), one can proceed as follows:

Start with an initial set of $\{w_{je} : j+e \leq J\}$, for example, $w_{je} \equiv 1$. Estimate β and σ^2 and calculate the residuals R_{je}. Inspect graphs of R_{je} to check whether (8.26) appears to be reasonably satisfied along the axes of

j (accident period), e (development period) and $j+e$ (valuation period). If the graphs reveal a violation of (8.26), adjust the w_{je} to reduce or remove the violation. Estimate β and σ^2 again and calculate new residuals R_{je}. Inspect graphs...

Needless to say, the w_{je} cannot be just any set of numbers, but should have a functional form that allows a credible extension to the future weights $\{w_{je} : j+e > J\}$. One could surmise, for example, that w_{je} form a function like $w_{je} = \exp(-\omega \cdot e)$, and try to find a suitable value of ω. For other examples, see Taylor (2000).

8.3 Bootstrapping

For most of the methods we have seen so far, we have been able to derive a formula for the mean squared error or predictions (MSEP) within the model that justifies the method. In those cases, the MSEP could be written as a function of unknown model parameters. In order to calculate an estimate of MSEP, we would replace unknown model parameters with point estimates. There are some problems connected with this procedure.

- Point estimates are subject to randomness, and could turn out different in a new data sample, even if the new data is generated by the same underlying mechanism as the original data.
- The MSEP does not specify the distribution of the prediction error. In cases where that distribution is important, for example to calculate a percentile of outstanding claims, that distribution is essential.
- In some models, it is possible to calculate the distribution of prediction error in principle, but in practice is is rather hairy.

Enter the bootstrap. In very simple terms, bootstrapping consists of random resampling of the data, in order to generate an empirical distribution of the quantity that is of interest.

This section outlines bootstrapping only for the log-linear model. For a more general presentation, the reader may consult Taylor (2000), Pinheiro et al. (2000), and the references in those sources.

Let us therefore continue with the log-linear model. We would like to generate, by simulation, a distribution of the prediction error $X_> - \bar{X}_>$. In order to do so, we consider first the distribution of the residuals (8.26), repeated here for ease of reference:

$$R_{je} = \frac{Z_{je} - \mathbf{m}_{je}'\hat{\boldsymbol{\beta}}}{\hat{\sigma}/\sqrt{w_{je}}} \qquad (8.27)$$

Now let us stack the observed residuals

$$\{R_{je} : j + e \le J\} \tag{8.28}$$

into a column vector that we denote by $\mathbf{R}^{n\times 1}$. We can write

$$\mathbf{R} \;=\; \hat{\sigma}^{-1}\mathbf{W}^{\frac{1}{2}}\left(\mathbf{Z}-\mathbf{M}\hat{\boldsymbol{\beta}}\right) \tag{8.29}$$

Using the model assumptions, one can verify that

$$\mathrm{E}\left(\mathbf{W}^{\frac{1}{2}}\left(\mathbf{Z}-\mathbf{M}\hat{\boldsymbol{\beta}}\right)\right)=\mathbf{0} \tag{8.30}$$

and

$$\mathrm{Cov}\left(\mathbf{W}^{\frac{1}{2}}\left(\mathbf{Z}-\mathbf{M}\hat{\boldsymbol{\beta}}\right)\right)=\sigma^2\left(I-\mathbf{W}^{\frac{1}{2}}\mathbf{M}\mathbf{Q}^{-1}\mathbf{M}\mathbf{W}^{\frac{1}{2}}\right)=\sigma^2\boldsymbol{\Lambda} \tag{8.31}$$

where

$$\mathbf{W}^{\frac{1}{2}}=\mathrm{diag}\left(\sqrt{w_{je}}\right) \tag{8.32}$$

and $\boldsymbol{\Lambda}$ denotes the matrix expression in the brackets following σ^2. The matrix $\boldsymbol{\Lambda}$ is easily calculated, as it only involves known, non-random quantities.

Ignoring the randomness in $\hat{\sigma}^2$, we conclude from (8.29)-(8.31) that

$$\mathbf{R} \overset{\text{approx.}}{\sim} [\mathbf{0},\boldsymbol{\Lambda}] \tag{8.33}$$

where $[\mathbf{0},\boldsymbol{\Lambda}]$ is shorthand notation for zero mean, covariance matrix $\boldsymbol{\Lambda}$.

In order to be able to simulate new samples of the empirical residuals, we transform the distribution of the residuals to one that is (approximately) standardised - i.e. uncorrelated with zero mean and unit variance. One can use the Cholesky decomposition to find a matrix $\mathbf{L}$ that satisfies

$$\mathbf{L}\boldsymbol{\Lambda}\mathbf{L}'=\mathbf{I} \tag{8.34}$$

Then vector of standardised residuals is

$$\mathbf{S}=\mathbf{L}\cdot\mathbf{R} \overset{\text{approx.}}{\sim} [\mathbf{0},\mathbf{I}] \tag{8.35}$$

We denote the empirical distribution of the components of $\mathbf{S}$ by $\hat{F}$.

To simulate N random replications of the past data $\{Z_{je} : j+e \leq J\}$, we proceed as follows:
For $i = 1, \cdots, N$:

1. Generate $\{S_{je}^{(i)} : j+e \leq J\}$ in such a way that $S_{je}^{(i)} \overset{\text{i.i.d.}}{\sim} \hat{F}$. This one achieves by sampling from the empirical distribution $\hat{F}$, with replacement. Denote the resulting vector by $\mathbf{S}^{(i)}$.

2. Calculate a set of pseudo-residuals by $\mathbf{R}^{(i)} = \mathbf{L}^{-1}\mathbf{S}^{(i)}$.

3. Calculate a set of pseudo-observations by $\mathbf{Z}^{(i)} = \mathbf{M}\hat{\boldsymbol{\beta}} + \hat{\sigma}\mathbf{W}^{-\frac{1}{2}}\mathbf{R}^{(i)}$or, in component notation: $Z_{je}^{(i)} = \mathbf{m}'_{je}\hat{\boldsymbol{\beta}} + \hat{\sigma}\, R_{je}^{(i)}/w_{je}$ for j, e such that $j+e \leq J$.

4. On the basis of the pseudo-observations, calculate estimates $\hat{\boldsymbol{\beta}}^{(i)}$ and $\hat{\sigma}^{(i)}$as well as predictors $\bar{X}_{je}^{(i)} = p_j \bar{\xi}_{je}^{(i)}$for j, e such that $j+e > J$.

In order to evaluate the distribution of the prediction error, we must also simulate N random replications of the future data $\{Z_{je} : j+e > J\}$. In order to do so, we first calculate matrices $\bar{\mathbf{M}}$, $\bar{\mathbf{W}}$ and $\bar{\mathbf{Q}} = \bar{\mathbf{M}}'\bar{\mathbf{W}}\bar{\mathbf{M}}$ that correspond to the future data. We then calculate $\bar{\boldsymbol{\Lambda}}$ as in (8.31), mutatis mutandis, and $\bar{\mathbf{L}}$ in such a way that $\bar{\mathbf{L}}\bar{\boldsymbol{\Lambda}}\bar{\mathbf{L}}' = \mathbf{I}$. Having done that, simulation of the future is analogeous to simulation of the past.
For $i = 1, \cdots, N$, we proceed as follows:

1. Generate $\{S_{je}^{(i)} : j+e > J\}$ so that $S_{je}^{(i)} \overset{\text{i.i.d.}}{\sim} \hat{F}$. Denote the resulting vector by $\bar{\mathbf{S}}^{(i)}$.

2. Calculate a set of future pseudo-residuals by $\bar{\mathbf{R}}^{(i)} = \bar{\mathbf{L}}^{-1}\bar{\mathbf{S}}^{(i)}$.

3. Calculate pseudo-observations $Z_{je}^{(i)} = \mathbf{m}'_{je}\hat{\boldsymbol{\beta}} + \hat{\sigma}\, R_{je}^{(i)}/w_{je}$ for j, e such that $j+e > J$.

4. Calculate $X_{je}^{(i)} = p_j \exp\left(Z_{je}^{(i)}\right)$for j, e such that $j+e > J$.

For each of the N simulations, one calculates and stores the values that are of interest. At the very simplest, one would calculate the aggregate estimation error, $E^{(i)} = \sum_{j+e>J}\left(X_{je}^{(i)} - \bar{X}_{je}^{(i)}\right)$.

An empirical estimate of MSEP is then $\text{MSEP}^* = N^{-1} \sum_{i=1}^{N} \left(E^{(i)}\right)^2$. This estimate enables one to avoid the cumbersome computations (8.20)-(8.25). Of course, implementing the bootstrap procedure also requires a bit of work.

Other properties of the simulated distribution of the estimation error, in particular percentiles, are also easy to find.

The empirical nature of the bootstrap procedure allows one to analyse the properties of other predictors as well. For example, the $\hat{\xi}_{je}$ or $\tilde{\xi}_{je}$ of the previous section are not unbiased, but they are easier to evaluate than $\bar{\xi}_{je}$. Using the bootstrap, one can analyse the increase in MSEP that replacing $\bar{\xi}_{je}$ by $\hat{\xi}_{je}$ or $\tilde{\xi}_{je}$ would engender.

9
General linear models

9.1 Introduction

In the previous chapter we considered a model where the logarithms of the claim observations X_{je} formed a normal linear model, with a mean that was linearly regressed on a smaller number of parameters. This allowed us to use the theory of normal linear models on the transformed variables.

Generalised linear models take a different approach, in that it is the mean itself that is assumed to be a transformation of a linear regression. The general formulation is

$$\mathrm{E}\left(X_{je}/p_j\right) = h\left(\mathbf{m}_{je}^{'}\boldsymbol{\beta}\right) \tag{9.1}$$

$$\mathrm{Var}\left(X_{je}/p_j\right) = \tau_{je}^2\varphi/w_{je} \tag{9.2}$$

$$\mathrm{Cov}\left(X_{je}/p_j, X_{j'e'}/p_{j'}\right) = 0 \text{ for } (j,e) \neq (j',e') \tag{9.3}$$

Here $\mathbf{m}_{je}^{'}$ denotes a row vector of knowable covariates associated with the distribution of X_{je}/p_j, and β denotes a column vector of unknown regression parameters. The function h is called the response function, and its inverse h^{-1} is called the link function.

Extensive theory and software exists for the analysis of generalised linear models when the distribution of the X_{je}/p_j belongs to the exponential distribution family. The software can be used to compute estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\varphi}$. Using those estimators one can compute fitted values and predictions,

$$\hat{X}_{je}/p_j = h\left(\mathbf{m}_{je}^{'} \cdot \hat{\boldsymbol{\beta}}\right) \tag{9.4}$$

We shall see some examples in what follows.

9.2 Exponential family of distributions

The exponential family of distributions if given by the density functions

$$f(y) = a(y, \varphi)\exp\left(\frac{y\theta - b(\theta)}{\varphi}\right) \tag{9.5}$$

The mean and variance of a distribution in the exponential family may be found by first observing that the moment generating function is given by

$$\begin{aligned} M(t) &= \mathrm{E}\left(e^{tY}\right) \\ &= \int a(y,\varphi)e^{ty}\exp\left(\tfrac{y\theta - b(\theta)}{a(\varphi)}\right)dy \\ &= \exp\left(\tfrac{b(\theta+ta(\varphi)) - b(\theta)}{a(\varphi)}\right) \end{aligned} \tag{9.6}$$

Thus the mean and variance are

$$\mathrm{E}\left(Y\right) = M'\left(0\right) = b'(\theta) \tag{9.7}$$

and

$$\mathrm{Var}\left(Y\right) = M''\left(0\right) - \left(b'(\theta)\right)^2 = \varphi b''(\theta) \tag{9.8}$$

A member of a given exponential family is fully characterized by its mean and variance. Note in particular that for a given family, the variance can be expressed as a function of the mean via the "natural parameter" θ:

$$\mathrm{Var}\left(Y\right) = \varphi b''\left(\left(b'\right)^{-1}\left(\mathrm{E}\left(Y\right)\right)\right) \tag{9.9}$$

One can also write $\mathrm{Var}(Y) = \varphi\tau^2(\theta)$, where $\tau^2(\theta) = b''(\theta)$.

Normal distribution

The normal distribution with mean μ and variance σ^2 has the density

$$\begin{aligned} f(y) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right) \qquad (9.10\text{a}) \\ &= \exp\left(-\frac{1}{2}\left(\ln(2\pi\sigma^2) + \frac{y^2}{\sigma^2}\right)\right)\exp\left(\frac{y\mu - \mu^2/2}{\sigma^2}\right) \end{aligned}$$

Thus it can be cast in the form (9.5) by defining $\theta = \mu$, $\varphi = \sigma^2, b(\theta) = \theta^2/2$, $b'(\theta) = \theta$, $b''(\theta) = 1$ and $a(y, \varphi) = \exp\left(-\frac{1}{2}\left(\ln(2\pi\varphi) + \frac{y^2}{\varphi}\right)\right)$.

Poisson distribution

The Poisson distribution with mean μ and has the point probabilities

$$p(y) = \frac{\mu^y}{y!}\exp(-\mu) = \frac{1}{y!}\exp(y\ln(\mu) - \mu) \tag{9.11}$$

Thus it can be cast in the form (9.5) by defining $\theta = \ln\mu$, $\varphi = 1, b(\theta) = b'(\theta) = b''(\theta) = e^\theta$ and $c(y, \varphi) = 1/y!$.

Gamma distribution

The gamma distribution with parameters (α, β) has the density

$$\begin{aligned} f(y) &= \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} \\ &= \frac{\alpha^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp\left(\frac{y(-\beta/\alpha) - \ln(\alpha/\beta)}{1/\alpha}\right) \end{aligned} \tag{9.12}$$

Thus it can be cast in the form (9.5) by defining $\theta = -\beta/\alpha$, $\varphi = 1/\alpha$, $b(\theta) = \ln(-1/\theta)$, $b'(\theta) = -1/\theta$, $b''(\theta) = 1/\theta^2$, and $c(y, \varphi) = \frac{\alpha^\alpha}{\Gamma(\alpha)} y^{\alpha-1}$.

Other distributions

There are a number of other distributions that can be cast in the form of an EDF, see Wüthrich & Merz (2008, ch. 6).

The theory of GLMs also allows for a distibution-free approach that employs only an assumed relationship between the mean and the variance of the observations and obtains parameter estimates by maximising the quasi-likelihood. The relationship is given by (9.9). In particular in the three example distributions above, the relationship between the mean and the variance is

- Normal distribution: $\mathrm{Var}(Y) = \varphi$
- Poisson distribution: $\mathrm{Var}(Y) = \mathrm{E}(Y)$
- Gamma distribution: $\mathrm{Var}(Y) = \varphi \mathrm{E}^2(Y)$

9.3 Applications

In loss reserving exercises, the link function is normally $h^{-1}(x) = \ln(x)$. The mean of the observed variables has the multiplicative form

$$\mathrm{E}\left(X_{je}/p_j\right) = h\left(\mathbf{m}_{je}^{'}\boldsymbol{\beta}\right) = \exp(\mathbf{m}_{je}^{'}\boldsymbol{\beta}) \tag{9.13}$$

For example, a model with multiplicative accident period effects and development period effects is obtained by defining

$$\mathbf{m}_{je}^{'}\boldsymbol{\beta} = \alpha_j + \beta_e \tag{9.14}$$

A model with a fixed accident period level, and free development period effects is obtained by defining

$$\mathbf{m}_{je}^{'}\boldsymbol{\beta} = \alpha + \beta_e \tag{9.15}$$

Similarly, all the model variations suggested for the log-linear model can also be combined with a generalised linear model. If p_j denotes a measure of risk exposed, then it seems semsible to define $w_{je} = p_j$.

The Poisson distribution enjoys great popularity because with that distribution, the maximum likelihood estimates of the GLM coincide with the estimates from the Chain-ladder method. As a result, a GLM with Poisson distributions is often cited as the model underlying the Chain-ladder method.

9.4 Bootstrapping the GLM

The Pearson standardised residuals are

$$R_{je} = \frac{X_{je}/p_j - \hat{X}_{je}/p_j}{\sqrt{\hat{\mathrm{Var}}\left(\hat{X}_{je}/p_j\right)}} = \frac{X_{je}/p_j - \hat{X}_{je}/p_j}{\hat{\tau}_{je}\sqrt{\hat{\varphi}/w_{je}}} \tag{9.16}$$

If the model is correctly specified, the mean and variance of R_{je} are approximately [0,1] in large samples. This property can be used for bootstrapping. A new sample of past and future pseudo-observations may be calculated from the formula

$$X_{je}^{(i)}/p_j = h(\mathbf{m}_{je}^{'}\hat{\boldsymbol{\beta}}) + \hat{\tau}_{je}\sqrt{\hat{\varphi}/w_{je}}\, R_{je}^{(i)} \tag{9.17}$$

If one is assuming a specific distribution, one can also simulate directly from that distribution, using the estimated $\hat{\boldsymbol{\beta}}$ and $\hat{\varphi}$.

9.5 Notes

The theory of generalised linear models is described in McCullagh & Nelder (1989). For worked examples, see Taylor (2000) and Pinheiro et al (2000). Wüthrich & Merz (2008) provide details about the estimation procedure.

10

Dynamic linear models

Remark 9 *This chapter will be reworked or eliminated. I am not convinced of the usefulness of dynamic linear models in the loss reserving context, although the theory is elegant.*

11
Miscellaneous topics

11.1 Introduction

This chapter is a collection of bits and pieces and will be reworked.

11.2 Simple model diagnostics

In assessing whether a given method provides reasonable estimates, it is often useful to calculate diagnostic quantities such as

- The implied ultimate claim frequency, i.e. reported claims plus estimated claims IBNR, divided by the risk volume. Unless there are explainable deviations or trends, the ultimate claim frequencies of successive accident periods should normally form a trendless sequence. If this is not the case, the model reporting pattern may be wrong.
- The implied average claim size for reported claims, including any allowance made for future revaluations. In the absence of very large claims, the average size of reported claims of successive accident periods should form a sequence that roughly follows an inflation trendline. A give-away sign of under-estimation is when the average claim size of immature accident periods is lower than that of more mature periods. In that case one must be prepared for revaluations.
- The implied average claim size for unreported claims. In many lines of business, there is a marked dependency between the average claim

size and the reporting delay. In property insurance the dependency is usually negative, while it is often positive in casualty insurance. The average claim size of unreported claims will give you an idea whether the estimated amount of claims IBNR is in harmony with their predicted number.

- The implied risk premium, i.e. the ultimate claim cost divided by the risk volume. Unless there are explainable deviations or trends, the ultimate risk premiums of successive accident periods should normally form a trendless or inflation-linked sequence.

11.3 Analysis of development

One should always keep track of the accuracy of previous estimates. In the detailed model of chapters 4 to 6, it is easy to split up the development into its different components: Number of new claims reported (actual vs. predicted), severity of new claims reported (actual vs. predicted) and revaluation of old claims (actual vs. predicted). In the more summary models, one can of course only keep track of the quantity whose evolution one has predicted.

11.4 The NP approximation

We have seen how one can calculate the mean squared error of the estimators. Unfortunately, mean squared error does not capture the skewness that is common to claim distributions. The Normal Power (NP) approximation is an extension of the normal approximation that makes allowance for the skewness in the distribution. In this section we will see very briefly how one can use the NP approximation to approximate percentiles in the probability distribution of claims IBNR.

Consider a random variable X with a probability distribution F. The NP approximation to the $(1-\epsilon)$ percentile of F is

$$F^{-1}(1-\epsilon) = E(X) + z_{1-\epsilon}\sqrt{E(X-EX)^2} + \left(\frac{z_{1-\epsilon}-1}{6}\right)\frac{E(X-EX)^3}{E(X-EX)^2} \tag{11.1}$$

where $z_{1-\epsilon} = \Phi^{-1}(1-\epsilon)$ is the $(1-\epsilon)$ percentile in the standard normal distribution. The NP approximation has been shown to work well in the right tail (small ϵ) of probability distributions that are not too skew.

From a practical point of view the NP approximation has a very attractive property. Let $X = \sum_{i=1}^{n} X_i$, with independent, but not necessarily identically distributed random variables $X_1, \cdots, X_n$. Then it is easy to verify that $E(X - EX)^k = \sum_{i=1}^n E(X_i - EX_i)^k$ for k=2,3. Therefore the NP approximation allows one to easily approximate the aggregate distribution of independent variables, if the moments of the constituent random variables are known.

Let us now calculate the necessary moments in a mixed, compound Poisson distribution. For the time being we drop all the indeces, but retain a notation that will indicate where the ultimate aim lies.

First, let a random variable Y have a compound Poisson distribution with frequency parameter $p\pi\theta$ and severity distribution G. Denote the non-central moments of G by $\mu^{(k)} = \int x^k G(dx)$ for k=1,2,3. Then it is well-known and easy to prove by moment generating functions, that

$$E_\theta(Y) = p\pi\theta\mu^{(1)} \tag{11.2}$$

and

$$E_\theta(Y - E_\theta Y)^k = p\pi\theta\mu^{(k)} \tag{11.3}$$

for k=2,3.

Now assume that the compound distribution is conditional on the random parameter Θ, and that Θ has a mixing distribution U. Denote the non-central moments of Θ by

$$\varphi^{(k)} = E(\Theta^k) \tag{11.4}$$

By tedious algebra one can verify that the unconditional moments of Y are

$$E(Y) = p\pi\varphi^{(1)}\mu^{(1)} \tag{11.5}$$

$$E(Y - EY)^2 = p\pi\varphi^{(1)}\mu^{(2)} + (p\pi\mu^{(1)})^2(\varphi^{(2)} - [\varphi^{(1)}]^2) \tag{11.6}$$

$$\begin{aligned} E(Y - EY)^3 \;=\;& p\pi\varphi^{(1)}\mu^{(3)} \\ +\;& 3(p\pi)^2\mu^{(1)}\mu^{(2)}(\varphi^{(2)} - [\varphi^{(1)}]^2) \\ +\;& (p\pi\mu^{(1)})^3(\varphi^{(3)} - 3\varphi^{(2)}\varphi^{(1)} + 2[\varphi^{(1)}]^3) \end{aligned} \tag{11.7}$$

In order to apply the NP approximation to the claims IBNR, we have to calculate the corresponding moments there.

Let us work within the Bayesian credibility model with $\Theta_1, \cdots, \Theta_J \sim \Gamma(\alpha, \beta)$and independent. The amount of claims IBNR for accident period j is $Y_{j,>J-j} = \sum_{d=J-j+1}^{\infty} Y_{jd}$, where Y_{jd} is the ultimate claim amount in respect of claims reported in calendar period $j+d$. Conditional on $\Theta_j = \theta_j$, $Y_{j,>J-j}$ has a compound Poisson distribution with frequency parameter

$p_j\theta_j\pi_{>J-j}$ and tail severity distribution $G_{\overline{>J-j}} = \sum_{d=J-j+1}^{\infty} \pi_d G_d / \pi_{>J-j}$. The moments of the tail severity distribution are:

$$\mu^{(k)}_{\overline{>J-j}} = \sum_{d=J-j+1}^{\infty} \pi_d \mu_d^{(k)} / \pi_{>J-j} \tag{11.8}$$

for $k = 1, 2, 3$.

The non-central moments in the à priori (or structural) distribution of Θ_j are

$$\varphi^{(k)} = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} \theta^{\alpha+k-1} e^{-\beta\theta} d\theta = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\beta^k} \tag{11.9}$$

The conditional distribution of Θ_j, given counts up to time $J-j$, is again a gamma distribution with updated parameters

$$\bar{\alpha} = \alpha + N_{j,\leq J-j} \tag{11.10}$$

and

$$\bar{\beta} = \beta + p_j \pi_{\leq J-j} \tag{11.11}$$

The corresponding non-central moments are of course

$$\bar{\varphi}^{(k)} = \frac{\Gamma(\bar{\alpha}+k)}{\Gamma(\bar{\alpha})\bar{\beta}^k} \tag{11.12}$$

If one now uses $p_j, \pi_{>J-j}, \mu^{(1)}_{\overline{>J-j}}, \mu^{(2)}_{\overline{>J-j}}, \mu^{(3)}_{\overline{>J-j}}, \bar{\varphi}^{(1)}, \bar{\varphi}^{(2)}, \bar{\varphi}^{(3)}$ to assemble (11.5)-(11.7), one has all the moments needed to calculate the NP approximation to the conditional probability distribution of the IBNR claim amount of accident period j.

Further, aggregating those moments across $j = 1, ..., J$, one gets the moments needed for the overall IBNR claim distribution. Note that this last step assumes independence of $\Theta_1, \cdots, \Theta_J$, an assumption which is not fulfilled in the dynamic linear model.

One could also use unconditional moments of $\Theta_1, \cdots, \Theta_J$. As $\Theta_1, \cdots, \Theta_J$ are more dispersed in the unconditional distribution than in the conditional distribution given the data, the resulting estimate would be somewhat on the safe side. In some situations this may be desirable.

11.5 The mean squared error of the one-period run-off result

Solvency II requires consideration of one-year run-off results. In this section we briefly consider how that concept could be realised when a credibility formula is used, and under the assumptions of the Bühlmann-Straub model.

Consider passing from development period e to development period $e+1$ while using a certain credibility formula. The run-off result in period $e+1$ is then the incremental claim development in period $e+1$, plus the change in the estimated outstanding claim development. Omitting the accident period subscript j, this becomes:

$$R_{e+1} = X_{e+1} + p\left(\bar{b}_{e+1}\pi_{>e+1} - \bar{b}_e\pi_{>e}\right) \tag{11.13}$$

It is easy to verify that, unconditionally, $\mathrm{E}(R_{e+1}) = 0$, so that $\mathrm{E}\big(R_{e+1}^2\big) = \mathrm{Var}(R_{e+1})$. Let us derive that variance the long way. We start by writing:

$$R_{e+1} = X_{e+1} + p\left(\left(z_{e+1}\hat{b}_{e+1} + (1-z_{e+1})\beta\right)\pi_{>e+1} - \left(z_e\hat{b}_e + (1-z_e)\beta\right)\pi_{>e}\right) \tag{11.14}$$

Ignoring the non-stochastic terms we find the unconditional variance:

$$\begin{aligned}
\mathrm{Var}(R_{e+1}) &= \mathrm{Var}\left(X_{e+1} + pz_{e+1}\hat{b}_{e+1}\pi_{>e+1} - pz_e\hat{b}_e\pi_{>e}\right) \\
&= \mathrm{Var}(X_{e+1}) + (pz_{e+1}\pi_{>e+1})^2\,\mathrm{Var}\left(\hat{b}_{e+1}\right) + (pz_e\pi_{>e})^2\,\mathrm{Var}\left(\hat{b}_e\right) \\
&+ 2pz_{e+1}\pi_{>e+1}\mathrm{Cov}\left(X_{e+1}, \hat{b}_{e+1}\right) \\
&- 2pz_e\pi_{>e}\mathrm{Cov}\left(X_{e+1}, \hat{b}_e\right) \\
&- 2p^2 z_e z_{e+1}\pi_{>e}\pi_{>e+1}\mathrm{Cov}\left(\hat{b}_e, \hat{b}_{e+1}\right)
\end{aligned} \tag{11.15}$$

We now calculate the variances and covariances in the above expression.

$$\mathrm{Var}(X_{e+1}) = (p\pi_{e+1})^2\lambda + p\pi_{e+1}\varphi \tag{11.16}$$

$$\mathrm{Var}\left(\hat{b}_{e+1}\right) = \lambda + \frac{\varphi}{p\pi_{\leq e+1}} \tag{11.17}$$

$$\mathrm{Var}\left(\hat{b}_e\right) = \lambda + \frac{\varphi}{p\pi_{\leq e}} \tag{11.18}$$

Note that $\hat{b}_e = X_{\leq e}/(p\pi_{\leq e})$ and $\hat{b}_{e+1} = (X_{\leq e} + X_{e+1})/(p\pi_{\leq e+1})$. Thus

$$\begin{aligned}
\mathrm{Cov}\left(X_{e+1}, \hat{b}_e\right) &= \tfrac{1}{p\pi_{\leq e}}\mathrm{Cov}(X_{e+1}, X_{\leq e}) \\
&= \tfrac{1}{p\pi_{\leq e}}\left(\mathrm{ECov}(X_{e+1}, X_{\leq e}|\Theta) + \mathrm{Cov}(\mathrm{E}(X_{e+1}|\Theta), \mathrm{E}(X_{\leq e}|\Theta))\right) \\
&= \tfrac{1}{p\pi_{\leq e}}\mathrm{Cov}(p\Theta\pi_{e+1}, p\Theta\pi_{\leq e}) \\
&= p\pi_{e+1}\lambda
\end{aligned} \tag{11.19}$$

and

$$
\begin{aligned}
\operatorname{Cov}\left(X_{e+1},\hat{b}_{e+1}\right) &= \operatorname{Cov}\left(X_{e+1},\tfrac{X_{\le e}+X_{e+1}}{p\pi_{\le e+1}}\right)\\
&= \tfrac{1}{p\pi_{\le e+1}}\left(\operatorname{Cov}\left(X_{e+1},X_{\le e}\right)+\operatorname{Cov}\left(X_{e+1},X_{e+1}\right)\right)\\
&= \left(p^2\pi_{\le e}\pi_{e+1}\lambda+\left(p\pi_{e+1}\right)^2\lambda+p\pi_{e+1}\varphi\right)/\left(p\pi_{\le e+1}\right)\\
&= \left(p^2\pi_{e+1}\pi_{\le e+1}\lambda+p\pi_{e+1}\varphi\right)/\left(p\pi_{\le e+1}\right)\\
&= p\pi_{e+1}\lambda+\left(\pi_{e+1}/\pi_{\le e+1}\right)\varphi
\end{aligned}
\tag{11.20}
$$

$$
\begin{aligned}
\operatorname{Cov}\left(\hat{b}_e,\hat{b}_{e+1}\right) &= \operatorname{Cov}\left(\tfrac{X_{\le e}}{p\pi_{\le e}},\tfrac{X_{\le e}+X_{e+1}}{p\pi_{\le e+1}}\right)\\
&= \tfrac{1}{p^2\pi_{\le e}\pi_{\le e+1}}\left(\operatorname{Cov}\left(X_{\le e},X_{\le e}\right)+\operatorname{Cov}\left(X_{\le e},X_{e+1}\right)\right)\\
&= \left(\left(p\pi_{\le e}\right)^2\lambda+p\pi_{\le e}\varphi+p^2\pi_{\le e}\pi_{e+1}\lambda\right)/\left(p^2\pi_{\le e}\pi_{\le e+1}\right)\\
&= \left(p^2\pi_{\le e}\lambda\left(\pi_{\le e}+\pi_{e+1}\right)+p\pi_{\le e}\varphi\right)/\left(p^2\pi_{\le e}\pi_{\le e+1}\right)\\
&= \lambda+\tfrac{\varphi}{p\pi_{\le e+1}}
\end{aligned}
\tag{11.21}
$$

Let us now put all this together:

$$
\begin{aligned}
\operatorname{Var}\left(R_{e+1}\right) &= \left(p\pi_{e+1}\right)^2\lambda+p\pi_{e+1}\varphi+\left(pz_{e+1}\pi_{>e+1}\right)^2\left(\lambda+\tfrac{\varphi}{p\pi_{\le e+1}}\right)+\left(pz_e\pi_{>e}\right)^2\left(\lambda+\tfrac{\varphi}{p\pi_{\le e}}\right)\\
&+ 2pz_{e+1}\pi_{>e+1}\left(p\pi_{e+1}\lambda+\left(\pi_{e+1}/\pi_{\le e+1}\right)\varphi\right)\\
&- 2pz_e\pi_{>e}\left(p\pi_{e+1}\lambda\right)\\
&- 2p^2z_ez_{e+1}\pi_{>e}\pi_{>e+1}\left(\lambda+\tfrac{\varphi}{p\pi_{\le e+1}}\right)\\
&= p^2\lambda L+p\varphi F
\end{aligned}
\tag{11.22}
$$

with

$$
\begin{aligned}
L &= \pi_{e+1}^2+\left(z_{e+1}\pi_{>e+1}\right)^2+\left(z_e\pi_{>e}\right)^2\\
&+ 2z_{e+1}\pi_{>e+1}\pi_{e+1}\\
&- 2z_e\pi_{>e}\pi_{e+1}\\
&- 2z_ez_{e+1}\pi_{>e}\pi_{>e+1}
\end{aligned}
\tag{11.23}
$$

$$
\begin{aligned}
F &= \pi_{e+1}+\left(z_{e+1}\pi_{>e+1}\right)^2/\pi_{\le e+1}+\left(z_e\pi_{>e}\right)^2/\pi_{\le e}\\
&+ 2z_{e+1}\pi_{>e+1}\pi_{e+1}/\pi_{\le e+1}\\
&- 2z_ez_{e+1}\pi_{>e}\pi_{>e+1}/\pi_{\le e+1}
\end{aligned}
\tag{11.24}
$$

To minimise the MSE of the run-off result one can try to differentiate L and F by z_{e+1} :

$$
\begin{aligned}
\tfrac{\partial L}{\partial z_{e+1}} &= 2z_{e+1}\pi_{>e+1}^2+2\pi_{>e+1}\pi_{e+1}-2z_e\pi_{>e}\pi_{>e+1}\\
&= 2\pi_{>e+1}\left(z_{e+1}\pi_{>e+1}+\pi_{e+1}-z_e\pi_{>e}\right)
\end{aligned}
\tag{11.25}
$$

DRAFT
09.04.14

$$\begin{aligned}\frac{\partial F}{\partial z_{e+1}} &= 2z_{e+1}\pi^2_{>e+1}/\pi_{\leq e+1} + 2\pi_{>e+1}\pi_{e+1}/\pi_{\leq e+1} - 2z_e\pi_{>e}\pi_{>e+1}/\pi_{\leq e+1} \\ &= \tfrac{2\pi_{>e+1}}{\pi_{\leq e+1}}\left(z_{e+1}\pi_{>e+1} + \pi_{e+1} - z_e\pi_{>e}\right)\end{aligned} \tag{11.26}$$

Note that the two derivatives differ only by a constant factor. One could see what happens if one tries to minimise the MSE of the one-period run-off result.

$$\begin{aligned}\frac{\partial \mathrm{Var}(R_{e+1})}{\partial z_{e+1}} &= p^2\lambda\tfrac{\partial L}{\partial z_{e+1}} + p\varphi\tfrac{\partial F}{\partial z_{e+1}} \\ &= 2\left(p^2\lambda\pi_{>e+1} + p\varphi\tfrac{\pi_{>e+1}}{\pi_{\leq e+1}}\right)\left(z_{e+1}\pi_{>e+1} + \pi_{e+1} - z_e\pi_{>e}\right) \\ &= 2p\pi_{>e+1}\left(p\lambda + \tfrac{\varphi}{\pi_{\leq e+1}}\right)\left(z_{e+1}\pi_{>e+1} + \pi_{e+1} - z_e\pi_{>e}\right) \\ &= 2p^2\lambda\tfrac{\pi_{>e+1}}{\zeta_{e+1}}\left(z_{e+1}\pi_{>e+1} + \pi_{e+1} - z_e\pi_{>e}\right)\end{aligned} \tag{11.27}$$

where ζ_{e+1} is the optimal credibility factor at time $e+1$. The derivative is zero is and only if

$$\begin{aligned}z_{e+1} &= \left(z_e\pi_{>e} - \pi_{e+1}\right)/\pi_{>e+1} \\ &= z_e - (1-z_e)\pi_{e+1}/\pi_{>e+1} \\ &\leq z_e \text{ if } 0 \leq z_e \leq 1\end{aligned} \tag{11.28}$$

That the optimal z_{e+1} with respect to run-off result is smaller than the previous is at odds with the credibility result that has z increasing as e grows. In particular we see that $z_e = 1 \Rightarrow z_{e+1} = 1$, that is once chain-ladder, always chain-ladder, if the run-off result is the governing criterion. We also see that $z_e = 0 \Rightarrow z_{e+1} < 0$, so that the optimal z_{e+1} is actually negative. This seems a little strange.

11.6 When is a credibility estimator worse than the apriori mean?

The mean squared error of the estimator $\bar{b}_j = z_j\left(\frac{X_{j,\leq J-j}}{p_j\pi_{\leq J-j}}\right) + (1-z_j)\beta$ is, as we know

$$Q\left(z_j\right) = \mathrm{E}\left(\bar{b}_j - b\left(\Theta_j\right)\right)^2 = z_j^2\frac{\varphi}{p_j\pi_{\leq J-j}} + \left(1-z_j\right)^2\lambda \tag{11.29}$$

We want to know when $Q\left(z_j\right) > \lambda$, i.e. the credibility estimator degrades precision.

$$
\begin{array}{llll}
z_j^2 \frac{\varphi}{p_j \pi_{\leq J-j}} + (1 - z_j)^2 \lambda & > & \lambda & \Leftrightarrow \\
z_j^2 \frac{\varphi}{p_j \pi_{\leq J-j}} + \left(z_j^2 - 2z_j\right) \lambda & > & 0 & \Leftrightarrow \\
z_j^2 \left(\frac{\varphi}{p_j \pi_{\leq J-j}} + \lambda \right) & > & 2 z_j \lambda & \Leftrightarrow \\
z_j \left(\frac{\varphi}{p_j \pi_{\leq J-j}} + \lambda \right) & > & 2\lambda & \Leftrightarrow \\
z_j & > & \frac{2\lambda}{\frac{\varphi}{p_j \pi_{\leq J-j}} + \lambda} & \Leftrightarrow \\
z_j & > & 2 \frac{\lambda p_j \pi_{\leq J-j}}{\lambda p_j \pi_{\leq J-j} + \varphi} & \Leftrightarrow \\
z_j & > & 2\zeta_j &
\end{array}
\qquad (11.30)
$$

Thus credibility estimation degrades precision if (and only if) the credibility factor used is more than twice the optimal credibility factor in the Bühlmann-Straub model. In particular this means that whenever $\varsigma_j < 0.5$, the chain ladder method degrades precision.

12
Inflation and discounting

12.1 Introduction

12.2 Inflation and discounting

In the three-dimensional model of chapters 3-5 it is easy to write down expressions for the inflated and possibly discounted value of future payments. Denote the rate of inflation by ϵ and the discount rate by δ. The inflated, discounted value of the estimated cost of claims IBNR is

$$\text{IBNR}_J^{(\text{ID})} = \sum_{j=1}^{J} \sum_{d=J-j+1}^{\infty} \sum_{t=0}^{\infty} \left(p_j \bar{\Theta}_j \pi_d \xi_d \upsilon_t\right) \left(\frac{1+\epsilon}{1+\delta}\right)^{(j+d+t)-J-0.5} \tag{12.1}$$

and the inflated, discounted value of the estimated future payments on claims RBNS is

$$\text{RBNS}_J^{(\text{ID})} = \sum_{j=1}^{J} \sum_{d=0}^{J-j} \sum_{t=J-(j+d)+1}^{\infty} \left(\bar{Y}_{jd} - U_{jd,\leq J-(j+d)}\right) \left(\frac{\upsilon_t}{\upsilon_{>J-(j+d)}}\right) \left(\frac{1+\epsilon}{1+\delta}\right)^{(j+d+t)-J-0.5} \tag{12.2}$$

By subtracting 0.5 in the exponent we have made allowance for the assumption that claim payments will be spread evenly over the payment period.

These equations can easily be extended to variable (non-stochastic) rates of inflation or interest.

12.3 Estimating inflation by the separation method

One can use the separation method (Taylor, 2000) to estimate past inflation from the data.

Assume that the expected amount of incremental claim payments in respect of accident period j in development period kan be written in the multiplicative form

$$E(U_{je}) = p_j \pi_e \lambda_{j+e} \tag{12.3}$$

In this equation, p_j denotes some measure of claim exposure, π_e denotes the proportion of claims paid in development period e in uninflated terms, and λ_{j+e} denotes the price index that applies in calendar period $j+e$.

Now introduce the calendar period $k = j + e$ and re-cast (12.3) in the following form:

$$E\left(\frac{U_{ke}}{p_{k-e}}\right) = \pi_e \lambda_k \quad k = 1, \cdots, J, e = 0, \cdots, k-1 \tag{12.4}$$

The transformation is easy to visualise if one imagines that the traditional development triangle is rotated so that its diagonals appear as horizontals.

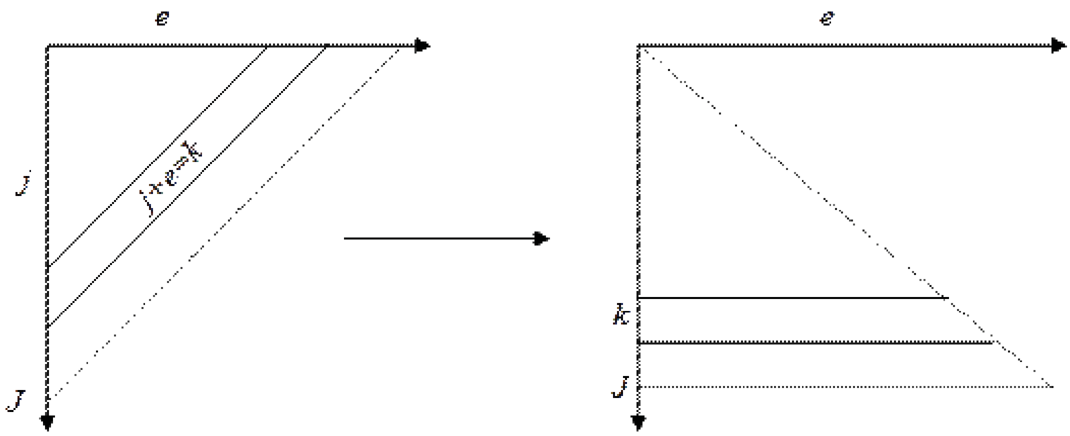
Write $X_{ke} = U_{ke}/p_{k-e}$. A heuristic estimator of the π_e and λ_k is given by the equations

$$\lambda_k^* \pi_{\leq k-1}^* = X_{k,\leq k-1} \text{ for } k = 1, \cdots, J \tag{12.5}$$

and

$$\lambda_{>e}^* \pi_e^* = X_{>e,e} \text{ for } e = 0, \cdots, J-1 \tag{12.6}$$

These equations are analogeous to the chain ladder equations (4.13)-(4.14) without the volume measures. The solution is of the same form, too. It can be calculated in the following way:



1. Set $\pi^*_{\leq J-1} = 1$ or some arbitrary $1 - \epsilon$.
2. Calculate the empirical development factors $\frac{\pi^*_{\leq e}}{\pi^*_{\leq e-1}} = \frac{\sum_{k=e+1}^{J} X_{k,\leq e}}{\sum_{k=e+1}^{J} X_{k,\leq e-1}}$ and thence $\pi^*_{\leq e}$ for $e = 0, \cdots, J-1$.
3. Calculate $\lambda^*_k = X_{k,\leq k-1}/\pi^*_{\leq k-1}$ for $k = 1, \cdots, J$.

If one sets $\pi^*_{\leq J-1} = 1-\epsilon$ instead of 1, all $\pi^*_{\leq e}$ change proportionally, and so do the price indices λ^*_k. However, the implied inflation rates $\lambda^*_k/\lambda^*_{k-1}$ remain unaffected. Thus in order to use the separation method to estimate inflation rates, one does not need to have fully developed accident periods.

13
Reinsurance recoveries

Remark 10 *This chapter must be extended, although I am not planning to produce a long text on this subject. Most reinsurance treaties do not lend themselves easily to an analytic approach, although stochastic simulation is always an option.*

Most direct insurance companies have a reinsurance program combined of proportional and non-proportional reinsurance. A simple reinsurance program with a retained quota share q and a non-proportional priority M would transform the gross ultimate claim amounts Y_{jdk} into net ultimate claim amounts:

$$Y_{jdk}^{(net)} = \min\left(M, q \cdot Y_{jdk}\right) \tag{13.1}$$

This can be used to calculate the severity distribution net of reinsurance and its moments:

$$\mu_d^{(net)} = q \int_0^{M/q} y G_d(dy) + M \cdot \left(1 - G_d(M/q)\right) \tag{13.2}$$

$$\rho_d^{(net)} = q^2 \int_0^{M/q} y^2 G_d(dy) + M^2 \cdot (1 - G_d(M/q)) \tag{13.3}$$

With a parametric distribution like the pareto and lognormal, this calculation is simple. Thus, in principle, one can carry out the same calculations for the net business as for the gross business. Unfortunately, most companies' reinsurance programs are not so simple.

A pragmatic approach is to

1. Calculate proportional reinsurance recoveries on RBNS claims, using the actuarial estimate of their outstanding cost.

2. Calculate any non-proportional reinsurance recoveries on RBNS claims, using case estimates of those individual claims that exceed the priority.

3. Calculate only proportional reinsurance recoveries on IBNR claims. The rationale is that any claim large enough to be covered by non-proportional reinsurance is likely to have been reported already.

Unless the non-proportional cover comes in at very low and high-frequent claim amounts or sets an overall limit to the retention of the company (like a stop loss contract), the actuary should normally not make allowance for recoveries from non-proportional contracts that cannot be linked to reported claims.

14
Data requirements

14.1 Introduction

Next to capital and staff, data is an insurer's most precious asset. Alas, very often the modelling efforts of the actuary are hindered by data that is inappropriately extracted or aggregated for his or her purpose. This section gives some general guidance about the data that the actuary should ask for. More specific requirements may of course be added.

Until now we have spoken about accident years, reporting years and so on. I will illustrate briefly why one should try to model at smaller time intervals wherever possible.

Imagine you are the actuary who has proudly presented his valuation results for the annual report. Six months on at the latest, your boss will say: “We've had so many new claims since last December, and besides, quite a substantial number of late reported claims from last year – and our case estimates have increased! Now what do you think?”. What you should think depends, of course, on whether the numbers quoted by your boss are in line with your expectations or not. Your problem is that a model based on yearly development patterns does not tell you what to expect during the year. This would not be a big issue if one could assume that every development during the year was spread evenly – but that is not the case either. Late reported claims incurred the year before tend to cluster at the start of the year. Claims incurred this year tend to reported later in the year, because they need to be incurred first. And so on.

Therefore, it is normally sensible to build a model with smaller time intervals – quarters, or even months. The initial investment in doing so is more than compensated by the facility with which you can calculate updated estimates at frequent time intervals and with a consistent set of assumptions. The data you collect must of course be able to support your modelling efforts.

Ideally, a data set should have individual claim records containing the following variables:

- Valuation date
- Accident date
- Reporting date
- Claim number
- Claim type
- Claim status
- Payments
- Outstanding case estimate

No claim must be omitted, whether it be old or settled or a "zero-claim". The last statement needs some qualification: claims which are immediately and irrevocably rejected can be omitted from the data, while claims that just turned out to be zero-claims after an assessment, should always be included in the data.

One should always ask for cumulative payment amounts, because the consequence of a valuation date missing is much less if one has cumulative data, than if one only had incremental data.

Many computer systems find it difficult to reconstruct outstanding case estimates on earlier valuation dates, therefore it is a good idea to retain copies of datasets that one has received, for future use.

Sometimes the volume of data makes it impractical to collect information on every individual claim, or the insurance company may be reluctant to hand over so detailed information. In that case, the variable "Claim number" should be omitted and replaced by "Number of claims", showing how many claims are included in each group. The variables "Payments" and "Outstanding case estimates" should of course show the aggregate for the group. In this context, "Group" means a combination of (valuation date, accident date, reporting date, claim type and claim status).

All aggregation makes it more difficult to estimate a severity distribution. If you only receive aggregate data, you may want to supplement it with information that allows you to assess the variability of the claim amounts. Such information could be:

- The largest claim in every group.
- The sum of squares of reported claim cost in every group. This will allow you to calculate a sample variance of reported claim cost.
- The sum and sum of squares of ln(total case estimate) in every group. This will allow you to fit a lognormal distribution to reported claim cost.

Or you could ask for a sample of individual claims to support your severity modelling.

Claims should be grouped into reasonably homogeneous groups prior to being analysed. Depending on the type of claims to be valued and the volume of data that is available, the actuary can decide to split claims into more subgroups.

In splitting claims by claim attributes, it is important that the attributes used are stable over the life of the claim. Attributes that may change over the life of a claim, will distort statistics by spurious changes and should be used with great caution. A particularly unpleasant side effect using of unstable claim attributes is that the number of claims reported in the past for a specific group may change in successive valuation dates.

One example of an attribute that is sometimes used to split claims, but which is not stable, is grouping into small claims and large claims. Regardless of the limit chosen for small claims, there will always be some migration between the two groups, causing apparent changes in the statistics.

Another unstable attribute is the distinction of claims into “zero claims” and “genuine claims”. If claimhandlers were clairvoyants who could predict for any claim whether or not it would generate a payment, then “genuine claims” would be a stable group. In reality, there is always a great deal of migration – zero claims turning out to be genuine claims and vice versa.

15

Appendix. Credibility estimation in the regression case

Remark 11 *This chapter was written for students with a knowledge of credibility theory, who only need an alignment of the notation. I am planning to extend it with some essential proofs, so that it can serve as a very quick introduction to the theory.*

The appendix gives a very brief overview of credibility theory in the regression case. No proofs will be given as this is not a course on credibility theory. However, most results can be verified using standard matrix algebra for covariance matrices.

Let us assume that we have a random vector of observations $\mathbf{X}^{n\times 1}$ (for example, past claims). Let us further assume the existence of an unobserved random element Θ (for example, underlying risk conditions). Now assume that in the conditional distribution given Θ the vector $\mathbf{X}$ is linearly regressed on a vector-valued function of Θ:

$$\mathrm{E}(\mathbf{X} \mid \Theta) = \mathbf{Y}^{n\times m}\mathbf{b}^{m\times 1}(\Theta) \tag{15.1}$$

with a known regression matrix $\mathbf{Y}$ with full column rank $m \leq n$.

Denote the first and second order moments of $\mathbf{b}(\Theta)$ by

$$\boldsymbol{\beta}^{m\times 1} = \mathrm{E}(\mathbf{b}(\Theta)) \tag{15.2}$$

and

$$\mathbf{\Lambda}^{m\times m} = \mathrm{Cov}(\mathbf{b}(\Theta)) \tag{15.3}$$

The expected conditional covariance matrix of $\mathbf{X}$ given Θ we denote by

$$\mathbf{\Phi}^{n\times n} = \mathrm{ECov}(\mathbf{X} \mid \Theta) \tag{15.4}$$

It is easy to verify that the unconditional covariance matrix of $\mathbf{X}$ is

$$\mathbf{\Sigma}^{n\times n} = \mathbf{\Phi} + \mathbf{Y\Lambda Y}' \tag{15.5}$$

We assume that $\mathbf{\Phi}$ and thereby $\mathbf{\Sigma}$ are of full rank.

We now set out to estimate the unknown regressor $\mathbf{b}(\Theta)$ by a linear function of the form

$$\mathbf{b}^* = \mathbf{g}^{m\times 1} + \mathbf{G}^{m\times n}\mathbf{X} \tag{15.6}$$

with fixed coefficient matrices $\mathbf{g}$ and $\mathbf{G}$. The mean squared error matrix of the estimator $\mathbf{b}^*$ is

$$\begin{array}{rcl} \mathbf{Q}^{m\times m}(\mathbf{g},\mathbf{G}) & = & \mathrm{E}(\mathbf{b}-\mathbf{b}^*)(\mathbf{b}-\mathbf{b}^*)' \\ & = & \mathbf{G\Phi G}' + (\mathbf{I}-\mathbf{GY})\mathbf{\Lambda}(\mathbf{I}-\mathbf{GY})' \\ & + & ((\mathbf{I}-\mathbf{GY})\boldsymbol{\beta}-\mathbf{g})\,((\mathbf{I}-\mathbf{GY})\boldsymbol{\beta}-\mathbf{g}) \end{array} \tag{15.7}$$

For any linear combination $\mathbf{a}'\mathbf{b}(\Theta)$ with a vector of fixed coefficients $\mathbf{a}^{m\times 1}$, the mean squared error of the estimator $\mathbf{a}'\mathbf{b}^*$ is trivially $\mathbf{a}'\mathbf{Q}\mathbf{a}$. The mean squared error is uniformly minimal for all choices of $\mathbf{a}^{m\times 1}$ if one chooses the following cofficients:

$$\mathbf{\Gamma} = \mathrm{Cov}(\mathbf{b}(\Theta),\mathbf{X}')\,\mathrm{Cov}^{-1}(\mathbf{X}) = \mathbf{\Lambda Y}'\mathbf{\Sigma}^{-1} \tag{15.8}$$

and

$$\boldsymbol{\gamma} = (\mathbf{I}-\mathbf{\Gamma Y})\boldsymbol{\beta} \tag{15.9}$$

The linear greatest accuracy credibility estimator is then

$$\bar{\mathbf{b}} = \boldsymbol{\gamma}^{m\times 1} + \boldsymbol{\Gamma}^{m\times n}\mathbf{X} \tag{15.10}$$

It can also be written in the form

$$\bar{\mathbf{b}} = \mathbf{Z}\hat{\mathbf{b}} + (\mathbf{I}-\mathbf{Z})\boldsymbol{\beta} \tag{15.11}$$

with

$$\hat{\mathbf{b}} = \left(\mathbf{Y}'\boldsymbol{\Phi}^{-1}\mathbf{Y}\right)^{-1}\mathbf{Y}'\boldsymbol{\Phi}^{-1}\mathbf{X} \tag{15.12}$$

and a credibility matrix

$$\begin{aligned}\mathbf{Z} &= \boldsymbol{\Gamma}\mathbf{Y}\\ &= \boldsymbol{\Lambda}\mathbf{Y}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}\\ &= \boldsymbol{\Lambda}\mathbf{Y}'\boldsymbol{\Phi}^{-1}\mathbf{Y}\left(\mathbf{I}+\boldsymbol{\Lambda}\mathbf{Y}'\boldsymbol{\Phi}^{-1}\mathbf{Y}\right)^{-1}\end{aligned} \tag{15.13}$$

Its mean squared error may be written as

$$\begin{aligned}\mathbf{Q}(\mathbf{Z}) &= \mathrm{E}(\mathbf{b}-\bar{\mathbf{b}})(\mathbf{b}-\bar{\mathbf{b}})'\\ &\quad\mathbf{Z}\left(\mathbf{Y}'\boldsymbol{\Phi}^{-1}\mathbf{Y}\right)^{-1}\mathbf{Z}' + (\mathbf{I}-\mathbf{Z})\boldsymbol{\Lambda}(\mathbf{I}-\mathbf{Z})'\end{aligned} \tag{15.14}$$

This equation holds for an arbitrary choice of $\mathbf{Z}$ (not just the one given in (15.13)).

A matrix identity that is often useful in calculating (15.13), is the following:

$$(\mathbf{I}+\mathbf{AB})^{-1} = \mathbf{I} - \mathbf{A}\left(\mathbf{I}+\mathbf{BA}\right)^{-1}\mathbf{B} \tag{15.15}$$

This holds whenever all the displayed inverses exist. Let $\mathbf{X}_2^{n_2\times 1}$be another random vector that is linearly regressed on $\mathbf{b}(\Theta)$, i.e.

$$\mathrm{E}(\mathbf{X}_2 \mid \Theta) = \mathbf{Y}_2^{n_2\times m}\mathbf{b}(\Theta) \tag{15.16}$$

The vector $\mathbf{X}_2$ could for example denote future claims that depend on the same underlying risk conditions as those already observed ($\mathbf{X}$). Assume that

$$\mathbf{\Phi}_2^{n_2 \times n_2} = \mathrm{ECov}(\mathbf{X}_2 \mid \Theta) \tag{15.17}$$

Finally assume that $\mathbf{X}_2$ and $\mathbf{X}$ are conditionally independent, given Θ. Then the linear predictor $\bar{\mathbf{X}}_2 = \mathbf{Y}_2\bar{\mathbf{b}}$ has a mean squared error matrix of

$$\mathrm{E}(\mathbf{X}_2 - \bar{\mathbf{X}}_2)(\mathbf{X}_2 - \bar{\mathbf{X}}_2)' = \mathbf{\Phi}_2 + \mathbf{Y}_2\mathbf{Q}(\mathbf{Z})\mathbf{Y}_2'. \tag{15.18}$$

Empirical credibility theory is the discipline where the parameters $(\beta, \mathbf{\Phi}, \mathbf{\Lambda})$ are replaced by estimates $(\beta^*, \mathbf{\Phi}^*, \mathbf{\Lambda}^*)$ prior to being used in (15.11)-(15.14).

16
Literature

- Arjas, E. (1989). The Claims Reserving Problem in Non-Life Insurance: Some Structural Ideas. ASTIN Bulletin Volume 19, No. 2 139-152.
- Ferguson, T.S. (1967). Mathematical Statistics: A Decision Theoretic Approach. Academic Press (New York).
- General Insurance Reserving Issues Taskforce (2006): A Change Agenda for Reserving. Presented to the Institute of Actuaries, 27 March 2006.
- Haastrup, S. and Arjas, E. (1996). Claims Reserving in Continuous Time; A Nonparametric Bayesian Approach. ASTIN Bulletin Volume 26, No. 2 139-164.
- Hertig, J. (1985). A Statistical Approach to IBNR-Reserves in Marine Reinsurance. ASTIN Bulletin Volume 15, No. 2 171-184.
- Hesselager, O. and Witting, T. (1988). A Credibility Model with Random Fluctuations in Delay Probabilities for the Prediction of IBNR Claims. ASTIN Bulletin Volume 18, No. 1 79-90.
- Hesselager, O. (1994). A Markov Model for Loss Reserving. ASTIN Bulletin Volume 24, No. 2 183-193.
- Hesselager, O. (1995). Modelling of Discretized Claim Numbers in Loss Reserving. ASTIN Bulletin Volume 25, No. 2 119-135.

- Jewell, W. S. (1989). Predicting IBNYR Events and Delays I. Continuous Time. ASTIN Bulletin Volume 19, No. 1 25-56.
- Jewell, W. S. (1990). Predicting IBNYR Events and Delays II. Discrete Time. ASTIN Bulletin Volume 20, No. 1 93-111.
- Mack, T. (1993). Distribution-free Calculation of the Standard Error of Chain Ladder Estimates. ASTIN Bulletin Volume 23, No. 2 213-225.
- Mack, T. (2000). Credible Claims Reserve: The Benktander Method. ASTIN Bulletin Volume 30, No. 2 333-347.
- Neuhaus, W. (1992). IBNR Models with Random Delay Distributions. Scandinavian Actuarial Journal 1992, No. 2.
- Neuhaus, W. (1992). Another Pragmatic Loss Reserving Method or Bornhuetter-Ferguson Revisited. Scandinavian Actuarial Journal 1992, No. 2.
- Norberg, R. (1986). A Contribution to Modelling of IBNR Claims. Scandinavian Actuarial Journal 1986, No. 3-4.
- Norberg, R. (1993). Prediction of Outstanding Liabilities in Non-Life Insurance. ASTIN Bulletin Volume 23, No. 1 95-115.
- Norberg, R. (1999). Prediction of Outstanding Liabilities. II. Model Variations and Extensions. ASTIN Bulletin Volume 29, No. 1 5-25.
- Pinheiro, P. J. R., Andrade e Silva, J. M.; Centeno, M. L. (2000). Bootstrap Methodology in Claim Reserving. Working Paper no. 11/00 from the Centre for Applied Mathematics to Forecasting and Economic Decision). URL http://pascal.iseg.utl.pt/∼cemapre/wpapers/0011.pdf.
- Schmidt, K.D. (2011). A Bibliography on Loss Reserving. URL http://www.math.tu-dresden.de/sto/schmidt/dsvm/reserve.pdf
- Taylor, G.C. (2000). Loss Reserving. An Actuarial Perspective. Kluwer Academic Publishers, Boston / Dordrecht / London.
- Wald, A. (1950). Statistical Decision Functions. John Wiley and Sons, New York; Chapman and Hall, London.
- Wüthrich, M.V. and Merz, M. (2008). Stochastic Claims Reserving Methods in Insurance. John Wiley & Sons, Ltd.