Experience rating of (re)insurance premiums under uncertainty about past inflation

Von der Fakultät für Mathematik und Naturwissenschaften der Carl von Ossietzky Universität Oldenburg zur Erlangung des Grades und Titels eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

angenommene Dissertation

von Herrn Michael Fackler,
geboren am 23.09.1967 in Memmingen

Gutachter:	Prof. Dr. Dietmar Pfeifer
Zweitgutachter:	Prof. Dr. Peter Ruckdeschel

Tag der Disputation:	06.07.2017

Lebenslauf

Persönliche Daten:

Name:	Michael Fackler
Geburtsdatum:	23.09.1967
Geburtsort:	Memmingen
Staatsangehörigkeit:	Deutschland

Bildungsweg:

1973 - 1986	Grundschule und Gymnasium in **Memmingen**, Abschluss Abitur
1986 - 1992	Diplomstudiengang Mathematik mit Nebenfach Theoretische Physik an der Universität **München** (1989/90 an der Universität **Pisa**), Abschluss Diplom
2012 - 2017	Promotion (extern, nebenberuflich) an der Carl von Ossietzky Universität **Oldenburg**

Erklärung

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst und nur die angegebenen Hilfsmittel benutzt habe.

München, den 15.05.2017,
Michael Fackler

Schadenerfahrungs-Tarifierung von (Rück)Versicherungsverträgen bei Unsicherheit über die vergangene Inflation

Zusammenfassung

Bei der Kalkulation von (Rück)Versicherungsprämien werden ältere Schadendaten häufig vor der statistischen Auswertung mittels eines Inflationsindex adjustiert, der exogen (nicht aus den Daten selbst) berechnet wurde. Dabei nimmt man implizit an, dass der verwendete Index die Inflation der versicherten Schäden exakt abbildet. In der Praxis dürfte es aber zumeist eine kleine Abweichung zwischen der „wahren" Inflation und dem Index geben.

In dieser Arbeit wird ein stochastischer Ansatz für dieses „Basisrisiko" vorgeschlagen und gezeigt, wie sich damit die statistischen Eigenschaften der Prämienkalkulation grundlegend, aber im Einklang mit der Intuition verändern. Insbesondere werden dadurch ältere Daten sukzessive weniger aussagekräftig als neuere.
Weiter wird ein umfassendes Modell für die Unsicherheit vergangener und zukünftiger Inflation entwickelt, dessen Kern eine Systematik aller die Tarifierung betreffenden Inflationsarten ist. Schließlich wird gezeigt, wie man die Prognosegenauigkeit der Prämienkalkulation durch geeignete Gewichtung der Daten nach Alter verbessern kann.

Hierfür wird ein Schadenzahlmodell für Risiken/Portefeuilles variabler Größe vorgeschlagen, das die Volumenabhängigkeit der Varianz sehr flexibel beschreibt und die Lücke zwischen zwei klassischen Modellen füllt: Unabhängigkeit der Einzelobjekte bzw. starke Abhängigkeit durch eine marktweite Frequenzschwankung.
Die bei nichtproportionalem Risikotransfer (Selbstbehalte, Deckungslimits, Layer) nichtlineare, gehebelte Wirkung der Inflation auf die Risikoprämie wird analytisch beschrieben. Es wird gezeigt, wie der Hebel von der Geometrie der Verteilungsfunktion der Schadenhöhe abhängt und durch einen neuen Verteilungsparameter quantifiziert werden kann: das regionale Pareto-alpha.

Diese mathematische Arbeit liefert auch qualitative Erkenntnisse für die Praxis: Sie zeigt auf, welche Inflationskomponenten am sensitivsten für die Tarifierung sind.

Experience rating of (re)insurance premiums under uncertainty about past inflation

Abstract

In (re)insurance premium rating, older loss data are often adjusted before the statistical evaluation by means of an “external” inflation index (calculated not from the data itself). By doing so, it is implicitly assumed that this index exactly reflects the inflation of the insured losses. However, in practice there should mostly be a slight deviation between the “true” loss inflation and the index.

In this thesis, we propose a stochastic approach for this “basis risk” and show how it changes the statistical properties of premium rating: fundamentally, but in accordance with intuition. In particular, older data becomes gradually less representative than recent data.
Then, a comprehensive model for the uncertainty of past and future inflation is developed, having at its heart a system of all types of inflation that matter for the rating. Finally, we show how to improve the forecast accuracy of premium rating by appropriately weighting the data by age.

For this purpose, we propose a loss count model for risks/portfolios of variable size that describes the dependence of variance on volume very flexibly, filling the gap between two classical models: independence of the single insured objects versus strong dependency via a market-wide loss frequency fluctuation.
To embrace non-proportional risk transfer (deductibles, limits, layers), we develop an analytical theory for the generally non-linear, leveraged impact of inflation on the risk premium. We show how the leverage depends on the geometry of the loss severity cdf, quantifying it by a novel distribution parameter: the regional Pareto alpha.

This mathematical thesis yields also qualitative insight for practitioners, revealing which inflation components are most sensitive for premium rating.

Contents

Foreword

The research on the topic presented here started more than 15 years ago from a seemingly simple question, emerging from my practical work as an actuary in charge of the pricing of non-proportional reinsurance treaties, a field where loss statistics are always scarce:

If we are in the (fortunate and rare) situation that we have a complete loss record of the past 40 years available, why don't we dare to use all that data?

At that time it was not clear that the quest for a mathematical, not just judgmental, answer to this question would lead from thoughts about the reliability of old data over inflation models to a comprehensive model for shifts affecting claim size and/or frequency over time, being apt to give finally a mathematical reason for the hesitation described above, and to address various other interesting issues.

It was not clear then, either, that I would carry on for so many years with this quest, while continuing to work as an actuary, nor that slow but steady progress would yield enough material to write a book about it.

I thank David Clark, Prof. Franz Merkl, Glenn Meyers, Prof. Peter Ruckdeschel, Prof. Peter Schuster, Dr. Björn Sundt, Prof. David Wilkie, and, in particular, Dr. Ulrich Riegel for having inspired this work with their work and very helpful suggestions.

I thank my parents for having always supported me, even when I did unusual things, or normal things in an unusual order.

Finally, I thank Prof. Dietmar Pfeifer for having initiated and supervised the latest of these unusual things: a doctorate in Mathematics after having spent 20 years in practitioners' world, far away from university.

Chapter 1

Introduction

1.1 Motivation

Although inflation at present (2017) is rather low in most developed countries, the insurance industry is increasingly concerned about it, in particular in the Non-Life lines of business. Periods of high inflation might occur in the future, as they have in the past. What is more, today's solvency rules and more generally a modern risk management attitude require much attention for systemic risk, and even without sophisticated mathematical modeling it is clear that insurance claims rising sharply due to inflation may adversely impact various insurance products at the same time. When affecting various lines of business and/or countries, even low inflation could create considerable correlation across seemingly well-diversified insurance portfolios.

The aim of this book is to contribute to the thinking about inflation, with special focus on the experience rating of premiums of insurance and reinsurance covers.

In particular, we want to draw the attention to past inflation. If we estimate the expected loss of an insurance cover from past loss experience, we have to project past events into the future world. It is clear that future inflation is unknown (thus a random variable), but do we really know past inflation? In fortunate cases the inflation that occurred across the years being reported in the loss record can be estimated from the loss data, however, this requires a large amount of consistent data covering many years, which is rare. It is much more common to use an inflation index that is not derived from the loss data itself. Fortunately, for many situations there are appropriate inflation indices available, being carefully developed from public or industry data sources. But can we really be sure that such an "external" index exactly describes the past inflation of the business being rated? Whatever large the data base of the external index, whatever sophisticated the statistical procedures applied to derive it, it might not exactly reflect "our" business. There could be a gap between the "true" inflation and the index, and we propose to model this gap as a random variable (*past inflation gap*). We will see that this adds a lot of complexity, but yields very intuitive results. All in all we shall come up with a loss model combining the randomness of both future and past inflation with the randomness of the losses stemming from sources other than inflation.

Further we provide a thorough study of the impact of inflation on insurance products involving non-proportional risk transfer, i.e. (re)insurance covers having either a retention or a limit being lower than the maximum possible loss, or both ("layers" of insurance). It is well known that the impact of inflation on such products is generally non-proportionate, however, the exact relationship seems technical. We develop elegant and intuitive formulae relating the ground-up inflation to that of the insurance cover, by looking at the geometry of the severity distribution function in a particular way.

The combination of non-proportional coverages and past inflation affects most sectors of Non-Life insurance premium rating, creating a huge challenge. Only personal lines insurance, the maybe most visible part of the industry, does not have to struggle with this. Here deductibles are very low or absent, while past inflation is secondary as one typically has tons of data available without any need to look back farther than a few years, such that whatever one does about past inflation, it cannot be a source of large errors/uncertainties. Instead, in commercial insurance (which in this book shall always embrace industrial insurance) and even more in reinsurance the data are such scarce that in order to have a sufficient data base for a sound experience rating one needs to assemble all losses occurred in the past 5 or 10, sometimes 30 years. What is more, coverages here usually have considerable retentions, ranging from say 10000 EUR to several millions in the case of industrial insurance layers or non-proportional reinsurance treaties. We will see that in this situation the inflation to the insurance cover can be very different from the ground-up inflation, and much more difficult to detect.

The aim of this work is to give insight into such rating situations relying on scarce data stretching over several years. As the introduction of the random past inflation gap, albeit a rather simple ingredient, leads to very complex stochastic properties, we have to restrict the analysis to the so-called short-tail lines of business having a rather quick and stable run-off like Property, Marine, Personal Accident, etc. However, the results should also give some orientation for the long-tail lines (which are essentially the various Liability lines).

The path towards the answer of the question raised in the foreword (how many years of data to use) will lead us to a more interesting question, which we will be able to tackle as well: Taking account of the stochastic properties of the proposed model, can we improve experience rating by altering the classical sample mean? The answer is yes: by assigning different weights to data of different age. Formulae for these weights vary a lot according to the rating situation: from simple and close to traditional approaches to surprisingly complex (think of Fibonacci numbers). On the way to all this we will get a lot of insight into what drives the volatility of insurance portfolios, so some results will be of value in themselves, independently of the rating problem studied here.

1.2 Scientific context

There is not too much literature paving the way. Actuarial research explores very advanced models and quite some papers address inflation in Non-Life insurance, see e.g. [Bohnert et al., 2016] and references therein. However, when it comes to applications, it is typically assumed that the data are correctly adjusted for inflation – a great oversimplification. If uncertainty is accounted for at all, this usually applies only to future inflation. Practitioners are well aware that we might not always exactly know past inflation, see [Brickmann et al., 2005], however, this concern apparently has not (yet) found its way into a lot of formal modeling. Nevertheless, this book is inspired from some very valuable sources, although most of them focus on somewhat different issues:

The basic structure of the indices used stems from the famous Wilkie model for retail prices, wages, etc., see [Wilkie, 1995], who uses discrete indices (typically one value per year) being (at least in the earlier generations of the model) basically chain products of lognormal random variables (RV's), albeit with possibly complex dependencies across subsequent years and indices. [Clark, 2006] applies such a model, an AR(1) process, to quantify dependencies in the future run-off of Liability losses. The emerging dependence structure reappears in our work, however, we have it across *past* years.

The stochastic past inflation gap, as we define it, works as if we had originally had the data correctly adjusted for inflation, but then got it somewhat distorted by the inflation gap. The same approach can be found in geophysics, where experts are concerned that the observed magnitudes of earthquakes that occurred decades ago might be somewhat inexact due to the less precise measuring devices available at that time. To account for this uncertainty a (normally distributed) random variable slightly distorting the "true" magnitude is introduced, see [Rhoades, 1996].

[Sundt, 1987] claimed as early as 1987 that, albeit being common practice, it is unnatural to regard the data of the past n years as wholly representative, while earlier years are not seen as representative at all – an ideal model should rather produce smoothly increasing weights than a jump from 0 to full weight at a certain point. His paper thereafter focuses on other issues, however, Sundt could have been first to clearly point this issue out.

[Gerber and Jones, 1975] introduced a recursive Credibility approach (heterogenous Credibility, i.e. with a grand mean) that can lead to geometric weights for the data: Say each year gets 90% of the weight of the subsequent year. Such weights seem indeed popular in finance and economy, bearing the name *exponentially weighted moving average*. However, the weights are apparently often set judgmentally, without there being a strict model assumption underneath. We will see that in our model geometric weights can be close to, but also far from, being optimal.

The research being probably closest to ours is that of Howard Mahler, who in the 1990ies published a series of papers on Credibility rating of Workers' Compensation insurance with "shifting parameters over time", see [Mahler, 1998] and references therein to the author's earlier papers. His approach is very general, combining the uncertainty stemming from non-constant risk parameters (this includes inflationary effects) with that of the long run-off inherent in this line of business. The model for the change of the risk parameters is different from ours, involving a Markov chain. So Mahler's results are not directly comparable, however, it shall be noted that the main goal of his research is the same: setting optimal weights to data from different years, according to a least square criterion (he uses homogeneous Credibility).

Finally, a recent paper by [Riegel, 2015] treats the rating of Motor quota shares (proportional reinsurance). Here the past inflation of the losses is assumed known, but the past inflation of the original premium income (which is a multiplier in the calculation of the expected loss) has a stochastic gap – we will include such a gap later in our analysis, too. Optimal weights for the years of the data set are calculated, according to a mean squared error criterion. Riegel's model is complex in that it is partly parametric, partly not, in order to get as much output as possible using as few assumptions as possible. We shall adopt an analogous semi-parametric setting, albeit deviating in many details.

1.3 Organization of this book

Chapter 2 gives a brief illustration of how experience rating of commercial insurance policies and reinsurance treaties is done in practice, involving in principle very simple statistics, namely the sample mean, albeit with considerable sophistication in the details.

Chapter 3 introduces our stochastic model for the uncertainty inherent in past inflation.

Chapter 4 explores a simplified rating example that provides a great deal of intuition for more complex real-world situations. By looking at the mean squared error of the straightforward estimate for the risk premium of a risk of constant size, we give a first answer to the initial question: which years of the data set should be used if many are available. Then the rating is further improved by assigning different weights to different years of the loss history. It turns out that in this simple example the optimal weights are related to the Fibonacci sequence.

Chapter 5 deals with a practical problem that adds a lot of complexity: the variable size of the insured risks. We propose a sophisticated model for the variance of the number of losses per year, embracing well-established models in one. This enables us to work with the collective model of risk theory, which is essential as non-proportional risk transfer lets frequency and severity effects intermingle.

Chapter 6 calculates the leveraged effect of ground-up inflation on non-proportional (re)insurance covers, coming up with very general and intuitive formulae relating said leverage to the geometry of the severity distribution function. The decisive geometric measure in this context turns out to be a generalized local Pareto alpha (*regional Pareto alpha*).

Chapter 7 designs a comprehensive model for random shifts over time in both frequency and severity, in both the past and the future, which lets us embrace all results developed so far.

Chapter 8 decomposes the effects of inflation and variable risk size into a system of 9 index series, making transparent their potential interactions.

Chapter 9 applies this system to derive all quantities needed to assess the stochastic properties of experience rating.

Chapter 10 calculates the mean squared error of the sample mean, optimizing it by assigning different weights to the loss records according to their age.

Chapter 11 concludes with suggestions about how the theoretical results of this book could inspire practitioners' work.

Chapter 2

Experience Rating

2.1 Insurance beyond personal lines

2.1.1 Cover structure

Commercial insurance and reinsurance are not entirely different from the more common personal lines insurance, however, there are some particular features.

Firstly, the insured risk is typically not a single object, but rather a number of things or people, think of all buildings belonging to a large firm (Fire insurance), a fleet of cars (Motor Third Party Liability, Motor Hull) or vessels (Marine Hull), or the employees of a company (Personal Accident group policy). Reinsurance treaties cover entire portfolios, say all Property Fire business held by an insurer, however, in principle this is the same kind of risk – a (very large) number of things. Obviously the size of such risks changes if the number of insured objects increases or decreases. To cater for this, the agreed premium is typically not flat, but a premium rate per *volume*, where the volume or *exposure* is a previously agreed measure for the size of the insured risk. To keep things simple for now, think of the number of insured objects/people. Exposure measures can be very sophisticated; we shall give some examples in the next section and discuss volumes in detail various times across the book.

Secondly, large risks produce far more losses than single objects. To avoid money swapping it is common to reduce the loss frequency via introduction of a *deductible* d. Of any loss Z, the insured bears the part up to d (*first risk*, mathematically $\min(Z, d)$), while the insurer only pays the exceeding part (*second risk*, mathematically $(Z-d)^+$). The deductible is typically chosen such high that the frequency of the losses exceeding d is rather low, possibly much less than one per year.

Further the *cover* or *liability* of the insurer is limited by a maximum per loss c (a feature common in personal lines insurance, too). Formally, the insurer pays $\min\Big((Z-d)^+, c\Big)$, thus the part of each loss exceeding $u := d + c$ is again to be beard by the insured. This kind of coverage is shortly labelled $c\,xs\,d$ (c *in excess of* d). In practice c is chosen such high that exhaustion of the cover is an extremely rare event.

A peculiar case is industrial insurance and reinsurance. Here it is not uncommon to split a cover in various adjacent so-called *layers* or *tranches* (term of French origin) $c_i\,xs\,d_i$, where $d_{i+1} = u_i := c_i + d_i$, which may be ceded to different (re)insurers specializing on coverages having high or low loss frequency. In lower layers total losses may occur more frequently, however, the top layer is chosen such high that it is exhausted very rarely.

As we will deal with layers throughout this book, we introduce the following compact notation:

Definition 2.1. If Z is a loss from the ground up, i.e. before application of the layer, the corresponding *loss to the layer* or *layer loss* is for $0 \leq d < u \leq \infty$

$$ {}_{d}^{u}Z := (Z-d)^{+} - (Z-u)^{+} = \min\Big((Z-d)^{+}, u-d\Big) $$

Terms used vary according to country and kind of insurance. Instead of deductible one also says *retention* or *attachment point*, in reinsurance sometimes *priority* (French origin). The maximum c is also called *limit*, however, this term occurs to be used for u as well, in particular in direct insurance when deductibles are small. Unequivocal words for u would be the French *plafond* (ceiling) and *detachment point*, which is used in finance. For a layer one also says *excess of loss* cover or shortly *XL*.

This simple structure of deductibles and limits is applied in a number of different ways. They may work per loss or per year (i.e. applying to the aggregate loss), however, the latter case is not treated in this book, as its mathematics is hard to treat analytically, at least in the case of risks with variable size. An important variant of the former case are catastrophe reinsurance treaties (*CatXL*), which apply the layer structure to accumulation losses (all losses stemming from e.g. a windstorm or earthquake event are aggregated). Thus, according to how the coverage is defined, Z models *single losses* or *accumulation losses*, respectively.

It shall be noted that reinsurance is not always non-proportional. Covers without deductible and limit do indeed exist, being called *proportional reinsurance* or *pro-rata reinsurance*. For example, if the portfolio of a Non-Life insurer is dominated by their Motor business, such that fluctuating Motor results largely affect the overall result, it could be an option to balance the portfolio by ceding say 60% of the aggregate Motor losses to a reinsurer (and paying them a reinsurance premium of about 60% of the aggregate pure Motor premiums). Such a treaty is called *quota share*, and when it effectively balances a portfolio, it is not considered money swapping.

We will treat pro-rata coverages and layers together wherever possible. Formally the former is a layer $\infty\, xs\, 0$, so we have the loss variable ${}_{0}^{\infty}Z = Z$. Another special case are the so-called *first-loss covers* with deductible 0 and limited liability. Normally the limit is chosen such high that its numerical impact is close to nil, however, there might be exceptions. Anyway, these variants are special layers and by providing the general mathematics for layers we embrace them as well.

Notice that we focus on the mathematical core structure of insurance covers here, neglecting certain occurring additional features that have much less impact on the risk premium than the basic layer structure, e.g. the common aggregate limit *per year* complementing the limit per loss. We thus treat slightly stylized layers here, however, their outcomes are numerically extremely close to real ones. More on the large variety of features used in reinsurance treaties and commercial insurance covers can be found in Section I.3 of [Parodi, 2014], a comprehensive book on Non-Life insurance pricing.

2.1.2 Data given

The data typically required for experience rating are the following, for a number of years, which may range from less than 10 to more than 30:

- Exposure measure (= volume): one aggregate figure per year; typically this is the *official exposure* the premium rate is based on.

- Losses:
  in case of proportional covers (and some coverages with very low deductibles and very high limits): the aggregate loss, i.e. one value per year;
  in case of typical layers: a list of all losses exceeding the *reporting threshold*, equaling e.g. 70% of

the retention. The number of data points can be anything from less than one loss per year (typical for CatXLs) up to dozens of losses per year for certain very low layers.

To learn more about exposures let us start with a commercial Property policy: Fire plus the typical allied perils like Water-pipe, Theft, and Natural Perils, which may be rated together with Fire or grouped into a few sections for the rating. The premium of a coverage for say several pubs of similar size could be based simply on the *number* of objects. If the pubs vary considerably in size, *square meters* could be a better option. The most common measure, however, is the *aggregate sum insured* (usually the replacement value, being periodically updated) of the pubs. Note that the latter is a fundamentally different kind of volume, as it includes the change of the cost level, i.e. inflation (or more precisely a certain inflation, namely that of the replacement values of pubs).

Both kinds of volume, those embracing inflation and those not, are common and can be useful, however, it is essential to not mix them up. It can be an advantage to have inflation included, notably if one wants to have all increases affecting the aggregate loss embraced in one figure. In fact insurers like to base their premium rates on such exposures, in order to get premium raises automatically whenever inflation occurs. However, for the exact modeling, as we will see, it is essential to keep frequency and severity effects separately and combine them just at the end. That means that if an available volume contains inflation, this latter has to be factored out to obtain an inflation-free volume.

In short, we have to distinguish the *official volume* (multiplier for the premium rate) from the volume used internally for the rating. This point shall be thoroughly explored in Chapter 8, for now we just illustrate the diversity of exposure measures by some further examples.

For a Personal Accident group policy protecting employees, the number of staff is a fair exposure in case the sum insured per employee is uniform and does not change over the years; if instead the employees have sums insured tied to their (yearly rising) salary, then aggregate *payroll* (*wage roll*) should be a better option. The latter is also common for Workers' Compensation and other Liability coverages protecting staff, although here the cost level can be affected by medical expenses (and their often very high inflation) as well. Other Liability policies use figures like *turnover*. Payroll and turnover include a certain inflation, however, it is questionable whether this exactly matches the inflation of the losses – such volume measures rather seem "sophisticated workarounds": simply the best available proxies. So one has to deal with two kinds of inflation, being possibly different, possibly correlated, and possibly not exactly known. This complex situation will be dealt with in the second half of this book.

Premiums for Motor fleets may be based on aggregate mileage. Another common option is *vehicle years*, which is a refinement of the car count assigning e.g. a weight of one third to a car that was part of the fleet for only 4 months. If it is deemed worth the effort, similar *pro-rata-temporis* refinements are applied in other business lines, too, say to the number of employees or objects. E.g., in group Health insurance one speaks of the *exposed population* or *population at risk*.

An interesting exposure in Marine Hull (covering ocean vessels) is the *tonnage* of the vessels, reflecting both variable number and size of vessels, but not embracing inflation.

In Aviation insurance many different exposures appear, according to what is seen as reflecting best the changes of the risk over time: number of flights, number of passengers, *passenger miles*, number of aircraft, *aggregate fleet value*.

Sometimes the insured individuals/objects are very heterogeneous within, so different weights may be introduced to distinguish exposures, e.g. young vs old drivers, motorcycles vs passenger cars vs trucks, or blue collar vs white collar workers. According to the situation, maybe a blue collar worker is counted as four white collar workers.

A particular case is reinsurance. The already stated vehicle years and tonnage are used in the respective business lines in some countries, however, the by far most common exposure is the *original premium income* that the ceding insurer collects for the reinsured portfolio. It is felt that the premium rating of the insurer should reflect best which parts of the portfolio bear a risk above average or below; this knowledge shall not be waisted. However, there is a significant problem: Even when the rating of the insurer is very sound, their final premiums will be affected by market pressure, which is very difficult to quantify but nevertheless has to be factored out. Here it is obvious that the inflation of the exposure measure and that of the losses often deviate, albeit they should in the (very) long run be similar, as market cycles balance out. In order to make the original premium inflation-free both the inflation contained in the rated premiums and distortions due to the market cycle have to be factored out. The result is called *on-level premium*; we will have a closer look at it later on.

Most of the mentioned examples, plus some more and further details, can be found across the book by [Parodi, 2014]. The essential take-away message from the possibly confusing variety of exposure measures is: They are not straightforward, not even those not containing inflation. Many are just approximations of the unknown "true" volume, however, mostly clever ones. It thus makes sense to model (some) uncertainty both on the *scale* (inflation) and the *frequency* (volume) side, and that is exactly what we are going to do.

2.2 The sample mean

2.2.1 Motivation

Large risks as described in the previous section are not only larger than personal lines risks, they are also far more heterogeneous. Thus, for (re)insurers it is not easy to create pools of similar large risks that can be rated together, as is routinely done in personal lines. Although methods like Credibility cater for heterogeneity within portfolios, there are many cases, in particular in industrial insurance and reinsurance, where pooling is impossible or embraces only a few risks. We assume in the following for simplicity that the rating is based on the data of a single large risk, bearing in mind that the same methods can be applied to pools.

Fortunately, large risks produce many more losses than small ones, so in many cases it is possible and reasonable to do classical experience rating: by just using the loss record of the risk itself. However, one has to use a longer loss history than the typical 1 to 3 years used in personal lines rating.

Notice that we treat only the case of sufficient loss data in this book, bearing in mind that very scarce loss records occur, where experience rating is inadequate. However, from practical experience it results that the commercial insurance and reinsurance risks having a loss record being sufficient for experience rating are a considerable share of all risks to be rated, and if we weigh them by premium, they become even more considerable. In short, sound experience rating, albeit not being applicable to all rating situations, is a major ingredient to the pricing process.

Experience rating can be done in various ways, parametric (with distributional models) or parameter free. The *sample mean* or *empirical mean*, i.e. the empirical first moment, is the basis for distribution free rating, but it also enters parametric methods: indirectly as benchmark or directly as input if parameters are estimated via the method of moments. In reinsurance rating and a lot of commercial rating the sample mean is arguably the most frequently applied rating method.

2.2.2 Structure

Let q be the year for which we do the rating (*quotation* year) – quotation is the traditional name for reinsurance pricing. We would typically have the data available from several years k:

$$k_{min} \leq k \leq k_{max} < q$$

We assume throughout the book that all available data are in principle representative, apart from uncertainties in scale and volume. Thus, if e.g. 15 years ago a fundamental change in law or in insurance conditions drastically altered the characteristics of the covered losses, we would ignore the data from the period prior to that change.

Typically the rating is done at some time in the year $q-1$ and $k_{max} = q-2$, the year before. However, it could be that the data of that year is incomplete and/or data of very recent years is left out due to still large uncertainty about the run-off of the losses. Then we could have more years between the end of the data set and the quotation year, e.g. $k_{max} = q-4$. Recall that we do not include run-off uncertainty in the models described in this book, so we assume that the loss data used is essentially ultimate. This is a substantial simplification, but we will see in the next subsection that more general models are far more intricate, too much so to embrace them in this analysis. Our restriction on rather short and/or predictable run-off shuts out a lot of Liability insurance, in particular many reinsurance layers, but for most Non-Life lines we can say that after a few years the losses are largely paid and the remaining claims reserves are quite precise predictions of the final payments due. This holds not only for Property and Hull insurance of most kinds, but also e.g. for Personal Accident and even for some Liability business.

Let X be the time series describing the (aggregated) loss affecting the (re)insurance cover we are rating. Then we want to predict X_q from the observed X_k. The sample mean in its most simple form would be the average of these observations. In case the volume of the risk varies over time things become more complex. The expected loss changes according to the volume, thus the X_k have to be rebased to the volume in the year q. Further, the average over the years is typically a weighted one, as it is well known that the linear combination of independent unbiased estimators is optimal (in terms of variance minimization) if the weights are inversely proportionate to the variances.

We now convert this reasoning in formulae, then sophistication begins. All the following is a procedure to transform X_k, the empirical loss, into a so-called *as-if loss*, taking into account how the situation changes from year k to year q.

Let us start with proportional coverages. Here X_k is the aggregate *ground-up loss* in the year k. We assume variable volume V_k, the case of constant volume follows as straightforward special case. If the volume measure is adequate, it is a standard assumption that the expected loss is proportionate to the volume. (Recall that the time series V must be, or be made, inflation-free.) Thus, in order to get an unbiased predictor, we have to replace X_k by $\frac{V_q}{V_k}X_k$. If in addition we have inflation and call the time series reflecting the *cost level* B, then we have to factor in the increase of the losses between the years k and q due to inflation. Altogether we get the (ideal) as-if loss

$$\frac{V_q B_q}{V_k B_k} X_k$$

Note that inflation here is intended to be specific for the risk in question; it may be very different from consumer price inflation. According to line of business and other risk characteristics, inflation could be influenced by things like wages (Liability; having often slightly larger increases than consumer prices), construction cost (Property), or steel prices (Marine; extremely volatile). We could even have that there is no monetary nor wages inflation at all, but that houses, cars, or vessels tend to be built using increasingly valuable components. This would create an inflationary effect on Property and Hull coverages and we would regard that as inflation, too – mathematically it is the same as "classical" inflation, as the loss sizes increase over the years.

A statistical (regression) approach to relate claims inflation, e.g. for a large Motor Liability portfolio, to some economic indices, is described in the recent paper by [Bohnert et al., 2016], which also provides a great deal of references to papers, from practice and academia, tackling various aspects of inflation in Non-Life insurance.

Whichever inflation we have, it will normally be somewhat uncertain, so we do not have B and must put up with a predictor $\widehat{B}$. From the above examples we have seen that the volume measure can be somewhat uncertain as well, so analogously we work with a predictor $\widehat{V}$. Combining the predictors for inflation and volumes we get the slightly different as-if loss

$$\frac{\widehat{V_q}\widehat{B_q}}{\widehat{V_k}\widehat{B_k}}X_k$$

Remark 2.2. Prediction may sound weird here – we in fact "predict" some *past* index values: If we see statistics as dealing with uncertainties in general, it is rather about *known vs unknown* than *past vs future*. So, if a past event is not (exactly) known, be it an index value or the magnitude of an earthquake occurred long ago, a reasonable approximation for the event is essentially a prediction, bearing some (stochastic) uncertainty just as predictions of future events.

In practice common predictors for inflation are certain indices, being often provided by external sources like public bureaus of statistics or insurance industry associations, which in several countries calculate e.g. construction cost indices to be used for Property insurance. Sometimes actuaries carefully adapt such indices, sometimes they must construct new indices on purpose from various data sources.

The volume (think of the number of insured objects) is specific per insurance cover, however, it can be seen as an index, too: essentially a time series whose changes between the years k and q describe shifts of the risk over time. This interpretation of the volume measure as an index will turn out to be very handy, emphasizing the mathematical analogy with inflation effects. However, for volumes there will usually be no official index available to predict them, instead one has to select/construct an appropriate "index" on a case by case basis. We will treat that in detail in Chapter 8.

To reflect the index interpretation in the as-if loss formula, we define:

Definition 2.3. We write IB for an index predicting B, and analogously IV predicting V. We will speak of B as the *true inflation index* vs the *picked inflation index* IB, and analogously for the volume. In this notation the classical real-world *as-if loss*, projecting the data of the year k to the year q, reads

$$S_k := \frac{\widehat{IV_q}\widehat{IB_q}}{IV_k IB_k}X_k$$

The rewriting with indices makes transparent that the values in the denominator are known, being past values of observable indices, hence here we could drop the hats. For the year q instead we know neither future inflation and volume, nor the index values predicting them. All these distinctions may seem fussy at the moment, but will pay out along the way to go.

Here's an overview of the evolution of the as-if loss:

$$X_k \longrightarrow \frac{V_q B_q}{V_k B_k}X_k \longrightarrow \frac{\widehat{V_q}\widehat{B_q}}{\widehat{V_k}\widehat{B_k}}X_k \longrightarrow \frac{\widehat{IV_q}\widehat{IB_q}}{IV_k IB_k}X_k = S_k$$

Recall that the second term is the unbiased predictor we would like to use but do not have available. The third and forth term are approximations thereof, which, if we manage to find adequate indices, are not too biased. This is often the best we can do when we are uncertain about inflation and/or volume.

Notice that only the product of inflation and volume is needed here, so one could combine both to get a unified time series reflecting shifts in both loss size and frequency, just as certain official volumes do. However, we will see later that the separation is needed for the calculation of the variance, further it is crucial in the non-proportional case following now.

For layers we use the collective model of risk theory. The aggregate layer loss equals

$$X_k = \sum_{i=1}^{N_k} {}^u_d Z_{k,i}$$

where N_k is the (stochastic) number of ground-up losses, while $Z_{k,i}$ is the i-th loss in the year k. (The losses are defined according to how the layer applies, single or accumulation losses.) Notice that we do not need all losses to calculate this sum, only the few ones exceeding *as-if* the retention d are required – this is matched by the typically provided list of all losses exceeding the somewhat lower reporting threshold. We give the scheme analogously to the proportional case:

$$\sum_{i=1}^{N_k} {}^u_d Z_{k,i} \longrightarrow \frac{V_q}{V_k} \sum_{i=1}^{N_k} {}^u_d \left(\frac{B_q}{B_k} Z_{k,i} \right) \longrightarrow \frac{\widehat{V_q}}{\widehat{V_k}} \sum_{i=1}^{N_k} {}^u_d \left(\frac{\widehat{B_q}}{\widehat{B_k}} Z_{k,i} \right) \longrightarrow \frac{\widehat{IV_q}}{IV_k} \sum_{i=1}^{N_k} {}^u_d \left(\frac{\widehat{IB_q}}{IB_k} Z_{k,i} \right)$$

The reasoning is as above, save that now only the volume acts proportionately, while inflation affects the losses *before* application of the layer, thus has a non-linear effect, which we will explore thoroughly in Chapter 6.

Definition 2.4. We call

$$S_k := \frac{\widehat{IV_q}}{IV_k} \sum_{i=1}^{N_k} {}^u_d \left(\frac{\widehat{IB_q}}{IB_k} Z_{k,i} \right)$$

the *(aggregate) as-if loss* of the year k to the layer $u - d \, xs \, d$.
d may equal 0, u may be infinite, such that the definition, in particular, embraces proportional coverages.

It must be noted that for accumulation losses the formula requires volume and inflation to be defined in a different manner than usually. Such losses, think of windstorms or earthquakes, behave in a particular way if the size of the risk changes: If the risk grows (more insured objects, or larger insured objects, or a mixture of both), it is not hit by *more* windstorms, but the impact of each windstorm will be *higher* (hitting more and/or larger objects than before). That means that *any* kind of change in risk size is an inflationary effect for accumulation losses and has accordingly to be incorporated in the index B and predictors thereof: As stated earlier, for the purpose of as-if adjustments, inflation is anything making losses larger, be it a monetary effect or not. See [Pfeifer, 1997] for an approach to estimate such inflation for accumulation losses from natural perils, being applicable in case one has a very long complete loss record available.

The volume instead must reflect changes in the frequency of *accumulation* events. While such changes are thinkable, due to say shifting storm or seismic activity, in practice it is arguably most common to assume constant activity, which means that the volume is constant over time.

It goes without saying that if e.g. a Property risk has a so-called *per risk and per event XL* cover, insuring it against both fire (mainly losses affecting single objects) and natural (accumulation) events, the experience rating has to be done separately for these two very different kinds of losses, employing, in particular, different indices: If IB and IV are the adequate indices for the fire losses, the inflation index for the NatCat losses would typically be $IB \, IV$ and the corresponding volume the trivial constant index. We shall treat accumulation losses in more depth in Chapter 8.

A hidden data issue shall be mentioned: If inflation is high and the reporting threshold not low enough, it could be that we miss some smaller losses that *as-if* would exceed the retention (albeit only by a small amount). We ignore this potential distortion in this book, as its numerical impact is typically very low. However, it shall be emphasized that if we are not certain about inflation, there is always the possibility that the reporting threshold, however low it is, cannot grasp every potential as-if loss with certainty. To keep the impact of this problem as low as possible, in practice one tries to keep a margin:

Say we think that a threshold of 80% of the retention is sufficient, but nevertheless we strive to get the loss data starting from 70%, at least for the oldest years where the aggregate impact of inflation is typically highest.

Having defined the aggregate as-if loss S_k for both cases, pro-rata and layer, we finally state the classical formula for the sample mean:

$$T\left(X_q\right) = \sum_{k=k_{min}}^{k_{max}} \frac{IV_k}{IV_+} S_k$$

where $IV_+ = \sum_{k=k_{min}}^{k_{max}} IV_k$. Notice that if we plug in $S_k = \frac{\widehat{IV_q}}{IV_k} \sum_{i=1}^{N_k} {}^u_d \left(\frac{\widehat{IB_q}}{IB_k} Z_{k,i}\right)$, the IV_k drop out of the formula, only their sum remains.

Mathematically, the classical sample mean is an average of the S_k, which are weighted according to their (predicted) volume. We motivate this weighting in detail for the pro-rata case, the non-proportional case is analogous. If the time series B and V are known, each $\frac{V_q B_q}{V_k B_k} X_k$ is an unbiased predictor for X_q. As for the variance, the most common assumption is that $\operatorname{Var}\left(X_k\right)$ be proportionate to V_k. It emerges from two plausible models, see [Mack, 1997], Sections 1.3.2 and 1.4.2: either from the individual model with X_k being an iid sum of RV's whose number is proportionate to V_k, or from the collective model with iid loss sizes and a Poisson distributed loss count. From this variance rule it follows that the variance of $\frac{V_q B_q}{V_k B_k} X_k$ is proportionate to $\frac{1}{V_k}$. As is well-known, the variance-minimizing average of unbiased estimators weighs them inversely proportionally to their variances, so to optimally combine the $\frac{V_q B_q}{V_k B_k} X_k$ we would have to use the weights $\frac{V_k}{V_+}$, where $V_+ = \sum_{k=k_{min}}^{k_{max}} V_k$. If we finally take into account that we do not know B and V and replace them by the respective picked indices, we get the above formula for $T\left(X_q\right)$.

It shall be noted that in practice the weights according to volume are often used even in case the Poisson model is not thoroughly believed in, in particular in case the data are not abundant and a two-parameter model for the loss count is hard to assess. As an alternative, equal weighting of the years occurs to be used, however, this is certainly the much less popular option. (Calculations with the Negative Binomial model show that equal weights would result as approximately optimal if the loss frequency were very high and the NegBin shape parameter very far away from the limiting Poisson case, see again Section 1.4.2 of [Mack, 1997].)

As for terminology, experience rating with the sample mean is also called *burning cost* (BC) calculation. From the name it is clear that it comes originally from Fire insurance, however, it is commonly used in other lines and, in particular, for the experience rating of non-proportional reinsurance treaties across all lines of business. S_k would be called the *as-if burning cost* of the year k.

As mentioned, we will from now on mainly work with the collective model introduced for non-proportional coverages, which embraces pro-rata covers as the special case of a layer $\infty\, xs\, 0$.

2.2.3 Simplifications and refinements

The basic assumptions underlying the calculation of the as-if losses are a simplification of reality, as always. However, they seem to trade off well between theoretical aspiration and what can be done with the data available in practice:

Firstly, we have one inflation index value per year. Of course, inflation is rather gradual, not stepwise increasing every New Year, thus a continuous inflation model would be theoretically more appealing. However, sometimes inflationary effects come stepwise due to (minor) changes in law or in insurance conditions, so to cater for both phenomenons one would need a very complex model combining smooth periods with jumps. More importantly, it is not clear how one could exploit the benefits of a daily cost level when faced with far less precise loss data: For proportional covers one typically gets the data aggregated per year, thus any allocation to single days would rather be an illusion of exactness. Very rarely the complete loss database with the loss dates is available for the rating, but practical experience tells that these dates are not always reliable: There could e.g. be a sharp peak at New Year, which is most probably not due to losses caused by fireworks, but to a default date set in case the loss record was entered incompletely into the data base. For layers instead, the (few) large losses reported come mostly with exact dates, but it seems questionable that using them would be worth the effort – data scarcity, as is common in layer pricing, does generally not admit a lot of model complexity.

Quarterly or monthly index values could be a frugal refinement; some indices used in practice disclose such values. However, for the exposures it will in practice be hard to get anything finer than one figure per year, and if say quarterly data are provided, they could happen to be a bit unreliable, coming from a quick and rough allocation, as they occur in practice when approximate figures fulfill the internal needs of the company. In short, yearly figures, albeit being a bit imprecise, are mostly the best option.

Secondly, if we account for inflation via the rule of three, i.e. by multiplying the historical losses by factors $\frac{B_q}{B_k}$ or predictors thereof, we assume that inflation affects all losses uniformly by the same factor. This is indeed a restrictive assumption, but weakening it would inevitably add a lot of complexity, and there is hardly any handy alternative in sight. (Power laws are studied occasionally, but are they really more plausible?) However, in practice a frugal workaround has emerged: segmentation of losses into classes of homogeneous loss types, data permitting. This can improve the quality of the rating considerably, certainly more than many thinkable mathematical refinements. The result are subsets (e.g. separating different loss causes in Property, material damage vs personal injury in Liability) having uniform inflation within but different inflation rates in between. It is straightforward how to adapt the above formulae for this procedure, however, for the sake of clarity we stay with the non-segmented case – formulae will get complex enough.

Thirdly, as mentioned, we do not model run-off uncertainty in this book. Avoiding over-complexity is a fair reason for this, however, there is a much more valid one. The common calculation of as-if losses via multiplication with $\frac{B_q}{B_k}$ (or its predictor) assumes that the size of losses in the year k corresponds to the cost level in that year, while losses in the year q have a size corresponding to the cost level in the quotation year. This assumption is indeed very plausible for the losses in many Non-Life lines. Even when the payment of say a complex Fire or Personal Accident loss extends across some years, it is basically the cost level of the accident year that defines the paid amounts. However, an important exception are large personal injury losses in the Liability lines, including MTPL and Workers' Compensation. In many countries a significant part of such losses are annuities, being paid to the victim (and/or his family) for many consecutive years, increasing with inflation: So if a loss occurred in the year k and an annuity is paid 7 years later, this payment will correspond to the cost level of the year $k+7$. Such payments should thus be as-if adjusted according to *calendar year*, not according to *accident year*. All in all, the total loss contains payments stretching over decades, corresponding each to one of the various cost levels of these years. A precise as-if calculation would require having the complete cash flow of the loss available, ideally segmented per head of damage in order to identify indexed annuities. Such detailed info occurs to be reported for reinsurance layers, however, it is far more common to have just the cash flow available, no finer segmentation. Nevertheless reinsurers aim to adjust losses per calendar year where deemed adequate, see [Lippe et al., 1991], bearing in mind that quite some imprecision remains. For proportional business instead, the loss info comes far more aggregated, without any chance to separate payments that should be adjusted per calendar year from those to be adjusted per accident year, a problem that cannot be compensated by the finest mathematical model.

In short, the risks with the longest and most volatile run-off in Non-Life insurance (those significantly paying annuities) cannot be included in this analysis for two reasons: large uncertainty about how to calculate as-if losses, and huge additional complexity due to the run-off. (These challenges might be worth another book, provided adequate models and adequate loss data are found.)

2.2.4 Number of years typically used

How many years are traditionally used in the sample mean (supposed that the loss history is long and deemed fairly representative)?

In practice it is tried to keep the standard error low, while the possibly fading representativity of older data is usually taken into account only qualitatively, in the following way: If a restriction on the more recent years leads to a tolerable standard error, the older years are left apart.

To illustrate this common reasoning, we look at a risk that does not vary (much) in size, such that (approximately) we could regard the loss record of the single years as iid, represented by Y, after having (supposedly correctly) adjusted for inflation. Then the variance of the (supposedly unbiased) sample mean equals $\frac{Var(Y)}{n}$, where n is the number of years used. Therefore, if we have in mind a maximum tolerable level $maxSE$ of the *relative* standard error, measured in percent of the true parameter $E(Y)$, we must use at least $n = \left(\frac{CV(Y)}{maxSE}\right)^2$ years.

E.g., if we do experience rating for a particularly stable risk having a coefficient of variation (CV) of 10%, then 4 years are sufficient to get the (relative) standard error of the sample mean down to 5%.

Many practitioners would see 5% as a desirable level: In insurance sectors with abundant data it is not uncommon to add a safety loading of about twice the (estimated) standard error to the point estimate for the risk premium, which in case of a normal distribution would yield an upper confidence bound at the 97.5% level. (To be exact, this level requires 1.96 standard deviations, see the discussion about (approximate) confidence intervals in Section 12.3 of [Klugman et al., 2008].) With a relative standard error of 5% this procedure yields a surcharge of 10% to the net premium – a figure largely seen as not excessive.

Of course, in case of very good data one would strive to get lower standard errors and surcharges, however, scarce-data situations requiring substantial safety loadings occur.

For a rather volatile risk having a CV of 50%, to get the relative standard error down to 5% we would need 100 (representative!) years, which in practice are hardly available. In fact in commercial and reinsurance rating, given the heterogeneity of the risks and the limited loss data, the maximum tolerated relative standard errors are rather high, so instead of a reassuring level of 5% pricing actuaries might content themselves with 20% or even a bit more. This extends the range of risks where experience rating is possible, bearing in mind that the single ratings will be much more uncertain than ratings of large personal lines portfolios.

The following table displays the *required number of years*, according to the CV of the risk and the tolerated relative standard error. Only values greater than 1 and up to 100 are listed.

CV loss	Tolerated relative standard error				
	5%	10%	20%	30%	50%
10%	4				
20%	16	4			
30%	36	9	2		
50%	100	25	6	3	
100%		100	25	11	4
200%			100	44	16
300%				100	36
500%					100

For what situations is our initial question (how many years are optimal) relevant anyway? Of course not if there is sufficient data to do a precise rating based on 3 years or less – here the impact of past inflation will be low anyway, whether we are somewhat uncertain about it or not. The trade-off between amount and up-to-dateness of the data probably starts to become interesting when more than say 5 years are available. At the long end, a pricing actuary will never see loss statistics stretching over 200 years, and if so, some sharp changes in environment or in insurance conditions will make very old years non-representative. Having 100 years would already be a very fortunate case, however, statistics over 30-50 years are used indeed, in particular for accumulation losses from natural catastrophes.

From the table we can see that the number of years relevant for our rating problem (5 - 50) corresponds to CV's ranging from about 20% to about 200%. For lower CV's extremely few years suffice, for higher CV's we anyway cannot get enough data to get the standard error down to a tolerable level.

Chapter 3

Past inflation gap

Now that we start the formal modeling, let us emphasize that throughout this book we assume that all relevant random variables are defined on an appropriate probability space $(\Omega, \mathcal{A}, \mathrm{P})$, being large enough to describe the phenomenons we are dealing with.

3.1 Definitions

To ease presentation we introduce some technical normalizations, which will pay out soon.

The picked indices for IB and IV are not known for the future, even for the present year the values are usually still unknown or provisional. We shift the time axis by defining that the last year with known picked index values be *the year zero*: Typically, from the perspective of the moment of the rating, this would be the year before, rarely the present year. We then have $k_{max} \leq 0 < q$, so the k are negative, while q is positive. When we, from now on, speak of *negative* vs *positive* years, or from *past* vs *future*, we mean according to this definition of the year 0.

Further, we set the base of both the true and the picked inflation index such that $B_0 = 1 = IB_0$. This means technically that we replace the original index $\underline{IB}$, where the underscore shall designate the index before normalization, by

$$IB_k := \frac{\underline{IB}_k}{\underline{IB}_0}$$

etc. In the moment of the rating the value $\underline{IB}_0$ is known, so it is possible to normalize the index history. As for the future, it is notably sufficient to predict normalized index values: Experience rating requires only ratios like $\frac{IB_q}{IB_k}$, which are invariant to base change.

In case the inflation rate is always positive, the normalized inflation index values of the past (negative years) are smaller than 1, while the future ones are greater than 1.

For the volume we normalize analogously such that $V_0 = 1 = IV_0$.

Finally, we define the *normalized losses*

$$\mathring{Z}_{k,i} := \frac{Z_{k,i}}{B_k}$$

We could think of the losses being first produced in an inflation-free world (corresponding to the world of the year 0), then the cost level comes in: $Z_{k,i} = B_k \mathring{Z}_{k,i}$. With inflation acting uniformly as we assume, the $\mathring{Z}_{k,i}$ are iid and independent of the B_k; let their common distribution be represented by the RV $\mathring{Z}$, which we call *normalized severity*. Notice that the exact as-if loss $\frac{B_q}{B_k} Z_{k,i}$ simplifies to $B_q \mathring{Z}_{k,i}$.

This framework looks technical, but can be interpreted easily. We shall give a thorough mathematical formulation of the interaction of indices and losses at the beginning of Chapter 7; for now an intuitive sketch suffices. It shall be emphasized that all quantities assembled here are regarded as RV's as follows: All indices are random time series, in particular B, V, and their picked counterparts IB, IV. The $\mathring{Z}_{k,i}$ are iid random losses and independent of all indices. The loss count N is a discrete random time series, being tied to V. If we had to simulate the interaction of their outcomes, it would work like this:

For each year k, simulate at first V_k. The result v_k defines the parameters of the distribution of N_k, which is simulated next. Then simulate n_k losses $\mathring{Z}_{k,i}$, $i = 1, ..., n_k$. These are the losses of the year k, however, having sizes corresponding to the state of the world in the year 0. Finally, simulate B_k and calculate the losses $Z_{k,i} = B_k \mathring{Z}_{k,i}$ of the year k in the money terms (cost level) of that year.

Let us now relate true and picked indices.

Definition 3.1. For any year k (positive ore negative) we define the *inflation gap* and analogously the *volume gap*:

$$\underline{\Delta B}_k := \frac{B_k}{\underline{IB}_k}, \qquad \underline{\Delta V}_k := \frac{V_k}{\underline{IV}_k}$$

For any kind of index the *gap* or *basis risk* shall be the true index divided by the corresponding picked index. For $k \leq 0$ we speak of the *past inflation gap*, etc.

Note that for the normalized indices the analogous formulae hold. In particular, all normalized gaps in the year 0 equal 1.

We emphasize that, like the true indices, we regard the picked indices as random. The essential difference is that the picked indices are observable for $k \leq 0$, while the true indices are not – they can only be approximately observed through the picked indices, "distorted" by the gap. Overall, for each kind of index we have a triple of RV's:

- true index: unobservable,
- picked index: past values observable,
- gap: unobservable, but after normalization taking on values (hopefully) rather close to 1.

The gaps "distort" true indices in the same way as the aforementioned imprecise devices distorted earthquake-magnitude measurements long ago. Gaps are also analogous to the basis risk in finance, which arises if a certain outcome cannot be hedged exactly and must be approximated by "picking" a mix of actually available investments. Finally, there is an analogy to structural models and state-space models, see [Brockwell and Davis, 2006]: If we take the logarithms in the formula $B_k = IB_k \Delta B_k$, we can write the true (logarithmized) time series as the sum of an observable time series and a random element, which is essentially the structure of the observation equation in the (one-dimensional) state-space representation.

With the gaps we can represent the as-if loss in another way, which will turn out useful soon:

$$S_k = \frac{\widehat{IV_q} \Delta V_k}{V_k} \sum_{i=1}^{N_k} {}_d^u \left(\frac{\widehat{IB_q} \Delta B_k}{B_k} Z_{k,i} \right) = \frac{\widehat{IV_q} \Delta V_k}{V_k} \sum_{i=1}^{N_k} {}_d^u \left(\widehat{IB_q} \Delta B_k \mathring{Z}_{k,i} \right) \tag{3.1}$$

One could alternatively define the gaps as the inverses of the above values, however, formulae turn out simpler with the chosen variant, which has the following intuitive interpretation: The inflation gap is the piece that has to be "added" to the picked inflation to become the true inflation, etc.

To get more insight let us look at two examples:

- If inflation is totally ignored in the rating, i.e. $IB_k \equiv 1$, then the inflation gap equals the inflation itself. This is undesirable but in practice as an extreme case could occur. Normally, however, practitioners feel that they make considerable effort to find/create good indices being close to the true values, such that one can expect the normalized gaps to take on values pretty close to 1.

- If the year-to-year inflation is always underestimated by the picked index, then ΔB is an increasing index, such that the values for negative k are smaller than 1. In practice it is often tried to be on the safe side with inflation, e.g. by adapting an available index via adding say 1% of extra inflation per year. This can, however, lead to systematic overestimation of inflation. Then ΔB is a decreasing index, such that the values for negative k are greater than 1.

In the following we will often work with the assumption that for certain indices there is no systematic error in the choice of the corresponding picked index, in other words, nothing systematic about the gap between true and picked index, meaning technically that gap and corresponding picked index are *independent*. This is somewhat restrictive an assumption, however, not too much, whenever we feel that the index choice is done pretty well.

For comparison we mention some other possible scenarios for the gaps. If e.g. the inflation rate is systematically underestimated by a tenth, this would in multiplicative terms mean $IB_k = B_k^{0.9}$ and $\Delta B_k = B_k^{0.1}$, yielding a total dependence of picked index and gap. Alternatively, one could think of the picked index being always behind the true one by one year. This means $IB_k = \frac{B_{k-1}}{B_{-1}}$ (with the normalizing constant B_{-1}) and $\Delta B_k = \frac{B_k}{B_{k-1}} B_{-1}$, which typically also leads to a strong dependence between IB and ΔB, according to the specific structure of the time series B. One could think of combinations of such dependencies, or approximate variants thereof, and some of these scenarios can possibly be studied on a case by case basis. However, we shall not treat such strong dependencies in this book.

3.2 Models

To explore the mathematical properties of the sample mean, we need a plausible model for the defined gaps (bearing in mind that we are only totally free in the model choice in case of no dependence between gap and corresponding picked index). As indicated above, we mainly focus on the benign case that picked indices are chosen/constructed carefully, such that we can expect the normalized gaps to be quite close to 1, with large deviations from 1 being impossible or very rare. The hope is that if the gaps are time series fluctuating few, the specific choice of model and parameters is not too sensitive.

We want to be frank here: Initially we *select* a model for each gap and *guess* its parameters. This seems arbitrary, but as a first step it is an accepted practice: In insurance and many other fields Bayesian statistics is popular, where so-called prior distributions are “selected” in such a way that the chosen parameters have not too sensitive an impact on the final outcome (“vague priors”). This is very different in the details from what we do, however, it is analogous in the selection of some model and its parameters a priori. Much closer to our approach is the earthquake model mentioned in the introduction, where potential errors of older measurement devices are catered for by introduction of a normally distributed error term with an estimated (where possible) or carefully selected (some call this procedure *educated guess*), plausible, rather small variance, see [Rhoades, 1996]. It could be that we get a chance to estimate some parameters later on (we are not optimistic), however, even if not, theoretical analysis of the model involving a number of scenarios for the chosen parameters will anyway give a lot of insight into how the various ingredients of the overall model interact and which parameters have the most sensitive impact on the properties of the model.

As the gaps are ratios of two indices, the true and the picked one, it is plausible that adequate models are similar to those used for the indices themselves, certainly not more complex, but rather less. We let us inspire from the well-established models introduced by Wilkie (see [Wilkie, 1995] and his earlier papers referenced therein) for a lot of yearly time series in the UK and other countries. His classical model for e.g. retail prices and wages is the AR(1) process with lognormal error. To be exact, if we call the time series reflecting the cost level R_k, the quantity $\ln\left(\frac{R_k}{R_{k-1}}\right)$, which approximately equals the inflation rate

$\frac{R_k}{R_{k-1}} - 1$ between subsequent years, is an autoregressive process of first order with a normally distributed random element. The basic recursive formula is

$$\ln\left(\frac{R_k}{R_{k-1}}\right) = a\ln\left(\frac{R_{k-1}}{R_{k-2}}\right) + (1-a)\,\mu + \tau e_k$$

where a is the autoregression coefficient, $0 \le a < 1$, μ is the long-term mean, τ is the standard deviation, and e_k is a standard normal error.

If we reduce complexity by dropping the autoregressive element ($a = 0$), the $\ln\left(\frac{R_k}{R_{k-1}}\right)$ become iid normal. Then $\ln(R_k)$ is a discrete random walk with normal error. R_k would be called a *geometric random walk*, being the product of iid lognormal variables. This is the model that we will preferably use in this book for inflation and volume gaps, however, many results can be stated much more generally.

Note that time series analysis also studies stochastic errors other than lognormal (or equivalently normal for the logarithmized series). We will see some examples, however, errors with infinite first/second moment are beyond the scope of this book.

Let us introduce some handy notation for geometric random walks with lognormal error. Let W be the *one-year inflation gap*, i.e. the lognormal variable generating $\underline{\Delta B}$. Then for arbitrary $k > l$ the ratio

$$\frac{\Delta B_k}{\Delta B_l} = \frac{\underline{\Delta B_k}}{\underline{\Delta B_l}}$$

is the product of $k - l$ independent copies of W.

Let ν and σ be expectation and standard deviation of $\ln(W)$ and $v := \exp(\nu)$, $w := \exp\left({}^{\sigma^2}\!/{}_2\right)$. Then one can easily see that

$$\mathrm{E}(W) = vw, \qquad \mathrm{E}\left(W^2\right) = v^2w^4, \qquad \mathrm{Var}(W) = v^2\left(w^4 - w^2\right), \qquad \mathrm{CV}^2(W) = w^2 - 1$$

More generally for real powers of W we have $\mathrm{E}(W^\alpha) = v^\alpha w^{\alpha^2}$. So for the inverse of W we have $\mathrm{E}\left(W^{-1}\right) = \frac{w}{v} =: y$, $\mathrm{E}\left(W^{-2}\right) = \frac{w^4}{v^2} = y^2w^2$, etc.

ΔB_q is a product of q independent copies of W, $\mathrm{E}(\Delta B_q) = v^q w^q$.

For negative k, ΔB_k is a product of $-k$ independent copies of W^{-1}, $\mathrm{E}(\Delta B_k) = \frac{w^{-k}}{v^{-k}} = y^{-k}$. This representation may seem odd, however, for the sake of clarity we preferably use positive exponents.

ΔB_q and ΔB_k are independent for any $k \le 0$. However, if $l \le k \le 0$, ΔB_l and ΔB_k are products containing $-k$ common copies of W^{-1}. One calculates easily that

$$\mathrm{Cov}(\Delta B_k, \Delta B_l) = \frac{w^{-3k-l} - w^{-k-l}}{v^{-k-l}} = y^{-k-l}\left(w^{-2k} - 1\right) > 0$$

Notice that all these covariances are strictly positive, as long as W is not a constant ($w > 1$). [Clark, 2006] gets such a covariance structure for the *future* run-off of Liability losses occurred in the past say 10 years, but not yet settled. His model is more complex, though, as he models inflation as a whole, using an AR(1) process.

This was the basic math for the inflation gap; for the volume it is analogous with the respective parameters. When dealing with various gaps at the same time, we will distinguish them by writing W_B, w_B, v_B, ... for the quantities belonging to $\underline{\Delta B}$; instead we write W_V etc. for those belonging to $\underline{\Delta V}$, and analogously for other gaps we possibly introduce later on.

3.3 Parameters

We have to "select" the parameters ν and σ of the inflation gap. To get an idea we at first look at various inflation indices being modeled by AR(1) processes in the way described above. The richest source is [Wilkie, 1995], providing parameters fitting the CPI (consumer price index) for over 20 developed countries in periods stretching from about 1970 into the 1990ies. While the long-term means vary a lot across those countries, the other parameters do not. The autoregression coefficients mainly range between 0.4 and 0.8, being often close to 0.6. The standard deviations mainly vary between 0.01 and 0.04, being often close to 0.02, tending to be larger if the mean is large. Only a few countries (including the UK) go beyond 0.04, up to 0.063.

[Clark, 2006] fits the US CPI, extending the period beyond the year 2000; he obtains figures in the same range. Wilkie further gives fits for UK wages, finding parameters quite close to retail prices and again in the above range. [Stephan, 2011] treats an interesting example, fitting three different kinds of Germany's (rather low) inflation across 50 years: consumer prices, producer prices, and labour costs. All of these could be relevant for Non-Life insurance products, notably their means deviate a great deal. The autoregression coefficients lie in the above range, the standard deviations between 0.01 and 0.026. For the variances this yields a range from 0.0001 to 0.0007, while Wilkie's (international) range would go up to 0.0016, or up to 0.004 if we include the very highest figures.

If the picked index IB is AR(1) as described with parameters a, μ, τ^2 and $\mathit{\Delta} B$ is a geometric random walk with parameters ν, σ^2, then one calculates quickly that $B = IB\mathit{\Delta} B$ is again AR(1) with parameters a, $\mu + \nu$, $\tau^2 + \left(1 + a^2\right)\sigma^2$. We feel that the true inflation is just another kind of inflation, deviating from the picked one, but belonging to the same country. Thus, we would expect it to have parameters more or less within the range of the common inflation measures in that country. In particular, the difference $\left(1 + a^2\right)\sigma^2$ of the variance terms cannot be too large, it should hardly exceed 0.0016, could indeed be much lower, and should only in extreme cases be as high as 0.004. To ensure this, given the above range for a, we need σ to be typically equal to or lower than 0.03, to reach 0.05 only in extreme cases. Even then we have $\sigma^2/2 \leq 0.0013$, such that w is anyway extremely close to 1. With $\mathrm{CV}(W) \approx \sigma \leq 0.05$ the inflation gap cannot be a very volatile random walk.

As for ν, if we assume that the index "pick" is done with care and is reviewed regularly, it would be implausible that an under- or overestimation of loss inflation by more than 1% per year would stay undetected over the years. (If the data are not too volatile, one would see that indirectly as an upward or downward trend in the as-if loss per year.) So we would expect ν to be mostly in the per mil range, say somewhere between -0.01 and 0.01. This is a magnitude greater than $\sigma^2/2$, such that the question whether $\mathrm{E}(W)$ is greater or smaller than 1, depends essentially on ν, or equivalently v, only.

Whatever the parameter constellation, in realistic situations ν and σ should generally be small, such that the normalized gap is typically a RV being pretty much concentrated about 1.

Chapter 4

Basic example

4.1 Structure

To understand the impact of the stochastic past inflation gap, we study a simplified example, which, however, will turn out to be very instructive, paving the way towards more general results. We call it the *basic example*:

- Proportional coverage.
- Constant volume. Think of the Fire insurance of an industrial plant that does not add new buildings, while the existing ones just grow in replacement value.
- $q = 0$. This is a bit optimistic, however, we could imagine that we are at the very end of the year $q - 1$, doing a last-minute rating, and already know the cost level of the following year.
- The inflation gap is independent of the picked inflation index
- ... and is a geometric random walk with lognormal error, which is nontrivial ($w > 1$).

For this risk the normalized volumes disappear (all equal 1) and the future inflation is not uncertain, as both B_q and IB_q equal 1. Equation 3.1 simplifies to

$$S_k = \sum_{i=1}^{N_k} \Delta B_k \mathring{Z}_{k,i} = \Delta B_k \mathring{X}_k$$

where $\mathring{X}_k := \frac{X_k}{B_k}$ is the *normalized aggregate loss* analogous to the normalized severity. With constant volume the $\mathring{X}_k$ are iid.

We set $e := \mathrm{E}\left(\mathring{X}_k\right) > 0$, $\eta := \mathrm{CV}^2\left(\mathring{X}_k\right) > 0$ (nontrivial RV), so $\mathrm{Var}\left(\mathring{X}_k\right) = e^2\eta$.

For ease of presentation we count backwards, so let now be $l > k \geq 0$. As the normalized losses are independent of each other and of the gaps, we quickly get

$$\mathrm{E}\left(S_{-k}\right) = \mathrm{E}\left(\Delta B_{-k}\right)\mathrm{E}\left(\mathring{X}_{-k}\right) = \frac{w^k}{v^k}e = y^k e \tag{4.1}$$

$$\mathrm{Cov}\left(S_{-k}, S_{-l}\right) = \mathrm{Cov}\left(\Delta B_{-k}, \Delta B_{-l}\right)\mathrm{E}\left(\mathring{X}_{-k}\right)\mathrm{E}\left(\mathring{X}_{-l}\right) = y^{k+l}\left(w^{2k} - 1\right)e^2 \tag{4.2}$$

$$\mathrm{Var}\left(S_{-k}\right) = \mathrm{Var}\left(\Delta B_{-k}\right)E^2\left(\mathring{X}_{-k}\right) + \mathrm{E}\left(\Delta B_{-k}^2\right)\mathrm{Var}\left(\mathring{X}_{-k}\right) = y^{2k}\left(w^{2k}\left(1+\eta\right) - 1\right)e^2 \tag{4.3}$$

4.2 Optimal number of years

With constant volume the sample mean weighs all years equally. We set $m := -k_{max} > 0$ and define the sample mean of the n most recent available years of the data set:

$$T_n(X_q) := \frac{1}{n} \sum_{k=m}^{m+n-1} S_{-k}$$

Which of these sample means is best? This is the question posed in the foreword. In the classical theory without uncertainty about inflation and volume the answer is clear: take as many years as available. Without the random $\varDelta B_k$ the S_{-k} would be iid and unbiased, hence the variance of $T_n(X_q)$ would decrease proportionally to $\frac{1}{n}$. With the inflation gap we might get a bias and the distributions of the S_{-k} will be different and not independent. To see this in detail we need to do some algebra.

Beforehand recall the distinction of the two common variants of the *mean squared error* (MSE): according to whether *process risk*, i.e. the stochastic deviation of the RV to be predicted from its expectation, is included (mean squared error of prediction *MSEP*) or not, quantifying only the precision of the parameter estimation (squared parameter estimation error *SPEE*). Both quantities are interesting, and we treat them in parallel. In the situations we investigate they turn out to have very similar properties, such that we can often speak of MSE without having to specify.

We want to predict $X_q = X_0 = \mathring{X}_0$, which has expectation e and is independent of the S_{-k} and thus, in particular, uncorrelated to linear combinations thereof like $T_n(X_q)$. We have

$$\mathrm{MSEP}(T(X_q)) = \mathrm{E}\left((T(X_q) - X_q)^2\right) = \mathrm{Var}(X_q) + \mathrm{Var}(T(X_q)) + \mathrm{Bias}^2(T(X_q)) - 2\mathrm{Cov}(X_q, T(X_q))$$

$$\mathrm{SPEE}(T(X_q)) = \mathrm{E}\left((T(X_q) - \mathrm{E}(X_q))^2\right) = \mathrm{Var}(T(X_q)) + \mathrm{Bias}^2(T(X_q))$$

These well-known formulae hold generally for predictors of X_q. However, in our setting the covariance term vanishes, hence

$$\mathrm{MSEP}(T_n(X_q)) = \mathrm{Var}(X_q) + \mathrm{Var}(T_n(X_q)) + \mathrm{Bias}^2(T_n(X_q)) \tag{4.4}$$

and by dropping the first summand, which is unrelated to $T_n(X_q)$, one gets the corresponding SPEE.

Notice that all the expectations and variances here and in the following formulae require strictly speaking that the parameters e and η of the normalized loss be fixed (which is standard practice in such MSE calculations), as well as the parameters w, v (or equivalently y) of the inflation gap. Note further that said moments are conditional on the past picked index values IB_k. All this seems natural, however, other choices are thinkable. We will generally discuss conditioning in MSE calculations in Section 7.4, treating (much) more complex models.

For the components of Formula 4.4 we get with Formulae 4.1, 4.2, 4.3

$$\mathrm{Bias}(T_n(X_q)) = \mathrm{E}(T_n(X_q)) - \mathrm{E}(X_q) = \frac{1}{n} \sum_{k=m}^{m+n-1} y^k e - e = \left(\frac{1}{n} \sum_{k=m}^{m+n-1} \left(y^k - 1\right)\right) e$$

$$\begin{aligned}
\mathrm{Var}(T_n(X_q)) &= \frac{e^2}{n^2} \sum_{k,l=m}^{m+n-1} \mathrm{Cov}(S_{-k}, S_{-l}) \\
&= \frac{e^2}{n^2} \left(2 \sum_{m \le k < l \le m+n-1} y^{k+l}\left(w^{2k} - 1\right) + \sum_{k=m}^{m+n-1} y^{2k}\left(w^{2k}(1+\eta) - 1\right) \right)
\end{aligned}$$

Let us interpret these terms as sequences in n. For $y > 1$ the bias tends to infinity, thus for large n the MSE must increase. Thus, there must be a *finite* n_0 yielding the optimal (smallest) MSE.

Notice that the variance tends to infinity even under the (slightly) weaker condition $yw > 1$. It has the lower bound $\frac{1}{n}\frac{e^2}{n}\sum_{k=m}^{m+n-1} y^{2k}w^{2k}\eta$, which is an arithmetic mean and thus greater than the geometric mean

$$\frac{e^2\eta}{n}\left(\prod_{k=m}^{m+n-1} y^{2k}w^{2k}\right)^{1/n} = \frac{e^2\eta}{n}\,(yw)^{2m+n-1}$$

For $y \leq 1$ the bias is bounded. It can be shown that if y is small enough the variance is bounded, too. We omit the details. (To see that the variance can get as low as you want, replace the sums by infinite ones and let y and yw be much smaller than 1; this yields arbitrarily small upper bounds.) This case is problematic anyway: It means negative bias, possibly a considerable one.

Small MSE's are obviously not desirable at any cost. For many quantities one can easily define predictors with very small MSE: Just select a fair predictor and use 95%, 90%, 85% ... of it. The bias will increase but the variance will decrease, and often the latter effect is initially stronger, such that one finds a predictor having a much lower MSE than the original one. But this comes at the expense of negative bias. The classical aim is in fact to optimize MSE among *unbiased* estimators. In our complex situation with the inflation gap we might not be able to ensure nonnegative bias, but we must trust that the picked inflation index is chosen in such a careful manner that strong negative bias is avoided.

In short, if premiums are rated far too cheap, one has indeed more serious problems than MSE optimization. There is no point in investigating this case further. If actuaries in practice suspect a potential picked index to underestimate inflation considerably, they will adjust or replace it. This does not mean that such errors do not occur, they do, however, MSE optimization for such situations is spurious. We can cater for this undesirable case only by qualitatively saying that if one has heavy doubts about being able to approximate (at least vaguely) the true inflation by a picked index, to avoid high negative bias it is best to use only very recent years, where the impact of a wrong index is lower. (This workaround will, however, lead to high random errors in the rating due to a small data base.)

Back to the main case of a bit more trust in the index choice. Extending the examples given after the definition of the inflation gap, we can state the following for the geometric-random-walk model:

- If the year-to-year inflation is on average overestimated by the picked index, then the ΔB_k decrease in expectation, i.e. $wv = \mathrm{E}(W) < 1$. Then $y = \frac{w}{v} > w^2 > 1$.
- On the contrary, if inflation is on average underestimated, we have $wv > 1$. However, if v is not too large, namely $\frac{1}{w} < v < w$, then we still have $y > 1$ and an ultimately increasing bias. Even a bit beyond, if $v < w^2$, we still have $yw > 1$ and an ultimately increasing variance. Thus, slight underestimation still leads to similar mathematical MSE properties as in the case of overestimation of the inflation.

In the following we assume at worst slight underestimation (on average) of the inflation. We want to look at the shape of the MSE sequence. If $y = 1 + \iota$ and $w = 1 + \delta$ are not too far from 1, each MSE can be well approximated by its Taylor series around $(1, 1)$ up to elements of first order. Let us look at the MSEP in Formula 4.4 element-wise.

- $\mathrm{Var}(X_q) = \mathrm{Var}(X_0) = e^2\eta$.
- The bias in first order equals $\left(m + \frac{n-1}{2}\right)e\iota$, so the squared bias equals 0 in first order terms.
- As for the variance, in first order the term in the double sum equals $2k\delta$, while the term in the second sum equals $2k\delta + \eta\,(1 + 2k\,(\iota + \delta))$. Note that for $y \geq 1$ these are not only first-order approximations, but also lower bounds.

Summing up the variance terms is lengthy but straightforward. The double sum turns out to equal

$$\left(\frac{n^3}{3} + (m-1)\,n^2 - \left(m - \frac{2}{3}\right) n\right)\delta$$

while the second sum equals

$$\left(n^2 + n\,(2m-1)\right)(\delta + \eta\,(\iota + \delta)) + n\eta$$

Collecting and rearranging all ingredients yields the MSEP, interpreted as a sequence or function in n:

$$\begin{aligned} \mathrm{MSEP}\left(T_n\left(X_q\right)\right) &= e^2\left(\frac{a}{n} + b + cn\right) \\ a &= \eta\left[1 + (2m-1)\,(\iota + \delta)\right] + \frac{1}{3}\delta \\ b &= (2m-1)\,\delta + \eta\,(1 + \iota + \delta) \\ c &= \frac{2}{3}\delta \end{aligned} \tag{4.5}$$

For the SPEE one simply has to replace b by

$$b^* = (2m-1)\,\delta + \eta\,(\iota + \delta)$$

In case of no gap ($\iota = 0 = \delta$), the MSEP would simplify to $e^2\eta\left(\frac{1}{n} + 1\right)$, which is the classical formula, decreasing in n. If the variance of the future year (process risk) is left out, we get the (much more common) SPEE formula $\frac{e^2\eta}{n}$.

Formula 4.5 instead yields a completely different behavior. The function is dominated for small n by the $\frac{1}{n}$ term, while for large n the term in n dominates. As long as $\iota + \delta$ is nonnegative (which is in first order equivalent to $yw \geq 1$) or negative but close to 0, the three coefficients a, b, c are positive. Then the MSEP decreases at first *sharply* to a minimum, rising then *slowly* towards infinity. The n_0 yielding the minimal MSEP is the optimal number of years for the rating, provided for this year (according to our parametrization it would be the year $k_{max} - n_0 + 1$) the data are still available. Otherwise all available years should be used.

We calculate the minimum of the MSEP as a function in *real* n; n_0 must be the adjacent integer below or above. Recall that for a function $\frac{a}{x} + b + cx$ in $x > 0$ with positive a, c, the global minimum is taken on at $\sqrt{\frac{a}{c}}$ and equals $b + 2\sqrt{ac}$. We could plug in the terms derived above, however, to get some intuition we simplify a bit. Unless the volatility of the losses from sources other than inflation, reflected by η, is not very low (an outcome which we do not expect too often in commercial insurance and reinsurance), we would expect η to dominate the coefficient a in Formula 4.5. Mathematically speaking, we assume that ι and δ be very small compared to 1 and δ be also much smaller than η. Then we get

$$a \approx \eta, \qquad n_0 \approx \sqrt{\frac{3\eta}{2\delta}}, \qquad \mathrm{MSEP}_{min} \approx e^2\left(b + 2\sqrt{\frac{2}{3}\delta\eta}\right)$$

How can we interpret n_0? If as assumed the inflation gaps are not too volatile, we have

$$\mathrm{CV}^2\,(W) = w^2 - 1 \approx 2\delta \approx \sigma^2$$

It follows that

$$n_0 \approx \frac{\sqrt{3\eta}}{\sigma} \approx \frac{\mathrm{CV}\left(\mathring{X}_k\right)}{\mathrm{CV}\,(W)}\sqrt{3} \tag{4.6}$$

The SPEE has the same properties, being just lower by the constant $e^2\eta$. Let us sum up.

Proposition 4.1. *In the basic example with realistic parameters, the optimal number of years for the sample mean (in terms of MSE minimization) approximately equals the ratio of the CV's of normalized yearly loss and one-year inflation gap, respectively, times* $\sqrt{3}$.

Notice that ι, the decisive parameter for the bias, does not enter this approximative formula. It only appears in the more precise variant working with the coefficients from Formula 4.5. This variant should be used if η is not large compared to δ.

The results found are even an indication for the case that the Taylor approximation of first order is not very good. Recall that it is, at least for $y \geq 0$, also a lower bound, notably a U-shaped one. Thus, we can expect the minimum of the exact MSE function to be not too far from the minimum of its lower bound.

The simple $\sqrt{3}$ formula 4.6 has charm, but at the same time reveals parameter sensitivity. If we misestimate the numerator or the denominator by say a factor of 2, we will have the same error in the estimate of the optimal number of years. Recall that we do not know $\mathrm{CV}\left(\mathring{X}_k\right)$; we could estimate it from the S_k, but these include the inflation gap, which will somewhat distort the estimate, albeit arguably not too much. Much worse, $\mathrm{CV}(W)$ was just guessed (albeit in a thoughtful manner), we could indeed get it wrong by a factor of 2, so we would come up with a n_0 equaling half or twice the true value. This is a bit unsatisfactory, however, the shape of the MSEP function in n (Formula 4.5) gives us a qualitative result, no matter whether we get the parameters right: The function first decreases sharply to the minimum, then increases much more slowly. Thus, a bit too many years will increase the MSE much less than a bit too few.

This is the first piece of a strategy: If you are uncertain about inflation gap parameters (but quite certain about having avoided strong negative bias in the picked inflation index), choose a scenario yielding a rather high n_0. If you somewhat overestimate the optimal number of years, this will not do too much harm.

Example 4.2. Say a large and rather stable Property risk, which does not change its volume over time, has a coefficient of variation of the normalized loss equaling 20%. If the inflation gap is volatile, say $\mathrm{CV}(W) = 3\%$, Formula 4.6 gives $n_0 \approx 11.5$, hence an optimum number of 11 or 12 years. For a moderate inflation gap, say $\mathrm{CV}(W) = 1\%$, we get about 35 years. An extremely volatile inflation gap, say $\mathrm{CV}(W) = 5\%$, would lead to 7 years.
It the risk were much more volatile, say $\mathrm{CV}\left(\mathring{X}_k\right) = 50\%$, we would get 2.5 times longer optimal periods: about 29 / 87 / 17 years.

If n_0 is very large, the optimal number of years is only of theoretical interest, as in practice the loss data set will rarely go back that far, such that one has to put up with using all available data. However, in this simplified example the only gap is about inflation. Once we introduce the volume gap, we might see higher overall uncertainties, which should lead to shorter optimal periods and to more situations where the oldest years are better dropped. We will illustrate this in Chapter 10.

4.3 Optimal weights

The question how many years to use, is exciting, but in the light of the findings leading to its answer a much more interesting question emerges: Is the classical sample mean the best available nonparametric statistics, or can we reduce the MSE further by assigning weights to the years according to age? Formulae 4.2 and 4.3 show that at least for $y \geq 0$ the variances and covariances increase with age of year, such that assigning lower weights to older years should reduce the overall variance. The situation is, of course, more complex than in the classical variance minimization case, as we have dependencies across the years. However, such problems were tackled successfully. We briefly sketch two examples.

- [Gerber and Jones, 1975] introduce a recursive formula in a heterogeneous Credibility setting, i.e. with a grand mean available. Their model yields weights that in a special case decrease geometrically (when counting years backwards as we are doing).
- [Mahler, 1998] treats a homogeneous Credibility model for Workers' Compensation insurance, i.e. without grand mean, thus similar to our situation. In the details his model is very complex, such that no analytical formulae for the weights are available. However, from his examples one can see that more recent years get higher weights, just as we would expect. (Subsequently, as Mahler introduces run-off volatility, which is very high in Workers' Compensation, the situation somewhat changes: Now the most recent years are just as uncertain as much older ones and the highest weight goes say to the third youngest year. This is again an intuitive result.)

Before optimizing weights in our example let us compare its covariance structure to that of the two referred models. In a compact representation the (co)variances read as follows.

Model	$k<l$	$k=l$
Gerber/Jones	a_k	a_k+b_k
Mahler	$a\,g\,(l-k)$	$a+b$
Inflation Gap	$a_k\,g\,(l-k)$	a_k+b_k

In our model we have

$$g\,(l-k)=y^{l-k},\quad a_k=y^{2k}\left(w^{2k}-1\right)e^2,\quad b_k=y^{2k}w^{2k}e^2\eta$$

Its basic structure is analogous to that of [Gerber and Jones, 1975] for the variances, where [Mahler, 1998] has a simpler one. As for the covariances, our structure is somewhat more complex than in both referred models.

Nevertheless, as our formulae for the variances and covariances are in a way highly symmetric, we are able to come up with analytical results. It is, however, a serpentine way to go, which starts with some linear algebra. Let $n>2$ from now on be fixed.

4.3.1 Linear algebra

First we calculate the optimal weights for a number of unbiased estimators, which do not need to be independent. We generalize a well-known result.

Lemma 4.3. *Let T_1, ..., T_n be unbiased estimators of a parameter t, having a positive definite covariance matrix $\Sigma := (Cov(T_k,T_l))$. Out of all unbiased linear combinations $T\,(g_1,...,g_n) := \sum_1^n g_k T_k$ with $\sum_1^n g_k = 1$, the one having minimal variance (SPEE) has the weights*

$$(g_1,...,g_n)^t=\frac{\Sigma^{-1}\vec{1}}{\vec{1}^t\Sigma^{-1}\vec{1}}=:\vec{x}_1$$

where $\vec{1}$ is a column vector of 1's. The same result comes about if the covariance matrix is replaced by its sum with a matrix of 1's

$$\Gamma_{kl} := (1+Cov(T_k,T_l))$$

Proof. For the first matrix this is a well-known result from portfolio theory, see e.g. Theorem 17.1 in [Härdle and Hlávka, 2007]. For an elementary proof by the method of Lagrange multipliers see Appendix D of [Riegel, 2015]. The second matrix is positive definite as a sum of a positive definite and a positive semidefinite matrix of 1's. One can see quickly that this "shifted" covariance matrix yields the same outcome:

Notice that the denominators $\vec{1}^t \Sigma^{-1} \vec{1}$ and $\vec{1}^t \Gamma^{-1} \vec{1}$ are just calibrations ensuring that the sums of the weights yield 1. If $(g_1, ..., g_n)^t$ fulfills the equation with Σ, then $\Sigma (g_1, ..., g_n)^t$ is a multiple of $\vec{1}$. One sees at a glance that then the same holds for $\Gamma (g_1, ..., g_n)^t$. Multiplying by the inverse of Γ from the left and calibrating yields the desired result. □

Note that instead of 1 we could have added any constant $c > 0$ to the covariance matrix. The matrix Γ has some advantages over Σ when we now calculate the minimal SPEE without the restriction to unbiased estimators. We will get a predictor with negative bias, which is not intended to be used, but provides some surprising insights.

Proposition 4.4. *Let T_1, ..., T_n be unbiased predictors of a RV X with expectation 1, having a positive definite covariance matrix $\Sigma := (Cov(T_k,T_l))$. Set $\Gamma_{kl} := (1 + Cov(T_k,T_l))$. Then out of all (not necessarily unbiased) linear combinations $T(g_1, ..., g_n) = \sum_1^n g_k T_k$, the one minimizing the SPEE has the coefficients*

$$(g_1, ..., g_n)^t = \Gamma^{-1} \vec{1} =: \vec{x}_0$$

Let s be the sum of the components of $\vec{x}_0$ (which must not exceed 1). We have

$$s = \vec{1}^t \Gamma^{-1} \vec{1}$$

i.e., s is the sum of the components of the inverse of Γ.

The vectors $\vec{x}_0$ (biased minimum) and $\vec{x}_1$ (unbiased minimum from the preceding lemma) are parallel and we have

$$\vec{x}_1 = \frac{\vec{x}_0}{s}, \qquad SPEE\left(T\left(\vec{x}_0^t\right)\right) = 1 - s, \qquad SPEE\left(T\left(\vec{x}_1^t\right)\right) = \frac{1}{s} - 1 = \frac{1-s}{s}$$

If X is uncorrelated to the T_k, the two vectors minimize the MSEP as well.

Proof. We have

$$\mathrm{SPEE}\left(\sum_1^n g_k T_k\right) = \mathrm{Var}\left(\sum_1^n g_k T_k\right) + \mathrm{Bias}^2\left(\sum_1^n g_k T_k\right)$$

which after some algebra yields

$$\mathrm{SPEE}\left(T\left(\vec{g}^t\right)\right) = \vec{g}^t \Sigma \vec{g} + \left(\vec{g}^t \vec{1} - 1\right)^2 = \vec{g}^t \Gamma \vec{g} - 2\vec{g}^t \vec{1} + 1$$

This is a quadratic form in $\vec{g}$, having a unique and positive global minimum, which is taken on at the root of the derivative (gradient): $\Gamma \vec{g} = \vec{1}$. This proves the first equation.

Now we apply the preceding lemma with $t = E(X) = 1$ and get the unbiased minimum $\vec{x}_1$. As $\vec{x}_0$ is the minimum without the additional unbiasedness condition, the sum of its components must not exceed 1, which is the sum of the components of $\vec{x}_1$. Comparing the formulae for $\vec{x}_0$ and $\vec{x}_1$, we see at a glance that they are parallel vectors and that $s = \vec{1}^t \Gamma^{-1} \vec{1}$.

By plugging $\vec{x}_0$ in the above quadratic form, one quickly sees that the corresponding SPEE equals $1 - s$. As this must be strictly positive, we must have that s is strictly smaller than 1. Plugging finally $\vec{x}_1$ in the quadratic form we get the last formula.

If X is uncorrelated to the T_k, we have

$$\mathrm{MSEP}\left(\sum_1^n g_k T_k\right) = \mathrm{Var}\left(X\right) + \mathrm{Var}\left(\sum_1^n g_k T_k\right) + \mathrm{Bias}^2\left(\sum_1^n g_k T_k\right)$$

where the first term is unrelated to the g_k, such that minimizing MSEP and SPEE are equivalent. □

To apply this proposition, we first transform the predictors S_k by dividing each by its expectation. More generally we introduce:

Definition 4.5. The *standardization* of a RV having positive expectation is the RV divided by the latter, e.g.

$$\dot{S}_k := \frac{S_k}{\mathrm{E}\left(S_k\right)}$$

From Formulae 4.1, 4.2, 4.3 we get for $l > k \geq 0$:

$$\dot{S}_{-k} = \frac{S_{-k}}{\mathrm{E}\left(S_{-k}\right)} = \frac{S_{-k}}{y^k e}$$

$$\mathrm{Cov}\left(\dot{S}_{-k}, \dot{S}_{-l}\right) = w^{2k} - 1$$

$$\mathrm{Var}\left(\dot{S}_{-k}\right) = w^{2k}\left(1+\eta\right) - 1$$

If we set $r := w^2 > 1$, we get a very concise representation for any l, $k \geq 0$:

$$1 + \mathrm{Cov}\left(\dot{S}_{-k}, \dot{S}_{-l}\right) = r^{\min(k,l)}\left(1 + \delta_{kl}\eta\right)$$

With $m = -k_{max}$ we have

$$\Gamma = \begin{pmatrix} r^m\left(1+\eta\right) & r^m & r^m & \cdots \\ r^m & r^{m+1}\left(1+\eta\right) & r^{m+1} & \\ r^m & r^{m+1} & r^{m+2}\left(1+\eta\right) & \\ & & & \cdots \\ \cdots & & & \cdots \end{pmatrix}$$

In order to get the global MSE minimum for linear combinations of the $\dot{S}_{-k}$, we must solve the equation

$$\Gamma \begin{pmatrix} g_1 \\ g_2 \\ \\ \\ g_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \\ \\ 1 \end{pmatrix} \tag{4.7}$$

We prove first that Γ is positive definite. Recall $r = w^2 > 1$ and $\eta > 0$. However, it is instructive to include here the limiting cases of either no inflation uncertainty ($w = r = 1$) or losses being deterministic apart from inflation ($\eta = 0$).

Lemma 4.6. *The matrix Γ is positive definite if $r \geq 1$, $\eta \geq 0$, and at least one of these inequalities is strict. For $\eta = 0$ we have*

$$\det \Gamma = r^{mn + \frac{(n-1)(n-2)}{2}}\left(r-1\right)^{n-1}$$

Proof. Starting with the case $\eta = 0$, we have for $n > 1$

$$\det \mathit{\Gamma} = r^{mn} \det \begin{pmatrix} 1 & 1 & 1 & & 1 \\ 1 & r & r & & r \\ 1 & r & r^2 & & r^2 \\ & & & \dots & \\ 1 & r & r^2 & & r^{n-1} \end{pmatrix} = r^{mn} \det \begin{pmatrix} 1 & 1 & 1 & & 1 \\ 0 & r-1 & r-1 & & r-1 \\ 0 & 0 & r^2-r & & r^2-r \\ & & & \dots & \\ 0 & 0 & 0 & & r^{n-1}-r^{n-2} \end{pmatrix}$$

where in the last step we have subtracted from each row the preceding one. Counting the factors r and $r-1$ on the main diagonal, we get the claimed result, which holds for arbitrary real r. For $r > 1$ the determinant is positive, and so are the determinants of all leading principal minors – these matrices have exactly the same structure. Thus, $\mathit{\Gamma}$ is positive definite. In the limiting case $r = 0$, $\mathit{\Gamma}$ is a matrix of 1's and positive semidefinite.

If $\eta > 0$, $\mathit{\Gamma}$ is positive definite as the sum of the matrix of the case $\eta = 0$ with a positive definite diagonal matrix. □

Now it is clear that we can apply the earlier results of this section and, in particular, that Equation 4.7 can be solved, yielding a unique root.

Proposition 4.7. *In the two limiting cases, the root* $(g_1, ..., g_n)^t = \vec{x}_0 = s\vec{x}_1$ *of Equation 4.7 and the resulting (normalized)* $SPEE(T(\vec{x}_1^t)) = \frac{1}{s} - 1$ *of the unbiased optimum, written in the terminology of Proposition 4.4, are as follows:*

$$\eta > 0,\ r = 1: \qquad \vec{x}_1^t = \left(\frac{1}{n}, ..., \frac{1}{n}\right), \quad s = \frac{1}{1+\frac{\eta}{n}}, \quad SPEE\left(T\left(\vec{x}_1^t\right)\right) = \frac{\eta}{n}$$

$$\eta = 0,\ r > 1: \qquad \vec{x}_1^t = (1, 0, ..., 0), \quad s = r^{-m}, \quad SPEE\left(T\left(\vec{x}_1^t\right)\right) = r^m - 1$$

Proof. In each of the two cases one checks immediately that the vector $\vec{x}_1$ fulfills Equation 4.7 up to a factor, namely $1 + \frac{\eta}{n}$ or r^m, respectively. This factor must be the inverse of s. So we get s and the SPEE. □

The interpretation of this result is straightforward and well in line with intuition:

- Without inflation basis risk, old years are as representative as recent ones, such that the years get equal weights. (For $w = 1 = y$ this still holds after going back to the non-standardized quantities, which we will do in Section 4.3.3.)
- With inflation gap, but no further randomness, it is sufficient to look at the most recent year, which has the lowest uncertainty about inflation.

4.3.2 Fibonacci algebra

Now we investigate the general case $\eta > 0$, $r > 1$. Intuitively one would say that it must yield weights somehow in between the above two limiting cases. This intuition will turn out to be true, however, we need some further detours.

At first we introduce the partial sums $s_k := \sum_{j=k}^{n} g_j$, including $s_{n+1} = 0$. Notice that $s_1 = s$ and $g_k = s_k - s_{k+1}$. Now we transform the LHS of Formula 4.7 in the following way (which is inspired by the row transform used in the last lemma):

$$\begin{pmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \dots & \dots & \\ & & & -1 & 1 \end{pmatrix} \Gamma \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & 1 & -1 & \\ & & & \dots & \dots \\ & & & & 1 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ \\ \\ s_n \end{pmatrix}$$

To the right of Γ we have expressed the vector of the g_k by the s_k with the help of an invertible transformation matrix. Further we have multiplied an invertible matrix from the left. If we multiply the RHS of Formula 4.7 with the same matrix, we get the vector $(1, 0, ..., 0)^t$.

The product of the above three matrices equals

$$\begin{pmatrix} r^m(1+\eta) & -r^m\eta & & & \\ -r^m\eta & r^{m+1}(1+\eta)+r^m(\eta-1) & -r^{m+1}\eta & & \\ & -r^{m+1}\eta & r^{m+2}(1+\eta)+r^{m+1}(\eta-1) & -r^{m+2}\eta & \\ & & & \dots & \\ & & & & -r^{m+n-2}\eta \\ & & & & r^{m+n-1}(1+\eta)+r^{m+n-2}(\eta-1) \end{pmatrix}$$

Now we divide the row k, for $k = 2, 3, ..., n$, by r^{m+k-2} and finally get the matrix equation

$$\begin{pmatrix} r^m(1+\eta) & -r^m\eta & & & \\ -\eta & r(1+\eta)+\eta-1 & -r\eta & & \\ & -\eta & r(1+\eta)+\eta-1 & -r\eta & \\ & & & \dots & \\ & & & -\eta & r(1+\eta)+\eta-1 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ \\ \\ s_n \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \\ \\ 0 \end{pmatrix}$$

Observe that now almost all rows look the same: Row k, for $k = 2, 3, ..., n-1$, yields the equation

$$-\eta s_{k-1} + [r(1+\eta) + \eta - 1]\, s_k - r\eta s_{k+1} = 0$$

The same equation holds for $k = n$, as the missing coefficient $-r\eta$ would be applied to $s_{n+1} = 0$, thus its missing does not make any difference. All in all we get a recursive formula for the s_k. If we divide by $-\eta$ and rearrange a bit, we get $s_{k-1} = bs_k - rs_{k+1}$, where $b := 1 + r + \frac{r-1}{\eta}$. This is a *backward recursive formula* of second degree for the s_k. We define $f_k := s_{n+1-k}$ and finally get

$$f_k = bf_{k-1} - rf_{k-2}$$

If we had $b = 1 = -r$, this would be the formula defining the Fibonacci sequence. Our formula is different, but can be solved analogously to the derivation of the closed-form formula for the Fibonacci sequence: It is obvious that the solutions of such recursive formulae constitute a two-dimensional vector space. Hence, if the quadratic equation

$$x^2 = bx - r$$

has two distinct roots p and q, then the sequences p^k and q^k are a basis for this vector space and all other solutions are linear combinations thereof. In our case with $D := b^2 - 4r$ we get $p = \frac{1}{2}\left(b + \sqrt{D}\right)$, $q = \frac{1}{2}\left(b - \sqrt{D}\right)$. One sees easily that $D > 0$, such that we have two distinct real roots, which are both positive. As their product pq equals $r > 1$, p must be greater than 1. A quick calculation shows that $q < 1$.

We have to find out which linear combination $f_k = ap^k + a'q^k$ of the two roots solves our problem.

Typically this is inferred from the first two elements of the sequence. Recall that for the starting value we have $f_0 = s_{n+1} = 0$. Thus, f_k must equal $a\left(p^k - q^k\right)$ with a constant a. To determine the latter we cannot use f_1, which we do not have, so we have to find another way. First we go back to the partial sums and observe that

$$s_k = f_{n+1-k} = a\left(p^{n+1-k} - q^{n+1-k}\right)$$

Now we plug this into the first row of the last matrix equation, which we have not used yet. We get

$$1 = r^m (1+\eta) s_1 - r^m \eta s_2 = a r^m \left[(1+\eta)(p^n - q^n) - \eta\left(p^{n-1} - q^{n-1}\right)\right]$$

which yields

$$a = \left\{r^m \left[(1+\eta)(p^n - q^n) - \eta\left(p^{n-1} - q^{n-1}\right)\right]\right\}^{-1}$$

As $0 < q < 1 < p$, we have $p^n > p^{n-1}$, $q^n < q^{n-1}$, thus a must be positive. For the sum of the coefficients we get

$$s = s_1 = a(p^n - q^n) = \left(r^m \left(1 + \eta - \eta \frac{p^{n-1} - q^{n-1}}{p^n - q^n}\right)\right)^{-1}$$

Finally, the coefficients of the global minimum equal

$$g_k = s_k - s_{k+1} = a\left[(p-1)\,p^{n-k} + (1-q)\,q^{n-k}\right]$$

This last formula can be interpreted easily. The sequence g_1, g_2, ..., which yields the weights of the years k_{max}, $k_{max} - 1$, ..., is a weighted sum of two strictly positive geometric sequences, having strictly positive weights (recall $0 < q < 1 < p$). The term with the powers of q in a way dampens the one with the powers of p. For small k the latter dominates, such that for the most recent years the sequence is similar to a geometric one. Easy algebra shows that the ratio of the weights g_k of subsequent years, which for very large n starts at a bit less than p, decreases at a decreasing rate. However, the weights do not get arbitrarily close – one sees quickly that $g_n = a(p-q)$, $g_{n-1} = a\left(p^2 - p - q^2 + q\right)$, such that $\frac{g_{n-1}}{g_n} = p + q - 1 = r + \frac{r-1}{\eta} > r > 1$. We emphasize that, as expected, the weights g_k decrease as the years get older.

This structure notably holds both for the global minimum $\vec{x}_0$ of the MSE, which we only introduced as a reference point, and the unbiased minimum $\vec{x}_1$, whose coefficients

$$g_k^* := \frac{g_k}{s}$$

are proportionate to the g_k. For the weight of the most recent year there is a simple representation. The first row of the last matrix equation reads $1 = r^m \left((1+\eta)\, s - \eta\,(s - g_1)\right)$, hence $g_1 = \frac{1}{\eta}\left(r^{-m} - s\right)$ and for the unbiased optimum we have

$$g_1^* = \frac{1}{\eta}\left(\frac{1}{r^m s} - 1\right) = 1 - \frac{p^{n-1} - q^{n-1}}{p^n - q^n}$$

Let us collect the main results:

Proposition 4.8. *If $\eta > 0$ and $r > 1$, the root $(g_1, ..., g_n)^t = \vec{x}_0 = s\vec{x}_1$ of Equation 4.7 and the resulting (normalized) SPEE of the unbiased optimum are as follows:*

$$\vec{x}_1^{\,t} = \left(\frac{p-1}{p^n - q^n} p^{n-k} + \frac{1-q}{p^n - q^n} q^{n-k}\right)_{(k=1,..,n)}$$

$$p = \frac{1}{2}\left(b + \sqrt{D}\right), \quad q = \frac{1}{2}\left(b - \sqrt{D}\right), \quad D = b^2 - 4r, \quad b = 1 + r + \frac{r-1}{\eta}$$

$$s=\left(r^m\left(1+\eta-\eta\frac{p^{n-1}-q^{n-1}}{p^n-q^n}\right)\right)^{-1}$$

$$SPEE\left(T\left(\vec{x}_1^t\right)\right)=\frac{1}{s}-1=r^m\left(1+\eta-\eta\frac{p^{n-1}-q^{n-1}}{p^n-q^n}\right)-1$$

Proof. It remains to prove the first and the last formula. The latter is obvious from the preceding formula, the former follows by dividing the above formula for the g_k by s, noting that $\frac{a}{s}=(p^n-q^n)^{-1}$. □

We did the whole analysis for fixed n. It is remarkable that for different n the formulae look so similar. In particular, the recursive formula does not depend on n.

For completeness we give some limits for large n:

Corollary 4.9. *For n tending to infinity we have*

$$g_1^*\searrow 1-\frac{1}{p},\qquad \frac{g_1}{g_2}=\frac{g_1^*}{g_2^*}\nearrow p,\qquad s\nearrow\left(r^m\left(1+\eta\left(1-\frac{1}{p}\right)\right)\right)^{-1}$$

$$SPEE\left(T\left(\vec{x}_1^t\right)\right)\searrow r^m\left(1+\eta\left(1-\frac{1}{p}\right)\right)-1$$

Proof. One sees easily that all sequences are monotonic; the limits are obvious. □

Remark 4.10. If $w=1+\delta$ is close to 1, we have $r=1+2\delta$ in first order terms and can see quickly that p and q approximately equal

$$1+\delta+\frac{\delta}{\eta}\pm\sqrt{\frac{2\delta}{\eta}}$$

Here the ratio of the CV's of the normalized yearly loss and of the one-year inflation gap, which approximately equals $\sqrt{\frac{\eta}{2\delta}}$, comes again into play, just as in the calculation of the optimal number of years in the traditional sample mean.

4.3.3 Final results

To conclude, we go back to the original, non-standardized predictors. Assorting terms we see that g_{k+1-m}^* is the coefficient belonging to $\dot{S}_{-k}$ in the formula for the unbiased MSE minimum.

The $\frac{S_{-k}}{y^k}=e\dot{S}_{-k}$ are multiples of the standardized as-if losses, estimating the desired figure e, such that the same weights yield an optimal unbiased estimation of e. However,

$$\sum_{k=m}^{m+n-1}g_{k+1-m}^*\frac{S_{-k}}{y^k}=\sum_{k=m}^{m+n-1}\frac{g_{k+1-m}^*}{y^k}S_{-k}$$

can also be seen as a linear combination of the original predictors S_{-k} with coefficients $\frac{g_{k+1-m}^*}{y^k}$, which again minimizes the MSE under the condition of unbiasedness. The fundamental difference is that here the coefficients do not add up to 1: They are chosen such that the predictor is unbiased, which involves dividing each S_{-k} by y^k. (If $y>1$, this implies that the sum of the weights is less than 1.) For the largest weight, which is applied to S_{-m}, we get

$$\frac{g_1^*}{y^m}=\frac{1}{y^m}\left(1-\frac{p^{n-1}-q^{n-1}}{p^n-q^n}\right)\longrightarrow\frac{1}{y^m}\left(1-\frac{1}{p}\right)$$

Summing up: If we know w, y, and η (more modestly: if we regard our estimates / educated guesses thereof as correct), we can, for fixed n, amend the sample mean by replacing the traditional equal weights by the ones derived here. Notice that the whole optimization is independent of y, which comes in only at the very end to calculate the final coefficients.

If we let n vary, coming back to the question "which years to use", it is clear that each optimization for a certain number of years n must yield a lower MSE than if we left the oldest year out: Recall that in the optimal case all coefficients are strictly positive, so any linear combination of n years that assigns weight 0 to some year is not optimal. All in all, the lowest MSE is obtained if we use *all* available years. (In practice, however, it could be that the oldest years have such low weights that it does hardly matter whether they are included or not.)

For numerical illustration we carry on Example 4.2.

Example 4.11. Risk having ...
$\mathrm{CV}\left(\mathring{X}_k\right) = 20\%$, thus $\eta = 0.04$;
volatile inflation gap $\mathrm{CV}(W) = 3\%$,
thus $r = 1 + \mathrm{CV}^2(W) = 1.0009$, $w = 1.00045$, $\sigma = 0.03$;
$m = 1$, i.e. very recent data available;
bias of W equaling 0.09%, thus $y = 1$.

The very benign choice of y (no bias in the classical sample mean) reveals the effect of decreasing vs equal weights best: Here we have "pure" variance, not overlaid by any bias. We calculate $p = 1.16$, $q = 0.86$. Notice that albeit w is very close to 1, such that one might have expected the impact of the inflation gap to be almost negligible, p is considerably greater than 1, such that the weights of the most recent years must be far from equal, as they would be without inflation gap.

The $\sqrt{3}$ formula and its more precise variant yield an optimum number of about 12 years for the classical sample mean, leading to a SPEE of 0.0074 (to be multiplied with e^2). The weighted sample mean over 12 years yields a SPEE of 0.0069, which is considerably less. The weight of the most recent year here is 14.8%, almost twice the average weight of 8.3%.

Using 16 years, which would ensure a relative standard error of 5% in the classical model, yields a SPEE of 0.0066 for the weighted sample mean. Thus, the standard error here equals 8.1% – the volatility of the inflation gap makes it considerably higher than in the classical model.

Using more years does not give any further substantial improvement: 50 years would yield a SPEE of 0.0065, the weight of the most recent year would be 14.0%.

Variants with a moderate bias, say 0.5%, yield higher differences between the optimal classical sample mean and the corresponding weighted one, being due to the fact that the latter is unbiased, the former not.

Appendix A displays detailed outputs for a number of variants. In particular, it is interesting to see how quickly the weights in the optimal sample mean decrease with age and how large the weight of the most recent year is.

4.4 Outlook

The basic example, albeit somewhat stylized, gives a lot of insight into what effects the inflation basis risk can have, according to the parameter constellation. The stochastic properties of the sample mean change fundamentally, no matter how closely about 1 the normalized gap fluctuates, as long as it does not equal 1 almost surely. Further the question whether the picked index over- or underestimates inflation, becomes crucial. The optimal number of years for the classical sample mean and the optimal weights for the weighted sample mean depend mainly on how the volatility of the inflation gap relates to that of the losses from sources other than inflation. Finally, we see that the optimal weights decrease with age of data, however, in a more complex way than just geometrically.

Towards more realistic models we have to climb three significant steps:

- Variable volume: This is not just about a new uncertainty, namely the volume gap, but at first about the question how in principle the variance grows with volume. We propose a very general model, to which we devote the following chapter.

- Non-proportional coverages: Inflation is leveraged to layers, but the details are only well understood for some special cases. We dedicate the chapter after next to this issue, coming up with a comprehensive generalization of known results.

- Future inflation/volume: The modeling of the future requires clarity about the interaction of picked indices and gaps, which will be treated in Chapters 7 to 9, constituting the core part of this book.

Chapter 5

Volume and variance

5.1 Contagion

Definition 5.1. Let N be a counting distribution (loss count) having finite first and second moment. We define:

Dispersion: $\mathrm{D}(N) := \frac{\mathrm{Var}(N)}{\mathrm{E}(N)}$

Overdispersion: $\mathrm{OD}(N) := \mathrm{D}(N) - 1 = \frac{\mathrm{Var}(N)}{\mathrm{E}(N)} - 1$

Contagion: $\mathrm{Ct}(N) := \frac{\mathrm{OD}(N)}{\mathrm{E}(N)} = \mathrm{CV}^2(N) - \frac{1}{\mathrm{E}(N)}$

While the first two concepts are common, the contagion so far has arguably only been introduced for the distributions of the Panjer (a,b,0) class (i.e. Poisson, Binomial, and Negative Binomial model; see [Klugman et al., 2008], Section 6.5), in order to indicate whether concentrations of many losses per year are more or less likely than in the Poisson case, see [Heckman and Meyers, 1983], as well as for mixed Poisson distributions, see [Meyers, 2007]. The contagion indicates overdispersion, having the same sign, and leads to some very handy and intuitive representations.

It is clear that we can express the first two moments of N in terms of mean and variance or equivalently in terms of mean and contagion – this is just a kind of parameter change. If we call the mean λ and the contagion c, we have

$$\mathrm{Var}(N) = \lambda + c\lambda^2$$

This representation is well-known for the Panjer (a,b,0) class, see [Fackler, 2011] for a unified treatment of all three distributions. However, having defined the contagion as general as above, the representation applies to any counting distribution with finite variance.

If we replace the variance by the contagion in Wald's equation

$$\mathrm{Var}(X) = \mathrm{Var}(N)\, E^2(Z) + \mathrm{E}(N)\,\mathrm{Var}(Z)$$

for the aggregate loss $X = \sum_{i=1}^{N} Z_i$ in the collective model of risk theory, we get after some algebra

$$\mathrm{CV}^2(X) = \mathrm{Ct}(N) + \frac{1 + \mathrm{CV}^2(Z)}{\mathrm{E}(N)} = c + \frac{1 + \mathrm{CV}^2(Z)}{\lambda} \tag{5.1}$$

which shows at a glance that if the contagion is positive, the volatility of X cannot be arbitrarily low, whatever high the loss frequency.

Finally, if we choose a subset $\tilde{N}$ of losses according to a criterion on the Z_i ("thinning"), say by a condition $Z_i > d$, then the resulting counting distribution, e.g. that of certain excess losses, is generated via iid Bernoulli variables O_i with some probability p: $\tilde{N} = \sum_{i=1}^{N} O_i$. One calculates easily

$$\mathrm{E}\left(\tilde{N}\right) = \mathrm{E}\left(N\right)\mathrm{E}\left(O\right) = p\lambda$$

and with Formula 5.1

$$\mathrm{Ct}\left(\tilde{N}\right) = \mathrm{CV}^2\left(\tilde{N}\right) - \frac{1}{\mathrm{E}\left(\tilde{N}\right)} = \mathrm{Ct}\left(N\right) + \frac{1+\mathrm{CV}^2\left(O\right)}{\mathrm{E}\left(N\right)} - \frac{1}{p\lambda} = c + \frac{1}{p}\frac{1}{\lambda} - \frac{1}{p\lambda} = c$$

Hence, the contagion is invariant to thinning. In particular, it is equal for the distributions of the loss counts of the various layers of a (re)insurance program.

5.2 Rate on line

We insert a short digression about the rate on line, a concept very common in layer pricing. It can, however, be defined a bit more generally and be applied in a very intuitive way, with the help of the contagion introduced above.

Definition 5.2. Assume that the loss severity in the collective model representing a risk $X = \sum_{i=1}^{N} Z_i$ is limited by a maximum Max. We call $RoL := \frac{\mathrm{E}(X)}{Max}$ the *rate on line* of the risk X.

The classical example are the losses to a limited layer $u - d\,xs\,d$, where we would have $Max = u - d$. Another example are ground-up losses that cannot exceed a certain amount, say the maximum sum insured in a portfolio of Property risks. RoL means expected loss per maximum loss, in the layer context it means expected loss per *line* (a word being used for liability).

If we use as above λ and c for expectation and contagion of the number of losses, we get from Equation 5.1 that

$$CV^2\left(X\right) = c + \frac{\mathrm{E}\left(Z^2\right)}{\lambda E^2\left(Z\right)} \leq c + \frac{\mathrm{E}\left(Z\right)Max}{\lambda E^2\left(Z\right)} = c + \frac{Max}{\lambda \mathrm{E}\left(Z\right)} = c + \frac{1}{RoL}$$

Here we have used the obvious inequality $Z^2 \leq Z\,Max$, which is anything but sharp, but in expectation it can be surprisingly so, namely if Z has a heavy-tailed distribution, as is often the case for layer losses.

To illustrate this, we calculate the factor $\frac{\mathrm{E}\left(Z^2\right)}{\mathrm{E}(Z)Max}$, on which the sharpness of the above inequality for $\mathrm{CV}^2\left(X\right)$ depends, for a layer with Pareto distributed losses, i.e. $\overline{F}_Z\left(x\right) = \left(\frac{d}{x}\right)^{\alpha}$, $d < x \leq u$. Said factor turns out to be a function of the Pareto alpha and the so-called *relative layer length* $rll := \frac{u}{d} > 1$ only, see [Schmutz and Doerr, 1998]. In fact one calculates quickly that $\mathrm{E}\left(Z\right) = \frac{d}{\alpha-1}\left(1 - rll^{1-\alpha}\right)$ and $\mathrm{E}\left(Z^2\right) = 2d^2\left(\frac{1-rll^{2-\alpha}}{\alpha-2} - \frac{1-rll^{1-\alpha}}{\alpha-1}\right)$, such that with $Max = u - d$

$$\frac{\mathrm{E}\left(Z^2\right)}{\mathrm{E}\left(Z\right)Max} = \frac{2}{rll-1}\left(\frac{\alpha-1}{\alpha-2}\frac{1-rll^{2-\alpha}}{1-rll^{1-\alpha}} - 1\right)$$

(To get the values for $\alpha = 1$ and $\alpha = 2$ take the limits according to L'Hôpital's rule, which are well defined and finite.) This term cannot exceed 1 and is quite close to its supremum, as long as the layer is not too long and α is not too large. See the following table:

rll \\ α	0.5	1	1.5	2	3
1.5	.97	.93	.90	.87	.80
2	.94	.89	.83	.77	.67
3	.91	.82	.73	.65	.50
5	.87	.74	.62	.51	.33

In reinsurance layer pricing this fact has made the approximative formula

$$\mathrm{CV}(X) \approx \frac{1}{\sqrt{RoL}} \tag{5.2}$$

very popular, in particular as the loss count is often assumed to be Poisson distributed ($c = 0$) or similar with slight overdispersion (c positive but small), which lets the approximation be much better than if the contagion were very large or negative.

The arguably most frequent application of Formula 5.2 is an approximative standard deviation loading. The classical standard deviation premium principle is indeed quite popular in reinsurance as a loading for profit (return on capital). Traditionally one would add to the net premium something in the range of $\kappa = 10\%$ of the standard deviation, see [Gerathewohl, 1979]. This seems a low percentage, but given the high volatility of reinsurance layers it can still yield a heavy loading.

Notice that, differently from the confidence bounds explained in Section 2.2.4, here usually not the standard error of the sample mean, but the (usually much larger) standard deviation of the *future* year is used. It thus reflects *process risk*, not *estimation error*, and is not tied to any specific rating method.

Generally, adding $\kappa\mathrm{Sta}(X)$ means in relative terms, i.e. in percent of $\mathrm{E}(X)$, adding $\kappa\mathrm{CV}(X)$, which is often replaced by the approximation $\frac{\kappa}{\sqrt{RoL}}$. This latter is notably a function of the expected loss only, not requiring any information about the distribution of the losses, so it is available whatever rating method was used. The resulting loading has been given the name *(square) root rate on line loading.*

Apart from loadings we can use Formula 5.2 to get a quick, but not too inexact, estimate of typical coefficients of variation of the yearly outcome of reinsurance layers. In practice the RoL's of such layers vary a great deal, from some hundred percent (which means several losses per year) for very low layers in some lines of business down to less than 1% in case of top layers. The corresponding approximate CV's for the RoL's 300%, 100%, 30%, 10%, 3%, 1% would be: 58%, 100%, 183%, 316%, 577%, 1000%. Recall from Section 2.2.4 that even if one is very tolerant about the relative standard error of a Burning Cost quotation, one should better not apply this method to risks with a CV of much more than 100%. In our approximative metric this means that Burning Cost rating is only suitable for layers with a RoL greater than say 30% – maybe in case of much tolerance and decades of good data one could go down to 10%.

From practical experience it can be said that the people doing layer pricing are aware of these limitations (partly intuitively, partly from mathematical analysis) – and that they exhaust the stated range of suitable RoLs: There is often no promising alternative to experience rating.

5.3 Diversification

Back to the main theme. If a portfolio grows, how does the variance of the loss count (presumed it is finite) increase?

Imagine a portfolio of v units, think of them as insured objects. The volume v shall for the moment be an integer and known. Let the loss count be N_v and let v vary over the years. If θ is the expected number of losses per unit, we have $\mathrm{E}(N_v) = v\theta$. Let the variance of a single unit be ε. Then the units have dispersion $\frac{\epsilon}{\theta}$, overdispersion $\frac{\epsilon-\theta}{\theta}$, and contagion $\frac{\varepsilon-\theta}{\theta^2}$.

What is $\mathrm{Var}(N_v)$? The two classical cases in the literature are the following:

- The single units are independent. Then the variance is additive and we get $\mathrm{Var}(N_v) = v\varepsilon$, $\mathrm{D}(N_v) = \frac{\epsilon}{\theta}$, $\mathrm{OD}(N_v) = \frac{\epsilon-\theta}{\theta}$, $\mathrm{Ct}(N_v) = \frac{\varepsilon-\theta}{\theta^2 v}$, $\mathrm{CV}^2(N_v) = \frac{\varepsilon}{\theta^2 v}$. Dispersion and overdispersion are invariant, while the contagion decreases with volume, just like the coefficient of variation. In a way we could say that the contagion “diversifies away”, shrinking like the CV.

- The units are independent only conditionally on a *common market factor* Q (think of social or weather conditions) having expectation 1 and variance β, i.e. we let N_v be a RV that conditionally, given Q, has expectation $v\theta Q$, see [Mack, 1997], Section 1.4.2. Typically the Poisson model (conditionally) is assumed, such that the variance equals the mean ($\epsilon = \theta$). Then we have

$$\mathrm{E}\left(N_v\right) = \mathrm{E}\left(\mathrm{E}\left(N_v \mid Q\right)\right) = \mathrm{E}\left(v\theta Q\right) = v\theta$$

$$\mathrm{Var}\left(N_v\right) = \mathrm{E}\left(\mathrm{Var}\left(N_v \mid Q\right)\right) + \mathrm{Var}\left(\mathrm{E}\left(N_v \mid Q\right)\right) = \mathrm{E}\left(v\theta Q\right) + \mathrm{Var}\left(v\theta Q\right) = v\theta + v^2\theta^2\beta$$

  Therefore, $\mathrm{D}\left(N_v\right) = 1 + v\theta\beta$, $\mathrm{OD}\left(N_v\right) = v\theta\beta$, $\mathrm{Ct}\left(N_v\right) = \beta$, $\mathrm{CV}^2\left(N_v\right) = \beta + \frac{1}{\theta v}$. Here the contagion of N_v equals the variance of the market factor. It is thus invariant, while dispersion and overdispersion grow linear with v. The fact that the contagion here does not shrink with increasing volume is due to the *common* market factor, which creates a strong dependence among the single units.

  If Q is Gamma distributed, the resulting distribution is Negative Binomial. In fact this mixed Poisson-Gamma model is the classical intuitive interpretation for NegBin.

Notice that although the volume v was originally intended to be integer-valued, these models work more generally for positive real v.

In practice it is felt that independence of units is often too optimistic an assumption, while the high correlation due to a dominant common market factor seems too pessimistic – it would be plausible to have something in between. We now design such a model, which embraces both examples and goes far beyond the Poisson case for the units.

5.4 A loss count model

The basic idea is analogous to [Meyers, 2007], who calculates the correlation of the loss count of two risks being affected by a market factor, which he calls *common shock*. We model a risk consisting of several units, which are conditionally iid, given the common market factor.

Proposition 5.3. *Let $\left(N_v\right)_v$ be a family of discrete loss distributions for risks having volume v, where the volumes may vary across a certain range of admissible values. Let Q be a nonnegative RV with mean 1 and (finite) variance β, such that for a certain $\theta > 0$ and a real γ, for all possible values of v and conditionally on Q, the distribution N_v has the following (finite) moments:*
$E\left(N_v \mid Q\right) = v\theta Q$,
$Var\left(N_v \mid Q\right) = v\theta Q + v\gamma\theta^2 Q^2$, such that, in particular,
$Ct\left(N_v \mid Q\right) = \frac{\gamma}{v}$ does not depend on Q.

Then we have $E\left(N_v\right) = v\theta$ and

$$Var\left(N_v\right) = v\left(\theta + \theta^2\gamma\left(1+\beta\right)\right) + v^2\theta^2\beta$$

$$D\left(N_v\right) = 1 + \theta\gamma\left(1+\beta\right) + v\theta\beta$$

$$OD\left(N_v\right) = \theta\left(\gamma\left(1+\beta\right) + v\beta\right)$$

$$Ct\left(N_v\right) = \beta + \frac{\gamma\left(1+\beta\right)}{v}$$

Proof. For the variance we have

$$\begin{aligned}\mathrm{Var}\left(N_v\right) = \mathrm{E}\left(\mathrm{Var}\left(N_v \mid Q\right)\right) + \mathrm{Var}\left(\mathrm{E}\left(N_v \mid Q\right)\right) = \mathrm{E}\left(v\theta Q + v\gamma\theta^2 Q^2\right) + \mathrm{Var}\left(v\theta Q\right) \\ = v\theta + v\gamma\theta^2\left(1+\beta\right) + v^2\theta^2\beta\end{aligned}$$

The rest follows immediately. □

Notice that $\mathrm{Var}\left(N_v \mid Q\right)$ grows linear with v, which corresponds to the idea that, given Q, N_v behaves like a sum of v iid units. After averaging over Q, the variance has instead a component in v and one in v^2. Notice how the model embraces the two classical ones: For $\beta = 0$ we get the independent example, while $\gamma = 0$ yields the conditional Poisson model with common market factor.

The resulting formula for the contagion has a component decreasing in v and a volume-invariant one. Reflecting this "mixed" structure, we define:

Definition 5.4. The discrete loss distribution specified in Proposition 5.3 shall be called *mixed contagion model.*

The two components of the contagion together with the mean yield a model in three parameters. (Notice that strictly speaking this is not a model, but rather a class of models, as only the first two moments are specified.) The essential question is: Do such distributions exist at all, beyond the two mentioned classical models? They do.

- The easiest example can be constructed from the Negative Binomial distribution. We use the common parametrization $p_k = \binom{\alpha+k-1}{k}\left(\frac{\alpha}{\alpha+\lambda}\right)^{\alpha}\left(\frac{\lambda}{\alpha+\lambda}\right)^{k}$ with mean λ and shape parameter α. From the well-known formula $\mathrm{Var}\left(N_v\right) = \lambda + \frac{\lambda^2}{\alpha}$ we see at a glance that the contagion is the inverse of the shape parameter. Now let N_v be conditionally, given the market factor Q, NegBin distributed with parameters $v\theta Q$ and $v\alpha$. This distribution is well defined for any positive real v, while for integer v it has the intuitive interpretation as a sum of v iid NegBin variables having parameters θQ and α. The conditions of the proposition are fulfilled with $\gamma = \frac{1}{\alpha}$.
  The limiting case Poisson (infinite α) works analogously, we already mentioned it – here we have $\gamma = 0$.

- Much more generally than NegBin, we can construct a mixed Poisson example, at least for integer v: Let M_i be a single unit, being conditionally Poisson distributed with parameter $\theta R_i Q$, given R_i and Q, where R_i is a mixing positive RV having mean 1 and variance γ, while Q is the market factor. Let the R_i be iid and independent of Q. The interpretation is that each unit has its individual parameter fluctuation, which is superimposed by a market-wide fluctuation – such a combination of local and global effects is very plausible in practice. Then the M_i are conditionally, given Q, mixed Poisson distributed with mean θQ and variance $\theta Q + \theta^2 Q^2\gamma$, such that the sum $N_v = \sum_{i=1}^{v} M_i$, conditionally on Q, is an iid sum. Therefore $\mathrm{E}\left(N_v \mid Q\right) = v\theta Q$ and $\mathrm{Var}\left(N_v \mid Q\right) = v\theta Q + v\gamma\theta^2 Q^2$ exactly as in the proposition.

- Analogously to NegBin above, we can construct a Binomial example: Let be m, v integers, $0 < p < 1$, and let the market factor Q be restricted to the open interval $(0, \frac{1}{p})$. Now let N_v be conditionally, given Q, Binomial with parameters mv and pQ. The interpretation is a sum of v copies of a RV that conditionally are independent and $\mathrm{Bin}\left(m, pQ\right)$ distributed each, thus have expectation mpQ and contagion $\frac{-1}{m}$ (underdispersion!). Then we have $\mathrm{E}\left(N_v \mid Q\right) = mvpQ$ and $\mathrm{Ct}\left(N_v \mid Q\right) = \frac{-1}{mv}$, where the latter is independent of Q. With $\theta = mp$ and $\gamma = \frac{-1}{m}$ we have the situation of the proposition. Here the range of Q matters: it was chosen such that all values pQ lie in the interval $(0, 1)$, ensuring well defined Binomial random variables. Notice that $\mathrm{Ct}\left(N_v\right) = \beta - \frac{1+\beta}{mv}$ can be negative like in the original Binomial model, however, for large v it is positive. We could say that N_v turns from underdispersion to overdispersion as the volume grows.

The analogy between Binomial and Negative Binomial is deep; many formulae can just be "translated" from one side to the other, see [Fackler, 2011] for details.

Special cases of the above NegBin and Binomial models were treated recently in [Li and Ferrara, 2015], based on the common shock approach, with invariable volume. The essential innovation of the mixed contagion model is indeed that the volume may vary over time.

How could the parameters of this model be estimated in practice? Looking at empirical mean and variance of the historical loss count is insufficient to infer three parameters. Bearing in mind that time series in practice are often such short that even the variance is hard to estimate, it is not a good idea to look at the empirical third moment, the more so as this would need further specification of the model with e.g. different outcomes for different variants of mixed Poisson. A way out could be portfolio segmentation: If we get the loss counts of several sub-portfolios separately (e.g. from a pool of various portfolios) and assume that they have the same characteristics as the overall portfolio, just a much smaller volume, we have in principle enough time series to estimate three parameters. Bear in mind that all this only works in a straightforward way if we have all volumes of all years available.

We will use this general loss count model from now, incorporating uncertainty about the volume soon. Real-world volumes (like the number of insured objects/people) are usually rather large, such that discrete volumes can mostly be approximated well by continuous ones. We will thus treat continuous volumes only, bearing in mind that at least for the well-established Negative Binomial distribution such volumes are exactly possible, not only approximately.

Notice further that in case of continuous volumes the units do not need to be small and indivisible – they could be large risks, having just the reference size with volume 1, on which the frequency per volume θ is based. In this case smaller objects with a volume of say $\frac{4}{257}$ or $\frac{1}{\sqrt{7}}$ could exist. Such constellations would still be covered by the above proposition, although the initial idea was about very small units. Note finally that, like θ, γ is a parameter related to the volume unit, thus must be adapted if the scale of the volume measure is changed. To be precise, the product of θ and γ is dimensionless, as is β.

5.5 Collective model

Now we apply the mixed contagion model in the collective model of risk theory. Let Z be the RV describing the loss severity, which, as explained in Chapter 2, could refer to losses affecting single objects or accumulation losses caused by events like windstorms, earthquakes, etc. The aggregate loss $X_v = \sum_{i=1}^{N_v} Z_i$ carries the subscript v indicating that the volume of the risk varies over time.

Recall that $\mathrm{E}(N_v) = \theta v$, $\mathrm{Var}(N_v) = \theta^2\left[\left(\frac{1}{\theta} + \gamma(1+\beta)\right)v + \beta v^2\right]$, and $\mathrm{Ct}(N_v) = \beta + \frac{\gamma(1+\beta)}{v}$. With
$\mu = \mathrm{E}(Z)$,
$\varepsilon = 1 + \mathrm{CV}^2(Z) = \frac{\mathrm{E}(Z^2)}{E^2(Z)}$, and
$e = \theta\mu$, we get

$$\mathrm{E}(X_v) = \theta\mu v = ev$$

and from Equation 5.1

$$\mathrm{Var}(X_v) = e^2 v^2 \left[\left(\beta + \frac{\gamma(1+\beta)}{v}\right) + \frac{\varepsilon}{\theta v}\right] = e^2\left[\left(\frac{\varepsilon}{\theta} + \gamma(1+\beta)\right)v + \beta v^2\right]$$

Notice how similar the respective variance formulae for N_v and X_v are: Simply replace the squared θ (frequency per volume) by the squared e (expected loss per volume) and the term $\frac{1}{\theta}$ by $\frac{\varepsilon}{\theta}$.

These formulae hold for any loss severity Z with its respective parameters μ, ε, and e. Let now be Z a ground-up loss and ${}^u_dZ$ the resulting excess loss to a layer $u-d\,xs\,d$. We could apply the formulae to

$\mu_{(d,u)} := \mathrm{E}\,({}^u_dZ)$,

$\varepsilon_{(d,u)} := 1 + \mathrm{CV}^2\,({}^u_dZ)$, and

$e_{(d,u)} := \theta\mu_{(d,u)}$,

yielding correct results for the aggregate layer loss.

However, as most ground-up losses will not hit the layer, such that ${}^u_dZ = 0$, we get something much more intuitive if we reparametrize with quantities describing the *proper* excess losses, by leaving aside the losses that do not hit the layer. Technically this means that, instead of the severity Z, we use a RV $\bar{Z}$ whose (unconditional) distribution equals the conditional distribution of Z, given $Z > d$ or equivalently ${}^u_dZ > 0$. It is indeed possible (albeit rarely found in the literature) to define $\bar{Z}$ or equivalently $\bar{Z}-d$ in a formally thorough manner, see Section 6.3.2 in [Heilmann and Schröter, 2013].

The proper excess losses are counted by the RV $\bar{N}_v := \sum_{i=1}^{N_v} \chi_{(Z_i>d)}$, while the sums

$$X_v = \sum_{i=1}^{N_v} {}^u_dZ_i, \qquad \bar{X}_v := \sum_{j=1}^{\bar{N}_v} {}^u_d\bar{Z}_{i_j}$$

(where i_j runs only over the losses exceeding d) have the same distribution: The first one has many more summands, but all these equal 0.

We have underlying parameters

$$\bar{\mu}_{(d,u)} := \mathrm{E}\left({}^u_d\bar{Z}\right) = \mathrm{E}\,({}^u_dZ \mid {}^u_dZ > 0) = \mathrm{E}\,({}^u_dZ \mid Z > d)$$

$$\bar{\varepsilon}_{(d,u)} := 1 + \mathrm{CV}^2\left({}^u_d\bar{Z}\right) = 1 + \mathrm{CV}^2\,({}^u_dZ \mid {}^u_dZ > 0) = 1 + \mathrm{CV}^2\,({}^u_dZ \mid Z > d)$$

and frequency $\theta_d := E\left(\bar{N}_v\right)$.

These quantities can be easily related to the original ones.

Lemma 5.5. *With $p := P(Z > d) = E\big(\chi_{(Z>d)}\big) > 0$ we have*

$\theta_d = p\theta$,

$\bar{\mu}_{(d,u)} = \frac{1}{p}\mu_{(d,u)}$,

$\bar{\varepsilon}_{(d,u)} = p\varepsilon_{(d,u)}$, *and* $\varepsilon_{(d,u)} \geq \frac{1}{p}$,

while the contagion of the number of proper excess losses is the same as for the ground-up loss number.

Proof. The first formula is clear. If we look at the two possible cases $Z > d$ and $Z \leq d$, we see that ${}^u_dZ$ has the same distribution as the aggregate loss in the collective model with the severity ${}^u_d\bar{Z}$ and the Bernoulli distributed loss count $\chi_{(Z>d)}$. For a formally strict presentation see again [Heilmann and Schröter, 2013]. So we have

$$\mu_{(d,u)} = \mathrm{E}\,({}^u_dZ) = \mathrm{E}\left(\chi_{(Z>d)}\right)\mathrm{E}\,({}^u_dZ \mid Z > d) = p\bar{\mu}_{(d,u)}$$

As the contagion in this collective model equals -1, Equation 5.1 yields

$$\varepsilon_{(d,u)} = 1 + \mathrm{CV}^2\,({}^u_d\,(Z)) = 1 - 1 + \frac{1 + \mathrm{CV}^2\,({}^u_dZ \mid Z > d)}{\mathrm{E}\left(\chi_{(Z>d)}\right)} = \frac{\bar{\varepsilon}_{(d,u)}}{p}$$

The subsequent inequality follows immediately from $\bar{\varepsilon}_{(d,u)} \geq 1$. The invariance of the contagion to thinning was proven right after its definition. □

Wrapping up we get

Proposition 5.6. *If the ground-up loss number is distributed according to the mixed contagion model, then for the aggregate excess loss we get the following formulae*

$$E(X_v) = \theta\mu_{(d,u)}v = \theta_d\bar{\mu}_{(d,u)}v = e_{(d,u)}v$$

$$Var(X_v) = e^2_{(d,u)}\left[\left(\frac{\varepsilon_{(d,u)}}{\theta} + \gamma(1+\beta)\right)v + \beta v^2\right] = e^2_{(d,u)}\left[\left(\frac{\bar{\varepsilon}_{(d,u)}}{\theta_d} + \gamma(1+\beta)\right)v + \beta v^2\right] \tag{5.3}$$

According to convenience we will work alternately with the parameters of the original and that of the proper excess loss. Original losses are technically much easier to deal with. Most calculations in the following chapter will use them, while the proper excess losses will later help interpret results.

The latter are indeed more intuitive and furthermore correspond to what is empirically observed: namely the frequency and severity of the losses actually hitting the layer, without considering the smaller ones, which in large part do not even come to the (re)insurer's knowledge.

If we look at the last formula and let X_v subsequently be the aggregate loss to various layers $u{-}d\,xs\,d$ belonging to the *same* ground-up risk, we see that the coefficients defining how variance depends on volume are the same for all these layers, except for the term $\frac{\bar{\varepsilon}_{(d,u)}}{\theta_d}$. This latter must be very high if d is large, as the numerator is not smaller than 1, while the denominator is proportionate to $\overline{F}(d)$. Hence, for very high layers $\frac{\bar{\varepsilon}_{(d,u)}}{\theta_d}$ dominates the variance formula, such that the variance grows almost linear with volume.

In short, high layers behave very much like the classical Poisson case, where γ and β equal 0. Conversely, the lower the layer, the larger the impact of the loss count parameters γ and β. They are particularly important if $d = 0$, i.e. for first risks and pro-rata (re)insurance. In order to have a chance to assess such coverages and, in particular, the underlying loss count model, one needs explicit loss count data, not just the aggregate losses traditionally provided for the rating of such business.

Chapter 6

Inflation leverage for layers

Some qualitative facts about inflation are well-known, see e.g. [Mack, 1997], Section 4.3.2. Let Z be a loss from the ground up and let the inflation rate be positive:

- Inflation is lower than proportionate for first risks ${}_0^u Z$.
- Inflation is higher than proportionate for second risks ${}_d^\infty Z$.

Both kinds of coverages exist in practice, so these effects can be observed. When practitioners speak of leveraged inflation, they mostly mean overproportional, however, we will see that less than proportionate occurs as well.

For proper layers with both nontrivial retention and liability there is no easy rule for the impact of the *ground-up inflation* on the *layer inflation*, as we shall call them. However, one example is well understood: the *Simple Pareto* distribution, being also called *Type I*, *Single-parameter* or *European Pareto* distribution. (We will simply say *Pareto* unless it could be mistaken for another Pareto variant.) Here the inflation to the layer can be easily quantified.

6.1 Special case Pareto

The Pareto distribution starts at a known positive threshold s, which we call *Pareto threshold.* Lower losses either do not exist or have a different distribution. The survival function in its most general form reads

$$\overline{F}_Z\left(x \mid Z > s\right) = \left(\frac{s}{x}\right)^\alpha$$

or equivalently

$$\overline{F}_Z\left(x\right) = \overline{F}_Z\left(s\right)\left(\frac{s}{x}\right)^\alpha, \quad x > s$$

where $\alpha > 0$ is the only parameter, the *Pareto alpha*, while s is known. That means we do not specify whether there are losses below the threshold, nor how they are distributed. We do not need this information if the layer is high enough, i.e. $d \geq s$.

For thresholds $t_{1,} t_2 \geq s$ the ratio of the loss frequencies equals

$$\frac{\overline{F}_Z\left(t_1\right)}{\overline{F}_Z\left(t_2\right)} = \left(\frac{t_2}{t_1}\right)^\alpha$$

This is the famous Pareto extrapolation formula.

If $d \geq s$, the tail starting at d is distributed as follows:

$$\overline{F}_Z (x \mid Z > d) = \frac{\overline{F}_Z (x)}{\overline{F}_Z (d)} = \left(\frac{d}{x}\right)^\alpha$$

This is again Pareto: one simply has to replace the original threshold by the new, higher one. One could call this a *memoryless property* of the Pareto distribution – the original threshold is "forgotten". This has implications on the impact of inflation:

Let Z^* be the same kind of loss a number of years later, after uniform *ground-up inflation* via a factor $g > 0$. Then Z^* is distributed like gZ and we have

$$\overline{F}_{Z^*} (x) = \overline{F}_Z \left(\frac{x}{g}\right) = \overline{F}_Z (s) \left(\frac{gs}{x}\right)^\alpha = \overline{F}_{Z^*} (gs) \left(\frac{gs}{x}\right)^\alpha$$

or equivalently

$$\overline{F}_{Z^*} (x \mid Z^* > gs) = \left(\frac{gs}{x}\right)^\alpha$$

which is again Pareto with the same alpha – only the threshold has changed due to inflation. However, if $d \geq gs$, we get for the higher tail again

$$\overline{F}_{Z^*} (x \mid Z^* > d) = \left(\frac{d}{x}\right)^\alpha$$

That means, if the retention d of the layer is such high that it exceeds the Pareto threshold both before and after inflation, the distribution of the upper tail starting at d is *invariant* to inflation. This is a consequence of the stated memoryless property, and while easing calculations it implies that in practice, if we only observe *left-censored* data, i.e. only see the losses being higher than a certain threshold (e.g. a deductible or a reporting threshold), we cannot detect inflation from the *shape* of the empirical distribution. In other words, if the losses in the tail are Pareto distributed, all variations of the empirical tail we might observe over time must be *random* fluctuations from sources *other* than inflation.

Inflation nevertheless affects the expected loss to the layer, however, only via the *frequency*, such that the changes in frequency and in expected loss must be the same. Assume at first that the volume of the risk producing the loss Z is constant. Then the change of the frequency between the situations before and after inflation, respectively, equals

$$\frac{\overline{F}_{Z^*} (d)}{\overline{F}_Z (d)} = \frac{\left(\frac{gs}{x}\right)^\alpha}{\left(\frac{s}{x}\right)^\alpha} = g^\alpha$$

Hence, if g is the ground-up inflation, the *layer inflation* equals g^α. If $g > 1$ (positive inflation rate), then according to whether α is greater, equal, or smaller than 1, the layer inflation is greater, equal, or smaller than the ground-up inflation. This applies to all layers with sufficiently large retention ($d \geq s$, gs) having finite expectation. For $\alpha > 1$ this includes layers with unlimited liability.

If the inflation rate is small, i.e. $g = 1 + h$ with small h, we have $g^\alpha \approx 1 + \alpha h$, such that the layer inflation rate is about α times the ground-up inflation rate. According to this approximation one could call α the leverage of the ground-up inflation to the layer. Notice that the exact leverage is often even higher, as for $g > 1$ and $\alpha > 1$ we have $g^\alpha > 1 + \alpha h$.

If the volume is variable but known, the stated effects hold for the frequency and the expected loss per volume unit, respectively. To get the overall outcome one has to combine the effects from inflation and from volume change.

However, in case the volume is not exactly known, problems arise. Say the losses have a Pareto tail with $\alpha = 2$ and $g = 1.1$ (inflation rate 10%). Then the loss frequency at each sufficiently high threshold, everything else unchanged, will increase by the factor $1.1^2 = 1.21$. But, the same increase could come from an undetected volume rise of 21% without any inflation at all, or (more realistically) from a combination of some inflation with some volume change, say an inflation rate of 6% plus a volume increase of 7.7%.

In short, in the Pareto case left-censored loss data, as it is typically reported for layer business, cannot distinguish between volume changes and ground-up inflation, unless the volume change over time is known. This is essentially why inflation, as was already stated in the introduction, is difficult and sometimes impossible to detect for (re)insurance covers with high retentions, where Pareto or similar distribution tails often occur. The problem is that nearly-Pareto is no easier to deal with than Pareto itself: If a tail is not exactly Pareto, but quite similar, one could in theory observe slight changes in tail shape due to inflation. In practice, however, these will mostly be hidden behind the random fluctuations in the empirical tail shape caused by the scarcity of the loss data.

While the impossibility to detect inflation from tail shape is a disadvantage of the Pareto distribution, it ironically incorporates an advantage: Potential misspecification of the inflation does not distort the estimation of the Pareto alpha. This is, in particular, relevant if, to assess the expected loss to a layer, instead of a Burning Cost calculation the Pareto model is used (together with a loss frequency estimate at a reference point).

To see this, let Z be the Pareto distributed losses in a certain year of the loss history and let Z^* be the same losses as-if, adjusted by the *presumed* inflation factor g. Then no matter whether g is correct an estimate or not, as long as d is high enough, the tail distribution beyond d will always be the same, preserving the α of the historical losses. So if all years in the loss history are Pareto distributed with the same α, which is what typically is assumed when the Pareto model is applied, the pool of these losses will again be Pareto distributed (beyond a certain threshold) with the same α, no matter whether the losses are as-if adjusted correctly, distorted by wrong inflation adjustments, or not adjusted at all. Hence, the classical estimate of the Pareto alpha from the pool of (possibly somewhat incorrectly) as-if adjusted observed losses is as good an estimate as if there were no inflation at all. The two problems, quantification of the inflation and estimation of the Pareto alpha (from possibly scarce data), are both hard to solve – but at least they are not intertwined.

The Pareto distribution is an important model; empirical distributions which the model fits fairly well are frequently observed, yielding, however, a wide range of different values for the parameter α. Reinsurers notice that certain layers in certain lines of business often yield similar alphas across many insurers in a country or even internationally, see [Schmutz and Doerr, 1998]. So it is well known that CatXL's covering accumulation losses from natural hazards like Windstorm, Flood, and/or Earthquake mostly yield very small values in the range of 1 or even less, such that here inflation to the layer is similar to or even lower than from the ground up. Layers protecting e.g. Fire losses often show alphas close to say 1.5 or 2, leading to quite some inflation leverage. The highest leverages are observed in certain medium/high Motor Liability layers in some countries, where alphas in the range of 3 are no exception. If in addition the ground-up inflation is high, which is non uncommon in MTPL (see [SwissRe, 2007], [Brickmann et al., 2005]), the leveraged effect can be dramatic. E.g., an inflation rate of 7% per year together with a Pareto alpha of 3 would lead to a layer inflation rate of 22.5%, notably per year.

It shall be noted that practitioners observe indeed that real-life distributions do not always look like Pareto, but often do so for a certain range of loss sizes. It could e.g. be that between 1 and 10 million USD a Pareto fit with $\alpha = 1.8$ works well, while above we see a Pareto curve having $\alpha = 2.6$. This would yield a piecewise Pareto distribution. For a layer $3\,xs\,2$ we could, however, calculate as if the whole distribution were Pareto with parameter 1.8, while for the layer $13\,xs\,12$ we could do the same with the parameter 2.6. This makes clear that different layers may experience different inflation rates despite the same underlying ground-up inflation: For the two layers mentioned here a ground-up inflation rate of 2% would result in layer inflation rates of 3.6% and 5.5%, respectively.

6.2 Local and regional Pareto alpha

Possibly even more likely than piecewise Pareto, practitioners observe empirical severities with rather smooth transitions from a loss size range similar to a certain Pareto curve to another area similar to a different Pareto curve, predominantly with the parameter alpha rising as the loss size increases. To cater for this, a smooth variant of piecewise Pareto has been introduced, informally by practitioners long ago, formally possibly only recently, see [Riegel, 2008].

Definition 6.1. Let F be a loss distribution function having a derivative (pdf) $f(x) > 0$ at $x > 0$. Let $\overline{F} = 1 - F$ be the survival function. Then we call

$$\alpha_x := -\left.\frac{d}{dt}\right|_{t=\ln(x)} \ln\left(\overline{F}\left(e^t\right)\right) = x\frac{f(x)}{\overline{F}(x)} = -x\left(\ln \overline{F}(x)\right)'$$

the *local Pareto alpha* at x.

Mathematically this is the (negative) slope of $\overline{F}$ at x on log-log scale. In fact on this scale the Pareto survival function becomes a straight line; a piecewise Pareto function looks piecewise linear. A function with a smooth transition from one Pareto curve to another one would look like a smooth curve connecting two linear pieces.

This definition can be generalized a bit. If F is only right-differentiable at x, one can replace the derivative by the right derivative. This could extend the range to $x = 0$, as long as the density is finite at 0. Conversely, one could define a local Pareto alpha from the left by replacing F by its left-continuous counterpart (by changing the values at the jumps of F accordingly) and working with left derivatives. However, the concept of the local Pareto alpha is most appealing for continuously differentiable cdf's. Discontinuities of the density do not distort it too much, but jumps of F do.

Now we define an extension of the local Pareto alpha that works for arbitrary cdf's. At first sight it looks totally unrelated, further it is not defined pointwise, but on intervals. Nevertheless it will turn out to be the *one* adequate measure for the impact of inflation on layers.

Definition 6.2. Let F be the cdf of a loss severity Z with survival function $\overline{F}$ and let (d, u) be an interval, $0 \leq d < u$. Then we call

$$\alpha_{(d,u)} := 1 + \frac{d\overline{F}(d) - u\overline{F}(u)}{\int_d^u \overline{F}(x)\, dx} = 1 + \frac{d\overline{F}(d) - u\overline{F}(u)}{\mathrm{E}\left({}_d^u Z\right)}$$

the *regional Pareto alpha* of the interval (d, u) and of the layer $u - d\, xs\, d$.

The *left-sided* regional Pareto alpha $\alpha_{(d,u)-}$ is defined analogously by replacing d and u by d_- and u_-, respectively.

The case $u = \infty$ can be included if F has finite expectation, which implies $\lim_{u\to\infty} u\overline{F}(u) = 0$.

This definition notably includes (in case of finite expectation) the layer $\infty\, xs\, 0$, i.e. proportional covers, where we always have $\alpha_{(0,\infty)} = 1$.

Notice further that both the local and the regional alpha are tail properties: In both definitions we could replace $F(x)$ by $F(x \mid Z > d)$, as in the layer area these functions are equal up to a factor. This yields the following alternative formula for the regional alpha:

$$\alpha_{(d,u)} = 1 + \frac{d - u\overline{F}(u \mid Z > d)}{\mathrm{E}\left({}_d^u Z \mid Z > d\right)}$$

We now show how the regional alpha emerges as a natural extension of the local alpha.

Proposition 6.3. *If for a loss distribution function F the local Pareto alpha exists in the (open) interval (d,u), we have*

$$\alpha_{(d,u)} = \int_d^u \alpha_x r(x)\,dx$$

with the weighting function

$$r(x) = \frac{\overline{F}(x)}{\int_d^u \overline{F}(z)\,dz}$$

Thus, the regional Pareto alpha is a frequency-weighted mean of the local Pareto alpha over the interval.

Proof. From the definition of the local alpha we get via partial integration

$$\int_d^u \alpha_x \overline{F}(x)\,dx = \int_d^u x f(x)\,dx = \int_d^u \left(\overline{F}(x) - \left(x\overline{F}(x)\right)'\right) dx = \int_d^u \overline{F}(x)\,dx + d\overline{F}(d) - u\overline{F}(u)$$

If we divide by $\int_d^u \overline{F}(z)\,dz$, we get the desired result. For infinite u take the limits. □

A few remarks:

- The result holds a bit more generally, namely if F is locally absolutely continuous on (d,u). Then its derivative exists almost everywhere and the fundamental theorem of calculus holds (for both the right and the left derivative, either can be used), see e.g. [Leoni, 2009]. In particular, we can do partial integration.
- If the density is continuous, the regional alpha must lie within the range of the values of the local alphas. Its exact value depends more on the local alphas in the lower part of the interval, due to the frequency weighting. If the local alpha varies few, which is what in practice is often (but not always) observed, the regional alpha must be located within a small range.
- From the definition of the regional alpha we see that it depends mainly on the values of the survival function at the endpoints of the interval, while the geometry of the distribution function inside the interval comes in in a less sensitive manner, only via the average layer severity.

6.3 Leverage on expectation

The regional Pareto alpha allows us to express the inflation leverage for layers in a surprisingly concise formula, which generalizes the well-known

$$\frac{\mathrm{E}\left({}^u_d(gZ)\right)}{\mathrm{E}\left({}^u_dZ\right)} = g^{\alpha}$$

for the Pareto case. Notice that here we always work with the original excess loss, not the proper one.

Theorem 6.4. *For any loss distribution function $F = F_Z$ and any layer $u-d\,xs\,d$, including unlimited layers in case of finite expectation, a ground-up inflation factor $g > 0$ leads to the layer inflation factor*

$$\frac{E\left({}^u_d(gZ)\right)}{E\left({}^u_dZ\right)} = g^{\overline{\alpha}} \tag{6.1}$$

where $\overline{\alpha}$ is a weighted mean of regional Pareto alphas across scaled intervals

$$\overline{\alpha} = \int_{g^{-1}}^1 \alpha_{(zd,zu)} w(z)\,dz$$

with the weighting function

$$w(z) = \frac{1}{z \ln g}$$

The formula holds as well if the regional Pareto alphas are replaced by their left-sided counterparts.

Proof. The RHS of Equation 6.1 can be simplified to $\exp\left(\int_{g^{-1}}^{1} \alpha_{(zd,zu)} \frac{1}{z} dz\right)$, where, however, the interpretation as a weighted mean does not work, while in the original version we see quickly that $w(z)$ integrates to 1 on the interval $\left(g^{-1}, 1\right)$. Notice that in case $g < 1$, to preserve the weight interpretation, one has to interchange the interval limits and work with $-w(x)$. Anyway, we can proof the theorem for any $g > 0$ without having to look at positive and negative inflation rates separately.

At first we show that for any fixed $0 \leq d \leq \infty$ the function in g

$$\int_0^{\frac{d}{g}} \overline{F}_Z(x)\, dx$$

is Lipschitz continuous on all closed subintervals of $(0, \infty)$. We leave the trivial cases of d equaling 0 or infinite aside. Let be $g_1 > g_2 \geq min > 0$, i.e., we relate two arbitrary points in an arbitrary closed interval with minimum min. Then we have

$$\int_0^{\frac{d}{g_2}} \overline{F}(x)\, dx - \int_0^{\frac{d}{g_1}} \overline{F}(x)\, dx = \int_{\frac{d}{g_1}}^{\frac{d}{g_2}} \overline{F}(x)\, dx \leq \left(\frac{d}{g_2} - \frac{d}{g_1}\right) = \frac{d\,(g_1 - g_2)}{g_2 g_1} \leq (g_1 - g_2)\, \frac{d}{min^2}$$

with the Lipschitz constant $\frac{d}{min^2}$.

Thus, said function is also absolutely continuous on closed subintervals of $(0, \infty)$, which means it is locally absolutely continuous, see e.g. [Leoni, 2009]. The latter property is inherited by

$$\Psi_d^u(g) := E\left({}_d^u(gZ)\right) = \int_d^u \overline{F}_Z\left(\frac{x}{g}\right) dx = g \int_{\frac{d}{g}}^{\frac{u}{g}} \overline{F}(x)\, dx = g \left(\int_0^{\frac{u}{g}} \overline{F}(x)\, dx - \int_0^{\frac{d}{g}} \overline{F}(x)\, dx \right)$$

such that we can apply the fundamental theorem of calculus to this function, and also to its logarithm:

$$\ln\left(\frac{\mathrm{E}\left({}_d^u(gZ)\right)}{\mathrm{E}\left({}_d^u Z\right)}\right) = \ln \Psi_d^u(g) - \ln \Psi_d^u(1) = \int_1^g \frac{\left(\Psi_d^u(z)\right)'}{\Psi_d^u(z)} dz$$

Here we can use the right or left derivative. Selecting the latter and applying product rule and chain rule for derivatives, we get for finite u

$$\begin{aligned}
\left(\Psi_d^u(z)\right)' = \left(z \int_{\frac{d}{z}}^{\frac{u}{z}} \overline{F}(x)\, dx \right)' &= \int_{\frac{d}{z}}^{\frac{u}{z}} \overline{F}(x)\, dx + z \left[\left(-\frac{u}{z^2}\right) \overline{F}\left(\frac{u}{z}\right) - \left(-\frac{d}{z^2}\right) \overline{F}\left(\frac{d}{z}\right) \right] \\
&= \int_{\frac{d}{z}}^{\frac{u}{z}} \overline{F}(x)\, dx + \frac{d}{z} \overline{F}\left(\frac{d}{z}\right) - \frac{u}{z} \overline{F}\left(\frac{u}{z}\right)
\end{aligned}$$

If we recall the definition of the regional Pareto alpha we get with $\Psi_d^u(z) = z \int_{\frac{d}{z}}^{\frac{u}{z}} \overline{F}(x)\,dx$ that

$$\ln\left(\frac{\mathrm{E}\left({}_d^u(gZ)\right)}{\mathrm{E}\left({}_d^uZ\right)}\right) = \int_1^g \frac{1}{z}\alpha_{\left(\frac{d}{z},\frac{u}{z}\right)}dz = \int_{g^{-1}}^1 \alpha_{(zd,zu)}\frac{1}{z}dz$$

where the last step follows by substitution.

The case $u = \infty$ is analogous, just the terms with $\overline{F}\left(\frac{u}{z}\right)$ are missing here. We are done, noting finally that, had we used the right derivative, this would have led via the left-sided limits of values of $\overline{F}$ to the analogous average of the left-sided regional alphas. □

A few remarks:

- The theorem includes pro-rata coverages, where all regional alphas equal 1.
- From the weights $\frac{1}{z\ln g}$ assigned to the $\alpha_{(zd,zu)}$ in the formula for $\overline{\alpha}$ we see that the lower intervals get somewhat more weight than the higher intervals.
- If the local alphas exist in the interval ranging from $\min\left(d, \frac{d}{g}\right)$ to $\max\left(u, \frac{u}{g}\right)$, at least to the extent that the fundamental theorem of calculus can be applied, then $\overline{\alpha}$ can be written as a (somewhat intricate) average of the local alphas in said interval:

$$\overline{\alpha} = \int_{g^{-1}}^1 \int_{zd}^{zu} \alpha_x r_z(x)\, w(z)\, dxdz$$

where

$$r_z(x) = \frac{\overline{F}(x)}{\int_{zd}^{zu}\overline{F}(y)\,dy}$$

- In case of a continuous pdf, we can say as above that if the local alphas vary few, $\overline{\alpha}$ must be located within their small range of values. However, here we have a much stronger result. The local alphas may vary a lot across the interval from $\min\left(d, \frac{d}{g}\right)$ to $\max\left(u, \frac{u}{g}\right)$, however, we can expect the regional alphas to vary much less as they already are averages, each over its interval (zd, zu). As long as g is not very far from 1, these intervals intersect a lot, so we can expect that the regional alphas entering the formula are often within a rather narrow range. Thus, although $\overline{\alpha}$ depends on g, this dependence should be very weak unless $\overline{F}$ has a very rough geometry. Even if the cdf is not differentiable, such that (some) local alphas do not exist, we would typically expect the regional alphas to vary not so much across the scaled intervals – the way they depend on attachment point, detachment point, and average severity of the layer prevents regional alphas from changing sharply unless there are high jumps in the cdf.
- If g is close to 1, we can well approximate $\overline{\alpha}$ by the regional Pareto alpha $\alpha_{(d,u)}$ of the layer. We do not need to do further calculus to see this, just look at $\alpha_{(zd,zu)}$ as a function of z. However, this function is only continuous at $z = 1$ if $\overline{F}$ is continuous at d and u. Otherwise we have different limits from the left and from the right. In this case we have to distinguish between positive $(g > 1)$ and negative $(g < 1)$ inflation rates. In case $g > 1$ the correct approximation for $\overline{\alpha}$ is notably $\alpha_{(d,u)-}$, while for $g < 1$ it is $\alpha_{(d,u)}$.

Corollary 6.5. *If g is close to 1 and $\overline{F}$ is continuous at d and u, we have*

$$\frac{E\left({}_d^u(gZ)\right)}{E\left({}_d^uZ\right)} \approx g^{\alpha_{(d,u)}} \tag{6.2}$$

All in all we have found an elegant and intuitive formula for the leverage of inflation to layers, working for arbitrary loss distributions, together with an approximation for small inflation rates that holds under a mild regularity condition, namely if the cdf is continuous at both the attachment and the detachment point of the layer. The above remarks make furthermore clear that Formula 6.2 is much more than just a convenient first order approximation – it can usually be expected to be fair even for values g being somewhat farther from 1.

Recall we did not need the cdf to have a density. However, in case the cdf is (somewhat) smooth, we can quantify the effects of inflation on frequency and severity of the layer separately, in an equally elegant way as done with the expected layer loss:

Definition 6.6. We will speak of *frequency inflation* and of *severity inflation* to quantify how a ground-up inflation factor g (multiplicatively) affects frequency and severity.

The effect of inflation on frequency can be seen easily if we replace the excess loss ${}^{u}_{d}Z$ by the function counting the losses that hit the layer, i.e. the indicator function $\chi_{(Z>d)}$.

Proposition 6.7. *For any locally absolutely continuous loss distribution function F_Z and any threshold $d>0$ we have that a ground-up inflation factor $g>0$ leads to the following frequency inflation factor*

$$\frac{E\left(\chi_{(gZ>d)}\right)}{E\left(\chi_{(Z>d)}\right)}=\frac{\overline{F}_{gZ}\left(d\right)}{\overline{F}_{Z}\left(d\right)}=g^{\widetilde{\alpha}}$$

where $\widetilde{\alpha}$ is a weighted mean of local Pareto alphas

$$\widetilde{\alpha}=\int_{g^{-1}}^{1}\alpha_{zd}w\left(z\right)dz$$

with the weighting function

$$w\left(z\right)=\frac{1}{z\ln g}$$

The formula holds as well if the local Pareto alphas are replaced by their left-sided counterparts.

Proof. The proof is analogous to that of the preceding theorem. The RHS of the first equation can be simplified to $\exp\left(\int_{g^{-1}}^{1}\alpha_{zd}\frac{1}{z}dz\right)$.

As $\Phi_d\left(g\right):=\overline{F}_{gZ}\left(d\right)=\overline{F}_{Z}\left(\frac{d}{g}\right)$ is locally absolutely continuous, we can apply the fundamental theorem of calculus:

$$\ln\left(\frac{\overline{F}_{gZ}\left(d\right)}{\overline{F}_{Z}\left(d\right)}\right)=\ln\Phi_d\left(g\right)-\ln\Phi_d\left(1\right)=\int_{1}^{g}\frac{\left(\Phi_d\left(z\right)\right)'}{\Phi_d\left(z\right)}dz=\int_{1}^{g}\frac{\frac{d}{z}f_Z\left(\frac{d}{z}\right)}{\overline{F}_Z\left(\frac{d}{z}\right)}\frac{1}{z}dz=\int_{1}^{g}\alpha_{\frac{d}{z}}\frac{1}{z}dz=\int_{g^{-1}}^{1}\alpha_{zd}\frac{1}{z}dz$$

where the last step follows by substitution. According to whether we use the left or the right derivative of the cdf, we involve the local Pareto alphas or their left-sided counterparts. □

From the proof we see that if the inflation rate is small (i.e. g close to 1) and if the pdf has limits from the left and the right at d, then $\widetilde{\alpha}$ approximately equals α_d for $g<1$ and α_{d-} for $g>1$.

Corollary 6.8. *If g is close to 1 and $\overline{F}_Z$ is differentiable at $d>0$, we have*

$$\frac{E\left(\chi_{(gZ>d)}\right)}{E\left(\chi_{(Z>d)}\right)}=\frac{\overline{F}_{gZ}\left(d\right)}{\overline{F}_{Z}\left(d\right)}\approx g^{\alpha_d} \tag{6.3}$$

This does notably not require that $\overline{F}_Z$ be locally absolutely continuous, apart from the stated regularity at d.

Proof. If the survival function is locally absolutely continuous, the formula follows immediately from the proposition, analogously to the preceding results for layer loss inflation. If more generally $\overline{F}_Z$ is differentiable (possibly only) at d, then $f_Z(d)$ and α_d are well defined and the function $\overline{F}_{(1+h)Z}(d) = \overline{F}_Z\left(\frac{d}{1+h}\right)$ is differentiable at $h = 0$. This means that $\frac{1}{h}\left(\overline{F}_Z\left(\frac{d}{1+h}\right) - \overline{F}_Z(d)\right)$ tends to $d\,f_Z(d)$ as h approaches 0. Thus, for small h we have

$$\frac{\overline{F}_Z\left(\frac{d}{1+h}\right)}{\overline{F}_Z(d)} \approx 1 + h\frac{d\,f_Z(d)}{\overline{F}_Z(d)} = 1 + h\alpha_d \approx (1+h)^{\alpha_d}$$

$\square$

Corollary 6.9. *For any locally absolutely continuous loss distribution function F and any layer u–d xs d, $d > 0$, including unlimited layers in case of finite expectation, we have that a ground-up inflation factor $g > 0$ leads to the severity inflation factor $g^{\overline{\alpha}-\widetilde{\alpha}}$.*

Proof. Severity is expected loss divided by frequency, which leads to the quotient of the two powers of g that appear in the preceding theorem and proposition, respectively. $\square$

This implies that if the local alphas vary few in the neighborhood of a layer (tail shape similar to Pareto), the severity inflation factor must be very close to 1, i.e. inflation hardly affects the layer severity.

Now recall that in the Pareto case with $g > 1$ (positive inflation rate), whether the layer inflation is greater / equal / smaller than the ground-up inflation, depends on whether the *global* alpha is greater / equal / smaller than 1. In the general case this transforms into a local or merely regional property:

- For the *local* alpha the connection to Pareto curves is obvious, as it is by definition the parameter of a local first-order approximation of the cdf with a Pareto curve. If e.g. a distribution has local alphas greater than 1 for a certain range of loss sizes, the leverage of a ground-up inflation $g > 1$ will be greater than 1 for frequencies and expected layer losses in that area, no matter whether the distribution function looks very similar to a Pareto curve or not.

- More generally, even where local alphas do not exist, there is an analogous rule for the *regional* alpha. If e.g. a distribution function has regional alphas greater than 1 for a certain range of loss sizes, the leverage of a ground-up inflation $g > 1$ will be greater than 1 for layers in that area, no matter whether the distribution function looks very similar to a Pareto curve or not. As for the connection with Pareto curves, if we look at the definition of the regional alpha, we see that whether it is greater, equal, or smaller than 1, depends on the sign of the term $d\overline{F}(d) - u\overline{F}(u)$, which is positive / zero / negative iff the cdf between d and u decays faster than / exactly as / more slowly than a Pareto distribution with $\alpha = 1$.

Like the global Pareto alpha, the regional and local alphas can be seen as measures for heavy-tailedness, with small alphas meaning *heavy* and large alphas *light*. At the same time these alphas emerge here as *the* crucial measure for the impact of ground-up inflation on layer inflation.

Ironically these two perspectives yield opposite perceptions of riskiness: In the traditional view heavy tails are the “dangerous” ones, involving high fluctuations and uncertainty. From this perspective the low alphas are the “bad” ones. If we focus on inflation leverage instead, the high alphas become the dangerous ones, creating the highest leverage, which may involve high fluctuations and uncertainty. Summing up, the perception of “bad” distribution tails becomes ambiguous here, such that the final judgment of what is dangerous, depends on how much weight is given to randomness due to inflation vs randomness from other sources.

6.4 Leverage on second moment

MSE calculations require variances, so we look at the squared layer loss.

Lemma 6.10. *For any loss Z and any layer $u - d\,xs\,d$ applied to it we have*

$$\left({}_d^u Z\right)^2 = {}_{d^2}^{u^2}\left(Z^2\right) - 2d\,{}_d^u Z$$

Proof. Simply check the three cases of Z being below, within, or beyond the layer area. □

Proposition 6.11. *For any loss distribution function F and any layer $u - d\,xs\,d$, including unlimited layers in case of finite expectation of Z and Z^2, we have that a ground-up inflation factor $g > 0$ leads to*

$$E\left(\left({}_d^u (gZ)\right)^2\right) = E\left(\left({}_d^u Z\right)^2\right) g^{2\breve{\alpha}} + 2dE\left({}_d^u Z\right)\left(g^{2\breve{\alpha}} - g^{\overline{\alpha}}\right)$$

$$1 + CV^2\left({}_d^u (gZ)\right) = \left\{\left(1 + CV^2\left({}_d^u Z\right)\right) g^{2\breve{\alpha}} + \frac{2d}{E\left({}_d^u Z\right)}\left(g^{2\breve{\alpha}} - g^{\overline{\alpha}}\right)\right\} g^{-2\overline{\alpha}}$$

where $\breve{\alpha}$ is the weighted mean of regional Pareto alphas analogous to $\overline{\alpha}$, but for Z^2 and the layer between d^2 and u^2.

Proof. Applying the preceding lemma twice we get with the help of Formula 6.1

$$\begin{aligned} \mathrm{E}\left(\left({}_d^u (gZ)\right)^2\right) &= \mathrm{E}\left({}_{d^2}^{u^2}\left(g^2 Z^2\right)\right) - 2d\mathrm{E}\left({}_d^u (gZ)\right) = \mathrm{E}\left({}_{d^2}^{u^2}\left(Z^2\right)\right)\left(g^2\right)^{\breve{\alpha}} - 2d\mathrm{E}\left({}_d^u Z\right) g^{\overline{\alpha}} \\ &= \left[\mathrm{E}\left(\left({}_d^u Z\right)^2\right) + 2d\mathrm{E}\left({}_d^u Z\right)\right] g^{2\breve{\alpha}} - 2d\mathrm{E}\left({}_d^u Z\right) g^{\overline{\alpha}} = \mathrm{E}\left(\left({}_d^u Z\right)^2\right) g^{2\breve{\alpha}} + 2d\mathrm{E}\left({}_d^u Z\right)\left(g^{2\breve{\alpha}} - g^{\overline{\alpha}}\right) \end{aligned}$$

The CV formula follows immediately if we divide by $\mathrm{E}^2\left({}_d^u (gZ)\right)$ and use again Formula 6.1. □

Notice that this proposition includes proportional covers, where all regional alphas equal one.

Again, for g being close to 1, we can approximate the average regional alphas by exact regional alphas being independent of g. For $\breve{\alpha}$ this would be the regional alpha of Z^2 on the interval $\left(d^2, u^2\right)$, which equals

$$\alpha^*_{(d,u)} := 1 + \frac{d^2 \overline{F}_{Z^2}\left(d^2\right) - u^2 \overline{F}_{Z^2}\left(u^2\right)}{\mathrm{E}\left({}_{d^2}^{u^2}\left(Z^2\right)\right)} = 1 + \frac{d^2 \overline{F}_Z(d) - u^2 \overline{F}_Z(u)}{\mathrm{E}\left({}_{d^2}^{u^2}\left(Z^2\right)\right)}$$

Here we have used $\overline{F}_{Z^2}(x) = \overline{F}_Z\left(\sqrt{x}\right)$. If $\overline{F}_Z$ or equivalently $\overline{F}_{Z^2}$ is locally absolutely continuous, this leads over $f_{Z^2}(x) = f_Z\left(\sqrt{x}\right)\frac{1}{2\sqrt{x}}$ to

$$\alpha_{Z^2}(x) = \frac{x f_{Z^2}(x)}{\overline{F}_{Z^2}(x)} = \frac{x f_Z\left(\sqrt{x}\right)}{2\sqrt{x}\,\overline{F}_Z\left(\sqrt{x}\right)} = \frac{1}{2}\frac{\sqrt{x} f_Z\left(\sqrt{x}\right)}{\overline{F}_Z\left(\sqrt{x}\right)} = \frac{1}{2}\alpha_Z\left(\sqrt{x}\right)$$

such that we can rewrite $\alpha^*_{(d,u)}$ as follows:

$$\alpha^*_{(d,u)} = \frac{\int_{d^2}^{u^2} \overline{F}_{Z^2}(x)\,\alpha_{Z^2}(x)\,dx}{\int_{d^2}^{u^2} \overline{F}_{Z^2}(x)\,dx} = \frac{1}{2}\frac{\int_{d^2}^{u^2} \overline{F}_Z\left(\sqrt{x}\right)\alpha_Z\left(\sqrt{x}\right)dx}{\int_{d^2}^{u^2} \overline{F}_Z\left(\sqrt{x}\right)dx} = \frac{1}{2}\frac{\int_d^u z\overline{F}_Z(z)\,\alpha_Z(z)\,dz}{\int_d^u z\overline{F}_Z(z)\,dz}$$

The last term follows by substitution and shows that $2\alpha^*_{(d,u)}$ can be written as a weighted average of the local alphas of Z (not of Z^2), however, with a different weighting function than that of $\alpha_{(d,u)}$, namely $z\overline{F}_Z(z)$ instead of $\overline{F}_Z(z)$, such that local alphas corresponding to larger loss sizes get higher weight. This motivates the following definition for arbitrary cdf's:

Definition 6.12. For a loss severity Z we call

$$\overset{\circ}{\alpha}_{(d,u)} := 2\left(1+\frac{d^2\overline{F}_Z(d)-u^2\overline{F}_Z(u)}{\mathrm{E}\left({}^{u^2}_{d^2}(Z^2)\right)}\right) = 2\alpha^*_{(d,u)}$$

the *second-moment regional Pareto alpha* of interval and layer between d and u.

The factor 2 ensures that for the Pareto distribution both regional alphas coincide. Recall they generally are different weighted averages of the local alphas (provided the latter exist) over the same interval. If the local alphas increase / decrease with loss size, then the second-moment regional alpha is greater / smaller than the regional alpha.

Now we can state the approximation for g being close to 1 in a compact manner.

Corollary 6.13. *If g is close to 1 and $\overline{F}$ is continuous at d and u, we have*

$$E\left(\left({}^{u}_{d}(gZ)\right)^2\right) \approx E\left(\left({}^{u}_{d}Z\right)^2\right) g^{\overset{\circ}{\alpha}_{(d,u)}} + 2dE\left({}^{u}_{d}Z\right)\left[g^{\overset{\circ}{\alpha}_{(d,u)}} - g^{\alpha_{(d,u)}}\right]$$

$$1+CV^2\left({}^{u}_{d}(gZ)\right) \approx \left\{\left(1+CV^2\left({}^{u}_{d}Z\right)\right) g^{\overset{\circ}{\alpha}_{(d,u)}} + \frac{2d}{E\left({}^{u}_{d}Z\right)}\left[g^{\overset{\circ}{\alpha}_{(d,u)}} - g^{\alpha_{(d,u)}}\right]\right\} g^{-2\alpha_{(d,u)}} \tag{6.4}$$

For $d>0$ the RHS of the two formulae may become negative, making them bad approximations for the strictly positive term on the respective LHS. We can, however, improve these approximations, sometimes even greatly, by appending a strictly positive lower bound. Let us detail this for the CV formula, being the one we will need in the following.

Proposition 6.14. *If g is close to 1 and $\overline{F}$ is continuous at d and u, we have, (slightly) improving the approximation provided by Formula 6.4:*

$$1+CV^2\left({}^{u}_{d}(gZ)\right) \approx \max\left(\left\{\left(1+CV^2\left({}^{u}_{d}Z\right)\right) g^{\overset{\circ}{\alpha}_{(d,u)}} + \frac{2d}{E\left({}^{u}_{d}Z\right)}\left[g^{\overset{\circ}{\alpha}_{(d,u)}} - g^{\alpha_{(d,u)}}\right]\right\} g^{-2\alpha_{(d,u)}}, 1\right)$$

If additionally $\overline{F}$ is differentiable at $d>0$, we have, (mostly much) improving the approximation further:

$$1+CV^2\left({}^{u}_{d}(gZ)\right) \approx \max\left(\left\{\left(1+CV^2\left({}^{u}_{d}Z\right)\right) g^{\overset{\circ}{\alpha}_{(d,u)}} + \frac{2d}{E\left({}^{u}_{d}Z\right)}\left[g^{\overset{\circ}{\alpha}_{(d,u)}} - g^{\alpha_{(d,u)}}\right]\right\} g^{-2\alpha_{(d,u)}}, \frac{1}{\overline{F}_Z(d)} g^{-\alpha_d}\right)$$

Proof. The first formula is simply a combination of Formula 6.4 with the (trivial, but general and exact) inequality $1+\mathrm{CV}^2\left({}^{u}_{d}(gZ)\right) \geq 1$, which is usually far from sharp: For large d the excess loss ${}^{u}_{d}(gZ)$ equals 0 with high probability, which makes it a highly volatile RV.

However, we can take advantage of this point, by using the inequality shown in Lemma 5.5. We get

$$1+\mathrm{CV}^2\left({}^{u}_{d}(gZ)\right) \geq \frac{1}{\mathrm{E}\left(\chi_{(gZ>d)}\right)} \approx \frac{g^{-\alpha_d}}{\overline{F}_Z(d)}$$

where the last step follows from Formula 6.3. This approximative inequality may not be sharp either, but for high d (and thus low $\overline{F}_Z(d)$) it is by far better than the first one. If we combine it with Formula 6.4, we get the second claimed formula. □

We have examined in detail the squared expected loss $\left({}^{u}_{d}(gZ)\right)^2$, which involved new quantities and notation, in addition to what was needed to analyze ${}^{u}_{d}(gZ)$. Luckily, for the squared loss count there is nothing analogous to do, as $\chi^2_{(gZ>d)} = \chi_{(gZ>d)}$. In particular, there is no need to look for a lower bound for the approximative variance/CV formulae of the loss count.

Remark 6.15. To make clearer how closely aggregate loss and loss count of a layer are connected, let us again go back to Formula 6.4

$$1+\mathrm{CV}^2\left({}^u_d\left(gZ\right)\right)\approx\left\{\left(1+\mathrm{CV}^2\left({}^u_dZ\right)\right)g^{\mathring{\alpha}_{(d,u)}}+\frac{2d}{\mathrm{E}\left({}^u_dZ\right)}\left[g^{\mathring{\alpha}_{(d,u)}}-g^{\alpha_{(d,u)}}\right]\right\}g^{-2\alpha_{(d,u)}}$$

Assume now that the layer is extremely thin, i.e. u is only a bit larger than $d>0$. Assume further that $\overline{F}$ is differentiable at d, such that the regularity conditions for the approximations of both layer losses and respective counts are fulfilled. Then one sees immediately that

$${}^u_dZ\approx\left(u-d\right)\chi_{(Z>d)},\qquad \mathrm{E}\left({}^u_dZ\right)\approx\left(u-d\right)\overline{F}_Z\left(d\right)$$

and infers quickly from their definitions that both regional alphas $\alpha_{(d,u)}$, $\mathring{\alpha}_{(d,u)}$ for such thin layers are close to α_d. Via lengthy but straightforward analysis it can be shown more strongly, at least for sufficiently smooth $\overline{F}$, that if u tends to d, on the RHS of Formula 6.4 the summand about the difference in square brackets vanishes, yielding

$$\left[\left(1+\mathrm{CV}^2\left(\left(u-d\right)\chi_{(Z>d)}\right)\right)g^{\alpha_d}\right]g^{-2\alpha_d}=\left(1+\mathrm{CV}^2\left(\chi_{(Z>d)}\right)\right)g^{-\alpha_d}=\frac{1}{\overline{F}_Z\left(d\right)}g^{-\alpha_d}$$

which is exactly the lower bound in the second (and better) approximation in Proposition 6.14, such that now both components of the maximum coincide. As the LHS of Formula 6.4 analogously tends to

$$1+\mathrm{CV}^2\left(\left(u-d\right)\chi_{(gZ>d)}\right)=\frac{1}{\overline{F}_{gZ}\left(d\right)}$$

we altogether get $\overline{F}_{gZ}\left(d\right)\approx\overline{F}_Z\left(d\right)g^{\alpha_d}$, which is Formula 6.3. We could say that, by “squeezing” the layer to the deductible, both Formula 6.4 and its second amendment “converge” to the same approximative formula for the excess loss count.

This remark emphasizes that the mathematics we are developing for aggregate loss and loss count of a layer, respectively, are deeply analogous. We will come back to this analogy various times throughout this book.

6.5 Examples: GPD variants

To illustrate the theory developed here, we explore local and regional alphas for the tail model classic: the *Generalized Pareto Distribution* (GPD). On the way we introduce less common parameterizations, which turn out to be very intuitive for the understanding of tail-behavior and the impact of inflation. By proceeding exactly as done with the Pareto model at the beginning of this chapter, we reveal a lot of analogies.

The GPD distribution function starts at a known threshold s, which is usually positive, but we will see that in some cases it can be 0, leading to ground-up models. The survival function in its common form, see e.g. [Embrechts et al., 2013], reads

$$\overline{F}_Z\left(x\mid Z>s\right)=\left(1+\xi\frac{x-s}{\sigma}\right)^{-\frac{1}{\xi}}$$

where the shape parameter ξ may be any real number, while the scale parameter σ is positive. (For $\xi=0$ take the limit, easily found via L’Hôpital’s rule.) For negative ξ the distribution is limited by the loss supremum $s+\frac{\sigma}{-\xi}$.

Now replace σ by the so-called *modified scale* $\omega := \sigma - \xi s > -\xi s$, see e.g. [Scarrott and MacDonald, 2012]. This complicates the parameter space a bit, but greatly simplifies the survival function:

$$\overline{F}_Z\left(x \mid Z > s\right) = \left(\frac{\omega + \xi s}{\omega + \xi x}\right)^{\frac{1}{\xi}}$$

or equivalently

$$\overline{F}_Z\left(x\right) = \overline{F}_Z\left(s\right)\left(\frac{\omega + \xi s}{\omega + \xi x}\right)^{\frac{1}{\xi}}, \quad x > s$$

This representation enables very quick calculations. For thresholds $t_1, t_2 \geq s$ we see at once that the ratio of the loss frequencies equals

$$\frac{\overline{F}_Z\left(t_1\right)}{\overline{F}_Z\left(t_2\right)} = \left(\frac{\omega + \xi t_2}{\omega + \xi t_1}\right)^{\frac{1}{\xi}}$$

which we could call the *GPD extrapolation formula*.

If $d \geq s$, we see immediately that the tail starting at d is distributed as follows:

$$\overline{F}_Z\left(x \mid Z > d\right) = \frac{\overline{F}_Z\left(x\right)}{\overline{F}_Z\left(d\right)} = \left(\frac{\omega + \xi d}{\omega + \xi x}\right)^{\frac{1}{\xi}}$$

This is again GPD with the same parameters ξ and ω: One simply has to replace the original threshold by the new, higher one. One could call this a *memoryless property* of the GPD distribution – the original threshold is "forgotten". Note that the traditional scale parameter σ is far from invariant when calculating a higher tail, transforming in an intricate manner.

Let Z^* be the same kind of loss as Z, say a number of years later, after uniform *ground-up inflation* via a factor $g > 0$. Then Z^* is distributed like gZ and we have

$$\overline{F}_{Z^*}\left(x\right) = \overline{F}_Z\left(\frac{x}{g}\right) = \overline{F}_Z\left(s\right)\left(\frac{\omega + \xi s}{\omega + \xi \frac{x}{g}}\right)^{\frac{1}{\xi}} = \overline{F}_{Z^*}\left(gs\right)\left(\frac{g\omega + \xi gs}{g\omega + \xi x}\right)^{\frac{1}{\xi}}$$

or equivalently

$$\overline{F}_{Z^*}\left(x \mid Z^* > gs\right) = \left(\frac{g\omega + \xi gs}{g\omega + \xi x}\right)^{\frac{1}{\xi}}$$

which is again GPD with the same exponent, while ω and the threshold are "inflated" by g. However, if $d \geq gs$, we get for the higher tail as before

$$\overline{F}_{Z^*}\left(x \mid Z^* > d\right) = \left(\frac{g\omega + \xi d}{g\omega + \xi x}\right)^{\frac{1}{\xi}}$$

where the only change from pre to post inflation is the replacement of ω by $g\omega$.

Everything else invariant, the change of the frequency between the situations before and after inflation, respectively, equals

$$\frac{\overline{F}_{Z^*}\left(d\right)}{\overline{F}_Z\left(d\right)} = \frac{\left(\frac{g\omega + \xi gs}{g\omega + \xi d}\right)^{\frac{1}{\xi}}}{\left(\frac{\omega + \xi s}{\omega + \xi d}\right)^{\frac{1}{\xi}}} = \left(\frac{g\omega + \xi gd}{g\omega + \xi d}\right)^{\frac{1}{\xi}}$$

Now we go through the three types of the GPD, being each an asymptotic tail model for the distributions belonging to the maximum domain of attraction of the respective extreme value distributions (see e.g. [Embrechts et al., 2013]).

6.5.1 Exponential (Gumbel)

If $\xi = 0$, we have $\omega = \sigma$ and get

$$\overline{F}_Z(x \mid Z > s) = \exp\left(-\frac{x-s}{\sigma}\right) = \exp\left(-\frac{x-s}{\omega}\right)$$

which means that we have a distribution with Exponential tail or, as $s = 0$ is possible, an Exponential distribution from the ground up. The memoryless property, which for this case (but interestingly not for the other GPD cases) is omnipresent in the literature, reads

$$\overline{F}_Z(x \mid Z > d) = \exp\left(-\frac{x-d}{\omega}\right)$$

After inflation we have $\overline{F}_{Z^*}(x \mid Z^* > gs) = \exp\left(-\frac{x-gs}{g\omega}\right)$ and $\overline{F}_{Z^*}(x \mid Z^* > d) = \exp\left(-\frac{x-d}{g\omega}\right)$. Inflation impacts linear on the only parameter of this model, which equals ω before inflation and $g\omega$ after, as well as on the average loss of the unlimited layer $\infty\, xs\, d$, which is well known to equal the Exponential parameter.

As for the frequencies, we get $\frac{\overline{F}_Z(t_1)}{\overline{F}_Z(t_2)} = \exp\left(\frac{t_2-t_1}{\omega}\right)$ and $\frac{\overline{F}_{Z^*}(d)}{\overline{F}_Z(d)} = \exp\left(\frac{d}{\omega}\left(1-\frac{1}{g}\right)\right)$.

Now we calculate the alphas. Up to a constant, we have $\ln \overline{F}(x) = -\frac{x}{\omega}$ for any point $x > s$ in the tail, hence

$$\alpha_x = \frac{x}{\omega}$$

The local alpha grows linear in x, thus changes rapidly and tends to infinity. That means that although Exponential is a (moderately) heavy-tail model, approximations with Pareto curves will only work well on rather small intervals.

For layers with $d \geq s$ we have $\mathrm{E}\left({}_d^u Z \mid Z > d\right) = \omega\left(1 - \exp\left(-\frac{u-d}{\omega}\right)\right)$, hence

$$\alpha_{(d,u)} = 1 + \frac{d - u\exp\left(-\frac{u-d}{\omega}\right)}{\omega\left(1-\exp\left(-\frac{u-d}{\omega}\right)\right)}$$

For unlimited layers this equals $1 + \frac{d}{\omega}$, such that $\alpha_{(d,\infty)} - \alpha_d$, the relevant exponent for the approximate severity inflation, equals 1, which means linear impact. Thus, for the severity the approximate impact of inflation equals the exact one.

Let us compare, for unlimited layers only, the approximate layer loss inflation to the exact one. Formula 6.2 becomes

$$g\exp\left(\frac{d}{\omega}\ln g\right) = g^{1+\frac{d}{\omega}} \approx \frac{\mathrm{E}\left({}_d^\infty (gZ)\right)}{\mathrm{E}\left({}_d^\infty Z\right)} = \frac{\overline{F}_{gZ}(d)}{\overline{F}_Z(d)}\frac{g\omega}{\omega} = g\exp\left(\frac{d}{\omega}\left(1-\frac{1}{g}\right)\right)$$

Here the precision of the approximative layer loss inflation depends on the precision of the approximation

$$\ln g \approx 1 - \frac{1}{g}$$

Both functions coincide at $g = 1$, having there the same slope; elsewhere $\ln g$, the approximation, is larger. Overall the approximation due to Formula 6.2 looks fair for the Exponential distribution, but is not excellent if g deviates considerably from 1 and the quotient of d and ω is large.

For the layer frequency the quality of the approximation is the same – recall that for the severity approximation and exact formula coincide.

The following table displays the deviation of the approximate from the exact layer loss inflation in percent. $h = g - 1$ is entered in percent, too.

$d/\omega \setminus h$	-20	-10	-2	0	2	10	20
0,2	0.5	0.1	0	0	0	0.1	0.3
1	2.7	0.6	0	0	0	0.4	1.6
10	30.8	5.9	0.2	0	0.2	4.5	16.9

The second-order calculation comparing the approximate to the exact formula for $\mathrm{E}\left({}_d^\infty\left((gZ)^2\right)\right)$ is analogous and yields similar results.

6.5.2 Proper GPD (Fréchet)

We call the case $\xi > 0$, allowing for arbitrarily large losses, *proper GPD*. It is closest to Pareto, which can be seen even better by another, closely related, parameter variant.

Define $\alpha := \frac{1}{\xi}$ and $\lambda := \frac{\omega}{\xi} = \alpha\sigma - s > -s$, see [Scollnik, 2007]. Now the GPD formulae look almost like Pareto, just λ comes in:

$$\overline{F}_Z\left(x \mid Z > s\right) = \left(\frac{\lambda + s}{\lambda + x}\right)^\alpha$$

or equivalently

$$\overline{F}_Z\left(x\right) = \overline{F}_Z\left(s\right)\left(\frac{\lambda + s}{\lambda + x}\right)^\alpha, \quad x > s$$

The two parameters can be interpreted in the following way: Start with a Simple Pareto distribution having parameter α and threshold $\lambda + s > 0$, then shift the x-axis by λ to the left (by subtracting λ from all losses), and you obtain the proper GPD. Via this perspective a lot of results, e.g. the moments of layer losses, can be deduced easily from the Pareto case without any calculations.

Note that if λ is positive, s may be 0, such that we have a ground-up model. This special case has various names, including Pareto (creating confusion with Simple Pareto). Unequivocal names are *Type II*, *Two-parameter*, *American*, or *Ballasted Pareto*, as well as *Lomax*.

$\lambda = 0$ is the Pareto case, which only works for proper tails starting at an $s > 0$. The same holds for negative λ, where we more strongly must have $s > |\lambda|$.

The GPD extrapolation formula reads $\frac{\overline{F}_Z(t_1)}{\overline{F}_Z(t_2)} = \left(\frac{\lambda + t_2}{\lambda + t_1}\right)^\alpha$. If $d \geq s$, we have the memoryless property and the invariance of the parameters α and λ to upward conditioning (higher tails):

$$\overline{F}_Z\left(x \mid Z > d\right) = \frac{\overline{F}_Z\left(x\right)}{\overline{F}_Z\left(d\right)} = \left(\frac{\lambda + d}{\lambda + x}\right)^\alpha$$

As for inflation, if Z^* is distributed like gZ, we have $\overline{F}_{Z^*}\left(x\right) = \overline{F}_{Z^*}\left(gs\right)\left(\frac{g\lambda + gs}{g\lambda + x}\right)^\alpha$ or equivalently $\overline{F}_{Z^*}\left(x \mid Z^* > gs\right) = \left(\frac{g\lambda + gs}{g\lambda + x}\right)^\alpha$ with the same exponent as before, while λ and the threshold are inflated. However, if $d \geq gs$, we get for the higher tail

$$\overline{F}_{Z^*}\left(x \mid Z^* > d\right) = \left(\frac{g\lambda + d}{g\lambda + x}\right)^\alpha$$

That means, if the retention d of the layer is such high that it exceeds the Pareto threshold both before and after inflation, the distribution of the upper tail starting at d is only slightly changed due to inflation: As a consequence of the stated memoryless property only λ changes, and if d is large compared to $|\lambda|$, the impact of this change on the tail shape is such low that in practice it will be hard to detect among the random fluctuations of the empirical tail. In other words, for high layers inflation is very hard to infer from tail data only, as long as the distribution is proper-GPD-like.

Inflation affects the expected loss to the layer much more via the *frequency*, such that the respective changes in frequency and in expected loss are similar.

Assume at first that the volume of the risk producing the loss Z is constant. Then the change of the frequency between the situations before and after inflation, respectively, equals

$$\frac{\overline{F}_{Z^*}(d)}{\overline{F}_Z(d)} = \left(\frac{g\lambda + gd}{g\lambda + d}\right)^{\alpha} = g^{\alpha}\left(\frac{\lambda + d}{g\lambda + d}\right)^{\alpha}$$

which for large d is very close to g^{α}.

If the volume is variable but known, the stated effects hold for the frequency per volume unit. To get the overall outcome one has to combine the effects from inflation and from volume change.

Now we calculate the alphas, which will turn out to behave much differently from the Exponential case. Up to a constant, we have $\ln \overline{F}(x) = -\alpha \ln(\lambda + x)$ for any point $x > s$ in the tail, hence

$$\alpha_x = \alpha \frac{x}{\lambda + x}$$

The local alphas converge to the parameter α, thus for large x change only slowly, making the distribution more and more similar to Pareto in the higher area of the tail. For positive λ the local alphas rise with x, which is what reinsurance practitioners often observe. This experience is also reflected in the recommendations for the Swiss Solvency Test provided by FINMA, the Swiss insurance supervisor, see [FINMA, 2006]. They give standard values for Pareto alphas by line of business, to be applied to distribution tails starting at 1 or 5 million CHF, respectively. The suggested alphas vary for some lines of business, being somewhat higher for the upper threshold. For such situations piecewise Pareto can be a good model, however, if the local alpha rises gradually, proper GPD may be a better one.

For negative λ the local alphas decrease, slowly out in the tail, fast for very small x.

In any case the local alphas are bounded between α_s and the parameter α, which can be interpreted as α_∞, the local Pareto alpha at infinity. Further the local alphas tend to change slowly (apart maybe for low losses in case of negative λ), such that we can hope that the approximations for layer loss/frequency inflation derived in this chapter be very close to the exact values.

For layers with $d \geq s$ we get immediately, by recalling the analogous Pareto formula and the shift explained above, that $\mathrm{E}\left({}^u_d Z \mid Z > d\right) = \frac{\lambda+d}{\alpha-1}\left(1 - \left(\frac{\lambda+d}{\lambda+u}\right)^{\alpha-1}\right)$, hence

$$\alpha_{(d,u)} = 1 + \frac{\alpha - 1}{\lambda + d} \frac{d - u\left(\frac{\lambda+d}{\lambda+u}\right)^{\alpha}}{1 - \left(\frac{\lambda+d}{\lambda+u}\right)^{\alpha-1}}$$

(For $\alpha = 1$ use L'Hôpital's rule.) For unlimited layers (with $\alpha > 1$) this can be written in various useful ways:

$$\alpha_{(d,\infty)} = 1 + \frac{d}{\lambda + d}(\alpha - 1) = \frac{d}{\lambda + d}\alpha + \frac{\lambda}{\lambda + d} = \alpha - \frac{\lambda}{\lambda + d}(\alpha - 1) = \frac{d}{\lambda + d}\alpha\left(1 - \frac{1}{\alpha}\right) + \alpha\frac{1}{\alpha}$$

The last expression is a weighted average of the respective local alphas at d and at infinite, reminding us that the regional alpha, being an average of the local alphas across the layer, must lie between these two values.

Let us compare, for unlimited layers only, the approximate layer loss inflation to the exact one. Recall $E\left({}^\infty_d Z \mid Z > d\right) = \frac{\lambda+d}{\alpha-1}$. Formula 6.2 becomes with $g = 1 + h$

$$g^{\alpha}\left((1+h)^{\frac{\lambda}{\lambda+d}}\right)^{1-\alpha}=g^{\alpha-\frac{\lambda}{\lambda+d}(\alpha-1)}\approx\frac{\mathrm{E}\left(\overset{\infty}{\underset{d}{}}(gZ)\right)}{\mathrm{E}\left(\overset{\infty}{\underset{d}{}}Z\right)}=\frac{\overline{F}_{gZ}(d)\,E\left(\overset{\infty}{\underset{d}{}}gZ\mid Z>d\right)}{\overline{F}_{Z}(d)\,\;E\left(\overset{\infty}{\underset{d}{}}Z\mid Z>d\right)}$$

$$=g^{\alpha}\left(\frac{\lambda+d}{g\lambda+d}\right)^{\alpha}\frac{g\lambda+d}{\lambda+d}=g^{\alpha}\left(\frac{\lambda+d}{g\lambda+d}\right)^{\alpha-1}=g^{\alpha}\left(1+h\frac{\lambda}{\lambda+d}\right)^{1-\alpha}$$

Thus, for proper GPD the precision of the approximative layer inflation formula depends on the precision of the approximation

$$\left((1+h)^{\frac{\lambda}{\lambda+d}}\right)^{1-\alpha}\approx\left(1+h\frac{\lambda}{\lambda+d}\right)^{1-\alpha}$$

As in the Exponential case, both sides are equal in first order; the precision of the approximation for not very small h depends on how d relates to λ and how far α is from 1. It turns out that for positive λ the approximation is excellent and always greater than the exact formula, while for negative λ it is always smaller and also quite precise, unless d is very small (close to $-\lambda$). The following table displays the deviation of the approximate from the exact layer loss inflation in percent. h is entered in percent, $\alpha = 3$.

$\frac{\lambda}{\lambda+d}$ \\ h	-20	-10	-2	0	2	10	20
-2	-19.7	-5.5	-0.2	0	-0.2	-6.3	-25.4
-1	-7.8	-2.0	-0.1	0	-0.1	-2.0	-7.8
-0.25	-1.4	-0.3	0	0	0	-0.3	-1.1
0.25	0.9	0.2	0	0	0	0.2	0.6
0.5	1.3	0.3	0	0	0	0.2	0.8
0.8	0.8	0.2	0	0	0	0.1	0.5

Let us now, in the same way, compare for $d > 0$ the approximate frequency inflation to the exact one. Formula 6.3 becomes with $g = 1 + h$

$$g^{\alpha}\left((1+h)^{\frac{\lambda}{\lambda+d}}\right)^{-\alpha}=g^{\alpha-\frac{\lambda}{\lambda+d}\alpha}\approx\frac{\overline{F}_{gZ}(d)}{\overline{F}_{Z}(d)}=g^{\alpha}\left(\frac{\lambda+d}{g\lambda+d}\right)^{\alpha}=g^{\alpha}\left(1+h\frac{\lambda}{\lambda+d}\right)^{-\alpha}$$

The precision of this approximation depends on the same circumstances as the preceding one, save that very small values for α are benign now, not those close to 1.

The second-order calculation comparing the approximate to the exact formula for $\mathrm{E}\left(\overset{\infty}{\underset{d^2}{}}\left((gZ)^2\right)\right)$ is analogous to the first-order calculation and yields similar results.

6.5.3 Power curve (Weibull)

We treat the case of negative ξ only shortly, as it is arguably the least important one for the insurance practice. The parametrization introduced for positive ξ formally works here as well, however, the negative parameters give much more intuitive formulae. With $\beta := -\alpha > 0$ and $\nu := -\lambda = s + \frac{\sigma}{-\xi} > 0$ we have

$$\overline{F}_{Z}(x\mid Z>s)=\left(\frac{\nu-x}{\nu-s}\right)^{\beta}$$

which illustrates immediately that this GPD variant is a power curve and that the parameter ν equals the loss supremum. For the local alphas we get

$$\alpha_x=\beta\frac{x}{\nu-x}$$

which is an increasing function, tending to infinity as x approaches the supremum.

Chapter 7

General setting

7.1 Model

Now we have all ingredients to write down the general setting of the sample mean with variability both in past and future index values. We have hinted at the underlying stochastic model in Chapter 3; now we give the formal definition. Albeit appearing intricate, it essentially just reflects the commonly assumed uniform multiplicative impact of inflation on ground-up losses, as we have already used it in Chapters 2 to 4. However, we combine this with a very general loss count model.

Definition 7.1. For a (re)insurance risk, let N_k be the number of (ground-up) losses it produces in the year k, let $Z_{k,i}$ be the respective loss sizes.

Let $\mathcal{W}$ be the σ-algebra generated by all *indices* (time series of positive real numbers) describing changes of the world over time, having one value per year, and being potentially relevant for the risk in question. This shall, in particular, include indices relating to various kinds of inflation and indices reflecting volume measures of risks.

We say the risk *is in* the *inflationary world structure* if the following assumptions hold:
Conditionally, given $\mathcal{W}$, the loss productions of different years are independent and the losses of each year constitute a collective model.
There are two $\mathcal{W}$-measurable indices $\underline{B}$ and $\underline{V}$ having normalizations $B_k := \frac{\underline{B}_k}{\underline{B}_0}$, $V_k := \frac{\underline{V}_k}{\underline{V}_0}$ rebasing them to a certain year 0;
there is further a positive RV $\mathring{Z}$ being independent of $\mathcal{W}$, such that, given $\mathcal{W}$, all $\frac{Z_{k,i}}{B_k}$ are distributed as $\mathring{Z}$; so given $\mathcal{W}$, each $Z_{k,i}$ is distributed as $B_k \mathring{Z}_{k,i}$, where the $\mathring{Z}_{k,i}$ constitute a set of iid RV's being independent of $\mathcal{W}$ and distributed as $\mathring{Z}$;
there are further parameters $\theta > 0$, $\beta \geq 0$, γ constituting a mixed contagion model, such that, given $\mathcal{W}$, each N_k is distributed according to this model with volume V_k, i.e.,

$$\mathrm{E}\left(N_k \mid \mathcal{W}\right) = V_k \theta, \qquad \mathrm{Ct}\left(N_k \mid \mathcal{W}\right) = \beta + \frac{\gamma\left(1+\beta\right)}{V_k}, \qquad \mathrm{Var}\left(N_k \mid \mathcal{W}\right) = V_k\left(\theta + \theta^2 \gamma\left(1+\beta\right)\right) + V_k^2 \theta^2 \beta$$

We call $\underline{B}$ the *true inflation* (or *cost level*) *index* and $\underline{V}$ the *true volume index* of the risk in question. $\mathring{Z}$ is named *normalized severity.*

Remark 7.2. The normalization of the indices defines a reference level, but the year 0 has no particular properties. In fact, rebasing to another year would mean multiplication of the normalized index values with a respective positive factor, one for the inflation and one for the volume index. By rescaling $\mathring{Z}$, θ, and γ accordingly, one would get identical rules based on the new year of reference.

Note that $\underline{B}$ and $\underline{V}$ might not be unique, but their normalizations obviously are.

Although we do not know all indices, and for observable indices only the values of the (more or less recent) past, we can exploit the properties of the inflationary world structure: For calculations it is not necessary to know all index values – it suffices to know the essential ones:

Proposition 7.3. *In the inflationary world structure we have for any smaller σ-algebra $\mathcal{U} \subseteq \mathcal{W}$:*

If $\sigma(B_k) \subseteq \mathcal{U}$, then $Z_{k,i}$ has the same distribution conditionally on $\mathcal{U}$ as conditionally on $\mathcal{W}$.

If $\sigma(V_k) \subseteq \mathcal{U}$, then N_k has the same first two moments conditionally on $\mathcal{U}$ as conditionally on $\mathcal{W}$.

Proof. B_k (for the second assertion V_k) is $\mathcal{U}$-measurable and $\mathcal{W}$-measurable, such that one can easily factor it in and out. Thus,

$$\mathrm{P}\left(\frac{Z_{k,i}}{B_k} > x \mid \mathcal{U}\right) = \mathrm{E}\left(\mathrm{P}\left(\frac{Z_{k,i}}{B_k} > x \mid \mathcal{W}\right) \mid \mathcal{U}\right) = \mathrm{E}\left(\mathrm{P}\left(\mathring{Z} > x\right) \mid \mathcal{U}\right) = \mathrm{P}\left(\mathring{Z} > x\right)$$

where in the last step follows from the independence of $\mathring{Z}$ and $\mathcal{U}$.

For the loss count we have

$$\mathrm{E}(N_k \mid \mathcal{U}) = \mathrm{E}(\mathrm{E}(N_k \mid \mathcal{W}) \mid \mathcal{U}) = \mathrm{E}(V_k\theta \mid \mathcal{U}) = V_k\theta$$

$$\begin{aligned}\mathrm{Var}(N_k \mid \mathcal{U}) &= \mathrm{E}(\mathrm{Var}(N_k \mid \mathcal{W}) \mid \mathcal{U}) + \mathrm{Var}(\mathrm{E}(N_k \mid \mathcal{W}) \mid \mathcal{U}) \\ &= \mathrm{E}\left(V_k\left(\theta + \theta^2\gamma(1+\beta)\right) + V_k^2\theta^2\beta \mid \mathcal{U}\right) + \mathrm{Var}(V_k\theta \mid \mathcal{U}) = V_k\left(\theta + \theta^2\gamma(1+\beta)\right) + V_k^2\theta^2\beta\end{aligned}$$

□

If we e.g. know $\underline{B}_k$, $\underline{B}_0$, $\underline{V}_k$, and $\underline{V}_0$, we can calculate the normalized index values B_k and V_k. Thus, the σ-algebra generated by the four former RV's is sufficient to exploit the inflationary world structure. We will in the following repeatedly use σ-algebras that are just as large as we need.

7.2 As-if loss

7.2.1 Moments

Let us apply the inflationary world structure to experience rating. In order to calculate expectation and variance of the as-if loss, let us start from Formula 3.1:

$$S_k = \frac{\widehat{IV_q}}{IV_k}\sum_{i=1}^{N_k} {}^u_d\left(\frac{\widehat{IB_q}}{IB_k}Z_{k,i}\right) = \frac{\widehat{IV_q}\Delta V_k}{V_k}\sum_{i=1}^{N_k} {}^u_d\left(\widehat{IB_q}\Delta B_k \mathring{Z}_{k,i}\right) = \frac{\widehat{IV_q}\Delta V_k}{V_k}\widehat{IB_q}\sum_{i=1}^{N_k} {}^{u^*}_{d^*}\left(\Delta B_k \mathring{Z}_{k,i}\right)$$

where $d^* := \frac{d}{\widehat{IB_q}}$, $u^* := \frac{u}{\widehat{IB_q}}$. In the last step we have scaled the layer down (or possibly up, in case of negative inflation rates) by the predicted future picked cost level $\widehat{IB_q}$. The resulting rescaled layer is applied to the normalized loss severity being "distorted" by the past inflation gap. If the latter is, as we would typically expect, close to 1, we have a chance to apply the approximations for layer inflation derived in the last chapter.

Before calculating expectations we should have a clear idea about what information we need to know, i.e. what to condition on. The two preceding chapters worked with known volumes and known inflation, respectively, thus to apply them we must – for the moment – assume that all index values we use are known, namely for both the picked indices and the gaps (which together give us the true indices). To evaluate S_k we need

$$\widehat{IV_q}, \widehat{IB_q}, IV_k, IB_k, \Delta V_k, \Delta B_k$$

The information on these six index values (i.e. the σ-algebra generated by them) is contained in $\mathcal{W}$, which embraces all potentially relevant indices for the risk we are rating. This clearly includes true and picked inflation, true and picked volume, thus also the corresponding gaps. But $\widehat{IV_q}$ and $\widehat{IB_q}$, the predictors of future picked volume and cost level, are (measurable) functions of the past values of the indices IV and IB, thus are embraced by $\mathcal{W}$ as well. Thus, conditioning on $\mathcal{W}$ lets us work with the required index values as known quantities.

A technical detail: As for the normalized severity, we fix (as is usual in MSE calculations, but not always explicitly mentioned) its distribution – or at least all we need about it (it will not be that much). Analogously we fix the parameters of the ground-up loss count distribution. The relevant info about the cdf of $\mathring{Z}$ and about the loss count shall be assembled in the set π.

Now we calculate the (conditional) expectation in the collective model

$$\mathrm{E}_\pi\left(S_k \mid \mathcal{W}\right) = \frac{\widehat{IV_q}\Delta V_k}{V_k}\widehat{IB_q}\,\mathrm{E}\left(N_k \mid \mathcal{W}\right)\mathrm{E}\left({}^{u^*}_{d^*}\left(\Delta B_k \mathring{Z}\right) \mid \mathcal{W}\right) = \widehat{IV_q}\widehat{IB_q}\Delta V_k\,\theta\mathrm{E}\left({}^{u^*}_{d^*}\left(\Delta B_k \mathring{Z}\right) \mid \mathcal{W}\right)$$

The subscript π of the expectation shall emphasize that the parameters of loss count and normalized loss severity are fixed. (We will, however, mostly drop such subscripts, as they are rarely needed for clarity.)

If the cdf of the normalized severity is continuous at d^* and u^*, Formula 6.2 yields, at least for inflation gaps being rather close to 1,

$$\mathrm{E}\left({}^{u^*}_{d^*}\left(\Delta B_k \mathring{Z}\right) \mid \mathcal{W}\right) \approx \mathrm{E}\left({}^{u^*}_{d^*}\mathring{Z} \mid \mathcal{W}\right)\Delta B_k^{\alpha_{(d^*,u^*)}} = \mathrm{E}\left({}^{u^*}_{d^*}\mathring{Z} \mid \widehat{IB_q}\right)\Delta B_k^{\alpha_{(d^*,u^*)}}$$

where in the last step, due to the independence of $\mathring{Z}$ and $\mathcal{W}$, we could drop all conditions in the expectation but that on $\widehat{IB_q}$, which defines the rescaled layer. This can be seen via the *weak union rule* for conditional independence, which states that if a RV is independent of a pair of RV's, one of the latter can be shifted into the condition: $\mathring{Z}$ is independent of $\mathcal{W} = \sigma\left(\widehat{IB_q}, \mathcal{W}\right)$, thus is independent of $\mathcal{W}$ conditionally on $\widehat{IB_q}$, such that

$$ {}^{u^*}_{d^*}\mathring{Z} = \left(\mathring{Z} - \frac{d}{\widehat{IB_q}}\right)^+ - \left(\mathring{Z} - \frac{u}{\widehat{IB_q}}\right)^+$$

which is notably a measurable function applied to $\left(\mathring{Z}, \widehat{IB_q}\right)$, is independent of $\mathcal{W}$ given $\widehat{IB_q}$. In order to have d^* and u^* not random, we need $\widehat{IB_q}$ given, but nothing more.

Notice that formally $\alpha_{(d^*,u^*)}$ is a conditional figure too, as a (measurable) function of the non-random d^* and u^*:

$$\alpha_{(d^*,u^*)} = 1 + \frac{d^*\overline{F}_{\mathring{Z}}\left(d^*\right) - u^*\overline{F}_{\mathring{Z}}\left(u^*\right)}{\mathrm{E}\left({}^{u^*}_{d^*}\left(\mathring{Z}\right) \mid \mathcal{W}\right)} = 1 + \frac{d^*\overline{F}_{\mathring{Z}}\left(d^*\right) - u^*\overline{F}_{\mathring{Z}}\left(u^*\right)}{\mathrm{E}\left({}^{u^*}_{d^*}\left(\mathring{Z}\right) \mid \widehat{IB_q}\right)}$$

The same holds for the other regional and local alphas relating to the rescaled layer, which we will use in the following.

Now we can finally work with the distribution of the normalized losses. With

$$\mu^* := \mu_{(d^*,u^*)} := \mathrm{E}\left({}^{u^*}_{d^*}\mathring{Z} \mid \widehat{IB_q}\right), \qquad e^* := \theta\mu^*$$

we get the following approximation:

$$\mathrm{E}\left(S_k \mid \mathcal{W}\right) \approx \widehat{IV_q}\widehat{IB_q}\Delta V_k \Delta B_k^{\alpha_{(d^*,u^*)}}\theta\mu^* = \widehat{IV_q}\widehat{IB_q}\Delta V_k \Delta B_k^{\alpha_{(d^*,u^*)}}e^*$$

Interpreting the outcome, we see that the two volume terms have linear impact on this approximate expectation, while the effect of the past inflation gap is leveraged by the regional alpha of the rescaled layer. The impact of the (predicted) future picked cost level is non-linear, too, as it, in addition to appearing as a linear factor, rescales the layer: from the interval (d,u) to (d^*,u^*). However, apart from this rescaling, all effects from indices in this approximation are multiplicative. This specific structure may not seem a big thing, but is a great step forward, which will turn out to be decisive for the calculation of *unconditional* moments later on.

Now we derive analogous formulae for the variance. Formula 5.3 applied to ${}_{d^*}^{u^*}\left(\Delta B_k \mathring{Z}\right)$ yields

$$\begin{aligned}
&\operatorname{Var}(S_k \mid \mathcal{W}) \\
&= \left(\frac{\widehat{IV_q}\Delta V_k}{V_k}\widehat{IB_q}\,\theta \mathrm{E}\left({}_{d^*}^{u^*}\left(\Delta B_k \mathring{Z}\right) \mid \mathcal{W}\right)\right)^2 \left[\left(\frac{1+\mathrm{CV}^2\left({}_{d^*}^{u^*}\left(\Delta B_k \mathring{Z}\right) \mid \mathcal{W}\right)}{\theta} + \gamma(1+\beta)\right) V_k + \beta V_k^2\right] \\
&= \widehat{IV_q}^2 \widehat{IB_q}^2 \theta^2 E^2\left({}_{d^*}^{u^*}\left(\Delta B_k \mathring{Z}\right) \mid \mathcal{W}\right) \left\{\beta \Delta V_k^2 + \frac{\Delta V_k}{IV_k}\left[\gamma(1+\beta) + \frac{1+\mathrm{CV}^2\left({}_{d^*}^{u^*}\left(\Delta B_k \mathring{Z}\right) \mid \mathcal{W}\right)}{\theta}\right]\right\}
\end{aligned}$$

If the cdf of the normalized severity is continuous at d^* and u^*, Formula 6.4 yields, at least for inflation gaps being rather close to 1,

$$\begin{aligned}
&1+\mathrm{CV}^2\left({}_{d^*}^{u^*}\left(\Delta B_k \mathring{Z}\right) \mid \mathcal{W}\right) \\
&\approx \left\{\left(1+\mathrm{CV}^2\left({}_{d^*}^{u^*}\mathring{Z} \mid \mathcal{W}\right)\right)\Delta B_k^{\mathring{\alpha}_{(d^*,u^*)}} + \frac{2d^*}{\mathrm{E}\left({}_{d^*}^{u^*}\mathring{Z} \mid \mathcal{W}\right)}\left[\Delta B_k^{\mathring{\alpha}_{(d^*,u^*)}} - \Delta B_k^{\alpha_{(d^*,u^*)}}\right]\right\}\Delta B_k^{-2\alpha_{(d^*,u^*)}}
\end{aligned}$$

With

$$\varepsilon^* := \varepsilon_{(d^*,u^*)} := 1+\mathrm{CV}^2\left({}_{d^*}^{u^*}\mathring{Z} \mid \mathcal{W}\right) = 1+\mathrm{CV}^2\left({}_{d^*}^{u^*}\mathring{Z} \mid \widehat{IB_q}\right)$$

where the final equality follows as above via the weak union rule, and the first-moment approximation we get the following approximative formula:

$$\begin{aligned}
\operatorname{Var}(S_k \mid \mathcal{W}) \approx \widehat{IV_q}^2 \widehat{IB_q}^2 e^{*2} \Bigg\{ & \beta \Delta V_k^2 \Delta B_k^{2\alpha_{(d^*,u^*)}} \\
& + \frac{\Delta V_k}{IV_k}\left[\gamma(1+\beta)\Delta B_k^{2\alpha_{(d^*,u^*)}} + \frac{\varepsilon^*}{\theta}\Delta B_k^{\mathring{\alpha}_{(d^*,u^*)}} + \frac{2d^*}{e^*}\left(\Delta B_k^{\mathring{\alpha}_{(d^*,u^*)}} - \Delta B_k^{\alpha_{(d^*,u^*)}}\right)\right]\Bigg\}
\end{aligned}$$

The structure of this variance formula is much more complex than for the expectation, however, essentially it is "just" a linear combination of powers of index values. The (predicted) future picked volume has quadratic impact, while the past inflation gap is leveraged in three different ways: The first two components have the double regional alpha of the rescaled layer, the remaining ones the regional alpha or its second-moment counterpart. The impact of the (predicted) future picked cost level is complex, as it, in addition to appearing as a squared factor, rescales the layer. The past volume gap appears squared in one summand and linear in the remaining ones, here in combination with a fifth element: the inverse of the past picked volume.

Applying analogously Proposition 6.14 instead of Formula 6.4, we can improve the approximation and ensure a strictly positive result: The first lower bound to $1+\mathrm{CV}^2\left({}_{d^*}^{u^*}\left(\Delta B_k \mathring{Z}\right) \mid \mathcal{W}\right)$ is 1, such that the sum in the square brackets in the above variance formula is greater or equal than

$$\left[\left(\gamma(1+\beta) + \frac{1}{\theta}\right)\Delta B_k^{2\alpha_{(d^*,u^*)}}\right]$$

The second lower bound requires a derivative at $d^* > 0$ and equals $\overline{F}_Z\left(d^* \mid \mathcal{W}\right)^{-1} \mathit{\Delta}B_k^{-\alpha_{d^*}}$, such that said sum is greater than or approximately equal to

$$\left[\gamma\left(1+\beta\right)\mathit{\Delta}B_k^{2\alpha_{(d^*,u^*)}} + \mathit{\Delta}B_k^{2\alpha_{(d^*,u^*)}}\frac{1}{\theta}\frac{1}{\overline{F}_Z\left(d^* \mid \mathcal{W}\right)}\mathit{\Delta}B_k^{-\alpha_{d^*}}\right] = \left[\gamma\left(1+\beta\right)\mathit{\Delta}B_k^{2\alpha_{(d^*,u^*)}} + \frac{1}{\theta_{d^*}}\mathit{\Delta}B_k^{2\alpha_{(d^*,u^*)}-\alpha_{d^*}}\right]$$

where we have used the frequency of the proper layer losses

$$\theta_{d^*} := \theta\overline{F}_{\mathring{Z}}\left(d^* \mid \mathcal{W}\right) = \theta\overline{F}_{\mathring{Z}}\left(d^* \mid \widehat{IB_q}\right)$$

The last equality follows again from the fact that $\mathring{Z}$ is independent of $\mathcal{W}$ given $\widehat{IB_q}$.

We want to reparametrize with the moments of the proper excess losses. Here and occasionally later (in particular in Appendix B) we need a particular case of iterated conditioning, combining a σ-algebra possibly generated by a RV, like $\sigma\left(\widehat{IB_q}\right)$, with another one generated by a binary RV like $\left\{\mathring{Z} > d^*\right\}$. The latter can be treated in terms of elementary conditional expectation. In this context, for clarity we occasionally write conditioning as subscript. So

$$\mathrm{E}_{\widehat{IB_q}}\left({}_{d^*}^{u^*}\mathring{Z} \mid \mathring{Z} > d^*\right) = \mathrm{E}_{\sigma\left(\widehat{IB_q}\right)}\left({}_{d^*}^{u^*}\mathring{Z} \mid \mathring{Z} > d^*\right)$$

means a conditional expectation belonging to the (abstract) regular conditional probability measure $\mathrm{P}_{\widehat{IB_q}}\left(.\right) = \mathrm{P}_{\sigma\left(\widehat{IB_q}\right)}\left(.\right)$, which for events would read

$$\mathrm{P}_{\widehat{IB_q}}\left(A' \mid A\right) = \frac{\mathrm{P}_{\widehat{IB_q}}\left(A' \cap A\right)}{\mathrm{P}_{\widehat{IB_q}}\left(A\right)}$$

With this notation we set

$$\bar{\mu}^* := \bar{\mu}_{(d^*,u^*)} := \mathrm{E}_{\widehat{IB_q}}\left({}_{d^*}^{u^*}\mathring{Z} \mid \mathring{Z} > d^*\right)$$

$$\bar{\varepsilon}^* := \bar{\varepsilon}_{(d^*,u^*)} := 1 + \mathrm{CV}^2_{\widehat{IB_q}}\left({}_{d^*}^{u^*}\mathring{Z} \mid \mathring{Z} > d^*\right)$$

which with Lemma 5.5 applied to the above probability measure yield $e^* = \theta\mu^* = \theta_{d^*}\overline{\mu}^*$ and $\frac{\varepsilon^*}{\theta} = \frac{\bar{\varepsilon}^*}{\theta_{d^*}}$.

Let us wrap up, interpreting the derived approximations essentially as functions of the two gaps. For a more compact notation we set $\tilde{e} := e^*\widehat{IV_q}\widehat{IB_q} = \widehat{IV_q}\widehat{IB_q}\theta\mu_{(d^*,u^*)}$.

Proposition 7.4. *For a risk in the inflationary world structure, if the survival function $\overline{F}$ of the normalized loss is continuous at d^* and u^*, and the normalized past inflation gap is not too far from 1, then the following approximations for the first two (conditional) moments of the as-if loss hold:*

$$E_\pi\left(S_k \mid \mathcal{W}\right) \approx e^*\widehat{IV_q}\widehat{IB_q}\mathit{\Delta}V_k\mathit{\Delta}B_k^{\alpha_{(d^*,u^*)}} = \tilde{e}\mathit{\Delta}V_k\mathit{\Delta}B_k^{\alpha_{(d^*,u^*)}}$$

$$Var\left(S_k \mid \mathcal{W}\right) \approx \tilde{e}^2\left\{\beta\mathit{\Delta}V_k^2\mathit{\Delta}B_k^{2\alpha_{(d^*,u^*)}} + \frac{1}{IV_k}\max\left(\sum_{i=1}^{4}a_i\mathit{\Delta}V_k\mathit{\Delta}B_k^{\alpha_i}, \sum_{i=1}^{j}a_i^{(j)}\mathit{\Delta}V_k\mathit{\Delta}B_k^{\alpha_i^{(j)}}\right)\right\}$$

$$Cov\left(S_k, S_l \mid \mathcal{W}\right) = 0, \quad k \neq l$$

where

$$a_1 = \gamma\left(1+\beta\right), \qquad a_2 = \frac{\varepsilon^*}{\theta} = \frac{\bar{\varepsilon}^*}{\theta_{d^*}}, \qquad a_3 = -a_4 = \frac{2d^*}{e^*} = \frac{2d^*}{\theta\mu^*} = \frac{2d^*}{\theta_{d^*}\overline{\mu}^*}$$

$$\alpha_1 = 2\alpha_{(d^*,u^*)}, \qquad \alpha_2 = \alpha_3 = \mathring{\alpha}_{(d^*,u^*)}, \qquad \alpha_4 = \alpha_{(d^*,u^*)}$$

while for the second sum in the maximum, which improves the variance approximation as a lower bound, there are two options:

$j = 1$, $a'_1 = a_1 + \frac{1}{\theta}$, $\alpha'_1 = \alpha_1$,
$j = 2$, $a''_1 = a_1$, $\alpha''_1 = \alpha_1$, $a''_2 = \frac{1}{\theta_{d^*}}$, $\alpha''_2 = 2\alpha_{(d^*,u^*)} - \alpha_{d^*}$
The first option holds generally, the second one only if $\overline{F}$ is differentiable at $d^* > 0$.

The formulae include unlimited layers and pro-rata coverages, as long as expectation and variance of the normalized loss $\mathring{Z}$ are finite.

In the following special cases the last two summands in the first sum in the variance formula vanish or cancel out, such that the sum only goes from 1 to 2 and the second sum is obsolete:
First risks, *i.e. $d = 0$.*
Proportional coverages. *Here both approximations are exact and we have $\alpha_1 = \alpha_2 = 2$.*
$\overline{F}$ has a Pareto tail *with parameter α starting below* $\min\left(d^*, \frac{d^*}{\Delta B_k}\right)$. *Here both approximations are exact and we have $\alpha_1 = 2\alpha$, $\alpha_2 = \alpha$.*

Proof. Assumed all indices as known, the years are independent, hence the covariance formula. It remains only to check the three special cases.

In the first and second one the last two summands in the first sum of the variance formula equal 0. As for the alphas, recall that for proportional covers regional alphas equal one, thus α_1 must equal 2, as must α_2: By definition the second-order regional alpha is twice a regional alpha.

Finally, in the Pareto case all local and regional alphas equal the global Pareto parameter, such that said last two summands cancel out. The assumption about the low-enough start of the Pareto tail makes sure that the regional alphas used in Theorem 6.4 refer completely to the Pareto tail.

In both the Pareto and the proportional case Approximations 6.2 and 6.4 are exact. This implies, in particular, that the lower bound in the variance formula never applies, such that it can be dropped. For first risks it could in theory occur that the first lower bound applies, but going through the calculations makes clear that this would require a remote coincidence of a strange severity geometry and an unrealistically low coverage limit, such that in practice nothing is lost by dropping the second sum. □

Notice how few details about the severity of $\mathring{Z}$ enter these approximative formulae for conditional expectation and (co)variance of the as-if loss: just the first two moments of the loss to the (rescaled) layer, plus two regional alphas (and a local alpha for the second lower bound).

The proposition makes clearer in what way the Pareto case is analogous to proportional covers: While for the latter the impact of the ground-up inflation is linear, for Pareto it is leveraged by the alpha. However, there is an important detail making a difference: The third variance component (having coefficient a_2) in the Pareto case has exponent α, not 2α.

7.2.2 Parameter discussion

To get some intuition for the parameters introduced here, in particular those of the variance, let us first look at S_0, the as-if loss of the last year of the observed picked index history. In practice the loss history often ends a bit earlier earlier ($k_{max} < 0$), such that we do not have S_0 available for the rating, but to understand the theory it is a useful reference point. The year 0 has no index basis risk, while normalized picked indices equal 1 as well, such that the formulae greatly simplify. Going through the calculations of the past section for $k = 0$ we see that no approximation is involved and for the variance the first sum shrinks to two summands, making the second sum obsolete. We have

$$\mathrm{E}\left(S_0 \mid \mathcal{W}\right) = \tilde{e}, \qquad \mathrm{Var}\left(S_0 \mid \mathcal{W}\right) = \tilde{e}^2 \left[\beta + a_1 + a_2\right]$$

So the (conditional) squared CV of S_0 is the sum of β, a_1, and a_2.

For earlier years ($k < 0$) the summands four and five come in, apart from Pareto tails and coverages without deductible. Which exponents (leverages) of indices in the variance formula dominate, depends on how its time-invariant coefficients relate to each other. We call them *variance coefficients* and now have a closer look at those of the first sum, including $a_0 := \beta$. The coefficients of the second sum are closely related, such that we can infer subsequently their impact from the constellation in the first sum.

Although normalized picked indices and gaps may somewhat deviate from 1, it is clear that, apart from extreme inflation or drastic volume changes, the size of the variance components depends mainly on the five a_i.

- $a_0 = \beta$ can be zero, namely if there is no varying common market factor.
- $a_1 = \gamma\,(1+\beta)$ can be zero, namely if, given the market factor, the loss count has no dispersion (Poisson).
- Both coefficients can be zero at the same time, namely in the Poisson case without fluctuating market factor, very popular a model, but not too realistic.

In practice some dispersion, mostly overdispersion, is often observed and frequently catered for by applying the Negative Binomial model, however, usually without employing complex models for the dependence of variance on volume, like the mixed contagion model presented in this book. Thus, the effects stemming from the parameters β and γ are typically not separated, only the combined effect of both is observed via the empirical variance. If the volume does not vary extremely over time, the resulting empirical contagion approximately equals $\beta + \gamma\,(1+\beta)$, which would be the contagion of the year 0 having normalized volume 1. For very large portfolios (see e.g. Section 1.4.2 of [Mack, 1997] for an analysis of German Fire market data using the Poisson-Gamma model) one can observe values in the range of 0.01. This figure seems small, however, if it approximates the variance of a market factor, the latter must have a CV of about 10%, which is a considerable fluctuation for a whole market.

- a_2 and $a_3 = -a_4$ depend heavily on the retention d, more precisely on the rescaled d^*, which is, however, mostly in the same range.

For very large risks without, or with a very small, retention, and in particular for proportional reinsurance, at least as long as it covers mainly personal lines or light commercial business (this is the typical quota share reinsurance product), we can expect

$$a_2 = \frac{\overline{\varepsilon}_{(d^*,u^*)}}{\theta_{d^*}}$$

to be tiny: Such risks typically produce tens of thousands of losses per year (θ_{d^*}), while the CV of the proper layer losses ($\overline{\varepsilon}_{(d^*,u^*)}$) in case of no, or just a small, retention is typically in the range of say 1 to 10. In such a situation a_2 would be (possibly much) smaller than 0.01. Instead,

$$a_3 = \frac{2d^*}{\theta_{d^*}\overline{\mu}_{(d^*,u^*)}}$$

can be expected to be smaller still: Typical average losses ($\overline{\mu}_{(d^*,u^*)}$) in many Non-Life personal lines / light commercial portfolios are in the range of some thousand or ten thousand EUR, which is much more than the common retentions (d^*). Thus, for such risks $\frac{2d^*}{\overline{\mu}_{(d^*,u^*)}}$ should mostly be smaller than $\overline{\varepsilon}_{(d^*,u^*)}$, such that equivalently a_3 is less than a_2. Further, if for such risks the impact of the terms belonging to a_3 and a_4 is low, the lower bound (second sum) does not apply, such that the first sum and its coefficients describe the whole picture.

On the converse, if d^* is very large (such that the layer frequency is very low), $a_2 \geq \frac{1}{\theta_{d^*}}$ is large, too. Notice that a_2 is the only coefficient in the first sum that can never equal 0.

a_3 for high layers tends again to be smaller than a_2 or of similar size. Look at

$$\frac{a_2}{a_3} = \frac{\overline{\varepsilon}_{(d^*,u^*)}\overline{\mu}_{(d^*,u^*)}}{2d^*} = \frac{\mathrm{E}_{\widehat{IB_q}}\left(\left({}^{u^*}_{d^*}\mathring{Z}\right)^2 \mid \mathring{Z} > d^*\right)}{2d^*\mathrm{E}_{\widehat{IB_q}}\left({}^{u^*}_{d^*}\mathring{Z} \mid \mathring{Z} > d^*\right)}$$

and recall from Section 5.2 that for heavy-tailed severities, as we often have them in high layers, the RHS of this formula is typically in the same range as (precisely: not much smaller than) $\frac{u^*-d^*}{2d^*} = \frac{u-d}{2d}$, as long as the layer is not too long. However, for very long layers we get a similar result, at least for proper GPD, the most common tail model. Recalling that such tails are shifted Pareto we get analogously to the formulae in Section 5.2 that

$$\mathrm{E}_{\widehat{IB_q}}\left({}^{u^*}_{d^*}\mathring{Z} \mid \mathring{Z} > d^*\right) = (\lambda + d^*)\,\frac{1 - srll^{1-\alpha}}{\alpha - 1}$$

$$\mathrm{E}_{\widehat{IB_q}}\left(\left({}^{u^*}_{d^*}\mathring{Z}\right)^2 \mid \mathring{Z} > d^*\right) = 2\,(\lambda + d^*)^2\left(\frac{1 - srll^{2-\alpha}}{\alpha - 2} - \frac{1 - srll^{1-\alpha}}{\alpha - 1}\right)$$

with the *shifted relative layer length*

$$srll := \frac{\lambda + u^*}{\lambda + d^*}$$

For unlimited layers (with $\alpha > 2$) this leads to

$$\frac{a_2}{a_3} = \frac{\lambda + d^*}{d^*}\frac{1}{\alpha - 2}$$

(The result for long limited layers is numerically close, albeit the formula looks much more complex.) Again we see that the two coefficients are in the same range. While a_3 can be the larger one, in the variance formula this coefficient is applied to the difference $\Delta B_k^{\mathring{\alpha}_{(d^*,u^*)}} - \Delta B_k^{\alpha_{(d^*,u^*)}}$. If the two exponents are close, this difference is small, such that overall for very high layers a_2 should mostly be the dominant coefficient in the variance formula.

How close the two regional alphas actually are, is very easy to see for proper GPD. Here the local alphas are monotonic, such that their weighted averages $\mathring{\alpha}_{(d^*,u^*)}$ and $\alpha_{(d^*,u^*)}$ must both lie between $\alpha_{d^*} = \frac{d^*}{\lambda+d^*}\alpha$ and $\alpha_\infty = \alpha$. Hence, their difference cannot be greater than $\frac{\lambda}{\lambda+d^*}\alpha$ and for large d^* the term $\Delta B_k^{\mathring{\alpha}_{(d^*,u^*)}} - \Delta B_k^{\alpha_{(d^*,u^*)}}$ is of order $\frac{1}{d^*}$.

Generally, the more similar to Pareto the geometry in the upper tail area is, the more the term belonging to a_2 dominates over all other terms of the first sum, and the less likely it is that the lower bound applies and distorts the picture described by the first sum alone.

Altogether there are three clear cases: Very large risks with no or a very low retention can be dominated either by β (variance grows approximately with squared volume) or by γ (variance depends about linear on volume). For such risks the three parameters of the loss count model have decisive impact on the variance. While retentions in industrial insurance and reinsurance are typically rather large, a real-world example for a very low but non-zero retention would be a quota share reinsurance treaty for a large personal lines portfolio where the single insured objects have a *uniform* deductible of say some hundred EUR. Motor Hull in some countries is an example, as is Householder Content.

The third clear case are high layers with heavy, Pareto-like tails, dominated by a_2, which depends on the geometry of the severity of the normalized losses, while from the loss count only the frequency is relevant. Here the variance is almost proportional to the volume, as in the second case, however, inflation is leveraged by a different alpha: The place of $2\alpha_{(d^*,u^*)}$ is taken over by $\mathring{\alpha}_{(d^*,u^*)}$, which in many cases is about half of the former.

It goes without saying that there are a lot of intermediate situations mixing the described relations, plus certainly some where the second sum comes into play. In particular, there seems to be no easy rule for large portfolios with a retention of intermediate size, where all components of the variance formula could play a significant role.

Considering in what combinations the indices for volume and inflation occur in the formulae for the moments of S_k, it becomes clear that only for proportional coverages, where all regional alphas equal 1, the separation of frequency and severity effects is secondary. A main ingredient is the product $\Delta V_k \Delta B_k^{\alpha(d^*,u^*)}$, where inflation is leveraged, but the volume is not, such that the two gaps must be known separately. Even if the regional alpha equals about 1, the split is important, as in the variance formula other exponents occur. Only in the particular case that the first summand in the variance formula by far dominates the four other ones, an incorrect split of frequency and severity effects would only have a low numerical impact on the results.

7.2.3 As-if loss count

What we have developed for the expected layer loss, we can do analogously for the loss count: One simply has to replace the loss to the layer by the (measurable) indicator function, which counts 1 if the loss hits the layer, else 0. The as-if adjustments for volume and inflation are the same.

Definition 7.5. We call

$$C_k := \frac{\widehat{IV_q}}{IV_k} \sum_{i=1}^{N_k} \chi_{\left(\frac{\widehat{IB_q}}{IB_k} Z_{k,i} > d\right)}$$

the *as-if loss count* of the year k of the layer with deductible d.

For this quantity, which predicts the loss count of the year q, we have

$$C_k = \frac{\widehat{IV_q}\Delta V_k}{V_k} \sum_{i=1}^{N_k} \chi_{\left(\widehat{IB_q}\Delta B_k \mathring{Z}_{k,i} > d\right)} = \frac{\widehat{IV_q}\Delta V_k}{V_k} \sum_{i=1}^{N_k} \chi_{\left(\Delta B_k \mathring{Z}_{k,i} > d^*\right)}$$

Let now be $d > 0$. This case is very similar to the as-if loss calculation, save that, when rescaling the retention, the factor $\widehat{IB_q}$ drops out (apart from defining d^*). It remains to factor out the past inflation gap. If the cdf of $\mathring{Z}$ is differentiable at $d^* > 0$, we can apply Formula 6.3, at least for inflation gaps being rather close to 1:

$$\mathrm{E}\left(\chi_{\left(\Delta B_k \mathring{Z} > d^*\right)} \mid \mathcal{W}\right) = \overline{F}_{\Delta B_k \mathring{Z}}\left(d^* \mid \mathcal{W}\right) \approx \mathrm{E}\left(\chi_{\left(\mathring{Z} > d^*\right)} \mid \mathcal{W}\right) \Delta B_k^{\alpha_{d^*}} = \overline{F}_{\mathring{Z}}\left(d^* \mid \mathcal{W}\right) \Delta B_k^{\alpha_{d^*}}$$

If we plug this into

$$\mathrm{E}\left(C_k \mid \mathcal{W}\right) = \frac{\widehat{IV_q}\Delta V_k}{V_k} \mathrm{E}\left(N_k \mid \mathcal{W}\right) \mathrm{E}\left(\chi_{\left(\Delta B_k \mathring{Z} > d^*\right)} \mid \mathcal{W}\right) = \widehat{IV_q}\Delta V_k\, \theta \mathrm{E}\left(\chi_{\left(\Delta B_k \mathring{Z} > d^*\right)} \mid \mathcal{W}\right)$$

we get

$$\mathrm{E}\left(C_k \mid \mathcal{W}\right) \approx \widehat{IV_q}\Delta V_k \Delta B_k^{\alpha_{d^*}} \theta \overline{F}_{\mathring{Z}}\left(d^* \mid \mathcal{W}\right) = \widehat{IV_q}\Delta V_k \Delta B_k^{\alpha_{d^*}} \theta_{d^*}$$

For the variance we apply Formula 5.3 to the indicator function $\chi_{\left(\Delta B_k \mathring{Z} > d^*\right)}$, which equals its square, such that

$$1 + \mathrm{CV}^2\left(\chi_{\left(\Delta B_k \mathring{Z} > d^*\right)} \mid \mathcal{W}\right) = \frac{1}{\mathrm{E}\left(\chi_{\left(\Delta B_k \mathring{Z} > d^*\right)} \mid \mathcal{W}\right)}$$

We get

$$
\begin{aligned}
&\operatorname{Var}\left(C_k \mid \mathcal{W}\right) \\
&= \left(\frac{\widehat{IV_q}\Delta V_k}{V_k}\theta \mathrm{E}\left(\chi_{\left(\Delta B_k \mathring{Z}>d^*\right)} \mid \mathcal{W}\right)\right)^2 \left[\left(\frac{1+\mathrm{CV}^2\left(\chi_{\left(\Delta B_k \mathring{Z}>d^*\right)} \mid \mathcal{W}\right)}{\theta} + \gamma\left(1+\beta\right)\right) V_k + \beta V_k^2\right] \\
&\qquad = \widehat{IV_q}^2 \theta^2 \mathrm{E}^2\left(\chi_{\left(\Delta B_k \mathring{Z}>d^*\right)} \mid \mathcal{W}\right)\left\{\beta \Delta V_k^2 + \frac{\Delta V_k}{IV_k}\left[\gamma\left(1+\beta\right) + \frac{1}{\theta \mathrm{E}\left(\chi_{\left(\Delta B_k \mathring{Z}>d^*\right)} \mid \mathcal{W}\right)}\right]\right\}
\end{aligned}
$$

which together with the first-moment approximation leads to

$$
\begin{aligned}
\operatorname{Var}\left(C_k \mid \mathcal{W}\right) &\approx \widehat{IV_q}^2 \theta^2 \overline{F}^2_{\mathring{Z}}\left(d^* \mid \mathcal{W}\right)\left\{\beta \Delta V_k^2 \Delta B_k^{2\alpha_{d^*}} + \frac{\Delta V_k}{IV_k}\left[\gamma\left(1+\beta\right)\Delta B_k^{2\alpha_{d^*}} + \frac{1}{\theta \overline{F}_{\mathring{Z}}\left(d^* \mid \mathcal{W}\right)}\Delta B_k^{\alpha_{d^*}}\right]\right\} \\
&= \widehat{IV_q}^2 \theta_{d^*}^2 \left\{\beta \Delta V_k^2 \Delta B_k^{2\alpha_{d^*}} + \frac{\Delta V_k}{IV_k}\left[\gamma\left(1+\beta\right)\Delta B_k^{2\alpha_{d^*}} + \frac{1}{\theta_{d^*}}\Delta B_k^{\alpha_{d^*}}\right]\right\}
\end{aligned}
$$

Altogether we have with $\tilde{\theta} := \theta_{d^*}\widehat{IV_q}$:

Proposition 7.6. *For a risk in the inflationary world structure, if the survival function $\overline{F}$ of the normalized loss is differentiable at $d^* > 0$ and the normalized past inflation gap is not too far from 1, the following approximations for the first two (conditional) moments of the as-if loss count hold:*

$$
\begin{aligned}
E_\pi\left(C_k \mid \mathcal{W}\right) &\approx \theta_{d^*}\widehat{IV_q}\Delta V_k \Delta B_k^{\alpha_{d^*}} = \tilde{\theta}\Delta V_k \Delta B_k^{\alpha_{d^*}} \\
Var\left(C_k \mid \mathcal{W}\right) &\approx \tilde{\theta}^2\left\{\beta \Delta V_k^2 \Delta B_k^{2\alpha_{d^*}} + \frac{1}{IV_k}\left[\gamma\left(1+\beta\right)\Delta V_k \Delta B_k^{2\alpha_{d^*}} + \frac{1}{\theta_{d^*}}\Delta V_k \Delta B_k^{\alpha_{d^*}}\right]\right\} \\
Cov\left(C_k, C_l \mid \mathcal{W}\right) &= 0, \quad k \neq l
\end{aligned}
$$

The analogy to the as-if-loss formulae is evident. The variance formula is simpler – it looks almost like the as-if loss variance for the Pareto case, save that instead of the global alpha there is the local one at d^*, and $\overline{\varepsilon}_{(d^*,u^*)}$ is replaced by 1.

What we found out about the sizes of the components of the as-if loss variance, applies to the loss count too, with $\frac{1}{\theta_{d^*}}$ taking the place of a_2 and the last two summands vanishing. And again we see that the split of frequency and severity effects matters, the only difference being that the frequency inflation gap is leveraged by the local alpha instead of the regional alphas.

It remains to treat the case $d = 0$ or equivalently $d^* = 0$. Here the as-if loss count simplifies to

$$
C_k = \frac{\widehat{IV_q}\Delta V_k}{V_k}\sum_{i=1}^{N_k}\chi_{\left(\mathring{Z}_{k,i}>0\right)} = \frac{\widehat{IV_q}\Delta V_k}{V_k}\sum_{i=1}^{N_k}\chi_{\left(\mathring{Z}_{k,i}>d^*\right)}
$$

where inflation has dropped out altogether. (In fact the ground-up loss count is unaffected by loss inflation.) However, we can formally treat this as the special (degenerate) case $\Delta B \equiv 1$ and repeat step by step the calculations made for $d > 0$, noting that now the approximation via Formula 6.3 is obsolete, such that we have exact formulae and need no additional assumptions. We get the above final formulae, but without ΔB_k, or formally equivalently, with the factor $1 = \Delta B_k^0$.

Corollary 7.7. *For $d = 0$ the formulae of the proposition hold with 0 taking the place of α_{d^*}. They are exact and do not require the assumptions on survival function and past inflation gap.*

7.3 Future loss

The moments of the future loss X_q can be calculated very much like those of the as-if loss, differing only in a few details. We have

$$X_q = \sum_{i=1}^{N_q} {}^{u}_{d}\left(Z_{q,i}\right) = \sum_{i=1}^{N_q} {}^{u}_{d}\left(B_q \mathring{Z}_{q,i}\right) = \sum_{i=1}^{N_q} {}^{u}_{d}\left(\widehat{IB_q}\frac{IB_q}{\widehat{IB_q}}\Delta B_q \mathring{Z}_{q,i}\right) = \widehat{IB_q}\sum_{i=1}^{N_q} {}^{u^*}_{d^*}\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q \mathring{Z}_{q,i}\right)$$

where we have again rescaled the layer with the predicted picked future cost level. This formula is simpler than that for the as-if loss, having no volume adjustment. At the same time it is more complex, as the normalized loss is distorted by two factors: the future inflation gap and the relative deviation of the future picked cost level from its predictor. The idea is that, as long as the latter is rather precise (i.e. close to the value it predicts), both factors, ΔB_q and $\frac{IB_q}{\widehat{IB_q}}$, are rather close to 1, such that we can again apply the approximations for layer inflation.

In the following the factor $\frac{IB_q}{\widehat{IB_q}}\Delta B_q$ will take the place that ΔB_k had in the calculations for the as-if loss, namely in the application of Formulae 6.1 and 6.4. Apart from this all steps are essentially the same.

Again we condition – for the moment – on $\mathcal{W}$, which embraces all index values we need, namely

$$\widehat{IV_q}, \widehat{IB_q}, IV_q, IB_q, \Delta V_q, \Delta B_q$$

For the expectation we get

$$\mathrm{E}\left(X_q \mid \mathcal{W}\right) = \widehat{IB_q}\,\mathrm{E}\left(N_q \mid \mathcal{W}\right)\mathrm{E}\left({}^{u^*}_{d^*}\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q \mathring{Z}\right) \mid \mathcal{W}\right) = \widehat{IB_q} IV_q \Delta V_q\, \theta \mathrm{E}\left({}^{u^*}_{d^*}\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q \mathring{Z}\right) \mid \mathcal{W}\right)$$

which, after plugging in

$$\mathrm{E}\left({}^{u^*}_{d^*}\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q \mathring{Z}\right) \mid \mathcal{W}\right) \approx \left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q\right)^{\alpha_{(d^*,u^*)}} \mathrm{E}\left({}^{u^*}_{d^*}\mathring{Z} \mid \mathcal{W}\right)$$

leads to

$$\mathrm{E}\left(X_q \mid \mathcal{W}\right) \approx IV_q \Delta V_q \widehat{IB_q}^{\,1-\alpha_{(d^*,u^*)}} \Delta B_q^{\alpha_{(d^*,u^*)}} IB_q^{\alpha_{(d^*,u^*)}} e^*$$

For the variance one has

$$\begin{aligned}\mathrm{Var}\left(X_q \mid \mathcal{W}\right) = {} & \widehat{IB_q}^{\,2} IV_q^2 \theta^2 \mathrm{E}^2\left({}^{u^*}_{d^*}\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q \mathring{Z}\right) \mid \mathcal{W}\right) \\ & \left\{\beta \Delta V_q^2 + \frac{\Delta V_q}{IV_q}\left[\gamma\left(1+\beta\right) + \frac{1}{\theta}\left(1 + \mathrm{CV}^2\left({}^{u^*}_{d^*}\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q \mathring{Z}\right) \mid \mathcal{W}\right)\right)\right]\right\}\end{aligned}$$

which, after plugging in the above approximation and

$$\begin{aligned}1 + \mathrm{CV}^2\left({}^{u^*}_{d^*}\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q \mathring{Z}\right) \mid \mathcal{W}\right) \approx {} & \left\{\left(1 + \mathrm{CV}^2\left({}^{u^*}_{d^*}\mathring{Z}\right) \mid \mathcal{W}\right)\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q\right)^{\mathring{\alpha}_{(d^*,u^*)}}\right. \\ & \left. + \frac{2d^*}{\mathrm{E}\left({}^{u^*}_{d^*}\mathring{Z} \mid \mathcal{W}\right)}\left[\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q\right)^{\mathring{\alpha}_{(d^*,u^*)}} - \left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q\right)^{\alpha_{(d^*,u^*)}}\right]\right\}\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q\right)^{-2\alpha_{(d^*,u^*)}}\end{aligned}$$

leads to

$$\operatorname{Var}(X_q \mid \mathcal{W}) \approx \widehat{IB_q}^2 IV_q^2 e^{*2}\Bigg\{\beta \Delta V_k^2 \left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q\right)^{2\alpha_{(d^*,u^*)}} + \frac{\Delta V_q}{IV_q}\Bigg[\gamma(1+\beta)\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q\right)^{2\alpha_{(d^*,u^*)}}$$

$$+\frac{\varepsilon^*}{\theta}\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q\right)^{\mathring{\alpha}_{(d^*,u^*)}} + \frac{2d^*}{e^*}\left(\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q\right)^{\mathring{\alpha}_{(d^*,u^*)}} - \left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q\right)^{\alpha_{(d^*,u^*)}}\right)\Bigg]\Bigg\}$$

$$= e^{*2}\Bigg\{\beta\,\widehat{IB_q}^{2-2\alpha_{(d^*,u^*)}} \Delta V_q^2 \Delta B_q^{2\alpha_{(d^*,u^*)}} IV_q^2 IB_q^{2\alpha_{(d^*,u^*)}}$$

$$+\Delta V_q IV_q\Bigg[\gamma(1+\beta)\,\widehat{IB_q}^{2-2\alpha_{(d^*,u^*)}} \Delta B_q^{2\alpha_{(d^*,u^*)}} IB_q^{2\alpha_{(d^*,u^*)}} + \frac{\varepsilon^*}{\theta}\widehat{IB_q}^{2-\mathring{\alpha}_{(d^*,u^*)}} \Delta B_q^{\mathring{\alpha}_{(d^*,u^*)}} IB_q^{\mathring{\alpha}_{(d^*,u^*)}}$$

$$+\frac{2d^*}{e^*}\left(\widehat{IB_q}^{2-\mathring{\alpha}_{(d^*,u^*)}} \Delta B_q^{\mathring{\alpha}_{(d^*,u^*)}} IB_q^{\mathring{\alpha}_{(d^*,u^*)}} - \widehat{IB_q}^{2-\alpha_{(d^*,u^*)}} \Delta B_q^{\alpha_{(d^*,u^*)}} IB_q^{\alpha_{(d^*,u^*)}}\right)\Bigg]\Bigg\}$$

Analogously we can carry the lower variance bounds from Proposition 6.14 through the calculations. Summing up we get

Proposition 7.8. *For a risk in the inflationary world structure, suppose the survival function $\overline{F}$ of the normalized loss is continuous at d^* and u^*, the normalized future inflation gap is not too far from 1, and the predictor of the normalized future picked inflation is quite precise. Then the following approximations for the first two (conditional) moments of the future loss hold:*

$$E_\pi(X_q \mid \mathcal{W}) \approx e^*\widehat{IB_q}^{1-\alpha_{(d^*,u^*)}}\left(\Delta V_q \Delta B_q^{\alpha_{(d^*,u^*)}}\right)\left(IV_q IB_q^{\alpha_{(d^*,u^*)}}\right) = \tilde{e}\Delta V_q \Delta B_q^{\alpha_{(d^*,u^*)}} \frac{IV_q IB_q^{\alpha_{(d^*,u^*)}}}{\widehat{IV_q}\widehat{IB_q}^{\alpha_{(d^*,u^*)}}}$$

$$Var(X_q \mid \mathcal{W}) \approx e^{*2}\Bigg\{\beta\,\widehat{IB_q}^{2-2\alpha_{(d^*,u^*)}}\left(\Delta V_q^2 \Delta B_q^{2\alpha_{(d^*,u^*)}}\right)\left(IV_q^2 IB_q^{2\alpha_{(d^*,u^*)}}\right)$$

$$+\max\left[\sum_{i=1}^4 a_i \widehat{IB_q}^{2-\alpha_i}\left(\Delta V_q \Delta B_q^{\alpha_i}\right)\left(IV_q IB_q^{\alpha_i}\right), \sum_{i=1}^j a_i^{(j)}\widehat{IB_q}^{2-\alpha_i^{(j)}}\left(\Delta V_q \Delta B_q^{\alpha_i^{(j)}}\right)\left(IV_q IB_q^{\alpha_i^{(j)}}\right)\right]\Bigg\}$$

$$= \tilde{e}^2\Bigg\{\beta\Delta V_q^2 \Delta B_q^{2\alpha_{(d^*,u^*)}} \frac{IV_q^2 IB_q^{2\alpha_{(d^*,u^*)}}}{\widehat{IV_q}^2\,\widehat{IB_q}^{2\alpha_{(d^*,u^*)}}}$$

$$+\frac{1}{\widehat{IV_q}}\max\left[\sum_{i=1}^4 a_i \Delta V_q \Delta B_q^{\alpha_i}\frac{IV_q IB_q^{\alpha_i}}{\widehat{IV_q}\widehat{IB_q}^{\alpha_i}}, \sum_{i=1}^j a_i^{(j)}\Delta V_q \Delta B_q^{\alpha_i^{(j)}}\frac{IV_q IB_q^{\alpha_i^{(j)}}}{\widehat{IV_q}\widehat{IB_q}^{\alpha_i^{(j)}}}\right]\Bigg\}$$

$$Cov(X_q, S_k \mid \mathcal{W}) = 0$$

The coefficients and exponents in the variance formula are the same as for the as-if loss, and we have the same special cases: condition for second lower variance bound; first risks, pro-rata, and Pareto, the latter holding for tails starting below

$$\min\left(d^*, \frac{d^*}{\Delta B_q}\frac{\widehat{IB_q}}{IB_q}\right) = \min\left(\frac{d}{\widehat{IB_q}}, \frac{d}{\Delta B_q IB_q}\right)$$

The formulae include unlimited layers and pro-rata coverages, as long as expectation and variance of the normalized loss $\mathring{Z}$ are finite.

Proof. Assuming indices as known, all years are independent, past and future ones, thus the covariance formula. The threshold for the Pareto tail is somewhat different from the as-if case, with $\frac{IB_q}{\widehat{IB_q}}\Delta B_q$ again taking the place of ΔB_k. The rest is just rearranging of terms, using $e^*\widehat{IB_q} = \tilde{e}\widehat{IV_q}^{-1}$. □

The proposition expresses expectation and variance in two ways, emphasizing firstly where and with what leverage the various indices appear, secondly the analogy to the as-if loss formulae, revealing that the main difference are the additional factors quantifying the relative precision of the predictors for IV_q and IB_q.

For the future layer loss count

$$\bar{N}_q := \sum_{i=1}^{N_q} \chi_{(Z_{q,i}>d)}$$

we can apply the same procedure as for the as-if loss count, which results in formulae very similar to those for the future layer loss, again with the separate case $d = 0$:

Corollary 7.9. *For a risk in the inflationary world structure, assume the survival function $\overline{F}$ of the normalized loss is differentiable at $d^* > 0$, the normalized future inflation gap is not too far from 1, and the predictor of the normalized future picked inflation is quite precise. Then the following approximations for the first two (conditional) moments of the future loss count hold:*

$$E_\pi\left(\bar{N}_q \mid \mathcal{W}\right) \approx \theta_{d^*} \widehat{IB_q}^{-\alpha_{d^*}} \left(\Delta V_q \Delta B_q^{\alpha_{d^*}}\right)\left(IV_q IB_q^{\alpha_{d^*}}\right) = \tilde{\theta} \Delta V_q \Delta B_q^{\alpha_{d^*}} \frac{IV_q IB_q^{\alpha_{d^*}}}{\widehat{IV_q}\widehat{IB_q}^{\alpha_{d^*}}}$$

$$\begin{aligned} Var\left(\bar{N}_q \mid \mathcal{W}\right) \approx \theta_{d^*}^2 \Bigg\{ & \beta\, \widehat{IB_q}^{-2\alpha_{d^*}} \left(\Delta V_q^2 \Delta B_q^{2\alpha_{d^*}}\right)\left(IV_q^2 IB_q^{2\alpha_{d^*}}\right) \\ & + \gamma(1+\beta)\, \widehat{IB_q}^{-2\alpha_{d^*}} \left(\Delta V_q \Delta B_q^{2\alpha_{d^*}}\right)\left(IV_q IB_q^{2\alpha_{d^*}}\right) + \frac{1}{\theta_{d^*}} \widehat{IB_q}^{-\alpha_{d^*}} \left(\Delta V_q \Delta B_q^{\alpha_{d^*}}\right)\left(IV_q IB_q^{\alpha_{d^*}}\right) \Bigg\} \\ = \tilde{\theta}^2 \Bigg\{ & \beta \Delta V_q^2 \Delta B_q^{2\alpha_{d^*}} \frac{IV_q^2 IB_q^{2\alpha_{d^*}}}{\widehat{IV_q}^2 \widehat{IB_q}^{2\alpha_{d^*}}} + \frac{1}{\widehat{IV_q}} \left[\gamma(1+\beta) \Delta V_q \Delta B_q^{2\alpha_{d^*}} \frac{IV_q IB_q^{2\alpha_{d^*}}}{\widehat{IV_q}\widehat{IB_q}^{2\alpha_{d^*}}} + \frac{1}{\theta_{d^*}} \Delta V_q \Delta B_q^{\alpha_{d^*}} \frac{IV_q IB_q^{\alpha_{d^*}}}{\widehat{IV_q}\widehat{IB_q}^{\alpha_{d^*}}} \right] \Bigg\} \end{aligned}$$

$$Cov\left(\bar{N}_q, C_k \mid \mathcal{W}\right) = 0$$

For $d = 0$ the formulae hold with 0 taking the place of α_{d^*}. They are exact and do not require the assumptions on survival function and indices.

What we have stated about the sizes of the components of the variances of as-if loss and respective count, applies to the corresponding future figures as well. The latter are more complex, however, essentially we again have linear combinations of powers of index values – involving different indices and exponents, but the same coefficients.

For comparison we assemble the involved indices, first for the as-if figures:

- predictors of future picked inflation and volume (typically calculated from observed past values)
- past picked inflation and volume (observable)
- past gaps for inflation and volume (arguably unobservable)

For the future figures:

- predictors of future picked inflation and volume (typically calculated from observed past values)
- future picked indices for inflation and volume (unobservable, corresponding past values observable)
- future gaps for inflation and volume (unobservable)

The latter two categories together yield the future true indices, however, keeping them separate is more transparent.

7.4 Conditions

We now have in principle assembled all ingredients to calculate MSEP and SPEE of the sample mean, however, only conditionally on all index values. We need unconditional figures, or more precisely *partly* unconditional ones, according to what we know when doing the pricing, and what not. The question what to condition on, in a complex environment like ours is not straightforward, as we have various sources of information. Directly observable are:

- the loss record, from year k_{min} to year $k_{max} \leq 0$, affected by changes in scale and volume.
- the picked inflation index up to year 0. (Recall that by definition this is the last year for which we have the picked indices available.) This index starts at the latest in the year k_{min}, often much earlier. If a publicly available index (or an amended version thereof) is used, it is usually known for decades – and such a long observation period is indeed necessary for sophisticated modeling: As a practical rule of thumb, AR(1) processes, or other well-established time series for inflation modeling having similar complexity, require at least 30 data points for a reliable parameter estimation.
- the picked volume index up to year 0. As volumes are typically provided together with losses, this "index" usually starts in the year k_{min}, sometimes a few years earlier. So it is rarely long enough a time series for very sophisticated models. To the worse, the volume may depend on business strategy and competition (the original premium income always does), which might change frequently, such that even in case of a long time series, in order to predict the future volume, it could be neither reasonable to use the whole history nor to apply complex time series models.

The gaps for inflation and volume are arguably not observable, at least not easily and in case $d > 0$ not directly. We will not discuss the (interesting) question whether there could be a chance to observe/estimate something about these indices, justifying our pessimism by recalling what was pointed out in the last chapter: If we have only tail data available and the tail is Pareto distributed, such that the tail shape is not affected by inflation, it is *a priori* impossible to discern changes in the layer loss frequency due to inflation from changes due to volume shifts. If the tail is similar to Pareto, but not equal, inflation inference is a priori possible from the varying tail shape, however, *a posteriori* rather not, as the (slow and slight) change in tail shape due to inflation will mostly be overwhelmed by other random fluctuations in the empirical tail.

In order to emphasize the diversity of existing approaches for MSE optimization, we recall two standard situations in insurance, which have led to quite different approaches:

- Parameter fit or expected-loss estimate, based on an iid loss record (no changes in scale and volume): The SPEE is defined as expected squared difference between the quantity to be estimated and its estimator, given the parameter, see e.g. [Klugman et al., 2008], Section 12.2. This means in a way averaging over all possible loss records coming from the same parameter constellation, although in each single inference situation the loss record is available.
- Reserve predictor in Chain Ladder models: Here the MSEP is defined as expected squared difference between the aggregate ultimate loss and the actual prediction for it, given the run-off triangle of the losses, see e.g. [Mack, 1993] or [Riegel, 2015]. Here one conditions on the loss record itself, without averaging.

The fundamental difference between these two situations is not SPEE vs MSEP. (Albeit less common, the first approach works analogously for MSEP.) Instead, it is averaging over empirical results belonging to a fixed parameter constellation vs conditioning on the specific empirical result.

Remark 7.10. It is interesting to note that in the time series analysis literature, namely in the context of *forecast uncertainty*, both MSE variants appear – and both are considered reasonable, see the introduction in [Ansley and Kohn, 1986] and the discussion in [Phillips, 1979] about *conditional* vs *unconditional* distributions of forecasts. In time series language, the prediction of an index value for the year q from observed values from several years $k \leq 0$ would be called *q-step(s)-ahead forecast*, while the MSE of the predictor (= forecast) takes various names, e.g. *PMSE* in [Kunitomo and Yamamoto, 1985], *MSFE* in the comprehensive survey paper by [Gonçalves Mazzeu et al., 2015]. Whatever the names, the distinction of conditional vs unconditional MSE is exactly as described above: The past index values are given in the first case, while in the second case they are random, but underlying model and parameters are fixed.

What is adequate in our situation?

As for the losses, it seems reasonable to fix as usual the underlying parameters, which are in our case the three parameters of the loss count model, plus the needed distributional properties of the normalized severity (moments, alphas).

As for the gaps, not having observations available, selecting in advance reasonable distributions and parameters seems to be the only viable approach. This is exactly what we have done with ΔB_k in the basic example of Section 4. It can be seen as a Bayesian approach and is indeed a common way to combine expert judgment and empirical data, in insurance and elsewhere – recall the example from geophysics in [Rhoades, 1996].

Now the picked indices come in. One could indeed think about applying distributional time series models and averaging over all index history outcomes (with fixed models and parameters), however, many different models are thinkable: for inflation possibly sophisticated ones being apt for long time series, for the volume rather simple ones to be assessed from a sometimes very short history. To embrace as many cases as possible we will, analogously to typical Chain Ladder calculations, condition on the observed past values of the picked indices.

This means that our approach is in way a hybrid, half parameter-fit-like (unconditional), half Chain-Ladder-like (conditional). This may appear odd, however, an analogous model has been successfully applied recently: [Riegel, 2015] optimizes the Burning Cost rating of proportional Motor reinsurance for a Chain Ladder variant with known loss inflation, but some uncertainty on the volume side (on-level premium). His MSEP is conditional on the *observed* loss triangle (as usual, no averaging) and keeps the *parameters* of the volume uncertainty fixed.

If we look at our preliminary results for S_k, X_q, and the corresponding loss counts, how is the setting going to change?

- We will keep fixed the information about loss count model and normalized severity.
- The conditions on the gaps will ultimately have to be dropped, but their models and parameters will be fixed; we collect the latter in the sets Π_B, Π_V and for the moment do not specify them further.
- We will keep the condition on the past picked indices, however, to be precise, this condition relates to the picked indices as they are observed, i.e. before normalization (rebasing to the year 0). We collect all observed values in $\sigma\left(\underline{IB}_{(j\leq 0)}, \underline{IV}_{(j\leq 0)}\right) \subseteq \mathcal{W}$. This is slightly imprecise a notation, not specifying where observations start, which might be different from index to index, but in any case will be in or before the year k_{min}.
- Conditioning on the past picked indices for *all* available past years implies conditioning also on $\widehat{IB_q}$ and $\widehat{IV_q}$, the predictors for the future picked index values, which are *functions* of the past index values. (Whatever functions this may be, it will in practice be measurable ones.)

- We will ultimately have to drop the condition on the future picked indices, which are not (yet) observed in the moment the rating is done.

7.5 Averaging

We want to calculate the *partly unconditional* moments of as-if loss and future loss, given the past observations $\left(\underline{IB}_{(j\leq 0)}, \underline{IV}_{(j\leq 0)}\right)$. More generally we select a possibly larger σ-algebra $\underline{I}$ such that

$$\sigma\left(\underline{IB}_{(j\leq 0)}, \underline{IV}_{(j\leq 0)}\right) \subseteq \underline{I} \subseteq \mathcal{W}$$

As its name indicates, it can be imagined as information on certain picked indices including the history of $\underline{IB}$ and $\underline{IV}$. This gives us the flexibility to integrate further information we may have observed. We will do so in the next chapter, for now we do not specify $\underline{I}$ further.

For the sake of concise notation we introduce a combination of indices that we will use in various variants.

Definition 7.11. For any real exponent α we set, for any year j,

$$\underline{D(\alpha)}_j := \underline{V}_j \underline{B}_j^{\alpha}$$

and analogously for gaps and picked indices. We name the straightforward predictors of the (normalized) future values of the latter accordingly

$$\widehat{ID(\alpha)}_q := \widehat{IV_q}\widehat{IB_q}^{\alpha}$$

We call the index $\underline{D(\alpha)} = \underline{V}\,\underline{B}^{\alpha}$ (defined element-wise) the *combined leveraged scale-frequency* impact.

This name reflects the following idea: Shifts in frequency or scale have multiplicative impact on the expected loss, the former linear, the latter leveraged by an exponent that depends on the geometry of the cdf of the normalized severity. If we take both kinds of shifts together, the overall effect is modeled by $\underline{D(\alpha)}$ and the corresponding picked index and gap.

Now we can restate the conditional moments of the as-if loss even more compactly:

$$\mathrm{E}\left(S_k \mid \mathcal{W}\right) \approx e^{*}\widehat{IV_q}\widehat{IB_q}\Delta V_k \Delta B_k^{\alpha_{(d^*,u^*)}} = \tilde{e}\Delta D\big(\alpha_{(d^*,u^*)}\big)_k$$

$$\begin{aligned}\mathrm{Var}\left(S_k \mid \mathcal{W}\right) &\approx \left(e^{*}\widehat{IV_q}\widehat{IB_q}\right)^2 \left\{\beta \Delta V_k^2 \Delta B_k^{2\alpha_{(d^*,u^*)}} + \frac{1}{IV_k}\max\left(\sum_{i=1}^{4} a_i \Delta V_k \Delta B_k^{\alpha_i}, \sum_{i=1}^{j} a_i^{(j)} \Delta V_k \Delta B_k^{\alpha_i^{(j)}}\right)\right\} \\ &= \tilde{e}^2 \left\{\beta \Delta D\big(\alpha_{(d^*,u^*)}\big)_k^2 + \frac{1}{IV_k}\max\left(\sum_{i=1}^{4} a_i \Delta D(\alpha_i)_k, \sum_{i=1}^{j} a_i^{(j)} \Delta D\Big(\alpha_i^{(j)}\Big)_k\right)\right\}\end{aligned}$$

$$\mathrm{Cov}\left(S_k, S_l \mid \mathcal{W}\right) = 0, \quad k \neq l$$

Recall that only the last formula is exact, while the first two formulae (apart from proportional and Pareto case) are approximative and require that ΔB_k be rather small. By combining these formulae we get

$$\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right) \approx \tilde{e}^2 \left\{(1+\beta)\,\Delta D\big(\alpha_{(d^*,u^*)}\big)_k^2 + \frac{1}{IV_k}\max\left(\sum_{i=1}^{4} a_i \Delta D(\alpha_i)_k, \sum_{i=1}^{j} a_i^{(j)} \Delta D\Big(\alpha_i^{(j)}\Big)_k\right)\right\}$$

while the third formula is equivalent to

$$\mathrm{E}\left(S_k S_l \mid \mathcal{W}\right) = \mathrm{E}\left(S_k \mid \mathcal{W}\right) \mathrm{E}\left(S_l \mid \mathcal{W}\right), \qquad k \neq l$$

Calculation of (partly) unconditional moments intuitively means averaging over $\mathcal{W}$ with $\underline{I}$ given. For this we will need that the first approximation above hold in expectation (conditionally, given $\underline{I}$):

$$\mathrm{E}\left(\mathrm{E}\left(S_k \mid \mathcal{W}\right) \mid \underline{I}\right) \overset{!}{\approx} \mathrm{E}\left(\tilde{e}\varDelta D\big(\alpha_{(d^*,u^*)}\big)_k \mid \underline{I}\right)$$

and analogously for S_k^2 and $S_k S_l$. This is only trivial if $\varDelta B_k$, as well as $\varDelta B_l$, is *almost surely* (with probability 1) close to 1, such that the approximation holds pointwise. If the $\varDelta B_j$, $j \leq 0$, can deviate considerably from 1, whatever small the probability for such events is, the approximations in expectation could become imprecise (in extreme cases some terms could even be infinite). It is plausible that in real world constellations this will hardly occur, however, it deserves some thought. We defer this (very technical) point to the appendix, where we will show that for many constellations with realistic settings, having e.g. lognormal normalized gaps being close to 1 with high probability, the required approximations hold in expectation as desired – notably with all appearing moments of the gaps being finite.

In the following, when we speak (a bit loosely) of RV's *fluctuating closely about 1*, this shall mean: closely enough to make the required approximations hold in expectation. To ease intuition one may for the moment think of fluctuations being close to 1 almost surely, bearing in mind that results can be extended to far more situations.

Notice that $\tilde{e} = \widehat{IV_q}\widehat{IB_q}\theta\, \mathrm{E}\left({}_{d^*}^{u^*}\mathring{Z} \mid \widehat{IB_q}\right)$ is $\underline{I}$-measurable, thus can be factored out of the expectation on the RHS of the last formula, as well as out of the formulae for the second moments. The variance coefficients are $\underline{I}$-measurable, too.

When now averaging over $\mathcal{W}$, we keep the parameters of gaps, loss count model and normalized severity fixed. We will occasionally emphasize this by a subscript, e.g. in the next formula.

Proposition 7.12. *For a risk in the inflationary world structure, suppose the survival function $\overline{F}$ of the normalized loss is continuous at d^* and u^*, and the normalized past inflation gap is a RV fluctuating closely about 1. Then with $\alpha = \alpha_{(d^*,u^*)}$ the following approximations for the first two moments of the as-if loss hold:*

$$E_{\varPi_V,\varPi_B,\pi}\left(S_k \mid \underline{I}\right) \approx \tilde{e}E\left(\varDelta D(\alpha)_k \mid \underline{I}\right)$$

$$\begin{aligned}
&Var\left(S_k \mid \underline{I}\right)\\
&\approx \tilde{e}^2\left\{Var(\varDelta D(\alpha)_k \mid \underline{I}) + \beta E\Big(\varDelta D(\alpha)_k^2 \mid \underline{I}\Big) + \frac{1}{IV_k}\max\left[\sum_{i=1}^{4} a_i E(\varDelta D(\alpha_i)_k \mid \underline{I}), \sum_{i=1}^{j} a_i^{(j)} E\Big(\varDelta D\Big(\alpha_i^{(j)}\Big)_k \mid \underline{I}\Big)\right]\right\}\\
&= \tilde{e}^2\Bigg\{(1+\beta)\, E\Big(\varDelta D(\alpha)_k^2 \mid \underline{I}\Big) - E^2\left(\varDelta D(\alpha)_k \mid \underline{I}\right) + \\
&\qquad \frac{1}{IV_k}\max\left[\sum_{i=1}^{4} a_i E(\varDelta D(\alpha_i)_k \mid \underline{I}), \sum_{i=1}^{j} a_i^{(j)} E\Big(\varDelta D\Big(\alpha_i^{(j)}\Big)_k \mid \underline{I}\Big)\right]\Bigg\}
\end{aligned}$$

$$E\left(S_k^2 \mid \underline{I}\right) \approx \tilde{e}^2\left\{(1+\beta)\, E\Big(\varDelta D(\alpha)_k^2 \mid \underline{I}\Big) + \frac{1}{IV_k}\max\left[\sum_{i=1}^{4} a_i E(\varDelta D(\alpha_i)_k \mid \underline{I}), \sum_{i=1}^{j} a_i^{(j)} E\Big(\varDelta D\Big(\alpha_i^{(j)}\Big)_k \mid \underline{I}\Big)\right]\right\}$$

$$Cov\left(S_k, S_l \mid \underline{I}\right) \approx \tilde{e}^2 Cov\left(\varDelta D(\alpha)_k, \varDelta D(\alpha)_l \mid \underline{I}\right), \quad k \neq l$$

$$E\left(S_k S_l \mid \underline{I}\right) \approx \tilde{e}^2 E\left(\varDelta D(\alpha)_k\, \varDelta D(\alpha)_l \mid \underline{I}\right), \quad k \neq l$$

Variant $j = 2$ of the variance formula requires in addition that $\overline{F}$ be differentiable at $d^* > 0$.

Proof. Using the tower property of conditional expectations, we get immediately

$$\mathrm{E}\left(S_k \mid \underline{I}\right) = \mathrm{E}\left(\mathrm{E}\left(S_k \mid \mathcal{W}\right) \mid \underline{I}\right) \approx \mathrm{E}\left(\tilde{e}\Delta D(\alpha)_k \mid \underline{I}\right)$$

where the second step is the "approximation in expectation" discussed above. The covariance formula, given the first formula, is equivalent to the formula for $S_k S_l$, so proving the latter is sufficient.

$$\mathrm{E}\left(S_k S_l \mid \underline{I}\right) = \mathrm{E}\left(\mathrm{E}\left(S_k S_l \mid \mathcal{W}\right) \mid \underline{I}\right) = \mathrm{E}\left(\mathrm{E}\left(S_k \mid \mathcal{W}\right)\mathrm{E}\left(S_l \mid \mathcal{W}\right) \mid \underline{I}\right) \approx \mathrm{E}\left(\tilde{e}\Delta D(\alpha)_k\, \tilde{e}\Delta D(\alpha)_l \mid \underline{I}\right)$$

The variance formula, given the first formula, is equivalent to the formula for S_k^2, so proving the latter is sufficient. We first calculate the weaker approximation without the lower bound

$$\begin{aligned}\mathrm{E}\left(S_k^2 \mid \underline{I}\right) = \mathrm{E}\left(\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right) \mid \underline{I}\right) &\approx \mathrm{E}\left(\tilde{e}^2\left\{(1+\beta)\,\Delta D(\alpha)_k^2 + \frac{1}{IV_k}\sum_{i=1}^{4} a_i \Delta D(\alpha_i)_k\right\} \mid \underline{I}\right)\\ &= \tilde{e}^2\left\{(1+\beta)\,\mathrm{E}\left(\Delta D(\alpha)_k^2 \mid \underline{I}\right) + \frac{1}{IV_k}\left[\sum_{i=1}^{4} a_i \mathrm{E}\left(\Delta D(\alpha_i)_k \mid \underline{I}\right)\right]\right\}\end{aligned}$$

which yields the claimed formula without the second sum. The (approximate) inequality with the latter is calculated analogously:

$$\begin{aligned}\mathrm{E}\left(S_k^2 \mid \underline{I}\right) = \mathrm{E}\left(\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right) \mid \underline{I}\right) &\gtrapprox \mathrm{E}\left(\tilde{e}^2\left\{(1+\beta)\,\Delta D(\alpha)_k^2 + \frac{1}{IV_k}\sum_{i=1}^{j} a_i^{(j)} \Delta D\left(\alpha_i^{(j)}\right)_k\right\} \mid \underline{I}\right)\\ &= \tilde{e}^2\left\{(1+\beta)\,\mathrm{E}\left(\Delta D(\alpha)_k^2 \mid \underline{I}\right) + \frac{1}{IV_k}\left[\sum_{i=1}^{j} a_i^{(j)} \mathrm{E}\left(\Delta D\left(\alpha_i^{(j)}\right)_k \mid \underline{I}\right)\right]\right\}\end{aligned}$$

We can recombine both formulae, just as in the conditional case. □

Remark 7.13. Instead of treating first and second sum in the variance formula separately, we could have carried the maximum of the two sums through the expectation, which would have led to the approximation

$$\mathrm{E}\left(S_k^2 \mid \underline{I}\right) \approx \mathrm{E}\left(\tilde{e}^2\left\{(1+\beta)\,\Delta D\left(\alpha_{(d^*,u^*)}\right)_k^2 + \frac{1}{IV_k}\max\left[\sum_{i=1}^{4} a_i \Delta D(\alpha_i)_k\,, \sum_{i=1}^{j} a_i^{(j)} \Delta D\left(\alpha_i^{(j)}\right)_k\right]\right\} \mid \underline{I}\right)$$

This approximation is larger and arguably more precise in quite some situations, but has the disadvantage that here we cannot split the expectation into separate terms of the kind $\mathrm{E}\left(\Delta D(\alpha_i)_k \mid \underline{I}\right)$. As such decompositions will be very useful in the following, we prefer to work with the former approximation variant. It is anyway plausible that in real-world cases the first sum will be mostly positive and even larger than the second sum, such that the latter is obsolete. For the rare cases where the second sum matters, it is essential that it be a handy formula that in the first place prevent the variance approximation from being negative. However, it does not have to be the best available lower bound.

In order to evaluate terms like $\mathrm{E}\left(\Delta D(\alpha)_k \mid \underline{I}\right)$ and respective second moments, it is tempting to assume that gaps are independent of picked indices, as in this case one could possibly drop the condition on $\underline{I}$. Whether and when this assumption is reasonable or too restrictive, will be explored in the next chapter.

Notice in particular that, after averaging, the covariances possibly differ from 0 – this depends on $\mathrm{Cov}\left(\Delta D\left(\alpha_{(d^*,u^*)}\right)_k, \Delta D\left(\alpha_{(d^*,u^*)}\right)_l \mid \underline{I}\right)$, i.e. on the interaction of the past gaps.

Let us now look at the other quantities of interest. For the loss count we get analogously, involving notably slightly different alphas,

Corollary 7.14. *For a risk in the inflationary world structure, suppose the survival function $\overline{F}$ of the normalized loss is differentiable at $d^* > 0$ and the normalized past inflation gap is a RV fluctuating closely about 1. Then with $\alpha = \alpha_{d^*}$ the following approximations for the first two moments of the as-if loss count hold:*

$$E_{\Pi_V,\Pi_B,\pi}\left(C_k \mid \underline{I}\right) \approx \tilde{\theta} E\left(\Delta D(\alpha)_k \mid \underline{I}\right)$$

$$\begin{aligned}
&Var\left(C_k \mid \underline{I}\right) \\
&\approx \tilde{\theta}^2 \left\{ Var\left(\Delta D(\alpha)_k \mid \underline{I}\right) + \beta E\left(\Delta D(\alpha)_k^2 \mid \underline{I}\right) + \frac{1}{IV_k}\left[\gamma\left(1+\beta\right) E\left(\Delta D(2\alpha)_k \mid \underline{I}\right) + \frac{1}{\theta_{d^*}} E\left(\Delta D(\alpha)_k \mid \underline{I}\right)\right]\right\} \\
&= \tilde{\theta}^2 \left\{ \left(1+\beta\right) E\left(\Delta D(\alpha)_k^2 \mid \underline{I}\right) - E^2\left(\Delta D(\alpha)_k \mid \underline{I}\right) + \frac{1}{IV_k}\left[\gamma\left(1+\beta\right) E\left(\Delta D(2\alpha)_k \mid \underline{I}\right) + \frac{1}{\theta_{d^*}} E\left(\Delta D(\alpha)_k \mid \underline{I}\right)\right]\right\}
\end{aligned}$$

$$Cov\left(C_k, C_l \mid \underline{I}\right) \approx \tilde{\theta}^2 Cov\left(\Delta D(\alpha)_k, \Delta D(\alpha)_l \mid \underline{I}\right), \quad k \neq l$$

For $d = 0$ the formulae hold with $\alpha = 0$. They are exact and do not require the assumptions on survival function and past inflation gap.

Now we proceed in the same way for the future loss, leaving again at first the lower bound for the variance aside. In abbreviated notation with $\alpha = \alpha_{(d^*,u^*)}$ the conditional moments read

$$\mathrm{E}\left(X_q \mid \mathcal{W}\right) \approx e^* \widehat{IB_q}^{1-\alpha} D(\alpha)_q = \tilde{e} \Delta D(\alpha)_q \frac{ID(\alpha)_q}{\widehat{ID(\alpha)_q}}$$

$$\begin{aligned}
&\mathrm{Var}\left(X_q \mid \mathcal{W}\right) \\
&\approx e^{*2} \left\{ \beta \widehat{IB_q}^{2-2\alpha} D(\alpha)_q^2 + \sum_{i=1}^{4} a_i \widehat{IB_q}^{2-\alpha_i} D(\alpha_i)_q \right\} = \tilde{e}^2 \left\{ \beta \Delta D(\alpha)_q^2 \frac{ID(\alpha)_q^2}{\widehat{D(\alpha)_q}^2} + \frac{1}{\widehat{IV_q}} \sum_{i=1}^{4} a_i D(\alpha_i)_q \frac{ID(\alpha_i)_q}{\widehat{ID(\alpha_i)_q}} \right\}
\end{aligned}$$

$$\mathrm{Cov}\left(X_q, S_k \mid \mathcal{W}\right) = 0$$

and lead to representations for the (partially) unconditional moments similar to those for the as-if figures. Note that here, with $\Delta D(\alpha)_q \frac{ID(\alpha)_q}{\widehat{ID(\alpha)_q}}$ taking the place that $\Delta D(\alpha)_k$ had above, in addition to $\Delta D(\alpha)_q$ having to fluctuate closely about 1, we need $\frac{ID(\alpha)_q}{\widehat{ID(\alpha)_q}}$ to behave similarly, such that their product fluctuates closely about 1. However, we must emphasize a technical detail here:

For $\frac{ID(\alpha)_q}{\widehat{ID(\alpha)_q}}$ it is in our context most natural to formulate this behavior *conditionally*, given the picked index history − recall that we do not want to average over the latter. This means that the numerator $ID(\alpha)_q$ shall be a RV fluctuating closely about the (conditionally known) denominator $\widehat{ID(\alpha)_q}$, which requires that the latter be a very good predictor of the former.

This desired property can notably be formulated equivalently for the respective indices before normalization: We have

$$\underline{ID(\alpha)_q} = \underline{ID(\alpha)_0} ID(\alpha)_q, \qquad \underline{\widehat{ID(\alpha)}_q} = \underline{ID(\alpha)_0} \widehat{ID(\alpha)_q}$$

with the same (conditionally known) normalizing factor $\underline{ID(\alpha)_0}$.

Proposition 7.15. *For a risk in the inflationary world structure, suppose the survival function $\overline{F}$ of the normalized loss is continuous at d^* and u^*, the normalized future inflation gap is a RV fluctuating closely about 1, and the future picked inflation fluctuates, conditionally given $\underline{I}$, closely about its predictor. Then with $\alpha = \alpha_{(d^*,u^*)}$ the following approximations for the first two moments of the future loss hold:*

$$E_{\Pi_V,\Pi_B,\pi}\left(X_q \mid \underline{I}\right) \approx e^* \widehat{IB_q}^{1-\alpha} E\left(D(\alpha)_q \mid \underline{I}\right) = \tilde{e}\frac{E\left(\Delta D(\alpha)_q \, ID(\alpha)_q \mid \underline{I}\right)}{\widehat{ID(\alpha)}_q}$$

$$\begin{aligned}
Var\left(X_q \mid \underline{I}\right) \approx e^{*2}\Bigg\{ & \widehat{IB_q}^{2-2\alpha}\left[Var\left(D(\alpha)_q \mid \underline{I}\right) + \beta E\left(D(\alpha)_q^2 \mid \underline{I}\right)\right] \\
& + \max\left[\sum_{i=1}^{4} a_i \widehat{IB_q}^{2-\alpha_i} E\left(D(\alpha_i)_q \mid \underline{I}\right), \sum_{i=1}^{j} a_i^{(j)} \widehat{IB_q}^{2-\alpha_i^{(j)}} E\left(D(\alpha_i)_q \mid \underline{I}\right)\right]\Bigg\} \\
= \tilde{e}^2 \Bigg\{ & (1+\beta)\frac{E\left(\Delta D(\alpha)_q^2 \, ID(\alpha)_q^2 \mid \underline{I}\right)}{\widehat{ID}_q^{\,2}} - \frac{E^2\left(\Delta D(\alpha)_q \, ID(\alpha)_q \mid \underline{I}\right)}{\widehat{ID(\alpha)}_q^{\,2}} \\
& + \frac{1}{\widehat{IV_q}} \max\left[\sum_{i=1}^{4} a_i \frac{E\left(\Delta D(\alpha_i)_q \, ID(\alpha_i)_q \mid \underline{I}\right)}{\widehat{ID(\alpha_i)}_q}, \sum_{i=1}^{j} a_i^{(j)} \frac{E\left(\Delta D\left(\alpha_i^{(j)}\right)_q ID\left(\alpha_i^{(j)}\right)_q \mid \underline{I}\right)}{\widehat{ID(\alpha_i)}_q}\right]\Bigg\}
\end{aligned}$$

$$Cov\left(X_q, S_k \mid \underline{I}\right) \approx e^{*2} \widehat{IV_q} \widehat{IB_q}^{2-\alpha} Cov\left(D(\alpha)_q, \Delta D(\alpha)_k \mid \underline{I}\right) = \tilde{e}^2 \frac{Cov\left(\Delta D(\alpha)_q \, ID(\alpha)_q, \Delta D(\alpha)_k \mid \underline{I}\right)}{\widehat{ID(\alpha)}_q}$$

Variant $j = 2$ of the variance formula requires in addition that $\overline{F}$ be differentiable at $d^* > 0$.

Proof. As for the approximate equalities, the steps are the same as in Proposition 7.12, including the amendment with the second sum. In the subsequent equalities the constants are resorted, further $D(\alpha)_q$ and variants thereof are written as products of the respective picked index and gap.

As for the as-if loss, "fluctuating closely about 1 / about its predictor" can be extended well beyond the easy case "almost surely": to a number of realistic models for indices having (rare) larger deviations from 1. This is deferred to Appendix B. □

For the future loss count we can apply the same procedure.

Corollary 7.16. *For a risk in the inflationary world structure, suppose the survival function $\overline{F}$ of the normalized loss is differentiable at $d^* > 0$, the normalized future inflation gap is a RV fluctuating closely about 1, and the future picked inflation fluctuates, conditionally given $\underline{I}$, closely about its predictor. Then with $\alpha = \alpha_{d^*}$ the following approximations for the first two moments of the future loss count hold:*

$$E_{\Pi_V,\Pi_B,\pi}\left(\bar{N}_q \mid \underline{I}\right) \approx \theta_{d^*} \widehat{IB_q}^{\alpha} E\left(D(\alpha)_q \mid \underline{I}\right) = \tilde{\theta}\frac{E\left(\Delta D(\alpha)_q \, ID(\alpha)_q \mid \underline{I}\right)}{\widehat{ID(\alpha)}_q}$$

$$
\begin{aligned}
Var\left(\bar{N}_q \mid \underline{I}\right) \approx \theta_{d^*}^2 \Bigg\{ & \widehat{IB_q}^{-2\alpha} \left[Var\Big(D(\alpha)_q \mid \underline{I}\Big) + \beta E\Big(D(\alpha)_q^2 \mid \underline{I}\Big)\right] \\
& + \widehat{IB_q}^{-2\alpha} \gamma \left(1+\beta\right) E\Big(D(2\alpha)_q \mid \underline{I}\Big) + \widehat{IB_q}^{-\alpha} \frac{1}{\theta_{d^*}} E\Big(D(\alpha)_q \mid \underline{I}\Big) \Bigg\} \\
= \tilde{\theta}^2 \Bigg\{ & \left(1+\beta\right) \frac{E\Big(\Delta D(\alpha)_q^2 \, ID(\alpha)_q^2 \mid \underline{I}\Big)}{\widehat{ID(\alpha)_q}^2} - \frac{E^2\left(\Delta D(\alpha)_q \, ID(\alpha)_q \mid \underline{I}\right)}{\widehat{ID(\alpha)_q}^2} \\
& + \frac{1}{\widehat{IV_q}} \left[\gamma \left(1+\beta\right) \frac{E\Big(\Delta D(2\alpha)_q \, ID(2\alpha)_q \mid \underline{I}\Big)}{\widehat{ID(2\alpha)_q}} + \frac{1}{\theta_{d^*}} \frac{E\Big(\Delta D(\alpha)_q \, ID(\alpha)_q \mid \underline{I}\Big)}{\widehat{ID(\alpha)_q}} \right] \Bigg\}
\end{aligned}
$$

$$
Cov\left(\bar{N}_q, C_k \mid \underline{I}\right) \approx \theta_{d^*}^2 \widehat{IV_q} \widehat{IB_q}^{-\alpha} Cov\Big(D(\alpha)_q, \Delta D(\alpha)_k \mid \underline{I}\Big) = \tilde{\theta}^2 \frac{Cov\Big(\Delta D(\alpha)_q \, ID(\alpha)_q, \Delta D(\alpha)_k \mid \underline{I}\Big)}{\widehat{ID(\alpha)_q}}
$$

For $d = 0$ the formulae hold with $\alpha = 0$. They are exact and do not require the assumptions on survival function and indices.

We have introduced a lot of notation in this chapter, which, however, makes clear that all (partly) unconditional moments are functions of the moments of the family of indices denoted with $D(\alpha)$, being each a product of a volume index with a power (leverage) of the corresponding inflation index.

We see that, just as before averaging, the formulae for the future loss (count), besides the gap $\Delta D(\alpha)$, involve the precision of the prediction of $ID(\alpha)_q$. For better insight it is desirable to know whether and how we can decompose the resulting terms into moments of gaps and moments containing only picked indices. This question requires some thought.

Overall we need clarity about the interaction of picked indices and gaps, both in past and future, and generally about the connection between scale and volume effects. For ease of calculations it is tempting (and at first sight plausible) to assume a lot of independence: between past and future uncertainty, between picked indices and gaps, and, in particular, between (ground-up) frequency and severity shifts. However, the devil is in the details. In order to state consistent and plausible assumptions, without inadvertently assuming to much real-world complexity away, we need a thorough analysis of the indices being relevant for experience rating. This deserves a separate chapter.

Chapter 8

World of indices

8.1 Primordial indices

In this chapter we make assertions for the whole index series, i.e. k runs across all years, positive and negative ones (past and future), and equations hold for any k. Further we look at indices in general, before normalization (rebasing) to the year 0. For ease of notation we will temporarily (in the first half of this chapter) drop the underscore ($\underline{B}_k$ etc.) indicating original indices before normalization.

In order to understand how the indices we need could interact, we try to decompose them into ultimate, or say *primordial* indices, which require no further decomposition. As we will see, both inflation and volume can embrace various effects, which are worth being considered separately. The same is true for a third index: the *official* volume, which we will denote M_k.

The official volume is needed if we want to model the loss or the loss frequency *per volume unit*. As explained in detail in Chapter 2, it is common to contractually define the premium of risks having variable size not flat, but per volume unit, and the volume used here is a publicly available one, which only rarely coincides with the volume V_k the loss frequency in the collective model is tied to – for the sake of clarity we shall from now on call the latter kind of volume *frequency volume*.

It is essential to be able to convert these two different types of volumes into each other, even when we do not need to rate the premium per official volume unit: In practice the official volume is mostly the only available volume, which means that the frequency volume has to be inferred somehow from the official one.

We at first look at true indices, bearing in mind that each such index might have to be approximated by a picked index, such that for each type of index we could finally come up with three indices: the true one, the picked one, and the gap in between.

For the moment we leave accumulation losses aside – we will be able to include them soon. For now think of a model for single losses affecting single insured units, say fire losses or minor accidents.

8.1.1 Inflation-free official volume

If the official volume does not contain any inflation (e.g. number of insured objects, vehicle years), things are rather easy. However, we should not oversimplify by assuming that the official volume in this case always equal the frequency volume. There could be (at least slight) changes in the frequency per volume unit, which could be due to trends or to the fact that the official volume is fair, but not perfect as an exposure measure – recall discussion and examples in Section 2.1.2. In general it makes sense to decompose

$$V_k = M_k A_k$$

where A_k is the index quantifying the *frequency per (official) volume unit.* Although one generally does not expect this index to fluctuate a lot, a real-world example illustrates how it can vary:

In Motor Liability *vehicle years* is a popular official volume, as it is readily available and certainly a fair exposure measure. However, the good data situation of this line of business enables insurers in many countries to monitor the frequency per vehicle year over time. So German market statistics revealed a slow but steady downward trend over decades (see [Cramer et al., 2005]), which was taken into account in the rating – then the trend abruptly reversed in 2011, which somewhat shook the industry. Such market-wide effects are not the only challenge. For a single insurer changes in the composition of the portfolio, e.g. an unexpectedly fast growing share of (risk-seeking) young drivers, can be even more influential.

All these issues do not rule out vehicle years as official volume – this might still be the best option: publicly available, easy to explain and monitor, reflecting market growth, etc. However, for the internal calculation, vehicle years, as well as other official exposures of similar kind, should be amended, and this is what the index A_k is introduced for. It has characteristics somewhat similar to inflation: While it might be hard to observe or very volatile for a single risk, in many cases one should be able to predict it via a picked index, say a market-wide frequency per volume unit, coming from whole-market statistics and ideally providing some segmentation, e.g. per driver age group.

Summing up, if the official volume is inflation-free, the frequency volume is the product of two primordial indices: the official volume and the frequency per volume unit. It is mostly plausible to assume that these two indices are independent.

8.1.2 Inflation-sensitive official volume

This case is much more complex. If the official volume is affected by inflation, this does not imply that here we have exactly the inflation affecting the losses – in general we would expect the two inflations to be similar (or say: far from independent), however, not necessarily identical. Recall some examples from Section 2.1.2: In the case of *aggregate sum insured* (Property business) practitioners feel indeed that the inherent inflation mostly reflects pretty well loss inflation, in particular in the common case that the yearly updating of sums insured is tied to the empirical market-wide loss inflation. Instead, in the case of *payroll* (Workers' Compensation, Employers' Liability), which embraces *wages inflation*, practitioners would expect loss inflation to be somewhat different, at least for personal injury losses, which largely depend on both wages and *medical inflation.* These two types of inflation are often different, but closely related, containing at least *monetary* inflation as a common factor.

Let us generally denote the two indices for inflation (or more exactly for the corresponding cost level) as follows:

- B_k: *loss inflation* – inflation contained in the ground-up losses
- $\widetilde{B}_k$: *(official) volume inflation* – inflation contained in the official volume

Now we can decompose the official volume into the part reflecting inflation (scale) and the remaining part, which shall be denoted $\bar{M}_k$, the *deflated (official) volume.* This yields the decomposition

$$M_k = \bar{M}_k \widetilde{B}_k$$

The deflated volume is inflation-free, so we have analogously to the first case

$$V_k = \bar{M}_k A_k$$

where A_k is now the *frequency per deflated (official) volume unit.*

This decomposition procedure may appear theoretical and cumbersome, but is common practice in quite a number of insurance rating situations, e.g. in reinsurance. Recall from Chapter 2 that here the most common official volume is *original premium income*. The inflation contained in this figure is a complex one: it largely reflects loss inflation, which (at least in case of sound pricing) is incorporated in the premiums. However, this inflationary effect is strongly superimposed by the forces of the market cycle, which may cause the yearly premium adjustments to be smaller or larger than loss inflation. So market pressure is an inflationary effect, too, which together with loss inflation constitutes the "inflation" $\widetilde{B}_k$ contained in the original premium. If this combined volume inflation is factored out, one gets what reinsurers call the on-level premium or *restated premium*, see e.g. [Riegel, 2015].

Exposures like aggregate sum insured and payroll are usually simpler to decompose. The inflation contained in the former is typically assessed by a *construction cost* index, in the latter case one would choose an appropriate *wages* index. Such indices are available in many countries, coming from public statistics or from the insurance industry itself. They are natural candidates for picked indices, which, however, might have to be amended, e.g. in case of risks whose underlying property values or wages, respectively, rise faster than the national average being measured by the index.

As for the connection between loss inflation and official-volume inflation, one could think of very intricate interrelations. To avoid overcomplexity we propose a special structure that should not be too restrictive. It requires, however, a *third* kind of inflation:

- $\bar{B}_k$: *basic inflation* – common part of the above two inflations

The basic idea is that in many cases the bulk of inflationary effects should affect both inflations in (about) the same way, while the remaining effects should (approximately) be neatly different. This can be described in a rather flexible way by the decompositions

$$B_k = H_k \bar{B}_k^{\tau}, \qquad \widetilde{B}_k = G_k \bar{B}_k^{\zeta}$$

Both formulae have the same structure: Inflation is a product of basic inflation, which is possibly leveraged by a positive exponent, with a second component, which one can in many cases interpret as *superimposed inflation*. The natural assumption would be that basic inflation and the superimposed inflations are independent. In practice often at least one of the exponents equals 1 and at least one of the superimposed inflations is a trivial index of 1's, yielding a much simpler model.

Notice that this *inflation decomposition* depends essentially on what the official volume is. For illustration we revisit the examples mentioned above.

- Aggregate sum insured (Property insurance): This case is benign. One would typically expect all three inflations to equal construction cost inflation, such that $\zeta = 1 = \tau$, $G \equiv 1 \equiv H$. In the case of layer business practitioners in some situations suspect large losses to have a somewhat higher inflation than average. This can be catered for either by leveraging basic inflation via say $\tau = 1.1$, or by say 0.4% extra inflation per year (which would be reflected by $H_k = 1.004^k$), or by a combination of both kinds of adjustments.

- Payroll: Typically basic and volume inflation coincide, reflecting the wages in the insured industry. Loss inflation could be the same, however, in case of layer business we already mentioned the considerable impact of medical inflation on personal injury losses, which are the main source of layer losses, at least for higher layers. One could come to the conclusion that here say 80% of the large-loss inflation is driven by wages, while 20% come from health cost increases that have nothing to do with wages. (A part of medical inflation comes directly or indirectly from wage increases, but there are many other sources.) Then we would set $\tau = 0.8$ and H_k would model the health cost component being unrelated to wages.

- Original premium. Think of Property reinsurance, a rather simple case, but complex enough. Here basic inflation will mostly equal construction cost inflation, while volume inflation splits in a natural way into basic inflation and the effects of the market cycle, so $\zeta = 1$ and G_k is an index reflecting *premium adequacy*. Such indices could be available on a countrywide basis (e.g. market loss ratios, being heavily smoothed to eliminate random effects), however, the challenge is the assessment of the premium adequacy for *individual* portfolios, which can deviate a lot from the market. Loss inflation would be modeled as described in the preceding example of direct Property insurance.

The last example motivates a slight correction, or rather generalization, of the model. A perfect rating of the original premium takes into account three kinds of changes over time:

- number of insured "units" (objects, people, etc.), reflected by $\bar{M}_k$
- loss frequency per unit, reflected by A_k
- loss inflation, reflected by B_k

Consequently we should ideally have

$$M_k = \bar{M}_k A_k B_k$$

such that $\widetilde{B}_k$ equals B_k and the indices G and H are trivial (equaling 1), as well as the exponents ζ and τ. However, in practice, even if the rating is perfect, adequate premiums often cannot be enforced, such that G is non-trivial. And even if the overall inflation is assessed correctly, inflation could be higher for the very large losses affecting layers, such that we may have a nontrivial H and/or τ. Overall we have

$$M_k = \bar{M}_k A_k \widetilde{B}_k$$

which compared to the initial formula for M_k contains the additional factor A_k.

To embrace all cases, whatever the official volume, in one formula, we set

$$M_k = \bar{M}_k A_k^{\delta} \widetilde{B}_k$$

where δ typically equals 0, unless M_k is the original premium, in which case δ typically equals 1.

All in all we have a model with five (!) *primordial indices* $\bar{M}$, A, G, H, $\bar{B}$ and four derived quantities M, V, $\widetilde{B}$, B. It does not seem too restrictive to assume the primordial indices as independent. This is no oversimplification. Albeit seeming a very easy case, it has a complex dependence among the quantities of interest, namely B, V, and M.

What complicates things is that one of the five primordial indices cannot be observed at all: $\bar{M}_k$ is "hidden" in M_k and has to be factored out:

$$\bar{M}_k = \frac{M_k}{A_k^{\delta} \widetilde{B}_k} = \frac{M_k}{A_k^{\delta} G_k \bar{B}_k^{\zeta}}$$

which leads to

$$V_k = \bar{M}_k A_k = \frac{M_k A_k^{1-\delta}}{G_k \bar{B}_k^{\zeta}}$$

This makes clear that although it is very plausible to assume the independence of the indices $V = \bar{M}A$ (frequency effects) and $B = H\bar{B}^{\tau}$ (scale effects), they are tied together by $\bar{B}$, which is needed to calculate the $\bar{M}_k$. What this means mathematically will be explained in Section 8.3.

Note that if, as done here, we write equations about indices, they are generally intended element-wise.

Our intertwined set of nine indices, which shall be called *comprehensive index system*, may appear over-complex, however, it is just a language for effects that appear in real-world situations and are far from remote. Further, as already stated, usually not all these effects appear together. E.g., if $H \equiv 1$, we have one index less, if further $\tau = 1$, loss inflation and basic inflation coincide, i.e. two indices less. The same simplification may occur for the volume inflation, albeit rarely in reinsurance with its both common and intricate volume, the premium income.

The following chart illustrates how the components of the comprehensive index system interact.

$$
\begin{array}{ccccccccccccccc}
H & & & & \bar{B} & & & & G & & & & & & \\
 & \searrow & & \swarrow & & \searrow & & \swarrow & & & & & & & \\
 & & B & & & & \widetilde{B} & & & & \bar{M} & & & & A \\
 & & & & & & & \searrow & & \swarrow & \dashleftarrow & \searrow & \dashleftarrow & \swarrow & \\
 & & & & & & & & M & & & & V & &
\end{array}
$$

This structure consists of two W's, being tied at $\widetilde{B}$. Each W describes the interrelation of two quantities: The upper left one relates loss vs official-volume inflation via $\bar{B}$, the lower right one relates official vs frequency volume via $\bar{M}$.

The comprehensive index system can elegantly embrace the previous case of an inflation-free official volume. We do so via the following parametrization: Set $\zeta = 0$, $\tau = 1$, and $G \equiv 1 \equiv H$. Then $\widetilde{B} \equiv 1$, which means exactly that the official volume does not contain any inflationary effects, and further we have $\bar{M} = M$ and $\bar{B} = B$.

We shall treat both kinds of official volumes together, in the language of the second one, bearing in mind that the formulae include the first case, despite his totally different characteristics – here scale and volume effects are neatly separated, with no intricate connection via the basic inflation.

8.1.3 Including Cat events

Now we derive an analogous model for accumulation losses. Such losses are aggregates over all losses stemming from the same event, say an earthquake or windstorm. As already explained in Section 2.2.2, scale and volume are fundamentally different for such losses: Not the single losses are counted, but the accumulation events. If a risk/portfolio grows due to an increasing number of insured units, this does not increase the number of catastrophes, but their size (affecting more units). Growth of the number of units is an *inflationary* effect here, just as growth of the unit size or pure monetary inflation.

Summing up, the loss frequency varies only if the frequency of catastrophic events varies, independently of portfolio size changes. As for frequency modeling, the portfolio is a single unit! On the other hand, all portfolio changes (in size and/or number of units) relate to the size of losses, i.e. to inflation, as does inflation in the classical sense.

If we express this in the terminology of the comprehensive index system, with B describing the inflation of the single losses constituting a Cat loss and A being the index of the Cat-loss frequency, we get

$$ {}^{cat}M = M $$

$$ {}^{cat}\bar{M} \equiv 1, \qquad {}^{cat}V = A $$

$$ {}^{cat}B = \bar{M}B $$

This leads us to a common representation of the models for single losses and accumulation losses, decomposing official volume, frequency volume, and loss inflation into primordial indices: If χ is an indicator variable equaling 0 for single-loss models and 1 for Cat-loss models, we have

$$M = \bar{M} A^{\delta} \widetilde{B}, \qquad \widetilde{B} = G \bar{B}^{\zeta}, \qquad V = \bar{M}^{1-\chi} A, \qquad B = \bar{M}^{\chi} H \bar{B}^{\tau}$$

The only difference between the two kinds of loss models in this representation is the position of the deflated volume, which in the Cat case is a factor of the loss inflation index, not of the frequency volume. In both cases, however, it is clear that if the five primordial indices are independent, so are B and V.

8.2 Observable indices

Now we adapt to the comprehensive index system what we have developed for inflation and frequency volume: the decomposition into picked index and gap.

The official volume M stands out in this context: It is always directly observable – otherwise it could not be used to contractually define the premium of (re)insurance covers.

Instead, for the four primordial indices A, G, H, and $\bar{B}$ we would typically expect the same situation as explained for the overall loss inflation B at the beginning of this book, before decomposing the indices: They will be impossible or at least very hard to observe directly, such that usually the best (or the only) option is the approximation by a picked index. This is exactly how they would be treated in practice: One carefully selects (or constructs) a picked index for each of them, and if this is done appropriately, it is reasonable to assume that the resulting gaps are time series that fluctuate rather few (thus after normalization are close to 1).

Having introduced the four picked indices IA, IG, IH, and $I\bar{B}$, which we call *primordial picked indices*, and the corresponding gaps, which we call *primordial gaps*, via

$$\Delta A = \frac{A}{IA}, \qquad \Delta G = \frac{G}{IG}, \qquad \Delta H = \frac{H}{IH}, \qquad \Delta \bar{B} = \frac{\bar{B}}{I\bar{B}}$$

the system is already complete: Recall that the official volume can be observed explicitly, without the help of a picked index. Technically this means that we use M as its own picked index IM, such that for the resulting gap we have $\Delta M \equiv 1$. Now five triples of true index, picked index, and gap, are specified and yield the remaining quantities.

We explain this in detail. The four equations at the end of the preceding section are the basic model assumptions connecting the five *primordial true indices* $\bar{M}$, A, G, H, and $\bar{B}$ with the four other true indices. The analogous equations

$$IM = I\bar{M} IA^{\delta} I\widetilde{B}, \qquad I\widetilde{B} = IG I\bar{B}^{\zeta}, \qquad IV = I\bar{M}^{1-\chi} IA, \qquad IB = I\bar{M}^{\chi} IH I\bar{B}^{\tau}$$

connect the five *observable picked indices* $IM = M$, IA, IG, IH, and $I\bar{B}$ with the four remaining picked indices, which means that the observable picked indices *define* the other ones, in particular

$$I\bar{M} := \frac{M}{IA^{\delta} I\widetilde{B}}$$

Remark 8.1. One could think of a variant where IB and $I\widetilde{B}$ are observed instead of IH and IG, such that the latter are calculated out of the former. This alternative model can be treated analogously, as long as the same independence structure holds.

Finally, from the definition of the gaps we get immediately

$$\Delta M = \Delta\bar{M} \Delta A^{\delta} \Delta\widetilde{B}, \quad \Delta\widetilde{B} = \Delta G \Delta\bar{B}^{\zeta}, \quad \Delta V = \Delta\bar{M}^{1-\chi} \Delta A, \quad \Delta B = \Delta\bar{M}^{\chi} \Delta H \Delta\bar{B}^{\tau}$$

These equations make clear that if the primordial gaps are time series fluctuating few over time, this holds for all other gaps, too.

Let us wrap up what we have introduced step by step.

Definition 8.2. The *comprehensive index system* consists of three kinds of indices, firstly nine *true indices* being connected by the four *generating equations*

$$M = \bar{M}A^{\delta}\widetilde{B}, \qquad \widetilde{B} = G\bar{B}^{\zeta}, \qquad V = \bar{M}^{1-\chi}A, \qquad B = \bar{M}^{\chi}H\bar{B}^{\tau}$$

having nonnegative exponents, with $\tau > 0$ and $\chi \in \{0,1\}$. The five indices $\bar{M}$, A, G, H, $\bar{B}$ are called *primordial* true indices; the other ones result as products of powers thereof.

Further there are nine corresponding *picked indices* IA, IG, etc., approximating the true ones, for which the analogous generating equations hold. IM and M coincide. The four indices IA, IG, IH, $I\bar{B}$ are called *primordial picked indices*. Together with $IM = M$ they are observable. The remaining picked indices result as products of powers of these five.

Finally, there are nine corresponding *gaps* or *basis risks*, defined by $\Delta A = \frac{A}{IA}$, $\Delta G = \frac{G}{IG}$, etc., such that the analogous generating equations hold. The gaps corresponding to the primordial picked indices are called *primordial gaps*. $\Delta M \equiv 1$, thus all gaps can be calculated out of the four primordial ones.

For an overview look at the following table, displaying the essential index properties.

Index	primordial	approximable by picked index	characteristic
B			scale
V			frequency
M		directly observable	possibly both
$\widetilde{B}$			scale
$\bar{B}$	x	x	scale
H	x	x	scale
G	x	x	scale
$\bar{M}$	x		frequency or scale
A	x	x	frequency

For further illustration we sketch how the experience rating procedure in practice typically works: As mentioned in Section 2.1.2, according to the kind of coverage, the per-year aggregates or the large losses, respectively, are provided for a number of years k, $k_{min} \le k \le k_{max} \le 0$, together with the corresponding official volumes M_k (usually provided at least until the year 0, $k_{min} \le k \le 0$) and mostly also a prediction $\widehat{M_q}$ of the future volume. If the latter is questionable, or not provided at all, it has to be (re)assessed by the pricing actuary, which can be done in two ways.

- A common option is to calculate $\widehat{M_q}$ directly from the (often short and erratic) time series $M_{(k\le 0)}$ (possibly taking additional information about growth perspectives etc. into account).

  Then the M_k are used a second time to get the frequency volume, firstly for the past, which involves the picked indices: $I\widetilde{B}_k = IG_k I\bar{B}_k^{\zeta}$ (proxy for $\widetilde{B}_k$) deflates $IM_k = M_k$, yielding (possibly together with IA_k)

$$I\bar{M}_k = \frac{M_k}{IA_k^{\delta} I\widetilde{B}_k} = \frac{M_k}{IA_k^{\delta} IG_k I\bar{B}_k^{\zeta}}$$

  which is then adjusted by IA_k (proxy for A_k) to get $IV_k = I\bar{M}_k^{1-\chi} IA_k$. (In the Cat case $\chi = 1$ we do not need $I\bar{M}_k$ here, but later for the loss inflation.)

To predict future index values, it is straightforward to stay with the developed multiplicative approach and construct all predictors according to the comprehensive index system. The future frequency volume would accordingly be predicted by first predicting $I\bar{B}_q$, IG_q, IA_q from the (often long and not too volatile) time series $I\bar{B}_{(k\le 0)}$, $IG_{(k\le 0)}$, $IA_{(k\le 0)}$, then calculating

$$\widehat{I\bar{M}_q} = \frac{\widehat{M_q}}{\widehat{IA_q}^{\delta}\,\widehat{IG_q}\,\widehat{I\bar{B}_q}^{\zeta}}, \qquad \widehat{IV_q} = \widehat{I\bar{M}_q}^{1-\chi}\,\widehat{IA_q}$$

Now all volumes needed to describe the frequency effects are available.

- A somewhat more sophisticated option is to calculate at first the time series $I\bar{M}_{(k_{min}\le k\le 0)}$, which is as short as $M_{(k_{min}\le k\le 0)}$, but does not contain inflationary effects and the like (as long as the picked indices are well chosen), thus should be less erratic and easier to use for prediction. From this one directly calculates $\widehat{I\bar{M}_q}$.
  Subsequently one calculates, with the help of the predictions of the other picked indices, according to the multiplicative structure:

  $$\widehat{IV_q} = \widehat{I\bar{M}_q}^{1-\chi}\,\widehat{IA_q}, \qquad \widehat{M_q} = \widehat{I\bar{M}_q}\,\widehat{IA_q}^{\delta}\,\widehat{IG_q}\,\widehat{I\bar{B}_q}^{\zeta}$$

  The advantage of this variant is that it should be easier to analyze and predict separately the picked indices whose product yields M, maybe each with a different time series model and prediction method, than to model directly the M_k, which constitute the product of various time series.

As for scale effects, the historical losses have to be adjusted by inflation. To do this, the time series IH_k is used to predict IH_q and we get $\widehat{IB_q} = \widehat{I\bar{M}_q}^{\chi}\,\widehat{IH_q}\,\widehat{I\bar{B}_q}^{\tau}$. Then the losses of the year k are as-if adjusted by the factor $\frac{\widehat{IB_q}}{IB_k}$.

Up to this point this is a standard procedure of real-world experience rating – albeit written in uncommon notation. In practice the indices are usually based on the year q, not normalized to the year we have defined as 0, but the outcome is the same, as for the final outcome only ratios of index values matter. As mentioned, our year 0 has the (mathematical) advantage to split the earlier years, where index uncertainty is just a gap, from the later years, where the whole index is unknown.

The essential ingredient of the comprehensive index system is notably the *explicit* decomposition of the two kinds of inflation (losses / official volume) into their common part $\bar{B}$ and the superimposed inflations. This is essential to understand the intricate dependence structure, as we will see now.

8.3 Dependence

Let us look at the gap of the deflated volume. In the equation $\Delta M = \Delta\bar{M}\Delta A^{\delta}\Delta\widetilde{B}$ the LHS is a constant index of 1's, hence

$$\Delta\bar{M} = \frac{1}{\Delta A^{\delta}\Delta\widetilde{B}} = \frac{1}{\Delta A^{\delta}\Delta G\Delta\bar{B}^{\zeta}}$$

This makes clear how frequency volume and loss inflation interact: The corresponding *true* indices may be assumed independent, but if we predict them in the outlined way via *picked* indices, involving the official volume, the resulting gaps

$$\Delta V = \Delta\bar{M}^{1-\chi}\Delta A = \Delta A^{1-(1-\chi)\delta}\Delta G^{\chi-1}\Delta\bar{B}^{(\chi-1)\zeta}$$

$$\Delta B = \Delta\bar{M}^{\chi}\Delta H\Delta\bar{B}^{\tau} = \Delta A^{-\chi\delta}\Delta G^{-\chi}\Delta H\Delta\bar{B}^{\tau-\chi\zeta}$$

can be independent, but in general are not. Here it makes sense to look at the two fundamental loss model types separately. For single losses ($\chi = 0$) we have

$$\Delta V = \Delta A^{1-\delta} \Delta G^{-1} \Delta \bar{B}^{-\zeta}, \qquad \Delta B = \Delta H \Delta \bar{B}^{\tau}$$

Inflation and frequency volume gap are independent if the official volume is inflation-free ($\zeta = 0$), otherwise both gaps are functions of $\Delta \bar{B}$.

The accumulation-loss formulae look quite different:

$$\Delta V = \Delta A, \qquad \Delta B = \Delta A^{-\delta} \Delta G^{-1} \Delta H \Delta \bar{B}^{\tau-\zeta}$$

Here the two gaps are independent if $\delta = 0$. While this holds for most official volumes, Cat losses are most relevant in reinsurance, where the official volume is often the original premium, such that $\delta = 1$. So we have dependence in most Cat reinsurance situations. Notice that in the (not uncommon) case $\tau = \zeta$ the basic inflation gap $\Delta \bar{B}$ drops out.

Coming back to plausible (and desirable) independence properties, recall that the natural setting was to assume the primordial true indices $\bar{B}$, A, G, H, and $\bar{M}$ as independent. Further we need to specify the interaction between true indices, picked indices, and gaps, notably only for primordial indices – this will specify the properties of all other indices. We aim to make only plausible and not too many assumptions. However, we would like to have that each primordial gap be independent of the corresponding picked index – this is a natural way to express mathematically that there is no systematic error in the way the primordial picked indices are chosen as approximations of the respective true indices.

If we want to reconcile the two desired independence structures, that for true indices and that between primordial gap and respective picked index, the straightforward way is as follows:

Definition 8.3. We say the *comprehensive index system* is *independent* if the nine indices $I\bar{B}$, $\Delta\bar{B}$; IA, ΔA; IG, ΔG; IH, ΔH; and $\bar{M}$ are independent. These indices are called the *primordial components* of the independent comprehensive index system.

Remark 8.4. It is natural to assume this property for the original indices, before rebasing them to the year 0. However, if the original index system is independent, the normalized index system inherits this property, as each normalized index is just a measurable function of the underlying original index.

Assorting the primordial components, we have: the four primordial picked indices, the four primordial gaps, and the true deflated volume. All true/picked indices and gaps can be written as products of (powers of) primordial indices. We call this *primordial decomposition* and will use it a myriad of times.

The following chart illustrates the structure of the independent index system for the non-Cat case and reveals most primordial decompositions at a glance.

$$\begin{array}{ccccccccccccccccc}
IH & & \Delta H & & I\bar{B} & & \Delta\bar{B} & & IG & & \Delta G & & & & IA & & \Delta A \\
& \curlyvee & & & & \curlyvee & & & & \curlyvee & & & \bar{M} & & & \curlyvee & \\
& H & & & & \bar{B} & & & & G & & & & & & A & \\
& & & & & & \searrow & & \swarrow & & & & & & & & \\
& & \searrow & & \swarrow & & & \widetilde{B} & & & & & \curlywedge & \swarrow & \swarrow & & \\
& & & & & & & & & \searrow & & & & & & & \\
& & & B & & & & & & & & M & & V & & &
\end{array}$$

In the independent comprehensive index system the five primordial true indices $\bar{B}$, A, G, H, and $\bar{M}$ are independent, as they can be written as disjoint products of the primordial indices: $A = IA\,\Delta A$, $H = IH\Delta H$, ..., $\bar{M}$. The five corresponding picked indices $I\bar{B}$, IA, IG, IH, and $I\bar{M}$ – we call them the *basic picked indices* – are independent, too, as

$$I\bar{M} = \frac{\bar{M}}{\Delta\bar{M}} = \bar{M}\Delta A^{\delta}\Delta G\Delta\bar{B}^{\zeta}$$

The last formula makes clear that, while in the independent comprehensive index system the (unobservable) *true* deflated volume $\bar{M}$ is independent of the primordial gaps (and thus of all gaps), the *picked* deflated volume $I\bar{M}$ is *not*. This is maybe not intuitive, but follows directly from the construction of $I\bar{M}$ and is essentially a (unwelcome) consequence of the unobservability of $\bar{M}$ and the need to approximate it indirectly via a number of picked indices. For the same reason (and as counterintuitive and unwelcome) we may have dependence between the (non-primordial) frequency volume gap and the corresponding picked index, and the same may occur for the (non-primordial) loss inflation.

To see this in detail, we again look at single-loss and accumulation-loss models separately, starting with the former. Here we have

$$IV = I\bar{M}IA = \bar{M}IA\,\Delta A^{\delta}\Delta G\Delta\bar{B}^{\zeta}, \qquad \Delta V = \Delta A^{1-\delta}\Delta G^{-1}\Delta\bar{B}^{-\zeta}$$

$$IB = IHI\bar{B}^{\tau}, \qquad \Delta B = \Delta H\Delta\bar{B}^{\tau}$$

δ usually equals 0 or 1, such that ΔA appears only in one of the volume equations. However, ΔG and (if $\zeta \neq 0$) also $\Delta\bar{B}$ appear on both sides, creating dependence. Loss inflation gap and respective picked index are instead independent.

In Cat-loss models we have

$$IV = IA, \qquad \Delta V = \Delta A$$

$$IB = I\bar{M}IHI\bar{B}^{\tau} = \bar{M}IHI\bar{B}^{\tau}\Delta A^{\delta}\Delta G\Delta\bar{B}^{\zeta}, \qquad \Delta B = \Delta A^{-\delta}\Delta G^{-1}\Delta H\Delta\bar{B}^{\tau-\zeta}$$

Here it is the other way round: independence on the volume side and dependence via ΔG and (possibly) $\Delta\bar{B}$ between loss inflation gap and respective picked index. In both cases we have a somewhat complex dependence structure, however, the primordial decomposition makes this transparent and will help us carry on.

To prepare the evaluation of the formulae developed in the preceding chapter, recall that these depend essentially on a number of variants of the index $D(\alpha)$. We rewrite the latter in the language of the comprehensive index system, by using, for the sake of concise notation, the exponent $\omega := 1 + (\alpha - 1)\,\chi$, which equals 1 for single-loss models and α for Cat models.

$$D(\alpha) = VB^{\alpha} = \bar{M}^{\omega}AH^{\alpha}\bar{B}^{\tau\alpha} = M^{\omega}A^{1-\delta\omega}G^{-\omega}H^{\alpha}\bar{B}^{\tau\alpha-\zeta\omega} = \bar{M}^{\omega}IA\,IH^{\alpha}I\bar{B}^{\tau\alpha}\Delta A\Delta H^{\alpha}\Delta\bar{B}^{\tau\alpha}$$

$$\begin{aligned} ID(\alpha) = IVIB^{\alpha} = I\bar{M}^{\omega}IA\,IH^{\alpha}I\bar{B}^{\tau\alpha} &= M^{\omega}IA^{1-\delta\omega}IG^{-\omega}IH^{\alpha}I\bar{B}^{\tau\alpha-\zeta\omega} \\ &= \bar{M}^{\omega}IA\,IH^{\alpha}I\bar{B}^{\tau\alpha}\Delta A^{\delta\omega}\Delta G^{\omega}\Delta\bar{B}^{\zeta\omega} \end{aligned}$$

$$\Delta D(\alpha) = \Delta V\Delta B^{\alpha} = \Delta A^{1-\delta\omega}\Delta G^{-\omega}\Delta H^{\alpha}\Delta\bar{B}^{\tau\alpha-\zeta\omega}$$

For each index more than one representation is useful, according to whether one is focused on observability or on independence. The respective last representations are the primordial decompositions.

Notice that for $\alpha = 1$ the formulae for single-loss and Cat models coincide.

To get more intuition we quickly look at the special case that basic inflation and loss inflation coincide and the volume inflation is not leveraged: $H \equiv 1$, $\tau = 1$, $\zeta = 1$. We see at a glance that now basic inflation enters the formula for $\Delta D(\alpha)$ with the leverage $\alpha - \omega$ (which can equal 0 for single losses and always equals 0 in the Cat case), while in case of an inflation-free official volume ($\zeta = 0$) the leverage is $\alpha > 0$. This makes clear that the impact of inflation basis risk is fundamentally different for the two types of official volumes, those containing inflation and those not.

Note finally that even the setting of the indices in the basic example is a special case of the comprehensive index system, however, a very particular one. Here the only shift over time is inflation, so the size of the risk does not vary, nor does the frequency per volume unit. Thus, $\bar{M}$ and A are trivial constant indices. Notably these invariances are *known*, such that $I\bar{M}$ and IA are the same trivial constant indices. Thus, $\Delta\bar{M} \equiv 1 \equiv \Delta A$. This seems to contradict the above formula $\Delta\bar{M}^{-1} = \Delta A^{\delta}\Delta G\Delta\bar{B}^{\zeta}$, but does not: The official volume is constant here and as such a special case of an inflation-free official volume. As mentioned earlier, this case enters the index system via $\zeta = 0$, $\tau = 1$, and $G \equiv 1 \equiv H$, which implies $\Delta G \equiv 1 \equiv \Delta H$. The only non-trivial index basis risk here is thus $\Delta\bar{B} = \Delta B$.

Definition 8.5. We call *trivial volume case* the special case of the comprehensive index system where official and frequency volume are both knowingly constant. Here the only random index category (with picked index and gap) is $\bar{B} = B$, while the remaining 7 true indices, as well as their picked indices and gaps, are trivial constant indices.

So far in this chapter, we have looked at indices in general, not at the particular ones rebased to the year 0. From now we again differentiate original and normalized indices, using the underscore for the original indices before normalization. For the latter we introduce a special case. Recall the normalized gaps in the basic example, which constituted a geometric random walk, such that, with the basis being the year 0, future and past values were independent. This property can (a bit more generally) be formulated for the comprehensive index system.

Definition 8.6. An independent *comprehensive index system* is called *strongly independent* if for each primordial gap the normalized future value is independent of the corresponding past time series. The normalized indices here split into a total of 13 independent RV's:
$I\bar{B}$, $\Delta\bar{B}_{(j\le 0)}$, $\Delta\bar{B}_q$; IA, $\Delta A_{(j\le 0)}$, ΔA_q; ..., ΔH_q; and $\bar{M}$.

8.4 Conditions

Having sorted out the complex interaction of indices, we want to specify the formulae developed in the last chapter, by applying the comprehensive index system instead of using $\underline{B}$ and $\underline{V}$. We will get pretty complex formulae, but this is unavoidable. It cannot be emphasized enough what the bottleneck in the observations is: the hidden deflated official volume, which we need to assess either the frequency volume (non-Cat case) or the overall loss inflation (Cat case). We typically only get (approximate) knowledge about $\bar{M}$ if we can calculate the past values $I\bar{M}_k$ from the $IM_k = M_k$, and this means knowing how to decompose the official volume. Further we cannot calculate any moments without specifying the interactions of the various inflationary effects, which means knowing how to decompose volume and loss inflation – ideally into independent components.

Formally, the fixed models/parameters and the conditions that we choose now are a special case of the model developed in the last chapter, as outlined at the end of Section 7.4.

- The exponents δ, ζ, and τ come in as new fixed values; they shall be assembled in: ι.
- The fixed gap parameter set $\Pi_V \cup \Pi_B$ is replaced by the corresponding one for the primordial gaps: $\Pi_{pr} := \Pi_A \cup \Pi_G \cup \Pi_H \cup \Pi_{\bar{B}}$. Each constellation of models and parameters for the primordial gaps

yields, together with the above three exponents, models and parameters for inflation and volume gap.

- As σ-algebra $\underline{I}$ we choose $\sigma\left(\Upsilon_{pr}, \underline{M}_{(j\leq 0)}\right)$, where

$$\Upsilon_{pr} := \left(\underline{IA}_{(j\leq 0)}, \underline{IG}_{(j\leq 0)}, \underline{IH}_{(j\leq 0)}, \underline{I\bar{B}}_{(j\leq 0)}\right)$$

is the vector of the observed primordial picked indices and $\underline{M}_{(j\leq 0)}$ is the observed official volume. As the past values of $\underline{IB}$ and $\underline{IV}$ can be calculated out of the respective observed values of the primordial picked indices and of $\underline{M}$, we have indeed $\sigma\left(\underline{IV}_{(j\leq 0)}, \underline{IB}_{(j\leq 0)}\right) \subseteq \underline{I}$. Intuitively, $\underline{I}$ is more detailed a knowledge of the index history than $\sigma\left(\underline{IV}_{(j\leq 0)}, \underline{IB}_{(j\leq 0)}\right)$.
Again $j \leq 0$ is a slightly imprecise notation – each index history in practice has its (finite) starting point. In the case of $\underline{M}$ this usually is k_{min}, for the other indices (fortunately) often much earlier.

Now we are able to rewrite/specify the results of Section 7.5, starting with the abbreviations introduced earlier:

$$d^* = \frac{d}{\widehat{I\bar{M}_q}^{\chi}\widehat{IH_q}\widehat{I\bar{B}_q}^{\tau}}, \qquad u^* = \frac{u}{\widehat{I\bar{M}_q}^{\chi}\widehat{IH_q}\widehat{I\bar{B}_q}^{\tau}}$$

$$\tilde{\theta} = \theta_{d^*}\widehat{I\bar{M}_q}^{1-\chi}\widehat{IA_q}, \qquad \tilde{e} = e^*\widehat{I\bar{M}_q}\widehat{IA_q}\widehat{IH_q}\widehat{I\bar{B}_q}^{\tau}$$

In the following summarizing theorem the terms $\widehat{IV_q}$ and IV_k appear. However, now they should not any more be seen as fundamental quantities, but instead as a compact notation for the products $\widehat{I\bar{M}_q}^{1-\chi}\widehat{IA_q}$ and $I\bar{M}_k^{1-\chi}IA_k$. Again for compactness, we write the future products $\Delta D(\alpha)_q\, ID(\alpha)_q$ as $D(\alpha)_q$, keeping in mind that ultimately we are interested in splitting picked indices and gaps.

Theorem 8.7. *Suppose a): A risk is in the inflationary world structure, the survival function $\overline{F}$ of the normalized loss is continuous at d^* and u^*, and the normalized primordial gaps referring to loss inflation are time series fluctuating closely about 1. Then the following approximations for the first two moments of the as-if loss hold ($k, l \leq 0$):*

$$E_{\Pi_{pr},\pi,\iota}\left(S_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \tilde{e}E\left(\Delta D\big(\alpha_{(d^*,u^*)}\big)_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)$$

$$Var\left(S_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \tilde{e}^2\Bigg\{(1+\beta)\,E\left(\Delta D\big(\alpha_{(d^*,u^*)}\big)_k^2 \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) - E^2\left(\Delta D\big(\alpha_{(d^*,u^*)}\big)_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)$$

$$+\frac{1}{IV_k}\max\left[\sum_{i=1}^{4} a_i E\left(\Delta D(\alpha_i)_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right), \sum_{i=1}^{j} a_i^{(j)} E\left(\Delta D\big(\alpha_i^{(j)}\big)_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)\right]\Bigg\}, \qquad j=1$$

$$Cov\left(S_k, S_l \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \tilde{e}^2\, Cov\left(\Delta D\big(\alpha_{(d^*,u^*)}\big)_k, \Delta D\big(\alpha_{(d^*,u^*)}\big)_l \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right), \qquad k \neq l$$

Suppose a) and b): $\overline{F}$ is differentiable at $d^* > 0$. Then in the preceding variance formula variant $j = 2$ holds as well. Further the following approximations for the first two moments of the as-if loss count hold:

$$E_{\Pi_{pr},\pi,\iota}\left(C_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \tilde{\theta}E\left(\Delta D(\alpha_{d^*})_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)$$

$$Var\left(C_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \tilde{\theta}^2 \Bigg\{ (1+\beta)\, E\left(\Delta D(\alpha_{d^*})_k^2 \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) - E^2\left(\Delta D(\alpha_{d^*})_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)$$

$$+ \frac{1}{IV_k}\left[\gamma\,(1+\beta)\, E\left(\Delta D(2\alpha_{d^*})_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) + \frac{1}{\theta_{d^*}} E\left(\Delta D(\alpha_{d^*})_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)\right]\Bigg\}$$

$$Cov\left(C_k, C_l \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \tilde{\theta}^2\, Cov\left(\Delta D(\alpha_{d^*})_k, \Delta D(\alpha_{d^*})_l \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right), \quad k \neq l$$

Suppose a) and c): The predictors of the future picked indices are constructed according to the multiplicative structure of the comprehensive index system and the future index values referring to loss inflation fluctuate, conditionally given $\left(\underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)$*, closely about their predictors. Then the following approximations for the first two moments of the future loss hold:*

$$E_{\Pi_{pr},\pi,\iota}\left(X_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \tilde{e}\,\frac{E\left(D\left(\alpha_{(d^*,u^*)}\right)_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)}{\widehat{ID\left(\alpha_{(d^*,u^*)}\right)}_q}$$

$$Var\left(X_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \tilde{e}^2 \Bigg\{ (1+\beta)\, \frac{E\left(D\left(\alpha_{(d^*,u^*)}\right)_q^2 \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)}{\widehat{ID\left(\alpha_{(d^*,u^*)}\right)}_q^{\,2}} - \frac{E^2\left(D\left(\alpha_{(d^*,u^*)}\right)_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)}{\widehat{ID\left(\alpha_{(d^*,u^*)}\right)}_q^{\,2}}$$

$$+ \frac{1}{\widehat{IV}_q} \max\left[\sum_{i=1}^{4} a_i \frac{E\left(D(\alpha_i)_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)}{\widehat{ID(\alpha_i)}_q}, \sum_{i=1}^{j} a_i \frac{E\left(D\left(\alpha_i^{(j)}\right)_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)}{\widehat{ID\left(\alpha_i^{(j)}\right)}_q}\right]\Bigg\}, \qquad j = 1$$

$$Cov\left(X_q, S_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \tilde{e}^2\,\frac{Cov\left(D\left(\alpha_{(d^*,u^*)}\right)_q, \Delta D\left(\alpha_{(d^*,u^*)}\right)_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)}{\widehat{ID\left(\alpha_{(d^*,u^*)}\right)}_q}$$

Suppose a), b), and c). Then in the preceding variance formula variant $j = 2$ holds as well. Further the following approximations for the first two moments of the future loss count hold:

$$E_{\Pi_{pr},\pi,\iota}\left(\bar{N}_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \tilde{\theta}\,\frac{E\left(D(\alpha_{d^*})_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)}{\widehat{ID(\alpha_{d^*})}_q}$$

$$Var\left(\bar{N}_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \tilde{\theta}^2 \Bigg\{ (1+\beta)\, \frac{E\left(D(\alpha_{d^*})_q^2 \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)}{\widehat{ID(\alpha_{d^*})}_q^{\,2}} - \frac{E^2\left(D(\alpha_{d^*})_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)}{\widehat{ID(\alpha_{d^*})}_q^{\,2}}$$

$$+ \frac{1}{\widehat{IV}_q}\left[\gamma\,(1+\beta)\, \frac{E\left(D(2\alpha_{d^*})_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)}{\widehat{ID(2\alpha_{d^*})}_q} + \frac{1}{\theta_{d^*}} \frac{E\left(D(\alpha_{d^*})_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)}{\widehat{ID(\alpha_{d^*})}_q}\right]\Bigg\}$$

$$Cov\left(\bar{N}_q, S_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \tilde{\theta}^2\,\frac{Cov\left(D(\alpha_{d^*})_q, \Delta D(\alpha_{d^*})_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)}{\widehat{ID(\alpha_{d^*})}_q}$$

For $d = 0$ *the loss count formulae hold with 0 taking the place of* α_{d^*}*. They are exact and do not require the assumptions on survival function and indices.*

Proof. The theorem essentially assembles the results of Section 7.5 for $\underline{I} = \sigma\left(\Upsilon_{pr}, \underline{M}_{(j\leq 0)}\right)$.

Recall that the proofs are only complete if "fluctuating closely about" means *almost surely.* Appendix B explores how the results can be extended to indices having (rare) larger deviations from 1. □

8.5 Rates on volume

Now that we have explored the ramifications of the comprehensive index system and, in particular, of the official volume, it is a good moment for the formal introduction of the *(aggregated) loss per (official) volume (unit)*, being also shortly called *loss rate* (provided the meaning is clear from the context):

$$Y_q := \frac{X_q}{M_q}$$

As mentioned earlier, for risks with variable size it is natural to look at this figure, which, when doing experience rating, is typically assumed to be constant over time (apart from the corrections for inflation etc. we are exploring in this book). The respective predictions of X_q and Y_q are closely related problems, but cannot be expected to be equivalent, as the uncertainty about M_q plays a different role. Both quantities are of interest – which one more, depends on the situation and is a rather practical than theoretical question. If in particular, as is common in both reinsurance and commercial insurance, the contractual (re)insurance premium is defined as a multiple (percentage, per mil, etc.) of the official volume, the loss per volume is typically more in the focus. It should generally be easier to handle than the future loss amount, as the uncertainty about the future volume can be expected to drop out, at least partly. Let us have a closer look.

8.5.1 Burning cost rate

For fixed k, S_k predicts X_q and is called the *as-if burning cost* of the year k. The straightforward corresponding predictor for the loss per volume $Y_q = \frac{X_q}{M_q}$ is

$$R_k := \frac{S_k}{\widehat{M_q}}$$

This quantity is called as-if burning cost, too, in particular in reinsurance. However, a more precise term is also in use: *as-if burning cost rate.* The above representation looks unfamiliar, however, some algebra transforms it into the common representation of burning cost rating. Recall that

$$S_k = \frac{\widehat{IV_q}}{IV_k} \sum_{i=1}^{N_k} {}^u_d \left(\frac{\widehat{IB_q}}{IB_k} Z_{k,i} \right)$$

If we plug in $\widehat{M_q} = \widehat{I\bar{M}_q}\,\widehat{IA_q}^{\delta}\,\widehat{I\widetilde{B}_q}$, $\widehat{IV_q} = \widehat{I\bar{M}_q}^{1-\chi}\,\widehat{IA_q}$, and subsequently $M_k = I\bar{M}_k IA_k^{\delta} I\widetilde{B}_k$, $IV_k = I\bar{M}_k^{1-\chi} IA_k$ we get

$$R_k = \frac{\widehat{IA_q}^{1-\delta}}{\widehat{I\bar{M}_q}^{\chi}\,\widehat{I\widetilde{B}_q}} \frac{\sum_{i=1}^{N_k} {}^u_d \left(\frac{\widehat{IB_q}}{IB_k} Z_{k,i} \right)}{IV_k} = \left(\frac{I\bar{M}_k}{\widehat{I\bar{M}_q}} \right)^{\chi} \left(\frac{\widehat{IA_q}}{IA_k} \right)^{1-\delta} \frac{I\widetilde{B}_k}{\widehat{I\widetilde{B}_q}} \frac{\sum_{i=1}^{N_k} {}^u_d \left(\frac{\widehat{IB_q}}{IB_k} Z_{k,i} \right)}{M_k}$$

Both representations are intuitive, albeit not at first glance:

- The RHS shows the aggregate excess loss of the year k (with loss inflation taken into account) divided by the *official* volume of that year. Three preceding correction factors take account for changes from year k to year q: in the deflated volume, in the frequency per volume unit, and in the cost level contained in the official volume, respectively. The first two of these factors can drop out, namely if $\chi = 0$ and/or $\delta = 1$.

- The preceding representation, instead of the official volume, uses IV_k, the picked *frequency* volume. This volume will be relevant in a moment when we consider the classical weighting of the observed years in the sample mean. Notice that in this representation the preceding correction factor does not depend on k.

If we write the respective correction factors in the denominator, we finally get the classical structure of the burning-cost-rate formula: the sum of excess losses divided by a certain volume, which we can write in four different ways:

$$\widehat{M_q}\frac{IV_k}{\widehat{IV_q}} = \widehat{M_q}\frac{I\bar{M}_k^{1-\chi} IA_k}{\widehat{I\bar{M}_q}^{1-\chi}\widehat{IA_q}} = \frac{\widehat{I\bar{M}_q}^{\chi}\widehat{I\widetilde{B}_q}}{\widehat{IA_q}^{1-\delta}} IV_k = \left(\frac{\widehat{I\bar{M}_q}}{I\bar{M}_k}\right)^{\chi} \left(\frac{IA_k}{\widehat{IA_q}}\right)^{1-\delta}\frac{\widehat{I\widetilde{B}_q}}{I\widetilde{B}_k} M_k \tag{8.1}$$

The last variant can be interpreted as the as-if volume according to the situation of the year q (instead of the year 0): Via correction factors M_k is transformed such that it can work as a frequency volume for experience rating. If in the case of reinsurance M_k is (as usual) the original premium, then the so-transformed premium is exactly what practitioners (in particular for non-Cat covers) call the *on-level premium*.

Now it becomes transparent when, and in what sense, R_k is easier to handle than S_k – such that Y_q should be easier to predict than X_q:

$$R_k = \frac{\sum_{i=1}^{N_k} {}^{u}_{d}\left(\frac{\widehat{IB_q}}{IB_k} Z_{k,i}\right)}{\left(\frac{\widehat{I\bar{M}_q}}{I\bar{M}_k}\right)^{\chi} \left(\frac{IA_k}{\widehat{IA_q}}\right)^{1-\delta}\frac{\widehat{I\widetilde{B}_q}}{I\widetilde{B}_k} M_k}$$

depends essentially on observable quantities, namely the past (large) losses and official volumes. As for future values, we "only" need the relative change of four picked indices between the years k and q: $I\bar{M}$, IA, $I\widetilde{B} = IGIB^{\zeta}$, and $IB = I\bar{M}^{\chi}IHI\bar{B}^{\tau}$.

In single-loss models the intricate $I\bar{M}$ drops out, such that one just needs the relative change of the four primordial picked indices. These inputs might not be easy to gather, however, in many real-world situations there is a chance to rely on long, established, maybe market-wide index series, which should make the prediction of future values quite reliable.

Notice that in case of proportional business the two inflation indices B and $\widetilde{B}$ cancel nearly out, only their ratio needs to be assessed, which should be easier still. Inflation can even drop out altogether. E.g., in case of direct Fire insurance, if the official volume is the aggregate sum insured and if the sums insured are regularly updated with the help of say an appropriate construction cost index, the picked indices for loss inflation and volume inflation typically both coincide with said index. Then, to calculate R_k, which in this case would be called the *(burning cost) rate per sum insured*, one only needs an assessment of eventual changes in the frequency per volume unit.

Instead, $S_k = \widehat{M_q}R_k$ cannot be calculated just from observations and some well-selected picked indices. Here one absolutely needs an assessment of the future volume M_q, which might be much more uncertain than the future values of the primordial picked indices, having to rely ultimately on the (possibly short and erratic) time series of past volumes M_k.

It remains to discuss the case of Cat losses. Here R_k requires the future deflated volume, which is a factor in the loss inflation needed to adjust the loss history *as-if*. Loss inflation is indeed complex here, but closely tied to the official volume:

$$IB = \frac{M\,IHI\bar{B}^{\tau-\varsigma}}{IA^{\delta}IG}$$

In case $\varsigma = 1 = \tau$ or more generally $\zeta = \tau$, this equals $M\,IA^{-\delta}$ times correction factors for superimposed inflation. As generally in Cat models the denominator of R_k (Formula 8.1) simplifies to $\widehat{M_q}\frac{IA_k}{\widehat{IA_q}}$, we can say that here the burning cost rate deals essentially with the official volume and the index IA. If the latter is assumed constant (which is not uncommon), only the volume remains and the denominator is independent of k; here the IV_k are all equal.

In fact many reinsurance practitioners feel that Cat-loss experience rating means essentially: as-if adjusting by taking the official volume as loss index, and equal weighting of the years. This emerges here as a special case ($\zeta = \tau$, $IA \equiv 1$, $IG \equiv 1 \equiv IH$), while at the same time we see the various sophistications that can be included through the systematic of the comprehensive index system.

To conclude the introduction of the burning cost rate, we finally combine the predictors referring to the single past years. From the formula of the classical weighted sample mean

$$T\left(X_q\right) = \sum_{k=k_{min}}^{k_{max}} \frac{IV_k}{IV_+} S_k$$

where $IV_+ = \sum_{k=k_{min}}^{k_{max}} IV_k$, we get the analogous sample mean in terms of R_k

$$T\left(Y_q\right) = \frac{T\left(X_q\right)}{\widehat{M}_q} = \sum_{k=k_{min}}^{k_{max}} \frac{IV_k}{IV_+} \frac{S_k}{\widehat{M}_q} = \sum_{k=k_{min}}^{k_{max}} \frac{IV_k}{IV_+} R_k$$

Recalling from above that

$$IV_k R_k = \frac{\widehat{IA_q}^{1-\delta}}{\widehat{I\bar{M}_q}^{\chi} \widehat{I\widetilde{B}_q}} \sum_{i=1}^{N_k} {}_d^u \left(\frac{\widehat{IB_q}}{IB_k} Z_{k,i} \right)$$

we calculate, using in the last step again Formula 8.1,

$$\begin{aligned} T\left(Y_q\right) &= \sum_{k=k_{min}}^{k_{max}} \frac{\widehat{IA_q}^{1-\delta}}{\widehat{I\bar{M}_q}^{\chi} \widehat{I\widetilde{B}_q}} \frac{1}{IV_+} \sum_{i=1}^{N_k} {}_d^u \left(\frac{\widehat{IB_q}}{IB_k} Z_{k,i} \right) = \frac{\widehat{IA_q}^{1-\delta}}{\widehat{I\bar{M}_q}^{\chi} \widehat{I\widetilde{B}_q}} \frac{\sum_{k=k_{min}}^{k_{max}} \sum_{i=1}^{N_k} {}_d^u \left(\frac{\widehat{IB_q}}{IB_k} Z_{k,i} \right)}{\sum_{k=k_{min}}^{k_{max}} IV_k} \\ &= \frac{\sum_{k=k_{min}}^{k_{max}} \sum_{i=1}^{N_k} {}_d^u \left(\frac{\widehat{IB_q}}{IB_k} Z_{k,i} \right)}{\sum_{k=k_{min}}^{k_{max}} \left(\frac{\widehat{I\bar{M}_q}}{I\bar{M}_k} \right)^{\chi} \left(\frac{IA_k}{\widehat{IA_q}} \right)^{1-\delta} \frac{\widehat{I\widetilde{B}_q}}{I\widetilde{B}_k} M_k} \end{aligned}$$

This is finally the classical representation of the burning cost rate: The numerator is the (double) sum of all as-if excess losses across all years of the observation period, while the denominator is the sum of the as-if volumes over the same years. The weights of the single years, i.e. the IV_k (times a factor), drop out of the formula, such that only their sum remains.

Notice that in practice sometimes, as an alternative to the volume-weighted sample mean, the non-weighted average of the burning cost rates is used:

$$T^{equal}\left(Y_q\right) = \frac{1}{k_{max} - k_{min} + 1} \sum_{k=k_{min}}^{k_{max}} R_k$$

It does, however, not emerge into a formula as simple as the former variant (apart from the Cat case with trivial IA discussed above, where weighted and unweighted burning cost rate coincide). More importantly, equal weighting is not optimal in the sense of variance minimization, while the weighted variant would be optimal under the common assumption that variance grows linear with volume (and that the picked indices are the true ones), see the discussion in Section 2.2.2.

For the optimal weights with index basis risk we have, however, still some way to go.

8.5.2 Properties

We would like to have for R_k and Y_q results analogous to those assembled in Theorem 8.7 for S_k and X_q. For $R_k = \frac{S_k}{\widehat{M_q}}$ this is easy, as it deviates from S_k by the factor $\widehat{M_q}$, which is a (possibly intricate) measurable function of the historical picked index values, thus a known value according to the conditions of the theorem. With the abbreviation

$$\mathring{e} := \frac{\tilde{e}}{\widehat{M_q}} = e^* \frac{\widehat{IV_q}\widehat{IB_q}}{\widehat{M_q}} = e^* \frac{\widehat{IA_q}^{1-\delta}\,\widehat{IH_q}\widehat{I\bar{B}_q}^{\tau-\zeta}}{\widehat{IG_q}}$$

we get

Proposition 8.8. *For a risk in the inflationary world structure, if the survival function $\overline{F}$ of the normalized loss is continuous at d^* and u^*, and the normalized primordial gaps referring to loss inflation are time series fluctuating closely about 1, then with $\alpha = \alpha_{(d^*,u^*)}$ the following approximations for the first two moments of the as-if loss rate hold:*

$$E_{\Pi_{pr},\pi,\iota}\left(R_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \mathring{e} E\left(\Delta D(\alpha)_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)$$

$$\begin{aligned} Var\left(R_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \mathring{e}^2 \Bigg\{ &(1+\beta)\, E\left(\Delta D(\alpha)_k^2 \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) - E^2\left(\Delta D(\alpha)_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \\ &+ \frac{1}{IV_k} \max\left[\sum_{i=1}^{4} a_i E\left(\Delta D(\alpha_i)_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right), \sum_{i=1}^{j} a_i^{(j)} E\left(\Delta D\left(\alpha_i^{(j)}\right)_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)\right]\Bigg\} \end{aligned}$$

$$Cov\left(R_k, R_l \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \mathring{e}^2\, Cov\left(\Delta D(\alpha)_k, \Delta D(\alpha)_l \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right), \quad k \neq l$$

Variant $j = 2$ of the variance formula requires in addition that $\overline{F}$ be differentiable at $d^* > 0$.

Proof. Simply divide the corresponding formulae of Theorem 8.7 by the $\sigma\left(\underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)$-measurable $\widehat{M_q}$ or its square, respectively. □

The corresponding formulae for $Y_q = \frac{X_q}{M_q}$ require some extra steps and cannot be written just in terms of $D(\alpha)$ and variants. However, we can again get concise and intuitive formulae with the help of some further notation. As explained after Definition 7.11, the family of indices about $D(\alpha) = VB^\alpha$ essentially quantifies the *combined* impact of scale and frequency volume on the various moments of the as-if / future losses. Now that we look at loss rates per (official) volume, it seems natural to work with an analogous family of indices, quantifying said impact *per volume*. According to this logic we define an index family $J(\alpha) := \frac{D(\alpha)}{M}$ (element-wise quotients), such that we have

$$J(\alpha) = \frac{D(\alpha)}{M} = \bar{M}^{\omega-1} A^{1-\delta} G^{-1} H^\alpha \bar{B}^{\tau\alpha-\zeta} = M^{\omega-1} A^{1-\delta\omega} G^{-\omega} H^\alpha \bar{B}^{\tau\alpha-\zeta\omega}$$

Notice that in single-loss models (where $\omega = 1$) $J(\alpha)$ is independent of the deflated volume $\bar{M}$, it only depends on the various inflation components and the frequency per volume.

Analogously we define picked indices $IJ(\alpha) := \frac{ID(\alpha)}{IM} = \frac{ID(\alpha)}{M}$ etc. As the official volume has no gap ($\Delta M \equiv 1$), $\Delta J(\alpha)$ and $\Delta D(\alpha)$ coincide. We will always use the latter, bearing in mind that it is adequate for dealing with both losses and loss rates.

For ease of notation we define a last index family $K(\alpha) := \frac{D(\alpha)}{M^2} = \frac{J(\alpha)}{M}$. Here again the gap $\Delta K(\alpha)$ coincides with $\Delta D(\alpha)$.

Analogously to $\widehat{ID(\alpha)}_q$, we name the straightforward predictors following the multiplicative structure of the comprehensive index system $\widehat{IJ(\alpha)}_q$ and $\widehat{IK(\alpha)}_q$.

In order to get simple and intuitive formulae, we have introduced step by step a lot of notation. For orientation see Appendix C providing an overview of index families and predictors.

Now recall the *conditional* moments of X_q, which are assembled before Proposition 7.15. We restate them in the language of the comprehensive index system, again leaving for the moment the lower variance bound aside, with $\alpha = \alpha_{(d^*,u^*)}$:

$$\mathrm{E}\left(X_q \mid \mathcal{W}\right) \approx \tilde{e} \varDelta D(\alpha)_q \frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q}$$

$$\mathrm{Var}\left(X_q \mid \mathcal{W}\right) \approx \tilde{e}^2 \left\{ \beta \varDelta D(\alpha)_q^2 \frac{ID(\alpha)_q^2}{\widehat{ID(\alpha)}_q^{\,2}} + \frac{1}{\widehat{IV}_q} \sum_{i=1}^{4} a_i \varDelta D(\alpha_i)_q \frac{ID(\alpha_i)_q}{\widehat{ID(\alpha_i)}_q} \right\}$$

$$\mathrm{Cov}\left(X_q, S_k \mid \mathcal{W}\right) = 0$$

The corresponding formulae for $Y_q = \frac{X_q}{M_q}$ result by dividing by M_q, M_q^2, and $M_q \widehat{M}_q$, respectively, which are $\mathcal{W}$-measurable.

$$\mathrm{E}\left(Y_q \mid \mathcal{W}\right) \approx \frac{\tilde{e}}{M_q} \varDelta D(\alpha)_q \frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q} = \frac{\tilde{e}}{\widehat{M}_q} \varDelta D(\alpha)_q \frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q} \frac{\widehat{M}_q}{M_q} = \mathring{e} \varDelta D(\alpha)_q \frac{IJ(\alpha)_q}{\widehat{IJ(\alpha)}_q}$$

$$\begin{aligned}\mathrm{Var}\left(Y_q \mid \mathcal{W}\right) &\approx \frac{\tilde{e}^2}{\widehat{M}_q^{\,2}} \left\{ \beta \varDelta D(\alpha)_q^2 \frac{ID(\alpha)_q^2}{\widehat{ID(\alpha)}_q^{\,2}} \frac{\widehat{M}_q^{\,2}}{M_q^2} + \frac{1}{\widehat{IV}_q} \sum_{i=1}^{4} a_i \varDelta D(\alpha_i)_q \frac{ID(\alpha_i)_q}{\widehat{ID(\alpha_i)}_q} \frac{\widehat{M}_q^{\,2}}{M_q^2} \right\} \\ &= \mathring{e}^2 \left\{ \beta \varDelta D(\alpha)_q^2 \frac{IJ(\alpha)_q^2}{\widehat{IJ(\alpha)}_q^{\,2}} + \frac{1}{\widehat{IV}_q} \sum_{i=1}^{4} a_i \varDelta D(\alpha_i)_q \frac{IK(\alpha_i)_q}{\widehat{IK(\alpha_i)}_q} \right\}\end{aligned}$$

$$\mathrm{Cov}\left(Y_q, R_k \mid \mathcal{W}\right) = 0$$

The notation with the newly introduced index families reveals the structural analogy between the respective conditional formulae for X_q and Y_q: The constant $\tilde{e}$ is replaced by $\mathring{e}$, while the factor $\frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q}$, which quantifies the relative deviation of the true index value $D(\alpha)_q$ from its predictor, is replaced by the corresponding factors for J and K. The latter appears only in the sum at the end of the variance formula.

Proposition 8.9. *For a risk in the inflationary world structure, suppose the survival function $\overline{F}$ of the normalized loss is continuous at d^* and u^*, the normalized primordial gaps referring to loss inflation are time series fluctuating closely about 1, the predictors of the future picked indices are constructed according to the multiplicative structure of the comprehensive index system, and the future index values referring to loss inflation fluctuate, conditionally given $\left(\underline{M}_{(j\le 0)}, \varUpsilon_{pr}\right)$, closely about their predictors. Then with $\alpha = \alpha_{(d^*,u^*)}$ the following approximations for the first two moments of the future loss per volume hold ($k, l \le 0$):*

$$E_{\varPi_{pr},\pi,\iota}\left(Y_q \mid \underline{M}_{(j\le 0)}, \varUpsilon_{pr}\right) \approx \mathring{e} \frac{E\left(\varDelta D(\alpha)_q IJ(\alpha)_q \mid \underline{M}_{(j\le 0)}, \varUpsilon_{pr}\right)}{\widehat{IJ(\alpha)}_q}$$

$$Var\left(Y_q \mid M_{(j\le 0)}, \Upsilon_{pr}\right) \approx \mathring{e}^2 \left\{ (1+\beta)\, \frac{E\left(\Delta D(\alpha)_q^2\, IJ(\alpha)_q^2 \mid \underline{M}_{(j\le 0)}, \Upsilon_{pr}\right)}{\widehat{IJ(\alpha)_q}^{\,2}} - \frac{E^2\left(\Delta D(\alpha)_q\, IJ(\alpha)_q \mid \underline{M}_{(j\le 0)}, \Upsilon_{pr}\right)}{\widehat{IJ(\alpha)_q}^{\,2}} \right.$$

$$\left. + \frac{1}{\widehat{IV_q}} \max\left[\sum_{i=1}^{4} a_i \frac{E\left(\Delta D(\alpha_i)_q\, IK(\alpha_i)_q \mid \underline{M}_{(j\le 0)}, \Upsilon_{pr}\right)}{\widehat{IK(\alpha_i)_q}}, \sum_{i=1}^{4} a_i^{(j)} \frac{E\left(\Delta D\left(\alpha_i^{(j)}\right)_q IK\left(\alpha_i^{(j)}\right)_q \mid \underline{M}_{(j\le 0)}, \Upsilon_{pr}\right)}{\widehat{IK\left(\alpha_i^{(j)}\right)_q}} \right] \right\}$$

$$Cov\left(Y_q, R_k \mid M_{(j\le 0)}, \Upsilon_{pr}\right) \approx \mathring{e}^2 \frac{Cov\left(\Delta D(\alpha)_q\, IJ(\alpha)_q\,, \Delta D(\alpha)_k \mid \underline{M}_{(j\le 0)}, \Upsilon_{pr}\right)}{\widehat{IJ(\alpha)_q}}$$

Variant $j=2$ of the variance formula requires in addition that $\overline{F}$ be differentiable at $d^* > 0$.

Proof. The calculation is as in Proposition 7.15, using $\underline{I} = \sigma\left(\Upsilon_{pr}, \underline{M}_{(j\le 0)}\right)$. □

In the same way as done here with R_k, one can use $\frac{C_k}{\widehat{M_q}}$ as predictor for the future *loss count per official volume unit* $\frac{\bar{N}_q}{M_q}$, where the basic constant would be

$$\mathring{\theta} := \frac{\tilde{\theta}}{\widehat{M_q}} = \theta_{d^*} \frac{\widehat{IV_q}}{\widehat{M_q}} = \theta_{d^*} \frac{\widehat{IA_q}^{\,1-\delta}}{\widehat{I\bar{M}_q}^{\,\chi}\, \widehat{IG_q}\, \widehat{I\bar{B}_q}^{\,\zeta}}$$

We do not outline this in detail, noting that the results are analogous. They will be included in the comprehensive theorem coming in the next chapter.

Chapter 9

Consolidation

Before combining all theory developed so far, we need a (very technical) chapter to specify the index uncertainties for past and future values. Our aim is to split the interactions between gaps and picked indices as far as possible.

9.1 Past uncertainty

If we look at the representations of the as-if quantities in Theorem 8.7 and Proposition 8.8, we see that we have to deal with conditional expectations of past gaps $\mathit{\Delta} D(\alpha)_k$, given $\left(\underline{M}_{(j\leq 0)}, \varUpsilon_{pr}\right)$; for the variances and covariances we further need analogous expectations of $\mathit{\Delta} D(\alpha)_k^2$ and $\mathit{\Delta} D(\alpha)_k\, \mathit{\Delta} D(\alpha)_l$, all this for a number of alphas. We aim to relate these expectations to the corresponding unconditional ones, which should be easier to specify, at least if, as we assume throughout this chapter, the comprehensive index system is independent. Let $k \leq 0$ be fixed.

Lemma 9.1. *In the independent comprehensive index system we have*

$$E\left(\mathit{\Delta} D(\alpha)_k \mid \underline{M}_{(j\leq 0)}, \varUpsilon_{pr}\right) = E\left(\mathit{\Delta} D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}, \varUpsilon_{pr}\right) = E\left(\mathit{\Delta} D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right)$$

Proof. From $\varUpsilon_{pr} = \left(\underline{IA}_{(j\leq 0)}, \underline{IG}_{(j\leq 0)}, \underline{IH}_{(j\leq 0)}, \underline{I\bar{B}}_{(j\leq 0)}\right)$ we see that the conditions $\left(\underline{M}_{(j\leq 0)}, \varUpsilon_{pr}\right)$ and $\left(\underline{I\bar{M}}_{(j\leq 0)}, \varUpsilon_{pr}\right)$ are equivalent, i.e. generate the same σ-algebra: For fixed j, $\underline{M}_j$ and $\underline{I\bar{M}}_j$ can be calculated out of each other with the help of $\underline{IA}_j$, $\underline{IG}_j$, and $\underline{I\bar{B}}_j$ via the formulae constituting the comprehensive index system. This proves the first equality (which does not require independence of the index system). As for the second equality, recall that the four picked indices assembled in $\varUpsilon_{pr}$ are primordial and independent of all gaps, such that the condition on them can be dropped. □

The remaining condition cannot be ignored, as there is dependence via the gaps. We have the primordial decompositions

$$\mathit{\Delta} D(\alpha)_k = \mathit{\Delta} A_k^{1-\delta\omega} \mathit{\Delta} G_k^{-\omega} \mathit{\Delta} H_k^{\alpha} \mathit{\Delta} \bar{B}_k^{\tau\alpha-\zeta\omega}$$

$$\underline{I\bar{M}}_j = \underline{\bar{M}}_j \underline{\mathit{\Delta} A}_j^{\delta} \underline{\mathit{\Delta} G}_j \underline{\mathit{\Delta} \bar{B}}_j^{\zeta}$$

for any $j \leq 0$. With k and j being both negative, we must expect some dependence between these terms, even when $j \neq k$. (For orientation recall $\mathit{\Delta} A_k = \frac{\underline{\mathit{\Delta} A}_k}{\underline{\mathit{\Delta} A}_0}$ etc., such that altogether we deal with index values from the years j, k, and 0.) Nevertheless we will see that the approximation

$$\mathrm{E}\left(\mathit{\Delta} D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right) \overset{!}{\approx} \mathrm{E}\left(\mathit{\Delta} D(\alpha)_k\right)$$

is a fair one. First we decompose both sides of this formula as far as possible. For the RHS this is easy: In the independent comprehensive index system the primordial gaps are independent, such that

$$\mathrm{E}\left(\mathit{\Delta}D(\alpha)_k\right) = \mathrm{E}\left(\mathit{\Delta}A_k^{1-\delta\omega}\right)\mathrm{E}\left(\mathit{\Delta}G_k^{-\omega}\right)\mathrm{E}\left(\mathit{\Delta}H_k^{\alpha}\right)\mathrm{E}\left(\mathit{\Delta}\bar{B}_k^{\tau\alpha-\zeta\omega}\right)$$

The LHS is more intricate, but at least one factor can be split off. The following lemma will help us here and throughout this chapter.

Lemma 9.2. *Let the pair of RV's (A, X) be independent of the pair of RV's (B, Y). Then A and B are conditionally independent, given (X, Y), and we have*

$$E(AB \mid X, Y) = E(A \mid X, Y)\, E(B \mid X, Y) = E(A \mid X)\, E(B \mid Y)$$

$$E(AB \mid Y) = E(A)\, E(B \mid Y)$$

Proof. The conditional independence follows by applying twice the *weak union* rule for conditional independence, which states that if a RV is independent of a pair of RV's, one of the latter can be shifted into the condition. Hence, if $\perp$ symbolizes (possibly conditional) independence, we have

$$(A, X) \perp (B, Y) \Rightarrow (A, X) \perp B \mid Y \Rightarrow B \perp (A, X) \mid Y \Rightarrow B \perp A \mid (X, Y) \Rightarrow A \perp B \mid (X, Y)$$

Independence implies multiplicativity of expectations (both intended conditionally on (X, Y) here), hence the first equality of the first formula. The second equality follows from $\mathrm{E}(A \mid X, Y) = \mathrm{E}(A \mid X)$ and $\mathrm{E}(B \mid X, Y) = \mathrm{E}(B \mid Y)$, where we could drop some conditions due to independence.

The second formula is a special case of the first with a trivial constant RV X. □

We now use this second formula. In the independent comprehensive index system $\underline{\mathit{\Delta}H}$ is independent of the other primordial gaps and of $\underline{I\bar{M}}$, such that we can always decompose

$$\mathrm{E}\left(\mathit{\Delta}D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right) = \mathrm{E}\left(\mathit{\Delta}H_k^{\alpha}\right)\mathrm{E}\left(\mathit{\Delta}A_k^{1-\delta\omega}\mathit{\Delta}G_k^{-\omega}\mathit{\Delta}\bar{B}_k^{\tau\alpha-\zeta\omega} \mid \underline{I\bar{M}}_{(j\leq 0)}\right)$$

In some special cases we can factor out one or two more terms, according to the exponents applied to $\underline{\mathit{\Delta}A}$ and $\underline{\mathit{\Delta}\bar{B}}$, respectively, in the formulae for $\mathit{\Delta}D(\alpha)$ and $\underline{I\bar{M}}$: If $\delta = 0$ or $1 - \delta\omega = 0$, then $\underline{\mathit{\Delta}A}$ vanishes in $\mathit{\Delta}D(\alpha)$ or in the condition, such that it can be factored out, just as $\underline{\mathit{\Delta}H}$. The same holds for the basic inflation if $\zeta = 0$ or $\tau\alpha - \zeta\omega = 0$. But, whatever the exponents, $\underline{\mathit{\Delta}G}$ cannot be factored out this way, such there will be a "remainder" that is not easy to deal with, which we call $\underline{\mathit{\Delta}L}$. In this notation $\mathrm{E}\left(\mathit{\Delta}D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right)$ equals $\mathrm{E}\left(\mathit{\Delta}L_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right)$ times the *unconditional* expectation of the further factors of $\mathit{\Delta}D(\alpha)_k$.

For $\mathit{\Delta}L_k$, be it the product $\mathit{\Delta}A_k^{1-\delta\omega}\mathit{\Delta}G_k^{-\omega}\mathit{\Delta}\bar{B}_k^{\tau\alpha-\zeta\omega}$ or have it some factor less, we can, however, state that conditional and unconditional expectation should usually be close, such that the same must hold for $\mathit{\Delta}D(\alpha)_k$. We give a number of reasons, with increasing solidity:

- Averaging: Replacing a difficult-to-assess term by an expectation thereof, via (possibly moderate) averaging, is a well-established workaround, which was successfully employed e.g. in stochastic Chain Ladder models, see e.g. [Mack, 1993] and [Riegel, 2015]. In fact, brute averaging of $\mathrm{E}\left(\mathit{\Delta}L_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right)$ over $\underline{I\bar{M}}_{(j\leq 0)}$ yields the desired approximation by $\mathrm{E}\left(\mathit{\Delta}L_k\right)$.

However, after having paid, throughout this book, so much attention on (possibly tiny) basis risks and other subtleties, relying just on such a reasoning would be a bit unsatisfactory. We can do more, by taking advantage of the specific setting of the problem. At first we show that $\mathrm{E}\left(\mathit{\Delta}L_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right)$ and $\mathrm{E}\left(\mathit{\Delta}L_k\right)$ are close to each other in terms of (average) squared difference.

Lemma 9.3. *For two RV's X, Y we always have*

$$E\left([E(X \mid Y) - E(X)]^2\right) \leq Var(X)$$

such that in case $E(X) > 0$ we have the average squared relative deviation

$$E\left(\left[\frac{E(X \mid Y)}{E(X)} - 1\right]^2\right) \leq CV^2(X)$$

Proof. As $\mathrm{E}(\mathrm{E}(X \mid Y)) = \mathrm{E}(X)$, we have

$$\mathrm{E}\left([\mathrm{E}(X \mid Y) - \mathrm{E}(X)]^2\right) = \mathrm{Var}(\mathrm{E}(X \mid Y)) = \mathrm{Var}(X) - \mathrm{E}(\mathrm{Var}(X \mid Y)) \leq \mathrm{Var}(X)$$

The second formula follows immediately. □

- Squared difference: If we apply the lemma to $\varDelta L_k$ and $\underline{I\bar{M}}_{(j\leq 0)}$, we see that the average squared relative deviation of $\mathrm{E}\left(\varDelta L_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right)$ and $\mathrm{E}(\varDelta L_k)$ cannot exceed $\mathrm{CV}^2(\varDelta L_k)$, which is a very small figure due to the generally low fluctuation of the gaps.

We can explore our setting further.

- Heuristics: If we have a close look at the index values $\underline{I\bar{M}}_j = \underline{\bar{M}}_j \underline{\varDelta A}_j^{\delta} \underline{\varDelta G}_j \underline{\varDelta \bar{B}}_j^{\zeta}$, $j \leq 0$, we see that each of them is a particular product. Its first factor $\underline{\bar{M}}_j$, the true deflated volume, in practice should mostly fluctuate materially over time, however, in the independent comprehensive index system it is independent of $\varDelta L_k$. Conversely, the remaining factors are in general not independent of $\varDelta L_k$, but can in practice be expected to vary much less than $\underline{\bar{M}}_j$, from which they are furthermore independent. Thus, in a lot of situations the condition on $\underline{I\bar{M}}_j$ is mainly influenced by the value taken by $\underline{\bar{M}}_j$, a random variable being independent of the gaps – in a way we could say that the conditions $\underline{I\bar{M}}_j = m_j$ and $\underline{\bar{M}}_j = m_j$ are similar, must at least lead to similar expectations. Then it is more than plausible that dropping the condition $\underline{I\bar{M}}_{(j\leq 0)}$ in $\mathrm{E}\left(\varDelta L_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right)$ does not alter the result significantly.

This heuristics can be made rigorous for an important special case.

- Lognormal case: If the time series appearing here, namely the primordial gaps (before normalization) and $\underline{\bar{M}}$, are standard time series like ARIMA with lognormal error, and if we, as is usual in time series analysis, condition on some remote (i.e. far in the past) initial values for each of these time series, then the yearly values of all appearing time series are lognormally distributed and we can apply properties of the multivariate normal distribution as follows.

Example 9.4. Let us calculate conditional expectation and variance of the normally distributed RV $\ln(\varDelta L_k)$, given $\underline{I\bar{M}}_{(j\leq 0)}$ or equivalently the (element-wise) logarithm thereof. We show that conditional and unconditional moments are very close, such that the same holds for the expected value of the lognormal RV $\varDelta L_k$.

We need some notation and define the vector $\overrightarrow{Y}$ by setting, for each $j \leq 0$ appearing in the condition,

$$Y_j := \ln \underline{I\bar{M}}_j = \sum_{i=0}^{3} X_{ji}$$

where $X_{j0} := \ln \underline{\bar{M}}_j$,
$X_{j1} := \ln \underline{\varDelta A}_j^{\delta} = \delta \ln \underline{\varDelta A}_j$,

$X_{j2} := \ln \underline{\Delta G}_j$,
$X_{j3} := \ln \underline{\Delta \bar{B}}_j^{\zeta} = \zeta \ln \underline{\Delta \bar{B}}_j$.

Then we have

$$Y_* := \ln \Delta L_k = \ln \underline{\Delta L}_k - \ln \underline{\Delta L}_0 = \sum_{i=1}^{3} c_i \left(X_{ki} - X_{0i}\right)$$

where $c_1 := \frac{1-\delta\omega}{\delta}$, $c_2 := -\omega$, $c_3 := \frac{\tau\alpha-\zeta\omega}{\zeta}$.

Notice that the sums are independent sums due to the independence of the index system.

If $\underline{\Delta L}$ has less than three factors, the last formula holds if we modify as follows:
If $\delta = 0$ or $1 - \delta\omega = 0$, we set X_{j1} and c_1 to 0.
If $\zeta = 0$ or $\tau\alpha - \zeta\omega = 0$, we set X_{j3} and c_3 to 0.

All appearing logarithmized time series elements are linear combinations of the underlying primordial quantities, which are independent and lognormally distributed. Hence, the RV $\begin{pmatrix} Y_* \\ \overrightarrow{Y} \end{pmatrix}$ is multivariate normal. We call its expectation $\begin{pmatrix} \mu_* \\ \overrightarrow{\mu} \end{pmatrix}$ and its variance $\begin{pmatrix} \sigma_*^2 & \Sigma_{01} \\ \Sigma_{10} & \Sigma_{11} \end{pmatrix}$. With the underlying time series being nontrivial and independent, this covariance matrix will have full rank. Then standard theory tells that Y_* is, conditionally given $\overrightarrow{Y} = \overrightarrow{y}$, normally distributed having
expectation $\mu_* + \Sigma_{01}\Sigma_{11}^{-1}\left(\overrightarrow{y} - \overrightarrow{\mu}\right)$ and
variance $\sigma_*^2 - \Sigma_{01}\Sigma_{11}^{-1}\Sigma_{10}$.

If we can show that these values are close to the unconditional ones, namely μ_* and σ_*^2, we are done. For the components of Σ_{01} we get, by taking account of the independences of variables with different second index i,

$$\begin{aligned}\mathrm{Cov}\left(Y_j, Y_*\right) = \mathrm{Cov}\left(\sum_{i=0}^{3} X_{ji}, \sum_{i=1}^{3} c_i \left(X_{ki} - X_{0i}\right)\right) &= \sum_{i=1}^{3} c_i \mathrm{Cov}\left(X_{ji}, X_{ki} - X_{0i}\right) \\ &= \sum_{i=1}^{3} c_i \left(\mathrm{Cov}\left(X_{ji}, X_{ki}\right) - \mathrm{Cov}\left(X_{ji}, X_{0i}\right)\right)\end{aligned}$$

All covariances appearing here relate to gaps, thus are very small, as long as the gaps fluctuate few. Consequently, Σ_{01} is a vector having very small components.

For the components of Σ_{11} we get analogously $(j, m \leq 0)$

$$\mathrm{Cov}\left(Y_j, Y_m\right) = \mathrm{Cov}\left(\sum_{i=0}^{3} X_{ji}, \sum_{i=0}^{3} X_{mi}\right) = \sum_{i=0}^{3} \mathrm{Cov}\left(X_{ji}, X_{mi}\right) \approx \mathrm{Cov}\left(X_{j0}, X_{k0}\right)$$

where in the last step we have used that the summand with second index $i = 0$ dominates the sum: It relates to the deflated volume $\underline{\bar{M}}$, which fluctuates much more than the gaps, whose covariances yield the rest of the sum. Thus, Σ_{11} equals approximately the covariance matrix of the deflated volume, a rather volatile time series. So Σ_{11} will be far from singular, such that its inverse cannot have arbitrarily large components.

Overall we can conclude that $\Sigma_{01}\Sigma_{11}^{-1}\Sigma_{10}$, the shift of the variance, must be extremely small (Σ_{01} has quadratic impact) and is notably independent of the condition. $\Sigma_{01}\Sigma_{11}^{-1}\left(\overrightarrow{y} - \overrightarrow{\mu}\right)$, the shift of the expectation, is usually very small too, albeit in the (highly unlikely) case of extreme conditions ($\overrightarrow{y}$ far away from the average $\overrightarrow{\mu}$) it could be larger.

Summing up, the specific setting of the condition that we have here ensures that the lognormal RV ΔL_k has about the same expected value conditionally as unconditionally, apart from very remote conditions.

At the same time the above calculations show how one could improve the approximate reasoning, if one wanted to: The observations of $\underline{I\bar{M}}_{(j\leq 0)}$ would lead via time series analysis to a model and estimated parameters for $\underline{I\bar{M}}$, which together with the (assumed) models and parameters for the primordial gaps lead to a model and parameters for $\underline{\bar{M}}$. Then one has all ingredients to calculate the (estimated) conditional moments of the RV Y_* given $\overrightarrow{Y} = \overrightarrow{y}$, which finally lead to an estimate of $\mathrm{E}\left(\Delta D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right)$.

We shall however, stay with the approximate approach. It works analogously for $\Delta D(\alpha)_k^2$ and the product $\Delta D(\alpha)_k\, \Delta D(\alpha)_l$. All in all we have

Proposition 9.5. *If the comprehensive index system is independent and the true deflated official volume is a time series that fluctuates considerably more than the primordial gaps, then*

$$E\left(\Delta D(\alpha)_k \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx E\left(\Delta D(\alpha)_k\right)$$

$$E\left(\Delta D(\alpha)_k^2 \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx E\left(\Delta D(\alpha)_k^2\right)$$

$$E\left(\Delta D(\alpha)_k\, \Delta D(\alpha)_l \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) \approx E\left(\Delta D(\alpha)_k\, \Delta D(\alpha)_l\right)$$

such that we can replace the final expressions for the moments of the as-if quantities in Theorem 8.7 and Proposition 8.8 by unconditional ones.

Notice that the approximations in the proposition hold as well (and are even exact) for the extreme opposite of a variable deflated volume: the trivial volume case. Here the intricate condition on $\underline{M}_{(j\leq 0)}$ or equivalently $\underline{I\bar{M}}_{(j\leq 0)}$ is nil.

9.2 Future uncertainty

The future-loss formulae are more complex than those referring to as-if figures, see Theorem 8.7 and Proposition 8.9. Nevertheless we can derive mainly exact results, as long as the comprehensive index system is *strongly* independent. To ease reading we will, whenever the meaning is clear from the context, from now on often write D for $D(\alpha)$ and analogously for J and K.

We need to calculate, conditionally on $\left(\underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)$, the expectations of $\Delta D_q ID_q = D_q$, D_q^2, $\Delta D_q IJ_q = J_q$, J_q^2, $\Delta D_q IK_q = K_q$, as well as the covariances of D_q and J_q with the past gap ΔD_k. Recall the ultimate goal is to split gaps from picked indices as far as possible. This is indeed possible, but requires a number of steps.

In order to treat D, J, K in parallel, we define a last set of indices.

Definition 9.6. Set

$$P(\alpha, z) := M^{-z} D(\alpha) = \frac{VB^\alpha}{M^z}$$

$z = 0, 1, 2$ yields the variables D, J, K, however, we admit arbitrary z.

$$\widetilde{G} := A^\delta \widetilde{B} = A^\delta G \bar{B}^\zeta$$

$$Q(\alpha, 0) := AH^\alpha \bar{B}^{\tau\alpha}, \qquad Q(\alpha, z) := Q(\alpha, 0)\, \widetilde{G}^{-z}$$

The corresponding picked indices and gaps are defined analogously.

Proposition 9.7. *We have*

$$M = \bar{M}\widetilde{G}$$

$$P(\alpha, z) = \bar{M}^{\omega - z} Q(\alpha, z)$$

$$Q(\alpha, z) = A^{1-\delta z} G^{-z} H^{\alpha} \bar{B}^{\tau\alpha - \zeta z}$$

Analogous formulae hold for the corresponding picked indices and gaps. For the latter we have the primordial decomposition

$$\Delta P(\alpha, z) = \Delta D(\alpha) = \Delta A^{1-\delta\omega} \Delta G^{-\omega} \Delta H^{\alpha} \Delta \bar{B}^{\tau\alpha - \zeta\omega} = \Delta Q(\alpha, \omega)$$

which notably does not depend on z.

Proof. Straightforward algebra, starting from the equations $M = \bar{M} A^{\delta} \widetilde{B}$ and $D(\alpha) = \bar{M}^{\omega} A H^{\alpha} \bar{B}^{\tau\alpha}$. □

The idea behind these time series is as follows:

- P embraces D, J, and K in one.
- $\widetilde{G}$ relates M and $\bar{M}$ in a compact manner; $\Delta\widetilde{G} = \Delta\bar{M}^{-1}$.
- Q is a compact notation for a complex product of the four "benign" primordial indices: those being observable via their picked index. It helps split P into an independent product of a power of the intricate $\bar{M}$ and the remaining part. This split is the first step of the primordial decomposition of the future value of P.

Appendix C gives a concise overview of all index families we have introduced across this book.

Lemma 9.8. *If the comprehensive index system is strongly independent, we have*

$$E\left(P(\alpha, z)_q \mid \underline{M}_{(j\le 0)}, \Upsilon_{pr}\right) = E\left(\bar{M}_q^{\omega - z} \mid \underline{I\bar{M}}_{(j\le 0)}\right) E\left(IQ(\alpha, z)_q \mid \Upsilon_{pr}\right) E\left(\Delta Q(\alpha, z)_q\right)$$

$$E\left(I\bar{M}_q^{\omega - z} \mid \underline{I\bar{M}}_{(j\le 0)}\right) = E\left(\bar{M}_q^{\omega - z} \mid \underline{I\bar{M}}_{(j\le 0)}\right) E\left(\Delta\widetilde{G}_q^{\omega - z}\right)$$

Proof. As explained in the past section, we can always replace the condition $\left(\underline{M}_{(j\le 0)}, \Upsilon_{pr}\right)$ by the equivalent $\left(\underline{I\bar{M}}_{(j\le 0)}, \Upsilon_{pr}\right)$, which in the independent index system consists of five independent indices. In particular, we have

$$\mathrm{E}\left(P(\alpha, z)_q \mid \underline{M}_{(j\le 0)}, \Upsilon_{pr}\right) = \mathrm{E}\left(\bar{M}_q^{\omega - z} Q(\alpha, z)_q \mid \underline{I\bar{M}}_{(j\le 0)}, \Upsilon_{pr}\right)$$

and can apply Lemma 9.2 three times. In the strongly independent index system past gaps and normalized future gaps are independent, such that $Q(\alpha, z)_q = IQ(\alpha, z)_q \, \Delta Q(\alpha, z)_q$ is independent of $\underline{I\bar{M}}_{(j\le 0)}$, which relates to past values of $\bar{M}$ and of some primordial gaps. More strongly, $\left(\bar{M}_q^{\omega - z}, \underline{I\bar{M}}_{(j\le 0)}\right)$ and $\left(Q(\alpha, z)_q, \Upsilon_{pr}\right)$ are independent, such that

$$\mathrm{E}\left(\bar{M}_q^{\omega - z} Q(\alpha, z)_q \mid \underline{I\bar{M}}_{(j\le 0)}, \Upsilon_{pr}\right) = \mathrm{E}\left(\bar{M}_q^{\omega - z} \mid \underline{I\bar{M}}_{(j\le 0)}\right) \mathrm{E}\left(Q(\alpha, z)_q \mid \Upsilon_{pr}\right)$$

The last factor splits due to the independence of $\left(IQ(\alpha, z)_q, \Upsilon_{pr}\right)$ and $\Delta Q(\alpha, z)_q$ into

$$\mathrm{E}\left(Q(\alpha, z)_q \mid \Upsilon_{pr}\right) = \mathrm{E}\left(IQ(\alpha, z)_q \mid \Upsilon_{pr}\right) \mathrm{E}\left(\Delta Q(\alpha, z)_q\right)$$

This proves the first formula.

$\Delta\widetilde{G} = \Delta\bar{M}^{-1}$ yields $I\bar{M}_q^{\omega - z} = \bar{M}_q^{\omega - z} \Delta\widetilde{G}_q^{\omega - z}$. With $\left(\bar{M}_q^{\omega - z}, \underline{I\bar{M}}_{(j\le 0)}\right)$ being independent of $\Delta\widetilde{G}_q^{\omega - d}$ we get

$$\mathrm{E}\left(I\bar{M}_q^{\omega - z} \mid \underline{I\bar{M}}_{(j\le 0)}\right) = \mathrm{E}\left(\bar{M}_q^{\omega - z} \mid \underline{I\bar{M}}_{(j\le 0)}\right) \mathrm{E}\left(\Delta\widetilde{G}_q^{\omega - z}\right)$$

□

Proposition 9.9. *If the comprehensive index system is strongly independent, we have*

$$E\left(P(\alpha,z)_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) = E\left(I\bar{M}_q^{\omega-z} \mid \underline{I\bar{M}}_{(j\leq 0)}\right) E\left(IQ(\alpha,z)_q \mid \Upsilon_{pr}\right) \frac{E\left(\Delta Q(\alpha,z)_q\right)}{E\left(\Delta\widetilde{G}_q^{\omega-z}\right)}$$

$$E\left(P(\alpha,z)_q^2 \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) = E\left(I\bar{M}_q^{2(\omega-z)} \mid \underline{I\bar{M}}_{(j\leq 0)}\right) E\left(IQ(\alpha,z)_q^2 \mid \Upsilon_{pr}\right) \frac{E\left(\Delta Q(\alpha,z)_q^2\right)}{E\left(\Delta\widetilde{G}_q^{2(\omega-z)}\right)}$$

Proof. The first formula follows immediately from the preceding Lemma. The second formula is proven analogously by replacing each index by its square. Squaring the indices does not alter their independence structure, so all steps are just the same. □

The proposition yields a neat decomposition of conditional expectations into terms relating to the picked indices only and terms relating to gaps only, which is the desired split of both "worlds". Conditions only appear with picked indices, not with gaps.

Remark 9.10. The final gap term $\frac{\mathrm{E}\left(\Delta Q(\alpha,z)_q\right)}{\mathrm{E}\left(\Delta\widetilde{G}_q^{\omega-z}\right)}$ does not look very intuitive. Analogously to the results for the as-if loss, one would have expected (or hoped in) getting the term $\mathrm{E}\left(\Delta P\left(\alpha,z\right)_q\right) = \mathrm{E}\left(\Delta D\left(\alpha\right)_q\right)$. However, both terms are strongly related and usually very close: From

$$\Delta Q(\alpha,z) = \Delta Q(\alpha,\omega)\,\Delta\widetilde{G}^{\omega-z} = \Delta D(\alpha)\,\Delta\widetilde{G}^{\omega-z}$$

one gets

$$\frac{\mathrm{E}\left(\Delta Q(\alpha,z)_q\right)}{\mathrm{E}\left(\Delta\widetilde{G}_q^{\omega-z}\right)} = \frac{\mathrm{E}\left(\Delta D(\alpha)_q\,\Delta\widetilde{G}_q^{\omega-z}\right)}{\mathrm{E}\left(\Delta\widetilde{G}_q^{\omega-z}\right)}$$

The latter expression can be seen as a "distorted" version of $\mathrm{E}\left(\Delta D(\alpha)_q\right)$. We will see in the following chapter, when calculating lognormal gaps, how slight this distortion typically is.

The above terms about the picked indices are split into one relating to the deflated volume, containing the respective condition, and a term collecting compactly the "rest". This latter can be decomposed further, yielding a complex representation that, however, will reveal in a moment that the primordial picked indices can be assessed totally independently of each other.

Lemma 9.11. *Let the n pairs of RV's (A_j, X_j) be independent, $j = 1, ..., n$. Then the A_i are conditionally independent, given the $(X_j)_{j=1,...,n}$, and we have*

$$E\left(\prod_1^n A_j \mid X_1,...,X_n\right) = \prod_1^n E(A_j \mid X_1,...,X_n) = \prod_1^n E(A_j \mid X_j)$$

Proof. This generalizes Lemma 9.2 and follows from it immediately. The independence of the pairs of RV's is equivalent to $(A_1, X_1) \perp (A_j, X_j)_{j>1}$, which implies $\left(A_1 \perp (A_j)_{j>1} \mid X_1, X_2, ..., X_n\right)$, which means that the n RV's A_j are independent, given the n RV's X_j. From this follows the first equality. The second one is clear: Each (A_i, X_i) is independent of $(X_j)_{j\neq i}$, such that the conditions on the latter can be dropped. □

Corollary 9.12. *In the independent comprehensive index system we have*

$$E\left(I\bar{M}_q^{\omega-z} \mid \underline{I\bar{M}}_{(j\leq 0)}\right) E\left(IQ(\alpha,z)_q \mid \Upsilon_{pr}\right) = E\left(IP(\alpha,z)_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right) =$$
$$E\left(I\bar{M}_q^{\omega-z} \mid \underline{I\bar{M}}_{(j\leq 0)}\right) E\left(IA_q^{1-\delta z} \mid \underline{IA}_{(j\leq 0)}\right) E\left(IG_q^{-z} \mid \underline{IG}_{(j\leq 0)}\right) E\left(IH_q^{\alpha} \mid \underline{IH}_{(j\leq 0)}\right) E\left(I\bar{B}_q^{\tau\alpha-\zeta z} \mid \underline{I\bar{B}}_{(j\leq 0)}\right)$$

Proof. The powers of the five basic picked indices appearing in the formula are independent and their product is $IP(\alpha, z)$. According to the lemma we can split and unite the conditional expectations and the corresponding conditions. □

We have thus found two representations for the expectations of future picked indices appearing in Proposition 9.9: a compact one using $IP(\alpha, z)_q$ and the whole condition, and a split where each picked index is conditioned by its own history only, without interrelation to other indices. (However, in all cases we could replace the index-specific condition by the overall one.)

The analogous split results for the predictors of future index values. If these are constructed according to the logic of the comprehensive index system, we have

$$\widehat{IP(\alpha, z)}_q = \widehat{I\bar{M}_q}^{\omega - z} \widehat{IA_q}^{1-\delta z} \widehat{IG_q}^{-z} \widehat{IH_q}^{\alpha} \widehat{I\bar{B}_q}^{\tau\alpha - \zeta z}$$

This enables use to relate predictions of future index values to the respective (conditional) expectations.

Definition 9.13. We call

$$\kappa (IA_q) := \frac{\mathrm{E}\left(IA_q \mid \underline{M}_{(j \leq 0)}, \Upsilon_{pr}\right)}{\widehat{IA_q}}$$

the *relative precision* of the predictor $\widehat{IA_q}$, and analogously for $\widehat{IG_q}$, ..., $\widehat{ID(\alpha)}_q$, $\widehat{IP(\alpha, z)}_q$, and other indices being generated as products of (powers of) picked indices of the comprehensive index system.

Notice that for powers of the five basic picked indices one can equivalently write

$$\kappa \left(IA_q^{1-\delta z}\right) = \frac{\mathrm{E}\left(IA_q^{1-\delta z} \mid \underline{IA}_{(j \leq 0)}\right)}{\widehat{IA_q}^{1-\delta z}}$$

etc., replacing the overall condition by the index-specific one.

Intuitively, $\kappa (IA_q)$ is an indicator for the quality of the forecast of IA_q. Note that in line with our systematic we define the relative precision *conditionally*, given the past index values, such that the predictor is a scalar. As stated in Remark 7.10, in time series analysis the corresponding *unconditional* figures are of interest as well: One would e.g. compare $\mathrm{E}(IA_q)$ to $\mathrm{E}\left(\widehat{IA_q}\right)$, averaging over all possible index histories belonging to the chosen model.

If we look at the expected future loss (count) in Theorem 8.7, we see that, up to a constant, it equals

$$\frac{\mathrm{E}\left(D(\alpha)_q \mid \underline{M}_{(j \leq 0)}, \Upsilon_{pr}\right)}{\widehat{ID(\alpha)}_q}$$

with α equaling $\alpha_{(d^*, u^*)}$ (or α_{d^*}), while Proposition 8.9 calculates the expected future loss rate from the analogous term about J_q. In the unified representation we get

Corollary 9.14. *If the comprehensive index system is strongly independent, we have*

$$\frac{E\left(P(\alpha, z)_q \mid \underline{M}_{(j \leq 0)}, \Upsilon_{pr}\right)}{\widehat{IP(\alpha, z)}_q} = \kappa\left(IP(\alpha, z)_q\right) \frac{E\left(\Delta D(\alpha)_q \, \Delta\widetilde{G}_q^{\omega - z}\right)}{E\left(\Delta\widetilde{G}_q^{\omega - z}\right)}$$

and can decompose the first factor of the RHS into five relative precisions

$$\kappa\left(IP(\alpha, z)_q\right) = \kappa\left(I\bar{M}_q^{\omega - z}\right) \kappa\left(IA_q^{1-\delta z}\right) \kappa\left(IG_q^{-z}\right) \kappa\left(IH_q^{\alpha}\right) \kappa\left(I\bar{B}_q^{\tau\alpha - \zeta z}\right) \tag{9.1}$$

Proof. Combine Proposition 9.9 and Corollary 9.12, taking into account that predictors and conditional expectations follow the same multiplicative logic. □

The decomposition of the relative precision makes clear that it is possible to assess this precision for each of the five basic picked indices separately. Once calculated the history of $I\bar{M}$ out of that of M with the help of the primordial picked indices, one can do time series analysis on each index regardless of their intricate connection.

Having found both short and long (but instructive) representations for the expectations of future index values, it remains to calculate the covariances between future and past values. The first step is easy.

Lemma 9.15. *If the comprehensive index system is strongly independent, we have*

$$\begin{aligned}
&Cov\left(P(\alpha,z)_q\,, \varDelta D(\alpha)_k \mid \underline{M}_{(j\leq 0)}, \varUpsilon_{pr}\right)\\
&\qquad\qquad = Cov\left(\bar{M}_q^{\omega-z}, \varDelta D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right) E\left(IQ(\alpha,z)_q \mid \varUpsilon_{pr}\right) E\left(\varDelta Q(\alpha,z)_q\right)
\end{aligned}$$

Proof. The LHS reads

$$\operatorname{Cov}\left(\bar{M}_q^{\omega-z} IQ(\alpha,z)_q\, \varDelta Q(\alpha,z)_q\,, \varDelta D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}, \varUpsilon_{pr}\right)$$

Note that due to the strong independence the two appearing gaps are independent. We apply Lemma 9.2 three times, splitting first into

$$\operatorname{Cov}\left(\bar{M}_q^{\omega-z}, \varDelta D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right) \operatorname{E}\left(IQ(\alpha,z)_q\, \varDelta Q(\alpha,z)_q \mid \varUpsilon_{pr}\right)$$

and then the second factor as in Lemma 9.8 into

$$\operatorname{E}\left(IQ(\alpha,z)_q \mid \varUpsilon_{pr}\right) \operatorname{E}\left(\varDelta Q(\alpha,z)_q\right)$$

To see the first split, write the covariance as difference

$$\begin{aligned}
&\operatorname{Cov}\left(\bar{M}_q^{\omega-z} IQ(\alpha,z)_q\, \varDelta Q(\alpha,z)_q\,, \varDelta D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}, \varUpsilon_{pr}\right)\\
&\qquad = \operatorname{E}\left(\bar{M}_q^{\omega-z} IQ(\alpha,z)_q\, \varDelta Q(\alpha,z)_q\, \varDelta D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}, \varUpsilon_{pr}\right)\\
&\qquad\quad - \operatorname{E}\left(\bar{M}_q^{\omega-z} IQ(\alpha,z)_q\, \varDelta Q(\alpha,z)_q \mid \underline{I\bar{M}}_{(j\leq 0)}, \varUpsilon_{pr}\right) \operatorname{E}\left(\varDelta D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}, \varUpsilon_{pr}\right)
\end{aligned}$$

and use

$$\left(\bar{M}_q^{\omega-z}, \varDelta D(\alpha)_k\,, \underline{I\bar{M}}_{(j\leq 0)}\right) \perp \left(IQ(\alpha,z)_q\, \varDelta Q(\alpha,z)_q\,, \varUpsilon_{pr}\right)$$

to split off $\operatorname{E}\left(IQ(\alpha,z)_q\, \varDelta Q(\alpha,z)_q \mid \varUpsilon_{pr}\right)$ on both sides of this difference. □

How can the remaining covariance

$$\begin{aligned}
&\operatorname{Cov}\left(\bar{M}_q^{\omega-z}, \varDelta D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right)\\
&\qquad = \operatorname{E}\left(\bar{M}_q^{\omega-z} \varDelta D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right) - \operatorname{E}\left(\bar{M}_q^{\omega-z} \mid \underline{I\bar{M}}_{(j\leq 0)}\right) \operatorname{E}\left(\varDelta D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right)
\end{aligned}$$

be assessed? Unconditionally, $\bar{M}_q^{\omega-z}$ and $\varDelta D(\alpha)_k$ are independent, but both RV's are tied to the conditions

$$\underline{I\bar{M}}_j = \underline{\bar{M}_j \varDelta A_j^{\delta} \varDelta G_j \varDelta \bar{B}_j^{\zeta}}$$

for a number of years $j \leq 0$. However, we can argue similarly as in the preceding section:

- Approximating by averaging over $\underline{I\bar{M}}_{(j\leq 0)}$ yields covariance 0.
- Heuristically, if the deflated volume $\underline{\bar{M}}$ is substantially more volatile a time series than the gaps, the conditions $\underline{I\bar{M}}_j = m_j$ are "similar" to the analogous $\underline{\bar{M}}_j = m_j$, such that replacing the conditions will not change the result a lot. As $\underline{\bar{M}}$ is independent of the gaps, we get, again via the weak union rule

  $$\left(\bar{M}_q^{\omega-z}, \underline{\bar{M}}_{(j\leq 0)}\right) \perp \varDelta D(\alpha)_k \Rightarrow \bar{M}_q^{\omega-z} \perp \varDelta D(\alpha)_k \mid \underline{\bar{M}}_{(j\leq 0)}$$

  such that

  $$\mathrm{Cov}\left(\bar{M}_q^{\omega-z}, \varDelta D(\alpha)_k \mid \underline{\bar{M}}_{(j\leq 0)}\right) = 0$$

  Thus, the covariance with the original condition should be close to 0.
- For lognormal primordial time series this reasoning can again be made rigorous, along the lines of Example 9.4, albeit a bit more lengthy.

We can conclude:

Proposition 9.16. *If the comprehensive index system is strongly independent and the true deflated official volume is a time series that fluctuates considerably more than the primordial gaps, then*

$$Cov\left(\bar{M}_q^{\omega-z}, \varDelta D(\alpha)_k \mid \underline{I\bar{M}}_{(j\leq 0)}\right) \approx 0$$

$$Cov\left(P(\alpha,z)_q, \varDelta D(\alpha)_k \mid \underline{M}_{(j\leq 0)}, \varUpsilon_{pr}\right) \approx 0$$

Notice two special cases where the approximations are exact and do not require the assumption on the variability of $\bar{M}$. The first (and less important) one is the trivial volume case, the second one is as follows:

If $\omega = z$, the intricate $\bar{M}_q^{\omega-z}$ is a trivial constant, such that the above covariances vanish exactly. This does not help with D_q (assessing loss amounts), which corresponds to $z = 0$. But J_q (assessing loss rates) is the case $z = 1$, and $\omega = 1$ is indeed possible, not to say prevalent:

- In the non-Cat case ω always equals 1. The vanishing of $\bar{M}_q^{\omega-z}$ also simplifies the expectation formulae developed in this chapter and confirms what was stated when the loss rates were introduced in the last chapter: Burning Cost rating of loss rates for single-loss models is particularly benign, as the future volume drops out, such that only the primordial picked indices need to be projected into the future.
- In the Cat case ω equals α. As in practice large accumulation losses often have local and regional alphas in the range of 1, one will frequently come across situations where $\bar{M}_q^{\omega-1}$ has no or just a very low impact, such that the above covariances vanish exactly or nearly.

9.3 Wrap up

Now we have all ingredients needed for a unified treatment of aggregate losses, loss counts, and respective rates, with particular attention to the parallels and subtle differences between the formulae for the past (as-if) quantities and those for the future realizations. For the sake of concise representation, we introduce a last piece of notation. Recall Definition 4.5, which introduced the standardization of a RV having positive expectation through dividing by the latter.

Definition 9.17. The *relative covariance* of two random variables having positive expectations is the covariance of their standardizations, e.g.

$$\mathrm{rCov}\left(S_k, S_l\right) := \mathrm{Cov}\left(\dot{S}_k, \dot{S}_l\right) = \frac{\mathrm{Cov}\left(S_k, S_l\right)}{\mathrm{E}\left(S_k\right)\mathrm{E}\left(S_l\right)}$$

Conditional standardization and *conditional relative covariance* are defined analogously, by replacing unconditional moments by conditional ones.

Notice that the relative covariance corresponds to the squared coefficient of variation, but is different from the correlation coefficient ϱ. The latter is normalized by the standard deviations, not the expectations.

Now we restate the properties of the future loss and its predictor in a way that embraces all quantities of interest.

Theorem 9.18. *Suppose:*
A risk is in the inflationary world structure.
The comprehensive index system is strongly independent.
The primordial gaps are time series fluctuating very few (such that all normalized gaps fluctuate closely about 1) and, in particular, much less than the true deflated official volume.
The survival function $\overline{F}$ of the normalized severity is sufficiently regular, as specified below.

Then the following approximations for the first two moments of the as-if loss hold ($k, l \leq 0$):

$$E_{\varPi_{pr},\pi,\iota}\left(S_k \mid M_{(j\leq 0)}, \varUpsilon_{pr}\right) \approx \tilde{e}E\left(\varDelta D(\alpha)_k\right)$$

$$\begin{aligned} &1 + CV^2\left(S_k \mid M_{(j\leq 0)}, \varUpsilon_{pr}\right) \\ &\quad\approx (1+\beta)\left(1 + CV^2\left(\varDelta D(\alpha)_k\right)\right) + \frac{1}{IV_k}\max\left(\sum_{i=1}^{4} a_i \frac{E\left(\varDelta D(\alpha_i)_k\right)}{E^2\left(\varDelta D(\alpha)_k\right)}, \sum_{i=1}^{j} a_i^{(j)} \frac{E\left(\varDelta D\left(\alpha_i^{(j)}\right)_k\right)}{E^2\left(\varDelta D(\alpha)_k\right)}\right) \end{aligned}$$

$$rCov\left(S_k, S_l \mid M_{(j\leq 0)}, \varUpsilon_{pr}\right) \approx rCov\left(\varDelta D(\alpha)_k, \varDelta D(\alpha)_l\right), \quad k \neq l$$

If additionally the predictors of the future picked indices are constructed according to the multiplicative structure of the comprehensive index system, and the future values of the indices referring to loss inflation fluctuate, conditionally given $\left(M_{(j\leq 0)}, \varUpsilon_{pr}\right)$, closely about their predictors, then the following approximations for the first two moments of the future loss hold ($k \leq 0 \leq q$):

$$E_{\varPi_{pr},\pi,\iota}\left(X_q \mid M_{(j\leq 0)}, \varUpsilon_{pr}\right) \approx \tilde{e}\frac{E\left(\varDelta D(\alpha)_q \varDelta\widetilde{G}_q^{\omega-z}\right)}{E\left(\varDelta\widetilde{G}_q^{\omega-z}\right)}\kappa\left(IP(\alpha, z)_q\right)$$

$$\begin{aligned} 1 + CV^2\left(X_q \mid M_{(j\leq 0)}, \varUpsilon_{pr}\right) &\approx (1+\beta)\frac{1 + CV^2\left(\varDelta D(\alpha)_q \varDelta\widetilde{G}_q^{\omega-z}\right)}{1 + CV^2\left(\varDelta\widetilde{G}_q^{\omega-z}\right)}\frac{\kappa\left(IP(\alpha, z)_q^2\right)}{\kappa^2\left(IP(\alpha, z)_q\right)} \\ &+ \frac{1}{\widehat{IV}_q}\max\left[\sum_{i=1}^{4} a_i \frac{E\left(\varDelta D(\alpha_i)_q \varDelta\widetilde{G}_q^{\omega-z'}\right)E^2\left(\varDelta\widetilde{G}_q^{\omega-z}\right)}{E\left(\varDelta\widetilde{G}_q^{\omega-z'}\right)E^2\left(\varDelta D(\alpha)_q \varDelta\widetilde{G}_q^{\omega-z}\right)}\frac{\kappa\left(IP\left(\alpha_i, z'\right)_q\right)}{\kappa^2\left(IP\left(\alpha, z\right)_q\right)},\right. \\ &\left.\sum_{i=1}^{j} a_i \frac{E\left(\varDelta D\left(\alpha_i^{(j)}\right)_q \varDelta\widetilde{G}_q^{\omega-z'}\right)E^2\left(\varDelta\widetilde{G}_q^{\omega-z}\right)}{E\left(\varDelta\widetilde{G}_q^{\omega-z'}\right)E^2\left(\varDelta D(\alpha)_q \varDelta\widetilde{G}_q^{\omega-z}\right)}\frac{\kappa\left(IP\left(\alpha_i^{(j)}, z'\right)_q\right)}{\kappa^2\left(IP\left(\alpha, z\right)_q\right)}\right] \end{aligned}$$

$$rCov\left(X_q, S_k \mid M_{(j\leq 0)}, \Upsilon_{pr}\right) \approx 0$$

The formulae for the other quantities of interest are almost identical, one just has to replace some variables and constants, according to the following table:

			$d>0$		$d=0$	
loss variable	X_q	Y_q	$\bar{N}_q$	${}^{\bar{N}_q}/{}_{M_q}$	N_q	${}^{N_q}/{}_{M_q}$
predictor	S_k	R_k	C_k	${}^{C_k}/{}_{\widehat{M}_q}$	C_k	${}^{C_k}/{}_{\widehat{M}_q}$
constant	$\tilde{e}$	$\mathring{e}$	$\tilde{\theta}$	$\mathring{\theta}$	$\tilde{\theta}$	$\mathring{\theta}$
z	*0*	*1*	*0*	*1*	*0*	*1*
z'	*0*	*2*	*0*	*2*	*0*	*2*
α	$\alpha_{(d^*,u^*)}$	$\alpha_{(d^*,u^*)}$	α_{d^*}	α_{d^*}	*0*	*0*
α_2	$\mathring{\alpha}_{(d^*,u^*)}$	$\mathring{\alpha}_{(d^*,u^*)}$	α_{d^*}	α_{d^*}	*0*	*0*
a_2	$\frac{\bar{\varepsilon}^*}{\theta_{d^*}}$	$\frac{\bar{\varepsilon}^*}{\theta_{d^*}}$	$\frac{1}{\theta_{d^*}}$	$\frac{1}{\theta_{d^*}}$	$\frac{1}{\theta}$	$\frac{1}{\theta}$
a_3	$\frac{2d^*}{e^*}$	$\frac{2d^*}{e^*}$	*0*	*0*	*0*	*0*

For all four loss variables we have $a_1 = \gamma\,(1+\beta)$, $a_4 = -a_3$, $\alpha_1 = 2\alpha$, $\alpha_3 = \alpha_2$, $\alpha_4 = \alpha$.
For the second sum in the variance formula, which improves the approximation given by the first sum as a lower bound, there are two options:
$j=1$, $a_1' = a_1 + \frac{1}{\theta}$, $\alpha_1' = \alpha_1$,
$j=2$, $a_1'' = a_1$, $\alpha_1'' = \alpha_1$, $a_2'' = \frac{1}{\theta_{d^*}}$, $\alpha_2'' = 2\alpha_{(d^*,u^*)} - \alpha_{d^*}$.

The appearing basis-risk indices have the primordial decompositions ($\omega = 1 + (\alpha - 1)\,\chi$)

$$\Delta D(\alpha) = \Delta A^{1-\delta\omega} \Delta G^{-\omega} \Delta H^{\alpha} \Delta \bar{B}^{\tau\alpha-\zeta\omega}$$

$$\Delta\widetilde{G} = \Delta A^{\delta} \Delta G \Delta \bar{B}^{\zeta}$$

$$\Delta D(\alpha)\, \Delta\widetilde{G}^{\omega-z} = \Delta Q(\alpha, z) = \Delta A^{1-\delta z} \Delta G^{-z} \Delta H^{\alpha} \Delta \bar{B}^{\tau\alpha-\zeta z}$$

while the above picked index can be written as independent product of basic picked indices

$$IP(\alpha, z) = I\bar{M}^{\omega-z} IA^{1-\delta z} IG^{-z} IH^{\alpha} I\bar{B}^{\tau\alpha-\zeta z}$$

Exponents other than α *have the analogous decompositions.*

The formulae include unlimited layers and pro-rata coverages, as long as expectation and variance of the normalized loss $\mathring{Z}$ *are finite.*

The theorem extends to the trivial volume case (where all volumes are constant).

In the following special cases the last two summands in the loss (rate) variance formulae vanish or cancel out, such that, like for the loss count (rate) variances anyway, the sum only goes from 1 to 2 and the second sum is obsolete:
First risks, i.e. $d=0$.
Proportional coverages. *Here* $\alpha_1 = \alpha_2 = 2$.
$\overline{F}$ *has a* Pareto tail *starting well below* d^*. *Here both* α *and* α_2 *equal the (global) Pareto alpha.*

The regularity conditions for the survival function are as follows:
Loss (rate): $\overline{F}$ *is continuous at* d^* *and* u^*.
Loss count (rate) for $d>0$: $\overline{F}$ *is differentiable at* d^*
Variant $j=2$ *of the loss (rate) variance: both.*

Proof. The theorem combines the assumptions and formulae of Theorem 8.7 and the analogous assertions for the loss-per-volume variables from Section 8.5.2 (including those for the loss count rate omitted there) with the moments calculated in this chapter, in particular Proposition 9.5, Corollary 9.14, and Proposition 9.16. For the sake of conciseness and intuitive understanding the (co)variance formulae are expressed in standardized manner, having divided by the respective expectation formulae.

The details on coefficients and special cases are inherited from the corresponding moments before averaging, as outlined in Proposition 7.4 (and the analogous Proposition 7.8). Recall that then the S_k formulae for the special case Pareto held if the Pareto tail started earlier than d^* and earlier than $\frac{d^*}{\Delta B_k}$. Now ΔB_k is random, such than it is more difficult to say whether $\frac{d^*}{\Delta B_k}$ is in the Pareto tail area. However, if this area starts low enough ("well below d^*"), $\frac{d^*}{\Delta B_k}$ is in the tail, as long as ΔB_k fluctuates closely about 1 almost surely. The analogous reasoning holds for X_q, with $\frac{IB_q}{\widehat{IB}_q}\Delta B_q$ taking the place of ΔB_k, see Section 7.3. As this is again a RV fluctuating quite closely about 1 in the same way, the Pareto case is analogous for the future loss formulae.

The approximative formulae in Theorem 8.7 were exact / extremely precise for pro-rata covers / Pareto tails. However, as many results of this chapter are approximate (apart from a few benign constellations), altogether we have mainly approximate results.

In Appendix B the range of RV's fluctuating closely about 1 is extended far *beyond* those being *almost surely* close to 1, namely to certain normalized time series being close to 1 *in probability.* It turns out that the (moderate and plausible) assumptions on $\overline{F}$ and the index structure required in this theorem enable us at the same time to make this extension, by using similar reasoning and the same kind of approximations as in this chapter. We will show that the theorem holds for a number of realistic time series models for gaps and picked indices, in particular if all primordial time series have lognormally distributed error terms and, in addition, the gaps are geometric random walks.

As for the Pareto case, if ΔB_k can take on arbitrarily high values (as a rare event), $\frac{d^*}{\Delta B_k}$ can be below the Pareto tail area. However, if this area starts low enough, this is extremely unlikely. In expectation, averaging over the gaps, the unlikely non-Pareto situations may create some distortion, but their impact is close to nil, such that the theorem's assertions for the as-if loss hold. For details see Appendix B.3, which covers the future loss, too. □

Notice that in the theorem all moments of gaps are unconditional ones, while those concerning picked indices are conditional on the observed picked-index history. We have, with some technical effort, managed to split gaps and picked indices in the moment formulae as desired.

9.4 Bias

Comparing the two expectation formulae in the above comprehensive theorem, we see that the bias of the as-if loss (count) (rate) depends on how much $\mathrm{E}\left(\Delta D(\alpha)_k\right)$ deviates from

$$\frac{\mathrm{E}\left(\Delta D(\alpha)_q\,\Delta\widetilde{G}_q^{\omega-z}\right)}{\mathrm{E}\left(\Delta\widetilde{G}_q^{\omega-z}\right)}\kappa\left(IP(\alpha,z)_q\right)$$

This formula appears intricate, however, its main input is the structure of the comprehensive index system. Looking at the decomposition of the appearing indices as stated in the theorem, we see that the only bit of information needed from the underlying models for normalized severity and ground-up loss count is the exponent α, which stands for the regional alpha of the rescaled layer (loss case) or the local alpha of the rescaled deductible (loss count case), respectively. Thus, if we know how the severity in that area is best approximated by a Pareto curve, we can assess (and avoid) the bias in the sample mean –

as long as we get the structure of the comprehensive index system right and are good at selecting and predicting picked indices.

If the picked indices are chosen well, we can expect the normalized gaps to be very close to 1 on average, such that the bias is mainly influenced by the relative precision of $IP(\alpha,z)_q$, which notably does not depend on the year k. Very roughly, one could say that, provided one is good at time series forecasting and has enough historical data, on should manage to get $\kappa\left(IP(\alpha,z)_q\right)$ close to 1. Let us have a closer look at this.

It is clear that the relative precision of a future index value depends on the kind of time series, the parameter inference method applied, and the data situation, so with the many components $IP(\alpha,z)$ embraces we have a huge variety of potential outcomes. However, Formula 9.1

$$\kappa\left(IP(\alpha,z)_q\right)=\kappa\left(I\bar{M}_q^{\omega-z}\right)\kappa\left(IA_q^{1-\delta z}\right)\kappa\left(IG_q^{-z}\right)\kappa\left(IH_q^{\alpha}\right)\kappa\left(I\bar{B}_q^{\tau\alpha-\zeta z}\right)$$

reduces this problem to five simpler ones. The basic picked indices (and powers thereof) should in practice often be tractable with standard models. Furthermore, usually for the basic inflation and often also for the superimposed inflations and the frequency per volume, one should have a long history available, possibly coming from market statistics and being thus unaffected from the data issues that a single rating task may have. For the deflated official volume $I\bar{M}$ the situation is more challenging; it is specific for the risk we are rating and its history is just as long (which often means: as short) as the provided list of official volumes. However, with inflationary and loss-per-volume effects factored out, we have reduced the complexity of the volume issue a great deal.

In some cases certain exponents in the above formula equal 0, which means that an index drops out and does not have to be forecasted at all:

- As for IJ_q $(z=1)$, we have already mentioned at the end of Section 9.2 that the intricate $I\bar{M}$ drops out in all single-loss models and quite some Cat models (namely if $\alpha=1$). If further $\delta=1$ (official volume is original premium), IA drops out, too. Basic inflation drops out if $\tau\alpha=\zeta$, which is not uncommon – at least for pro-rata covers usually all three "greeks" equal 1. Altogether we can have three picked indices less due to vanishing exponents, with only IG and IH remaining.
- By contrast, the prediction of ID_q $(z=0)$ bears less potential for simplification, but never requires to predict IG_q. (We need IG, however, for the past, to calculate the $I\bar{M}_j$.)
- IH can never drop out due to its exponent.

The gaps may simplify in the same way, "loosing" factors due to vanishing exponents. However, this occurs in partly different cases.

The future involves the two gaps

$$\mathit{\Delta}D(\alpha)\,\mathit{\Delta}\widetilde{G}^{\omega-z}=\mathit{\Delta}Q(\alpha,z)=\mathit{\Delta}A^{1-\delta z}\mathit{\Delta}G^{-z}\mathit{\Delta}H^{\alpha}\mathit{\Delta}\bar{B}^{\tau\alpha-\zeta z}$$

$$\mathit{\Delta}\widetilde{G}^{\omega-z}=\left(\mathit{\Delta}A^{\delta}\mathit{\Delta}G\mathit{\Delta}\bar{B}^{\zeta}\right)^{\omega-z}$$

While the first one has the same exponents as $IP(\alpha,z)$, the exponents of $\mathit{\Delta}\widetilde{G}$ are different and do not vanish at the same time. However, in the important special case $\omega=z$ the distortion $\mathit{\Delta}\widetilde{G}^{\omega-z}$ reduces to the trivial index 1.

For comparison, the past involves the gap

$$\mathit{\Delta}D(\alpha)=\mathit{\Delta}A^{1-\delta\omega}\mathit{\Delta}G^{-\omega}\mathit{\Delta}H^{\alpha}\mathit{\Delta}\bar{B}^{\tau\alpha-\zeta\omega}$$

having partly different exponents than above. Here $\mathit{\Delta}H$ and $\mathit{\Delta}G$ never drop out due to their exponents.

- For single-loss models ($\omega = 1$) we have almost the same constellation as for IJ_q above: $\mathit{\Delta}A$ and/or $\mathit{\Delta}\bar{B}$ drop out if $\delta = 1$ and/or $\tau\alpha = \zeta$.

- For Cat models ($\omega = \alpha$) basic inflation drops out independently of α, namely if $\tau = \zeta$. $\mathit{\Delta}A$ drops out if $\delta = 1 = \alpha$.

Disappearing indices make the assessment of past and/or future uncertainty much easier. Overall it turns out that, among all experience rating variants, assessing *rates per volume* in the *single-loss* case ($z = 1 = \omega$) has the best chances for simplification due to vanishing indices. Proportional business is (not surprisingly) most benign, even in the Cat case. (The latter is, however, less relevant in practice, as most accumulation-loss covers are non-proportional.)

It is maybe surprising in how many constellations basic inflation drops out. One could be tempted to conclude that inflation is less critical for experience rating than expected. However, it must be noted that there is no chance to "get rid" of G, the part of the volume inflation being not described by $\bar{B}$. In many constellations this seems to be the most problematic inflation component.

Bear in mind that we are focusing on bias here. The variances involve different alphas, thus do not yield the same simplifications.

Speaking of alphas, recall that, apart form the pro-rata and the Pareto case, local and regional alphas vary with loss size. For a layer, the regional $\alpha_{(d^*,u^*)}$ and the local α_{d^*} are usually not the same, albeit often close. This means that aggregate loss and loss count, despite their very similar formulae, may show somewhat different effects due to deviating alphas leveraging indices differently. Even more variety arises if we have a program consisting of several adjacent (re)insurance layers, stretching across a wide range of loss sizes. Here we could have rising local/regional alphas (very common), say from 1 to 3 from the bottom to the top layer, or alternatively decreasing ones, or alphas going up and down, etc. This can lead to constellations where an index has a highly leveraged impact on the rating of a layer, while it does hardly matter for another layer due to a nearly vanishing exponent.

To gather some more intuition, let us elaborate two examples, starting with the old one having a very benign index world:

Example 9.19. Experience rating of *rates per volume* for
single-loss Fire (re)insurance, using the
aggregate sum insured as official volume, having
equal loss and volume inflation.

We have $G \equiv 1 \equiv H$, $\zeta = 1 = \tau$, $\delta = 0$, $\omega = 1 = z$. The past gap is based on the index

$$\mathit{\Delta}D(\alpha) = \mathit{\Delta}A\mathit{\Delta}\bar{B}^{\alpha-1}$$

while the future uncertainty is modeled by the indices

$$IP(\alpha, z) = IA\,I\bar{B}^{\alpha-1}$$

$$\mathit{\Delta}Q(\alpha, z) = \mathit{\Delta}A\mathit{\Delta}\bar{B}^{\alpha-1}, \qquad \mathit{\Delta}\widetilde{G}^{\omega-z} \equiv 1$$

Here $I\bar{M}$ drops out and the distortion $\mathit{\Delta}\widetilde{G}^{\omega-z}$ is trivial, such that the remaining three indices all have the same structure $A\bar{B}^{\alpha-1}$.

If we have, in particular, a proportional coverage or a layer hit by Pareto distributed losses with $\alpha = 1$, then inflation drops out of the formula altogether. However, the uncertainty about the index A remains, reflecting shifts in the frequency per volume unit. While in other cases this uncertainty is arguably often overwhelmed by inflationary effects of various kind, here it could take a prominent role.

Equally simple are inflation-free volumes, for all lines of business, both for Cat and non-Cat losses. Recall that the comprehensive index system embraces these volumes elegantly (see end of Section 8.1.2).

Example 9.20. *Inflation-free official volumes* are parametrized as follows:

$\zeta = 0$, $\tau = 1$, $G \equiv 1 \equiv H$, $\delta = 0$. Here we have $\bar{B} = B$ and $\bar{M} = M$, such that $\widetilde{G} \equiv 1$. As in the preceding example most indices of interest look the same. The past gap is based on the index

$$\Delta D(\alpha) = \Delta A \Delta \bar{B}^{\alpha}$$

while the future uncertainty is modeled by the indices

$$IP(\alpha, z) = I\bar{M}^{\omega-z} IA\, I\bar{B}^{\alpha} = M^{\omega-z} IA\, I\bar{B}^{\alpha}$$

$$\Delta Q(\alpha, z) = \Delta A \Delta \bar{B}^{\alpha}, \qquad \Delta \widetilde{G}^{\omega-z} \equiv 1$$

The resulting formulae are the same as in the preceding example, save that the basic inflation has exponent α, not $\alpha - 1$, and $IP(\alpha, z)$ has the additional factor $I\bar{M}^{\omega-z}$, which here equals $M^{\omega-z}$.

To conclude this chapter, let us look once more at the picked indices and in particular at the relative precision of a single one. Let IU be the power of one of the five basic picked indices, as they appear in the decomposition of $\kappa\left(IP(\alpha, z)_q\right)$.

Proposition 9.21. *For powers of basic picked indices we have*

$$\kappa\left(IU_q\right) = \frac{E\left(IU_q \mid \underline{M}_{(j\leq 0)}, \Upsilon_{pr}\right)}{\widehat{IU_q}} = \frac{E\left(IU_q \mid \underline{IU}_{(j\leq 0)}\right)}{\widehat{IU_q}} = \frac{E\left(\underline{IU}_q \mid \underline{IU}_{(j\leq 0)}\right)}{\widehat{\underline{IU}_q}}$$

Proof. As shown earlier, in the expectation of future basic picked indices (or of powers thereof) we can replace the overall condition by the one on the specific picked index, which can equivalently be replaced by a power thereof. This proves the second equality. The third one follows from $\underline{IU}_q = \underline{IU}_0 IU_q$ and $\widehat{\underline{IU}_q} = \underline{IU}_0 \widehat{IU_q}$, where $\underline{IU}_0$ notably is (conditionally) a known factor. □

This proposition makes clear that for the assessment of the relative precision one can work with the original or the normalized index, whichever be more handy. Time series analysis usually treats non-normalized indices, i.e. would compare $\mathrm{E}\left(\underline{IU}_q \mid \underline{IU}_{(j\leq 0)}\right)$ with $\widehat{\underline{IU}_q}$, while we stated most results using the normalization tailor-made four our specific setting.

In practice, for each index $\underline{IU}$ of interest, one would use the history $\underline{IU}_{(j\leq 0)}$ to choose a suitable model and an appropriate predictor $\widehat{\underline{IU}_q}$. The latter depends on the assumed model, such that the relative precision is model-dependent. What is more, to assess it, we need an estimate of $\mathrm{E}\left(\underline{IU}_q \mid \underline{IU}_{(j\leq 0)}\right)$, which is again model-dependent.

It would be interesting to go through a variety of suitable models for index series and assess the relative precision wherever possible. However, time series analysis is not the core of this book, so we will not dig deeper into this. For illustration, however, we look at the relative precision of the arguably simplest time series model, the geometric random walk, which we have used for the inflation gap in the basic example. This does not mean that we see this model as generally suitable for picked indices – true and picked indices will often require more complex models than gaps. Nevertheless the geometric random walk is a possibility, at least an instructive one.

Example 9.22. If $\ln(\underline{IU})$ is a random walk with normal error, the random variables

$$L_j := \ln\left(\frac{\underline{IU}_j}{\underline{IU}_{j-1}}\right) \sim \mathrm{N}\left(\mu, \sigma^2\right)$$

are independent. Let $\underline{IU}_{-n}, \underline{IU}_{1-n}, ..., \underline{IU}_0$ be the known observations of $\underline{IU}$.

The common unbiased parameter estimates are

$$\widehat{\mu} = \frac{1}{n}\sum_{j=1-n}^{0} L_j, \qquad \widehat{\sigma^2} = \frac{1}{n-1}\sum_{j=1-n}^{0}\left(L_j - \widehat{\mu}\right)^2$$

however, alternative estimators might be thinkable.

For the logarithmized normalized future index we have

$$\ln\left(IU_q\right) = \ln\left(\frac{\underline{IU}_q}{\underline{IU}_0}\right) = L_1 + ... + L_q \sim \mathrm{N}\left(q\mu, q\sigma^2\right)$$

which would usually be predicted by $q\widehat{\mu}$, such that $\widehat{IU_q} = \exp\left(q\widehat{\mu}\right)$.

The true value IU_q is lognormally distributed (even unconditionally) and we have

$$\mathrm{E}\left(IU_q \mid \underline{IU}_{(j\leq 0)}\right) = \mathrm{E}\left(e^{L_1+...+L_q} \mid \underline{IU}_{(j\leq 0)}\right) = \mathrm{E}\left(e^{L_1+...+L_q}\right) = \exp\left(q\mu + \frac{q}{2}\sigma^2\right)$$

Thus, $\kappa\left(IU_q\right) = \exp\left(q\mu + \frac{q}{2}\sigma^2 - q\widehat{\mu}\right)$. If we estimate this, using the estimates for μ and σ, we get

$$\widehat{\kappa\left(IU_q\right)} = \exp\left(\frac{q}{2}\widehat{\sigma^2}\right)$$

While replacing unknown parameters by their estimates is a standard procedure (being often without alternative), statisticians are well aware that this may underestimate prediction errors, see e.g. [Stine, 1987] stating that having to use estimates instead of true parameters increases the MSE in autoregressive forecasts. Accordingly, in our situation we might get an estimate for the relative precision being too close to 1.

For the sake of a good estimate it could – in the above example and in similar cases – be a possibility to deviate for a moment from the conditional setting by looking at the corresponding *unconditional* figure, i.e. instead of calculating a (possibly unsatisfactory) estimate for

$$\frac{\mathrm{E}\left(\underline{IU}_q \mid \underline{IU}_{(j\leq 0)}\right)}{\widehat{\underline{IU}_q}} = \mathrm{E}\left(\frac{\underline{IU}_q}{\widehat{\underline{IU}_q}} \mid \underline{IU}_{(j\leq 0)}\right)$$

one could try to approximate this figure by an (ideally quite precise) estimate of

$$\mathrm{E}\left(\frac{\underline{IU}_q}{\widehat{\underline{IU}_q}}\right)$$

However, whether this or any alternative approach improves the estimate, depends on the specific situation.

Comparisons of predictors with average realizations are treated in the time series literature under the label *forecast uncertainty,* see Remark 7.10. Besides simulations, various analytical asymptotic results are derived, being exacter the longer the index history. Many (mainly autoregressive) models are looked at, including even attempts to address model risk, however, the by far most investigated model is apparently the AR(1) process with zero mean. The bulk of research seems to address the *unconditional* variant only, and as no clearly distinguishing names for the two problems have been established, there is a certain risk to mix them up.

The *conditional* case analogous to our approach is treated by [Phillips, 1979], [Fuller and Hasza, 1981], and [Ansley and Kohn, 1986]. A lot more literature on forecast uncertainty has been published, see the survey paper [Gonçalves Mazzeu et al., 2015] for a vast overview.

The focus of the literature about forecast uncertainty is predominantly the MSE, which is closely tied to the relative precision introduced above, at least if the future years of the time series are conditionally, given the past years, lognormally distributed.

Example 9.23. To see the connection between relative precision and MSE, let now $\underline{IU}$ be such a "lognormal" time series and $\underline{IY} := \ln\left(\underline{IU}\right)$ its logarithm. If $\underline{IU}_q$ has a lognormal distribution conditionally on $\underline{IU}_{(j\le 0)}$, $\underline{IY}_q$ is normally distributed given $\underline{IY}_{(j\le 0)}$. (The conditions on the past time series $\underline{IU}_{(j\le 0)}$ and $\underline{IY}_{(j\le 0)}$ are obviously equivalent, generating the same σ-algebra.)

So we have $\widehat{\underline{IU}_q} = \exp\left(\widehat{\underline{IY}_q}\right)$ and the lognormal expectation formula

$$\mathrm{E}\left(\underline{IU}_q \mid \underline{IU}_{(j\le 0)}\right) = \exp\left(\mathrm{E}\left(\underline{IY}_q \mid \underline{IY}_{(j\le 0)}\right) + \frac{1}{2}\mathrm{Var}\left(\underline{IY}_q \mid \underline{IY}_{(j\le 0)}\right)\right)$$

from which we get

$$\begin{aligned}\ln\kappa\left(\underline{IU}_q\right) = \ln\mathrm{E}\left(\underline{IU}_q \mid \underline{IU}_{(j\le 0)}\right) - \ln\widehat{\underline{IU}_q} &= \mathrm{E}\left(\underline{IY}_q \mid \underline{IY}_{(j\le 0)}\right) + \frac{1}{2}\mathrm{Var}\left(\underline{IY}_q \mid \underline{IY}_{(j\le 0)}\right) - \widehat{\underline{IY}_q} \\ &= \frac{1}{2}\mathrm{Var}\left(\underline{IY}_q \mid \underline{IY}_{(j\le 0)}\right) - \mathrm{Bias}\left(\underline{IY}_q \mid \underline{IY}_{(j\le 0)}\right)\end{aligned}$$

Thus, the relative precision is a function of (conditional) variance and bias of $\underline{IY}_q$. These two quantities at the same time constitute

$$\mathrm{Var}\left(\underline{IY}_q \mid \underline{IY}_{(j\le 0)}\right) + \mathrm{Bias}^2\left(\underline{IY}_q \mid \underline{IY}_{(j\le 0)}\right) = \mathrm{SPEE}\left(\underline{IY}_q \mid \underline{IY}_{(j\le 0)}\right)$$

If both are very small, we have the desirable situation that the SPEE is close to 0 and the relative precision is close to 1. However, in the relative precision formula a positive bias can somewhat cancel out with the variance even if both are not close to 0, such that a relative precision close to 1 may come about despite of a large SPEE.

Chapter 10

Optimal experience rating

Let us finally apply our theory, proceeding very much in the spirit of the basic example of Chapter 4, but far more generally. Recall that the latter treated the proportional coverage of a risk having (knowingly) constant frequency volume, which was rated for the year zero, such that loss inflation was the only index, having as sole uncertainty the basis risk of past years. Now we are able to treat

- a pro-rata cover, first risk, or layer
- for a risk with variable volume
- that requires a rating for the year $q \geq 0$ (typically $q = 1$ or $q = 2$).

Thus, we take into account uncertainty *both* about inflation and volume, *both* towards the past (basis risk) and the future (overall randomness). As in the basic example, the preferred model for the inflation gap is the geometric random walk with lognormal errors, however, a number of results hold more generally. Recall that our theory is applicable to lognormal settings (see Appendix B), albeit it was initially proven only for normalized time series with almost surely low deviation from 1.

10.1 Preliminaries

We formulate results for the loss amount in this chapter. With the structural analogy outlined in Theorem 9.18, it is straightforward how to extend them to loss count and rates per volume.

For any predictor of X_q we generally have

$$\mathrm{MSEP}\left(T\left(X_q\right)\right) = \mathrm{E}\left(\left(T\left(X_q\right) - X_q\right)^2\right) = \mathrm{Var}\left(X_q\right) + \mathrm{Var}\left(T\left(X_q\right)\right) + \mathrm{Bias}^2\left(T\left(X_q\right)\right) - 2\mathrm{Cov}\left(X_q, T\left(X_q\right)\right)$$

$$\mathrm{SPEE}\left(T\left(X_q\right)\right) = \mathrm{E}\left(\left(T\left(X_q\right) - \mathrm{E}\left(X_q\right)\right)^2\right) = \mathrm{Var}\left(T\left(X_q\right)\right) + \mathrm{Bias}^2\left(T\left(X_q\right)\right)$$

This results from easy algebra and holds, in particular, if $T\left(X_q\right)$ is a linear combination of the S_k, which is the kind of predictor we study in this book. What is more (and is proven by exactly the same algebra), the formulae hold conditionally too, in particular for the condition $\left(M_{(j\leq 0)}, \Upsilon_{pr}\right)$ we want to apply.

If one has a set of predictors available and aims to optimize (i.e. minimize) MSEP and/or SPEE, these two problems are related, but in general not the same. The respective formulae differ by two terms:

- $\mathrm{Var}\left(X_q\right)$ is not related to the predictor – process risk is always part of the MSEP, irrespective of the rating method. While being interesting in itself, it does not matter for the comparison of different rating methods and rating optimization. This simplifies things a lot – among all formulae in Theorem 9.18, that for the future variance is by far the most intricate one.

- The covariance term connects future loss and predictor. However, if they are uncorrelated, both optimizations are equivalent and we can speak of MSE optimization, embracing both MSEP and SPEE.

This is the situation that we have in our setting: The second covariance formula in Theorem 9.18 states that, conditionally on $\left(M_{(j\leq 0)}, \Upsilon_{pr}\right)$, the S_k, and thus linear combinations thereof, are (approximately) uncorrelated to X_q, as long as the comprehensive index system is strongly independent.

Recall that the latter means that, in addition to the independence of the index system, past and normalized future gaps are independent. While this is undeniably a strong assumption, it holds for the geometric random walk, the model that we chose for the inflation basis risk in the basic example. We shall, throughout most of this chapter, use the geometric random walk model with lognormal errors for *all* gaps.

Let us first motivate this model choice. In the comprehensive index system, all gaps are products of powers of the four primordial gaps (independent decomposition). The corresponding true indices are A, G, H, and $\bar{B}$. We interpret each as product of a picked index and a gap. If we feel that the respective picked index is chosen well, we can argue as done in Chapter 3 for the overall loss inflation B: The corresponding gap should be rather less complex than the index itself, such that the geometric random walk is a reasonable model, as is the use of the very common lognormal error structure. Finally, if each primordial basis risk is a lognormal geometric random walk, so are gaps in general – they are all products of powers of primordial gaps.

The independence of the primordial gaps makes it extremely easy to relate parameters and moments of any gap to those of the underlying primordial gaps. We shall detail this later on. For now we work with $\Delta D(\alpha)$ as if it were a simple index; in most of the following $\Delta D(\alpha)$ just takes the role that ΔB had in the basic example of Chapter 4, from which we borrow parametrization and notation, with adequate adaptations for the other gaps we need.

If all gaps are geometric random walks with lognormal errors, $\Delta D(\alpha)_q$ is a product of q independent copies of a lognormal RV W_D, whose moments can be elegantly parametrized by constants v_D, w_D, y_D. (We here again write D for $D(\alpha)$ and will even drop subscripts, as long as there is no risk of confusion with other quantities.)

In compact notation we have

$$\mathrm{E}\left(W\right) = vw, \qquad \mathrm{E}\left(W^2\right) = v^2 w^4$$

For positive k (counting backwards) ΔD_{-k} is a product of k independent copies of W^{-1} with moments

$$\mathrm{E}\left(W^{-1}\right) = \frac{w}{v} = y, \qquad \mathrm{E}\left(W^{-2}\right) = \frac{w^4}{v^2} = y^2 w^2$$

Now we can state the moments of the gaps appearing in Theorem 9.18 in an extremely concise manner, counting again backwards ($k \leq 0$). In particular we have

$$\mathrm{E}\left(\Delta D_{-k}\right) = y^k, \qquad 1 + \mathrm{CV}^2\left(\Delta D_{-k}\right) = w^{2k}$$

$$\mathrm{rCov}\left(\Delta D_{-k}, \Delta D_{-l}\right) = w^{2k} - 1, \qquad l > k \geq 0$$

$\mathrm{E}\left(D(\alpha_i)_{-k}\right)$ can in general not be represented in terms of the parameters of $\Delta D(\alpha)$, as it belongs to the (slightly different) world of $D(\alpha_i)$. We write

$$\mathrm{E}\left(D(\alpha_i)_{-k}\right) = y_i^k, \qquad y_i \coloneqq y_{D(\alpha_i)}$$

The future terms are more complex, but the one we immediately need can be written compactly as

$$\frac{\mathrm{E}\left(\Delta D(\alpha)_q\, \Delta\widetilde{G}_q^{\omega-z}\right)}{\mathrm{E}\left(\Delta\widetilde{G}_q^{\omega-z}\right)} = \frac{\mathrm{E}\left(\Delta Q(\alpha,z)_q\right)}{\mathrm{E}\left(\Delta\widetilde{G}_q^{\omega-z}\right)} = \bar{v}^q\bar{w}^q, \quad \bar{v} := \frac{v_{Q(\alpha,z)}}{v_{\widetilde{G}^{\omega-z}}},\ \bar{w} := \frac{w_{Q(\alpha,z)}}{w_{\widetilde{G}^{\omega-z}}}$$

Notice that $\bar{v} = v_{\Delta D(\alpha)}$, such that typically, with all parameters w being very close to 1, $\bar{v}^q\bar{w}^q$ is pretty close to $\mathrm{E}\left(\Delta D(\alpha)_q\right)$, which means that the distortion via $\Delta\widetilde{G}_q^{\omega-z}$ is only very slight.

Before going into the calculations we give an intuitive interpretation of $\Delta D(\alpha)$. As mentioned various times, the index $D(\alpha) = VB^{\alpha}$ quantifies the combined impact of scale and frequency volume on the aggregate loss (count), with the scale being leveraged due to the non-proportional effect of ground-up inflation on layers. (Details depend on the specific α: With $\alpha_{(d^*,u^*)}$ we quantify the impact on the expectation of the aggregate loss, with α_{d^*} the impact on the expected loss count; other exponents enter some variance components.) ID is the combined picked index used as approximation for this combined impact. The gap ΔD is what one needs to "add" to this picked index to get the true one. Thus, if ID on average underestimates the (combined leveraged) growth of scale/frequency over time, then ΔD grows on average, i.e. $\mathrm{E}(W) = vw > 1$. Conversely, if the picked index overestimates growth (one would intuitively see that as a conservative index prediction), then ΔD decreases on average.

We illustrate how this relates to the risk of being systematically too cheap in the rating altogether: Theorem 9.18 yields for the lognormal-random-walk-gap setting:

$$\mathrm{E}\left(S_{-k} \mid M_{(j\le 0)}, \Upsilon_{pr}\right) \approx \tilde{e}\, y^k$$

$$\mathrm{E}\left(X_q \mid M_{(j\le 0)}, \Upsilon_{pr}\right) \approx \tilde{e}\, \kappa \left(IP(\alpha,z)_q\right) \bar{v}^q\bar{w}^q$$

Whatever value $\kappa\left(IP(\alpha,z)_q\right)$ take (ideally close to 1, but we can imagine it to be on either side): If $y < 1$, the S_{-k} for large k (remote past) will be too cheap as predictors. Thus, one should strive for a picked-index choice yielding $y \ge 1$. Overestimating on average the growth of D yields $\mathrm{E}(W) = vw \le 1$, which is a bit stronger a property than the desired $y = \frac{w}{v} \ge 1$.

Overall, if ID overestimates or just slightly underestimates the growth of D over time, then to rate a sufficient premium one can use the "old" S_{-k} as far as available and representative. If not, one either finds a way to adjust (i.e. increase) the old predictors or better leaves them aside.

Note that the reasoning here is essentially the same as outlined for the overall-loss-inflation gap ΔB after its definition in Chapter 3 and in Section 4.2. Thus, ΔD takes the role that ΔB had in the basic example, not only mathematically, but also materially.

10.2 Classical sample mean

We proceed as in Section 4.2. The common burning cost calculated out of the n most recent years of the loss record can be written as

$$T_n(X_q) := \sum_{k=m}^{m+n-1} g_{n,k} S_{-k}$$

where $-m = k_{max}$ is the most recent year (we count backwards) and $g_{n,k}$ is the weight of the predictor S_{-k}. This notation embraces two variants, the less common equal weighting $g_{n,k} = \frac{1}{n}$ and the traditional volume weighting

$$g_{n.k} = \frac{IV_{-k}}{\sum_{j=m}^{m+n-1} IV_{-j}}$$

which, as already stated in Section 2.2.2, minimizes variance/MSE in some standard models, e.g. if the loss count is Poisson distributed and (as is commonly assumed) past indices are known without basis risk and projected correctly to the future, such that the as-if losses are unbiased predictors.

As done in the basic example, we want to prove that in many real-world situations the MSE of the classical sample mean diverges with increasing n, such that there is an optimal finite number of years to be used. The expected burning cost for geometric-random-walk $\mathit{\Delta} D(\alpha)$ with lognormal errors yields

$$\mathrm{E}\left(T_n\left(X_q\right) \mid M_{(j\leq 0)}, \Upsilon_{pr}\right) \approx \tilde{e} \sum_{k=m}^{m+n-1} g_{n,k} y^k \geq \tilde{e} \prod_{k=m}^{m+n-1} \left(y^k\right)^{g_{n,k}} = \tilde{e} y^{\sum_{k=m}^{m+n-1} k g_{n,k}}$$

For equal weights the sum $\sum_{k=m}^{m+n-1} k g_{n,k}$ in the exponent equals $\frac{n+1}{2}$, which is a divergent sequence in n. In the volume-weighted case the sum equals

$$\frac{\sum_{k=m}^{m+n-1} k \, IV_{-k}}{\sum_{k=m}^{m+n-1} IV_{-k}}$$

It can be shown easily that this sequence diverges, unless the IV_{-k} decrease such quickly that $\sum_{k=m}^{\infty} k \, IV_{-k}$ is finite. However, under realistic circumstances – being inflation-free volumes, the IV_{-k} should rather be quite stable than decrease very quickly – we can mostly expect the sequence to tend to infinity.

All in all, in real-world constellations $\sum_{k=m}^{m+n-1} k g_{n,k}$ usually diverges. Thus, in line with the basic example we have that for $y > 1$ expected Burning Cost, bias, and MSE tend to infinity as n increases.

Like in Section 4.2, it should be possible to prove divergence of the variance under a slightly weaker condition for y (again not for all, but many volume constellations). However, this would be intricate and technical without arguably adding a lot of value. The main message is already clear and the same as found for the basic example:

Classical weighting in the sample mean limits the number of years to be used; to keep the MSE low, one has to stay with the more recent data. To be able to use all available representative data and/or to reduce the MSE, one has to find better weights. This is addressed in the following section.

10.3 Optimal weights

Now we generalize the sample mean, for a *fixed* number $n > 2$ of years of empirical data, by defining

$$T\left(X_q\right) \coloneqq \sum_{k=m}^{m+n-1} g_{k+1-m} S_{-k} = \sum_{k=1}^{n} g_k S_{1-k-m}$$

with arbitrary weights for the S_{-k}. The aim is to find optimal weights in terms of MSE, where optimal shortly means: no bias, minimum MSE.

At first we collect results holding well beyond our specific setting.

10.3.1 General properties

Assume we have n not necessarily unbiased predictors T_1, ..., T_n of a RV X which, conditionally on a certain σ-algebra $\mathcal{U}$, have a positive definite covariance matrix. We want to find their optimal linear combination

$$T\left(\vec{g}\right) \coloneqq \sum_{k=1}^{n} g_k T_k$$

where *optimal* shall mean: unbiased and minimizing the SPEE, conditionally on $\mathcal{U}$.

To this end, we just have to rewrite some findings from Section 4.3 a bit more generally. The first step is a parameter change by defining *rescaled weights*

$$h_k := g_k \frac{\mathrm{E}(T_k \mid \mathcal{U})}{\mathrm{E}(X \mid \mathcal{U})}$$

Then we have

$$T(\vec{g}) = \sum_{k=1}^{n} h_k \mathrm{E}(X \mid \mathcal{U}) \dot{T}_k$$

with the *conditionally standardized* RV's $\dot{T}_k = \frac{T_k}{\mathrm{E}(T_k|\mathcal{U})}$. As the latter have conditional expectation 1, the conditional unbiasedness of $T(\vec{g})$ is equivalent to $\sum_{k=1}^{n} h_k = 1$.

The $\dot{T}_k$ are conditionally unbiased predictors of the conditionally standardized RV $\dot{X} = \frac{X}{\mathrm{E}(X|\mathcal{U})}$, while their covariance matrix $\Sigma := \mathrm{Cov}\left(\dot{T}_k, \dot{T}_l \mid \mathcal{U}\right)$ is positive definite, being just a recalibration of the original covariance matrix.

Now we can apply Proposition 4.4, which obviously holds conditionally as well: It was proven by simple algebra about expectations and variances, which works exactly the same for conditional probability measures. With

$$\Gamma := \left(1 + \mathrm{Cov}\left(\dot{T}_k, \dot{T}_l \mid \mathcal{U}\right)\right) = (1 + \mathrm{rCov}(T_k, T_l \mid \mathcal{U}))$$

we get the optimal linear combination $\sum_{k=1}^{n} h_k \dot{T}_k$ through the coefficients

$$\left(h_1^{'}, ..., h_n^{'}\right)^t = \frac{\Gamma^{-1}\vec{1}}{\vec{1}^t \Gamma^{-1}\vec{1}}$$

Therefore, $\sum_{k=1}^{n} h_k^{'} \mathrm{E}(X \mid \mathcal{U}) \dot{T}_k$ is the optimal predictor of $X = \dot{X}\,\mathrm{E}(X \mid \mathcal{U})$. Expressed in terms of the equivalent original parametrization, the best predictor $\sum_{k=1}^{n} g_k T_k$ has the weights

$$g_k = h_k^{'} \frac{\mathrm{E}(X \mid \mathcal{U})}{\mathrm{E}(T_k \mid \mathcal{U})}$$

Summing up, the optimal rescaled weights are a matrix function of the conditional relative covariances $\mathrm{rCov}(T_k, T_l \mid \mathcal{U}) = \mathrm{Cov}\left(\dot{T}_k, \dot{T}_l \mid \mathcal{U}\right)$ of the predictors T_k; the optimal weights result via multiplication with the quotients of the conditional expectations of RV and respective predictors. Once all needed expectations and relative covariances are found, the two steps, *weights* (optimize) and *level* (ensure unbiasedness), are totally unrelated tasks.

Proposition 10.1. *If n not necessarily unbiased predictors $T_1, \ldots, T_n$ of a RV X have, conditionally on a certain knowledge (σ-algebra) $\mathcal{U}$, a positive definite covariance matrix, then the unbiased linear combination $\sum_{k=1}^{n} g_k T_k$ that conditionally minimizes the SPEE has the coefficients*

$$g_k = h_k \frac{E(X \mid \mathcal{U})}{E(T_k \mid \mathcal{U})}, \qquad (h_1, ..., h_n)^t = \frac{\Gamma^{-1}\vec{1}}{\vec{1}^t \Gamma^{-1}\vec{1}}, \qquad \Gamma = (1 + rCov(T_k, T_l \mid \mathcal{U}))$$

If the T_k are conditionally uncorrelated to X, this linear combination conditionally minimizes the MSEP as well.

Proof. The proposition summarizes the preceding results, adding the case of equivalence of SPEE and MSEP minimization as explained at the beginning of this chapter. □

10.3.2 Special results

Now we apply the last Proposition to the n predictors S_{-m}, ..., S_{-m-n+1}, in the conditional setting, given $\mathcal{U} = \sigma\left(M_{(j\leq 0)}, \Upsilon_{pr}\right)$.

Proposition 10.2. *In the situation of Theorem 9.18, the approximate* level factors *for the rescaling of the weights are, for* $k = 1, ..., n$*,*

$$\frac{g_k}{h_k} = \frac{E\left(X_q \mid M_{(j\leq 0)}, \Upsilon_{pr}\right)}{E\left(S_{1-k-m} \mid M_{(j\leq 0)}, \Upsilon_{pr}\right)} \approx \frac{E\left(\Delta Q(\alpha,z)_q\right)\kappa\left(IP(\alpha,z)_q\right)}{E\left(\Delta\widetilde{G}_q^{\omega-z}\right)E\left(\Delta D(\alpha)_{1-k-m}\right)}$$

while the rescaled weights *minimizing the MSE across all unbiased linear combinations are*

$$(h_1, ..., h_n)^t = \frac{\Gamma^{-1}\vec{1}}{\vec{1}^t\Gamma^{-1}\vec{1}}$$

with the (approximate) matrix

$$\Gamma = \left(1 + rCov\left(S_{1-k-m}, S_{1-l-m} \mid M_{(j\leq 0)}, \Upsilon_{pr}\right)\right) \approx 1 + rCov\left(\Delta D(\alpha)_{1-k-m}, \Delta D(\alpha)_{1-l-m}\right)(1+\delta_{kl}\beta)$$
$$+ \delta_{kl}\frac{1}{IV_{1-k-m}}\max\left(\sum_{i=1}^{4} a_i \frac{E\left(\Delta D(\alpha_i)_{1-k-m}\right)}{E^2\left(\Delta D(\alpha)_{1-k-m}\right)}, \sum_{i=1}^{j} a_i^{(j)} \frac{E\left(\Delta D\left(\alpha_i^{(j)}\right)_{1-k-m}\right)}{E^2\left(\Delta D(\alpha)_{1-k-m}\right)}\right)$$

Proof. Simply collect the respective terms from Theorem 9.18. □

The proposition holds generally, as long as the assumptions of the theorem hold. Now we specify them for our preferred distribution model, having lognormally distributed random walk gaps.

Corollary 10.3. *Specifically for geometric-random-walk gaps, as long as they, as well as all other primordial indices, have lognormal errors, we have*

$$\frac{g_k}{h_k} \approx \frac{\kappa\,\bar{v}^q\bar{w}^q}{y^{m+k-1}} \tag{10.1}$$

where we have shortly written κ *for* $\kappa\left(IP(\alpha,z)_q\right)$*. For* $n \geq l \geq k \geq 1$ *we have*

$$\Gamma_{kl} \approx w^{2(m+k-1)} + \delta_{kl}\left\{\beta w^{2(m+k-1)} + \frac{1}{IV_{1-k-m}}\max\left(\sum_{i=1}^{4} a_i \frac{y_i^{m+k-1}}{y^{2(m+k-1)}}, \sum_{i=1}^{j} a_i^{(j)} y_{D\left(\alpha_i^{(j)}\right)}^{m+k-1} y^{-2(m+k-1)}\right)\right\} \tag{10.2}$$

Proof. As is shown in Appendix B.2.3, Theorem 9.18 applies to the case that all time series, gaps and picked indices, are generated by lognormal error terms. □

We see that the main ingredient of the *level-factor* formula is $\kappa = \kappa\left(IP(\alpha,z)_q\right)$, which is only slightly distorted by some gap parameters being rather close to 1. Thus, ensuring unbiasedness means essentially index prediction. Conversely, the *weight* formula has as only ingredient from the picked indices the past frequency volume, further it depends on several gap parameters and the variance coefficients. Of course, both formulae require the exponents of the index system and some alphas.

If we compare Formula 10.2 to the corresponding matrix

$$\Gamma_{kl} = w^{2(m+k-1)} + \delta_{kl}\eta\, w^{2(m+k-1)}, \qquad l \leq k$$

of the basic example, we see that, apart from β taking the place of η, now the diagonal is much more complex, such that we cannot generally apply the Fibonacci algebra that in Section 4.3.2 led us to concise explicit formulae for the optimal weights.

Nevertheless some pretty general properties hold.

Proposition 10.4. *In the situation of Theorem 9.18, applied to time series having lognormal errors throughout and geometric-random-walk gaps, the matrix $\Xi = \Theta + \Lambda$, where $\Theta_{kl} = w^{2(m+\min(k,l)-1)}$ and Λ is the diagonal matrix with strictly positive diagonal elements*

$$\Lambda_{kk} := \Lambda_k := \beta w^{2(m+k-1)} + \frac{1}{IV_{1-k-m}} \max\left(\sum_{i=1}^{4} a_i \frac{y_i^{m+k-1}}{y^{2(m+k-1)}}, \sum_{i=1}^{j} a_i^{(j)} y_{D\left(\alpha_i^{(j)}\right)}^{m+k-1} y^{-2(m+k-1)}\right)$$

approximates the relative covariances of the sample mean.

Ξ is positive definite.

The components of the vector $\Xi^{-1}\vec{1}$ are strictly positive, such that the same holds for the resulting (approximate) optimal rescaled weights

$$(h_1, ..., h_n)^t = \frac{\Xi^{-1}\vec{1}}{\vec{1}^t \Xi^{-1}\vec{1}}$$

For the latter we have

$$\frac{h_{k+1}}{h_k} < \frac{\Lambda_k}{\Lambda_{k+1}}$$

If in particular the Λ_k increase in k, the h_k decay faster than the inverses of the Λ_k.

Proof. Ξ is defined exactly such that it emerges from the last corollary as relative covariance matrix of the as-if quantities.

That the Λ_k are strictly positive, is ultimately (for the problematic case that the first sum in the maximum has a dominant fourth negative summand) ensured by the second sum (which complicated the variance calculations so much). So Λ is positive definite. The same holds for Θ, which is the matrix in Proposition 4.6 consisting of powers of $r = w^2 > 1$ (case $\eta = 0$). Thus, $\Xi = \Theta + \Lambda$ is positive definite, too.

That the components of $\Xi^{-1}\vec{1}$ are nonnegative, was proven by [Riegel, 2015], Appendix C, by induction, even more generally for matrices $\Theta_{kl} = \Theta_{\min(k,l)}$, where $0 < \Theta_1 < ... < \Theta_n$, and nonnegative diagonal elements Λ_k. (Notice that the referred paper has the inverse order of rows and columns.)

In the vector equation

$$\Xi\,(h_1, ..., h_n)^t = \frac{\vec{1}}{\vec{1}^t \Xi^{-1}\vec{1}}$$

all rows are equal, such that for subsequent rows k and $k+1$ we have

$$\sum_{l=1}^{n} \Theta_{kl} h_l + \Lambda_k h_k = \sum_{l=1}^{n} \Theta_{k+1,l} h_l + \Lambda_{k+1} h_{k+1}$$

or equivalently

$$\Lambda_k h_k = \sum_{l=1}^{n} \left(\Theta_{k+1,l} - \Theta_{kl}\right) h_l + \Lambda_{k+1} h_{k+1} \tag{10.3}$$

As both the h_l and the $(\Theta_{k+1,l} - \Theta_{kl})$ are nonnegative, we have $\Lambda_k h_k \geq \Lambda_{k+1} h_{k+1}$.

If we can prove that this inequality is strict with all h_k being positive, we are done. Assume that conversely a certain h_k equals 0. Then this must hold as well for h_{k+1} and all subsequent ones. However, not all h_k can equal 0, such that we must have a certain index such that

$$h_k > 0 = h_{k+1} = ... = h_n$$

Plugging this in Equation 10.3 we get

$$0 < \Lambda_k h_k = \sum_{l=1}^{n} \left(\Theta_{k+1,l} - \Theta_{kl}\right) h_l + \Lambda_{k+1} h_{k+1} = \sum_{l=1}^{k} \left(\Theta_{k+1,l} - \Theta_{kl}\right) h_l = \sum_{l=1}^{k} \left(\Theta_l - \Theta_l\right) h_l = 0$$

where the third equality follows from the special structure of the matrix Θ, while to the very left we have used (for the first time) that the diagonal of Λ has strictly positive elements. We have a contradiction, which implies $h_k > 0$ for all k. Thus, the sum in Equation 10.3 cannot vanish, such that we must have $\Lambda_k h_k > \Lambda_{k+1} h_{k+1}$ or equivalently $\frac{h_{k+1}}{h_k} < \frac{\Lambda_k}{\Lambda_{k+1}}$. □

This proposition enables us to draw a number of conclusions.

Firstly, it states that despite all the approximations used on the way to the matrix Ξ we have not lost the essential property of the relative covariance matrix: positive definiteness. So we can work with Ξ as if it described the exact relative covariance between the as-if losses. Of course we aim for (and trust in) good approximations, however, should a case of less precise approximation arise, at least this does not entail technical issues.

Secondly, whichever values the highly complex diagonal may take, we get some general properties, having an intuitive interpretation and being in line with what one would have expected: All years have positive weights, reflecting that an optimal sample mean does not discard any representative loss history, however old it be. What is more, it turns out that the weights in typical cases decrease with rising index k. Recall that larger k, as we count backwards, means older years, for which we expect more basis risk in the indices. Thus, lower weights for older years are perfectly in line with intuition, as long as no extreme volume shifts change the picture. Now look at the definition of the Λ_k. The first summand (slightly) rises with k. The second one is complex, but if the $IV_{1-k-m} = I\bar{M}_{1-k-m}^{1-\chi} IA_{1-k-m}$ are not much greater for older years than for more recent ones (which in practice hardly occurs, unless $\chi = 0$ and the portfolio size $\underline{I\bar{M}}$ shrinks quickly; sharp decreases of the frequency per unit $\underline{IA}$ are largely implausible), altogether the Λ_k should rise in k or decrease only slightly. With the $\frac{\Lambda_k}{\Lambda_{k+1}}$ being close to 1 or smaller, the $\frac{h_{k+1}}{h_k}$ are mostly less than 1, and so are the

$$\frac{g_{k+1}}{g_k} = \frac{h_{k+1}}{y h_k}$$

at least if y is in the range of 1 or greater. Thus, in typical real-world constellations the weights decrease with age.

Notice that the Λ_k are not just components of a complex (partly conditional) covariance term – they are variances themselves, namely the conditional relative variances (squared CV's) of the predictors *before* averaging over the unobserved index values. To see this, apply Proposition 7.4 to lognormal random-walk gaps.

Last but not least, we can approximately calculate the h_k for some special cases. What if this matrix is very close to a simpler one, say the matrix of the basic example or a diagonal matrix? Approximations of complex matrix products are tricky, in particular if their inverses are involved, however, we are in the benign situation that Ξ is the sum of two positive definite (thus invertible) matrices, such that for quite fair approximations Ψ of Ξ, and ultimately of Γ, we can claim that

$$(h_1, ..., h_n)^t = \frac{\Gamma^{-1}\vec{1}}{\vec{1}^t \Gamma^{-1}\vec{1}}$$

is well approximated by $\frac{\Psi^{-1}\vec{1}}{\vec{1}^t \Psi^{-1}\vec{1}}$.

10.3.3 Examples

Now recall the parameter discussion in Section 7.2.2, where we studied which of the coefficients appearing in the variance formulae can be the largest and in what real-world situations this occurs. Out of the situations discussed then, two emerge as being instructive and realistic, while yielding explicit (approximate) rescaled weights. These two cases are as different as we can imagine.

Example 10.5. $\beta > 0$ dominates. The classical real-world example is the pro-rata coverage of a large portfolio producing thousands of non-Cat losses per year, while the loss count has a considerable market fluctuation, but conditionally on that, is nearly undispersed (Poisson or very close to). Here first and second sum have no substantial impact, such that

$$\varGamma_{kl} \approx \varXi_{kl} \approx w^{2(m+\min(k,l)-1)} + \delta_{kl}\beta w^{2(m+\min(k,l)-1)} = r^{m+\min(k,l)-1}\left(1+\delta_{kl}\beta\right)$$

where $r = w^2$. This is the relative covariance matrix of the basic example, save that β (the main variance component) takes the place of η, which then was the total squared CV of the normalized loss. So we get the h_k from Proposition 4.8: by simply replacing η by β. The g_k result from Formula 10.1, which contains the same powers of y as in the basic example, but additionally the factor $\kappa\bar{v}^q\bar{w}^q$ comes in, reflecting the projection into the future year q, which was not necessary in the basic example, where we had $q = 0$.

Numerically, the decrease of the weights can be expected to be even more marked than in the basic example, as now we should have more basis risk, but less fluctuation from sources other than indices:

For the single say Fire risk studied in the basic example we regarded a loss CV of say 20% to 50% as plausible, see Example 4.11. This yields an η in the range of 0.04 to 0.25. Instead, β is the variance of the common market factor, creating a whole-market fluctuation in the ground-up loss frequency. As outlined in Section 7.2.2, this variance should in practice rarely be larger than 0.01. (β dominates over the other variance coefficients, but is nevertheless rather small. The other coefficients are much smaller still or vanish.)

On the other hand, now $w = \exp\left(\sigma^2/2\right)$ does not just reflect the one-year gap of some single-loss inflation $B = H\bar{B}^\tau$ like Fire loss inflation, but embraces the overall basis risk of the combined leveraged scale-frequency effect $D(\alpha)$, which could be considerably larger. To see this, let us calculate σ^2. In the independent comprehensive index system this is very easy, as the multiplicative structure

$$\varDelta D(\alpha) = \varDelta A^{1-\delta\omega}\varDelta G^{-\omega}\varDelta H^{\alpha}\varDelta\bar{B}^{\tau\alpha-\zeta\omega}$$

becomes additive if we take the logarithms. Due to the independence of the primordial gaps we get a sum of independent normally distributed RV's. The log one-year gap has the same structure and its variance equals

$$\sigma^2_{D(\alpha)} = \left(1-\delta\omega\right)^2\sigma^2_A + \omega^2\sigma^2_G + \alpha^2\sigma^2_H + \left(\tau\alpha-\zeta\omega\right)^2\sigma^2_{\bar{B}}$$

For comparison, $\varDelta B = \varDelta H\varDelta\bar{B}^\tau$ leads analogously to

$$\sigma^2_B = \sigma^2_H + \tau^2\sigma^2_{\bar{B}}$$

At least for $\alpha \geq 1$, $\sigma_{D(\alpha)}$ should be larger than σ_B, despite some mitigation for the basic inflation through the term $-\zeta\omega$. We anyway have additional components stemming from A and G. In particular for reinsurance with the original premium as official volume, we can expect the basis risk about G (reflecting here the difficult-to-assess market cycle) to play a mayor, if not the dominant role.

For illustration we look at the instructive (and not unrealistic) case $\tau = 1$, $H \equiv 1$. Here σ_B coincides with $\sigma_{\bar{B}}$, while

$$\sigma^2_{D(\alpha)} = \left(1-\delta\alpha\right)^2\sigma^2_A + \alpha^2\sigma^2_G + \left(1-\zeta\right)^2\alpha^2\sigma^2_{\bar{B}}$$

in the Cat case and

$$\sigma_{D(\alpha)}^2 = (1-\delta)^2 \sigma_A^2 + \sigma_G^2 + (\alpha - \zeta)^2 \sigma_{\bar{B}}^2$$

for single-loss models.

In real-world Cat covers the official volume mostly contains inflation (original premium, aggregate sum insured), such that $\zeta = 1$ and basic inflation drops out of the above formula. Here $\sigma_{D(\alpha)}$ equals σ_G, leveraged with α, plus the additional component stemming from $\varDelta A$.

For non-Cat σ_G is not leveraged, while $\sigma_{\bar{B}}$ is, namely by α or $\alpha - 1$, according to whether the official volume is inflation-free or not.

Altogether we can expect $\sigma_{D(\alpha)}$ to be larger than $\sigma_{H\bar{B}^\tau}$ in most constellations, possibly much larger. Numerical examples with various values β and $\sigma_{D(\alpha)}$ are displayed in Appendix A, together with the analogous ones for the basic example with parameters η and σ_B. Note in particular the case of 12 years of data, coming with 3 different scenarios for σ and β (η): The larger the former and/or the smaller the latter, the steeper the curve of the optimal weights and the larger the weight of the most recent year – it can exceed 25%, notably for parameter constellations that are not extreme.

A much different approximation becomes applicable if the diagonal matrix $\varLambda$ has much larger values than in this example. Recall that $\varTheta$ is a matrix of powers of w which, albeit typically reflecting more basis risk than in the basic example, are usually still quite close to 1. Thus, $\varTheta$ is close to the (singular) matrix of 1's, and if the invertible matrix $\varLambda$ has very high entries, the deviation of the $\varTheta_{kl}$ from 1 has a very low numerical impact, such that we get

$$\varGamma_{kl} \approx \varXi_{kl} \approx 1 + \delta_{kl} \varLambda_k$$

As in the proof of Lemma 4.3 we can see easily that

$$\frac{\varGamma^{-1}\vec{1}}{\vec{1}^t \varGamma^{-1} \vec{1}}$$

yields the same vector as when we replace $\varGamma$ by the matrix $(\varGamma_{kl} - 1)$. This latter matrix here approximately equals the diagonal matrix $\varLambda$, such that we are (approximately) back in the classical situation of uncorrelated predictors with a diagonal covariance matrix. Here the optimal weights are well known: they are proportionate to the inverse variances, which is indeed the output of

$$\frac{\varLambda^{-1}\vec{1}}{\vec{1}^t \varLambda^{-1} \vec{1}}$$

Let us look at the most realistic special case of this situation.

Example 10.6. a_2 is very large and dominates the variance. The classical real-world example is a high layer with a loss frequency well below 10% and a severity rather similar to Pareto. If we rate the loss count (rate), we do even not require this assumption on tail geometry. In any case, the first sum in the variance formula is dominated by the second summand and the improvement through the second sum is obsolete.

$$a_2 = \frac{\overline{\varepsilon}(d^*, u^*)}{\theta_{d^*}} > \frac{1}{\theta_{d^*}}$$

is in the range of 100 for such layers. With a_2 dominating, the impact of the parameters β and γ of the mixed contagion loss count model is approximately nil, only the frequency comes in: Here we essentially have no model risk from the loss count side, as one is anyway numerically close to the Poisson model.

Things simplify further. We have

$$\Lambda_k \approx \frac{1}{IV_{1-k-m}} a_2 \frac{y_2^{m+k-1}}{y^{2(m+k-1)}}, \qquad y_2 = y_{D(\alpha_2)}$$

where α_2 equals $\alpha = \alpha_{d^*}$ for loss-count rating (such that $y_2 = y$), while for aggregate-loss rating α_2 is the second-order regional alpha of the (rescaled) layer. But, if the severity is similar to Pareto, the two regional alphas are close, such that $y_2 \approx y$. So we generally have

$$\Lambda_k \approx \frac{a_2}{IV_{1-k-m} y^{m+k-1}}$$

which implies that the optimal rescaled weights h_k are proportionate to the $IV_{1-k-m} y^{m+k-1}$. These are the (picked) frequency volumes, being slightly distorted by powers of y.

But simplification does not end here. Formula 10.1 yields the original weights as

$$g_k \sim \frac{h_k}{y^{m+k-1}} \sim IV_{1-k-m}$$

Recalling that g_k belongs to the predictor S_{1-k-m}, we can state in the original notation with negative k: The optimal weight of S_k in the sample mean is IV_k, which is the classical weight of the standard model without basis risk in picked indices. The stochastic effects due to index basis risk cancel (approximately) out here.

Summing up, pro-rata coverages of large portfolios and high layers protecting risks having a Pareto-like severity tail are two benign endpoints in a complex universe of constellations. Both require only a few model characteristics. They yield remarkably different optimal weights: the former Fibonacci-like as in the basic example, the latter proportionate to the (frequency) volume as in traditional models ignoring past index uncertainty.

Myriads of other constellations exist, being mainly driven by the proportions among the variance parameters. Many will lead to weights somewhere in between the two above examples, some might look much different and be inaccessible for easy qualitative description. However, if the parameters of the (ground-up) loss count and the regional/local alphas of the normalized severity (in the layer area) are approximately known (i.e. inferred somehow), one can do the bias correction and calculate the optimal weights by simple algebra from Formulae 10.1 and 10.2. The remaining ingredients come from the world of indices: the exponents of the index system, the parameters of the primordial gaps, and the relative precision of the combined index $IP(\alpha, z)$.

Chapter 11

Final words

11.1 What we have done

The initial question, *to use or not to use old data* in experience rating, has led us on a long journey. Some of the results found on the way are interesting in themselves, independently of the context.

- The three-parameter *mixed contagion model* for the loss count (Chapter 5) generalizes two classical models: variance proportional to volume due to independent insured units vs Poisson-Gamma with strong dependence through a global fluctuation. It fills the gap between them, providing a model for all situations that are felt being somewhat intermediate. The *contagion* helps infer loss count parameters for a layer from those of another layer or even from the ground-up losses. The latter, if available, should be many enough for a sound parameter estimation.

- The *regional Pareto alpha* (Chapter 6) enables us to describe the nonlinear impact of ground-up inflation on layers in a general and elegant way, bridging the gap between the two well-understood cases: proportional (re)insurance vs layers with Pareto distributed loss severity. An easy-to-handle multiplicative effect of inflation emerges, coming with a leverage depending on the geometry of the loss severity cdf in the layer area. The general concept does not even require a smooth cdf.

- The *inflationary world structure* (Section 7.1) may be useful in other contexts, too. Its assumptions on the severity are a straightforward formalization of the commonly assumed uniform impact of inflation on ground-up losses independently of their size. For the loss count one could think of modifications and/or specifications, e.g. if one needs to specify more than just first and second moment.

- The *comprehensive index system* (Chapter 8, being inspired by the variety of real-world examples from Chapter 2) is a decomposition of all indices describing shifts over time in the loss experience. This embraces not only (various kinds of) inflation, but also the (various kinds of) volume of the insured risk. The interpretation of the volume as an index is in fact key for the understanding of the analogies and intricate interactions of all shifts over time affecting insurance risks. This understanding should be helpful far beyond our specific setting with gaps, picked indices, and a number of independence assumptions.

- The formulae for the *SPEE-minimizing weights* for n possibly biased, possibly correlated predictors (Section 10.3.1, being based on Section 4.3.1) are general, neither restricted to the sample mean nor to our specific covariance structure.

All other results are closely interweaved, which justifies the decision to assemble them in a monograph instead of writing several articles.

- The *inflation gap* and the *basic example* (Chapters 3, 4) at first sight appeared stylized, but opened the road to generalizations reflecting the complexity of real-world (re)insurance. This required a lot of small steps (Chapters 7, 9), including the thinking about adequate conditioning, however, at the end (Chapter 10) the basic example reappears as one of the easy-to-grasp special cases having an explicit formula for the optimal weights of the loss data according to their age.

- Speaking of *optimal weights*, this task emerges as the better question than: which years to use. If we can assess the quantities needed to answer the initial question, we can use them to debias the as-if losses and to give them optimal weights, decreasing with age. This enables us to use all available representative years of the loss record and, more importantly, to reduce the MSE.

- Albeit the resulting formulae are complex and tie a myriad of pieces together, they make clear that experience rating consists of a number of tasks that can be addressed quite independently of each other. In particular, the *debiasing* of the as-if losses is essentially a classical time series task which, apart from a minor correction via the gaps, is the same for all years: Five indices (at most) have to be projected to the future year q. The only ingredient from the insured losses is the geometry of the severity cdf (regional/local alpha). Conversely, the optimal weights can be calculated from the time-invariant gap parameters and variance coefficients – as for empirical time series, only the history of the picked frequency volume comes in.

- In all this complexity we find a number of cases where some *uncertainties cancel out* (or nearly out), simplifying the rating task. Most interesting are the situations where the uncertainty about basic inflation, the core of the whole exercise, does not matter ... while less prominent uncertainties come to light.

- Besides, the results reveal deep analogies, and subtle differences, between the rating of: aggregate loss vs loss count, amount vs rate on volume. En passant we get a mathematical, not just pragmatic, justification for charging *rates per volume* to risks of variable size – the rating of rates bears much lower uncertainties than the rating of dollar amounts.

11.2 What to do in practice about it

This book is theoretical work. It can be applied most easily in an ideal world, where in addition to a great data history we:

- easily identify the indices and exponents of the comprehensive index system,

- easily find adequate primordial picked indices,

- know the parameters of the respective gaps.

However, the intention of this book was to inspire *real-world* actuarial work. Theory can indeed help practice even when the world is less ideal than described above, contributing to questions like:

- To solve a difficult quantification task like insurance premium rating, which problems do we have to solve with great accuracy? Which issues are less sensitive, such that rough answers suffice?

- One step further: environment change. If different options make sense, e.g. for the payment terms of (re)insurance covers, which ones could we boost to ease premium rating?

From this standpoint, the main lessons learned from the theory developed in this book are the following:

- We must put a lot of effort into the specification of the index system, in particular in case the official volume contains inflation. We absolutely need to understand clearly which parts of the various inflations can be explained by common sources.
- We must put a lot of effort into finding or constructing primordial picked indices. They should have a pretty long history, however, it is even more important that they be such closely related to the business in question that we can trust them to be quite similar to the unknown corresponding true indices. Otherwise the assumptions made about the primordial gaps (low volatility, simple structure, independence of corresponding picked index) are not plausible.
- For pro-rata coverages and low layers the loss count model has a high impact on the variance. This is an arguably underestimated model risk. One should try to infer the parameters of the mixed contagion model (or of better loss count models yet to be discovered) from adequate ground-up loss count data having such high loss numbers that the inference of three parameters makes sense. Note that this data in practice is rarely provided for the rating, one usually gets reported only aggregate losses or large losses, respectively. Thus, one should ask for the ground-up loss count or try to gather other suitable data, e.g. market-wide statistics.
- To know the leverages we need regional and/or local alphas. We do not know the severity, but quite some cases can be tackled easily: First risks should have parameters very similar to pro-rata coverages, while layers, apart from very low ones, usually attach in the loss size range where the severity is rather heavy-tailed. Local alphas may vary in this area, but often rather slowly, such that they can be assessed by observing the slope of the empirical survival function on log-log scale. (To do that perfectly, one would have to correct losses by the true inflation, however, for not-too-variable local alphas the method is rather insensitive to somewhat wrongly inflation-adjusted losses.) From the local alphas we can approximate the regional ones. Once we have the relevant alphas, we know all exponents appearing as leverage of gaps and picked indices.
- For the gap parameters, at least for now, the only option is arguably expert judgment (more modestly: educated guessing). The discussion about the inflation gap in Chapter 3 might give some orientation about parameter size.
- If one feels uncomfortable about the choice of picked indices and gap parameters, the main concern is probably not optimal weighting, but the fear of being systematically too cheap. Somewhat conservative picked indices may help and are indeed not uncommon in practice. However, even in such a situation the (supposedly) optimal weights, being based on quite uncertain parameters, steer into the right direction: Lower weights for the remote past reduce the impact of wrong index choice.

To end with, let us discuss a potential environment change:

Now that we have written down it formally, it is evident how intricate the as-if adjustment of data via indices for experience rating really is. The main source of complexity is the inflation contained in the official volume, which is a business-driven choice, not an inseparable mathematical parameter of the risk. In other words, it can possibly be changed:

In some situations one could come to the conclusion that a well-established official volume is tied to loss inflation in such a difficult-to-assess manner that one better uses an alternative, maybe even an inflation-free volume. Such volumes do not lead to automatic premium increases due to inflation, however, this disadvantage could in many cases be outweighed by the much simplified experience rating.

Particularly in reinsurance with its traditional volume, the *original premium*, there might be room for making the life of pricing actuaries easier. Despite its undeniable advantages (transparent, tied to the risk assessment of the ceding insurer) we have seen clearly that this official volume leads to a highly problematic superimposed premium inflation index G. Firstly G here often has an even greater impact than basic inflation has, secondly it essentially reflects the market cycle, which has proved to be often erratic and hard to predict even short-term, thirdly this market cycle usually affects many reinsured portfolios at the same time, increasing the correlation across reinsurance treaties.

This “accumulation” of issues could motivate reinsurers to think about alternative official volumes (aggregate sum insured, vehicle years, tonnage, ...), as they are occasionally used in some business lines.

Appendix A

Numerical examples

The numerical scenarios for the basic example (variants of Example 4.11) and the analogous Example 10.5 are assembled in a comprehensive table, which has four areas:

- The top frame displays η (or β) and the parameters of the one-year gap W, representing them in various useful ways.
- Below the key figures to optimize the sample mean are shown: optimal number of years for the classical sample mean; the power bases p and q needed to calculate the optimal weights for the generalized sample mean.
- The third frame compares the SPEE and its square root S.E. (standard error) for the two optimizations studied: classical sample mean using the (rounded) optimal number of years; optimally weighted sample mean for various numbers of years n.
- At the bottom the optimal weights are displayed, starting with the most recent year.

For ease of orientation several cells are left empty; the missing values are equal to the corresponding ones of the preceding column. For ease of reading some figures are displayed in percent.

$\sqrt{\eta}$, $\sqrt{\beta}$ [%]	20					20	50	10		10		
η, β	0.04					0.04	0.25	0.01		0.01		
CV(W) [%]	3					1.2	3	3		5		
$r-1$	.0009					.0001	.0009	.0009		.0025		
$w-1$	.0004					.0001	.0004	.0004		.0012		
σ	0.03					0.012	0.03	0.03		0.05		
E$(W)-1$ [%]	0.09				0.5	0.01	0.09	0.09		0.25		
y	1				1.004	1	1	1		1		
m	1				1	1	1	1		1		
p	1.16				1.16	1.06	1.06	1.35		1.64		
q	0.86				0.86	0.94	0.94	0.74		0.61		
n_0 approx.	11.55				11.55	28.87	28.87	5.77		3.47		
n_0	11.57				11.60	28.88	28.89	5.82		3.54		
n	12	16	30	50	12	30	30	6	12	3	6	12
average weight [%]	8.3	6.3	3.3	2.0	8.3	3.3	3.3	16.7	8.3	33.3	16.7	8.3
SPEE n_0	.0074				.0076	.0028	.0179	.0039		.0072		
S.E. n_0 [%]	8.6				8.7	5.3	13.4	6.3		8.5		
SPEE opt. weight	.0069	.0066	.0065	.0065	.0069	.0026	.0166	.0037	.0035	.0070	.0065	.0064
S.E. opt. wgt. [%]	8.3	8.1	8.1	8.1	8.3	5.1	12.9	6.1	5.9	8.4	8.0	8.0
g_k^* of year:												
1 (latest)	14.8	14.2	14.0	14.0	14.8	6.2	6.2	27.6	25.9	44.7	39.4	39.1
2	12.9	12.3	12.0	12.0	12.9	5.8	5.9	21.1	19.2	30.8	24.2	23.8
3	11.2	10.6	10.3	10.3	11.2	5.5	5.5	16.4	14.3	24.6	15.0	14.5
4	9.9	9.2	8.9	8.9	9.9	5.2	5.2	13.3	10.6		9.6	8.8
5	8.7	8.0	7.7	7.7	8.7	4.9	5.0	11.3	7.9		6.6	5.4
6	7.7	6.9	6.6	6.6	7.7	4.7	4.7	10.4	5.9		5.3	3.3
7	6.9	6.0	5.7	5.7	6.9	4.4	4.4		4.5			2.0
8	6.3	5.3	4.9	4.9	6.3	4.2	4.2		3.4			1.2
9	5.8	4.7	4.2	4.2	5.8	4.0	4.0		2.7			0.8
10	5.5	4.2	3.6	3.6	5.5	3.8	3.8		2.1			0.5
11	5.2	3.7	3.1	3.1	5.2	3.6	3.6		1.8			0.3
12	5.1	3.4	2.7	2.7	5.1	3.4	3.4		1.7			0.3
13		3.1	2.3	2.3		3.3	3.3					
14		2.9	2.0	2.0		3.1	3.1					
15		2.8	1.7	1.7		3.0	3.0					
16		2.7	1.5	1.5		2.9	2.9					
17			1.3	1.3		2.7	2.7					
18			1.1	1.1		2.6	2.6					
19			1.0	0.9		2.5	2.5					
20			0.8	0.8		2.5	2.4					
n (oldest)			0.3	0.02		2.0	2.0					

Appendix B

Random-inflation leverage

We want to extend the theory for leveraged layer inflation developed in Chapter 6 to the case of random ground-up inflation, as it is investigated from Section 7.5 onwards. In principle we are interested in knowing whether and how the fundamental approximative Formulae 6.2 and 6.3 for layer inflation, which hold under mild regularity conditions for the loss severity cdf, can be preserved if, instead of a known fixed ground-up inflation g being close to 1, we have a positive RV Q being close to 1. What "Q close to 1" in this context exactly means, has to be specified: The aim is to embrace far more RV's than the trivial case of Q being almost surely restricted to a small interval about 1, in order to make the results of Section 7.5 and the following chapters applicable to models for inflation basis risk and other gaps that are more interesting and realistic than nearly-constant RV.

The straightforward generalizations of the basic pointwise approximations would be:

$$\mathrm{E}\left({}_d^u\left(QZ\right)\right) \overset{!}{\approx} \mathrm{E}\left({}_d^u Z\right)\mathrm{E}\left(Q^{\alpha(d,u)}\right)$$

$$\mathrm{E}\left(\chi_{(QZ>d)}\right) \overset{!}{\approx} \mathrm{E}\left(\chi_{(Z>d)}\right)\mathrm{E}\left(Q^{\alpha_d}\right)$$

However, to be able to exploit the formulae developed in Sections 7.2 and 7.3, we need something even more general. If we look at the proofs of Propositions 7.4 and 7.12, we see that in the first place we would like the approximation

$$\begin{aligned}\mathrm{E}\left(S_k \mid \mathcal{W}\right) = \widehat{IV_q}\widehat{IB_q}\varDelta V_k\,\theta\mathrm{E}\left({}_{d^*}^{u^*}\left(\varDelta B_k \mathring{Z} \mid \mathcal{W}\right)\right) &\approx \widehat{IV_q}\widehat{IB_q}\varDelta V_k \varDelta B_k^{\alpha(d^*,u^*)}\theta\mathrm{E}\left({}_{d^*}^{u^*}\mathring{Z} \mid \widehat{IB_q}\right)\\ &= e^*\widehat{IV_q}\widehat{IB_q}\varDelta V_k \varDelta B_k^{\alpha(d^*,u^*)} = \tilde{e}\varDelta V_k \varDelta B_k^{\alpha(d^*,u^*)} = \tilde{e}\varDelta D\big(\alpha_{(d^*,u^*)}\big)_k\end{aligned}$$

to hold in expectation, conditionally given $\underline{I}$ with $\sigma\left(\underline{IV}_{(j\leq 0)}, \underline{IB}_{(j\leq 0)}\right) \subseteq \underline{I} \subseteq \mathcal{W}$. Thus, in addition to the normalized severity $\mathring{Z}$ and the inflation gap $\varDelta B_k$ (which shall be close to 1 in an appropriate way), as third RV the volume gap $\varDelta V_k$ comes in. What is more, the two gaps can be dependent, only their independence of $\mathring{Z}$ is clear due to the inflationary world structure.

B.1 Basic inequalities

Before investigating this setting in detail, let us shortly look at the key figures of layers in general. Average layer loss and respective loss count have upper bounds that are trivial and no sharp, but nevertheless useful.

Lemma B.1. *For a ground-up loss Z and a layer $u - d\,xs\,d$ we have the following upper bounds:*

$$E\big(\chi_{(Z>d)}\big) \le 1, \qquad E\big({}^{u}_{d}Z\big) \le \min\left(u, E(Z)\right), \qquad E\Big(\big({}^{u}_{d}Z\big)^2\Big) \le \min\left(u^2, E\left(Z^2\right)\right)$$

Proof. The loss count $\chi_{(Z>d)}$ is pointwise bounded by 1, while the loss to the layer

$${}^{u}_{d}Z = \min\Big((Z-d)^+, u-d\Big)$$

is pointwise bounded by both u and Z. So we can take the expectations. □

As for the second and the third formula, the upper bounds u^n and $\mathrm{E}\left(Z^n\right)$, $n = 1, 2$, can in practice both be infinite – but usually are not so at the same time.

B.1.1 As-if loss

To assess the as-if loss, let $h > 0$ be such small that for $\Delta B_k \in [1-h, 1+h]$ the pointwise approximation

$$\mathrm{E}\left({}^{u^*}_{d^*}\left(\Delta B_k \mathring{Z} \mid \mathcal{W}\right)\right) \approx \Delta B_k^{\alpha_{(d^*,u^*)}} \mathrm{E}\left({}^{u^*}_{d^*}\mathring{Z} \mid \widehat{IB_q}\right)$$

or equivalently

$$\mathrm{E}\left(S_k \mid \mathcal{W}\right) \approx \tilde{e}\Delta D\big(\alpha_{(d^*,u^*)}\big)_k$$

is very precise. With the abbreviation $\alpha = \alpha_{(d^*,u^*)}$ and the notation from Section 7.2 for combinations of a regular conditional probability measure (here given $\underline{I}$) with a binary condition (here splitting whether ΔB_k is close to / far from 1), we get via the law of total probability

$$\begin{aligned}
\mathrm{E}\left(S_k \mid \underline{I}\right) - \tilde{e}\mathrm{E}\left(\Delta D\big(\alpha_{(d^*,u^*)}\big)_k \mid \underline{I}\right) &= \mathrm{E}_{\underline{I}}\left(S_k\right) - \tilde{e}\mathrm{E}_{\underline{I}}\left(\Delta D(\alpha)_k\right) = \mathrm{E}_{\underline{I}}\left(\mathrm{E}_{\mathcal{W}}\left(S_k\right) - \tilde{e}\Delta D(\alpha)_k\right) \\
&= \mathrm{P}_{\underline{I}}\left(|\Delta B_k - 1| \le h\right)\left[\mathrm{E}_{\underline{I}}\left(\mathrm{E}_{\mathcal{W}}\left(S_k\right) - \tilde{e}\Delta D(\alpha)_k \mid |\Delta B_k - 1| \le h\right)\right] \\
&\quad + \mathrm{P}_{\underline{I}}\left(|\Delta B_k - 1| > h\right)\left[\mathrm{E}_{\underline{I}}\left(\mathrm{E}_{\mathcal{W}}\left(S_k\right) - \tilde{e}\Delta D(\alpha)_k \mid |\Delta B_k - 1| > h\right)\right]
\end{aligned}$$

In the second step we have used that $\tilde{e}$ is $\underline{I}$-measurable and $\mathrm{E}_{\underline{I}}\left(S_k\right) = \mathrm{E}_{\underline{I}}\left(\mathrm{E}_{\mathcal{W}}\left(S_k\right)\right)$ (tower property). Now we easily get an upper bound for the error of the approximation in question:

$$\begin{aligned}
&\left|\mathrm{E}\left(S_k \mid \underline{I}\right) - \tilde{e}\mathrm{E}\left(\Delta D\big(\alpha_{(d^*,u^*)}\big)_k \mid \underline{I}\right)\right| \\
&\qquad \le \mathrm{P}\left(|\Delta B_k - 1| \le h \mid \underline{I}\right)\left|\mathrm{E}_{\underline{I}}\left(\mathrm{E}_{\mathcal{W}}\left(S_k\right) - \tilde{e}\Delta D(\alpha)_k \mid |\Delta B_k - 1| \le h\right)\right| \\
&\qquad + \mathrm{P}\left(|\Delta B_k - 1| > h \mid \underline{I}\right)\left|\mathrm{E}_{\underline{I}}\left(\mathrm{E}_{\mathcal{W}}\left(S_k\right) - \tilde{e}\Delta D(\alpha)_k \mid |\Delta B_k - 1| > h\right)\right| \\
&\qquad \le \mathrm{E}_{\underline{I}}\left(\left|\mathrm{E}\left(S_k \mid \mathcal{W}\right) - \tilde{e}\Delta D(\alpha)_k\right| \mid |\Delta B_k - 1| \le h\right) \\
&\quad + \mathrm{P}\left(|\Delta B_k - 1| > h \mid \underline{I}\right)\left\{\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(S_k \mid \mathcal{W}\right) \mid |\Delta B_k - 1| \le h\right) + \mathrm{E}_{\underline{I}}\left(\tilde{e}\Delta D(\alpha)_k \mid |\Delta B_k - 1| \le h\right)\right\}
\end{aligned}$$

The final expression of this inequality has a structure that we will see repeatedly in the following, so we discuss it heuristically, as a guidance for the mathematics to follow. Its first summand (second to last row) is very small, as this holds *pointwise* for $\left|\mathrm{E}\left(S_k \mid \mathcal{W}\right) - \tilde{e}\Delta D(\alpha)_k\right|$, as long as $|\Delta B_k - 1| \le h$. We could say that this term is as small as the underlying pointwise approximation is precise.

Can the terms in the final row yield a small figure too? This is indeed the case if the probability $\mathrm{P}\left(|\Delta B_k - 1| > h \mid \underline{I}\right)$ is very small and the two terms in braces are bounded in an appropriate way – it would not suffice to find RV's ΔB_k making said probability extremely small if the subsequent terms could become arbitrarily large for such ΔB_k.

The last term equals $\mathrm{E}_{\underline{I}}\left(\Delta D(\alpha)_k \mid |\Delta B_k - 1| > h\right)$ times the positive scalar $\tilde{e} = \theta\mu^*\widehat{IV_q}\widehat{IB_q}$.

The preceding term equals

$$\mathrm{E}_{\underline{I}}\left(\widehat{IV_q}\widehat{IB_q}\Delta V_k\theta\mathrm{E}\left({}_{d^*}^{u^*}\left(\Delta B_k\mathring{Z}\mid\mathcal{W}\right)\right)\mid|\Delta B_k-1|>h\right)$$

From the above lemma we get

$$\mathrm{E}\left({}_{d^*}^{u^*}\left(\Delta B_k\mathring{Z}\right)\mid\mathcal{W}\right)\le\min\left(u^*,\mathrm{E}\left(\Delta B_k\mathring{Z}\mid\mathcal{W}\right)\right)=\min\left(u^*,\Delta B_k\mathrm{E}\left(\mathring{Z}\mid\mathcal{W}\right)\right)=\min\left(u^*,\Delta B_k\mathrm{E}\left(\mathring{Z}\right)\right)$$

where the last step follows from the independence of $\mathring{Z}$ and $\mathcal{W}$ in the inflationary world structure. For the case $u^*=\infty$ we have, throughout this book, supposed that $\mathrm{E}\left(\mathring{Z}\right)$ be finite (otherwise the risk would be uninsurable). Thus, in any case we have a finite upper bound for $\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(S_k\mid\mathcal{W}\right)\mid|\Delta B_k-1|\le h\right)$, namely $\theta\widehat{IV_q}\widehat{IB_q}u^*\mathrm{E}_{\underline{I}}\left(\Delta V_k\mid|\Delta B_k-1|>h\right)$ or $\theta\widehat{IV_q}\widehat{IB_q}\mathrm{E}\left(\mathring{Z}\right)\mathrm{E}_{\underline{I}}\left(\Delta V_k\Delta B_k\mid|\Delta B_k-1|>h\right)$, or both.

These upper bounds can, as well as the term discussed before, be written as a positive constant times an expectation

$$\mathrm{E}_{\underline{I}}\left(\Delta D(a)_k\mid|\Delta B_k-1|>h\right)=\mathrm{E}_{\underline{I}}\left(\Delta V_k\Delta B_k^a\mid|\Delta B_k-1|>h\right)$$

where a equals 0, 1, or $\alpha_{(d^*,u^*)}$. Thus, we are done if we can, for certain ΔB_k yielding a very low $\mathrm{P}\left(|\Delta B_k-1|>h\mid\underline{I}\right)$, give a common upper bound for said expectations $\mathrm{E}_{\underline{I}}\left(\Delta D(a)_k\mid|\Delta B_k-1|>h\right)$.

Before investigating this, we show that the other approximations of interest lead to closely related mathematical problems, such that we will be able to treat all of them together.

B.1.2 Other first moments

For the *as-if loss count* we proceed analogously, selecting an h such that for $\Delta B_k\in[1-h,1+h]$

$$\mathrm{E}\left(C_k\mid\mathcal{W}\right)=\widehat{IV_q}\Delta V_k\theta\mathrm{E}\left(\chi_{\Delta B_k\mathring{Z}>d^*}\mid\mathcal{W}\right)\approx\tilde{\theta}\Delta D(\alpha_{d^*})_k$$

is a very precise approximation. Then we calculate exactly as above

$$\begin{aligned}&\left|\mathrm{E}\left(C_k\mid\underline{I}\right)-\tilde{\theta}\mathrm{E}\left(\Delta D(\alpha_{d^*})_k\mid\underline{I}\right)\right|\\&\qquad\le\mathrm{E}_{\underline{I}}\left(\left|\mathrm{E}\left(C_k\mid\mathcal{W}\right)-\tilde{\theta}\Delta D(\alpha_{d^*})_k\right|\mid|\Delta B_k-1|\le h\right)\\&+\mathrm{P}\left(|\Delta B_k-1|>h\mid\underline{I}\right)\left\{\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(C_k\mid\mathcal{W}\right)\mid|\Delta B_k-1|>h\right)+\mathrm{E}_{\underline{I}}\left(\tilde{\theta}\Delta D(\alpha_{d^*})_k\mid|\Delta B_k-1|>h\right)\right\}\end{aligned}$$

and get again a small term in the second to last row, while below we see the same probability as above. As for the two following terms, the final one is analogous to the corresponding term for the as-if loss, with α_{d^*} taking the place of $\alpha_{(d^*,u^*)}$. For the preceding term we use the first formula of Lemma B.1 and get

$$\mathrm{E}_{\underline{I}}\left(\widehat{IV_q}\Delta V_k\theta\mathrm{E}\left(\chi_{\Delta B_k\mathring{Z}>d^*}\mid\mathcal{W}\right)\mid|\Delta B_k-1|>h\right)\le\theta\widehat{IV_q}\mathrm{E}_{\underline{I}}\left(\Delta V_k\mid|\Delta B_k-1|>h\right)$$

As above we have to find an upper bound for terms $\mathrm{E}_{\underline{I}}\left(\Delta D(a)_k\mid|\Delta B_k-1|>h\right)$, this time for a equaling 0 or α_{d^*}.

The future quantities require working with $\frac{IB_q}{\widehat{IB_q}}\Delta B_q$, such that the formulae are a bit more complex. However, structurally the reasoning is the same as for the as-if quantities. For the *future loss* we select an h such that for $\frac{IB_q}{\widehat{IB_q}}\Delta B_q\in[1-h,1+h]$

$$\mathrm{E}\left({}_{d^*}^{u^*}\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q\mathring{Z}\right)\mid\mathcal{W}\right)\approx\left(\frac{IB_q}{\widehat{IB_q}}\Delta B_q\right)^{\alpha_{(d^*,u^*)}}\mathrm{E}\left({}_{d^*}^{u^*}\mathring{Z}\mid\widehat{IB_q}\right)$$

or equivalently with the abbreviation $\alpha = \alpha_{(d^*,u^*)}$

$$\begin{aligned}\mathrm{E}\left(X_q \mid \mathcal{W}\right) &= \widehat{IB}_q IV_q \mathit{\Delta} V_q \theta \mathrm{E}\left({}^{u^*}_{d^*}\left(\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q \mathring{Z}\right) \mid \mathcal{W}\right)\\ &\approx \widehat{IB}_q\widehat{IV}_q\left(\frac{IV_q}{\widehat{IV}_q}\mathit{\Delta} V_q\right)\left(\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q\right)^{\alpha}\theta\mathrm{E}\left({}^{u^*}_{d^*}\mathring{Z} \mid \widehat{IB}_q\right) = \tilde{e}\mathit{\Delta} D(\alpha)_q\frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q}\end{aligned}$$

is a very precise approximation. Then we get, by splitting the expectation into the cases of $\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q$ being close to / far from 1:

$$\begin{aligned}&\left|\mathrm{E}\left(X_q \mid \underline{I}\right) - \tilde{e}\mathrm{E}\left(\mathit{\Delta} D\big(\alpha_{(d^*,u^*)}\big)_q\frac{ID\big(\alpha_{(d^*,u^*)}\big)_q}{\widehat{ID\big(\alpha_{(d^*,u^*)}\big)}_q} \mid \underline{I}\right)\right| = \left|\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(X_q \mid \mathcal{W}\right) - \tilde{e}\mathit{\Delta} D(\alpha)_q\frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q}\right)\right|\\ &\quad\le \mathrm{E}_{\underline{I}}\left(\left|\mathrm{E}\left(X_q \mid \mathcal{W}\right) - \tilde{e}\mathit{\Delta} D(\alpha)_q\frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q}\right| \mid |\mathit{\Delta} B_k - 1| \le h\right) + \mathrm{P}\left(\left|\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q - 1\right| > h \mid \underline{I}\right)\\ &\qquad\left\{\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(X_q \mid \mathcal{W}\right) \mid \left|\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q - 1\right| > h\right) + \mathrm{E}_{\underline{I}}\left(\tilde{e}\mathit{\Delta} D(\alpha)_q\frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q} \mid \left|\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q - 1\right| > h\right)\right\}\end{aligned}$$

Again the last expression starts with a small first summand. The remainder is small if the probability $\mathrm{P}\left(\left|\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q - 1\right| > h \mid \underline{I}\right)$ is very small and if we can give low enough an upper bound for the final two terms in braces. The latter of these, up to a positive constant, equals

$$\mathrm{E}_{\underline{I}}\left(\mathit{\Delta} D(a)_q\frac{ID(a)_q}{\widehat{ID(a)}_q} \mid \left|\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q - 1\right| > h\right)$$

where $a = \alpha_{(d^*,u^*)}$. The former term has, analogously to the as-if loss, an upper bound of the same kind, where $a = 0$ or $a = 1$.

For the *future loss count* we proceed in the same way, selecting an h such that for $\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q \in [1-h, 1+h]$

$$\begin{aligned}\mathrm{E}\left(\bar{N}_q \mid \mathcal{W}\right) &= IV_q\mathit{\Delta} V_q\theta\mathrm{E}\left(\chi_{\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q\mathring{Z}>d^*} \mid \mathcal{W}\right)\\ &\approx \widehat{IV}_q\left(\frac{IV_q}{\widehat{IV}_q}\mathit{\Delta} V_q\right)\left(\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q\right)^{\alpha_{d^*}}\theta\mathrm{E}\left(\chi_{\mathring{Z}>d^*} \mid \widehat{IB}_q\right) = \tilde{\theta}\mathit{\Delta} D(\alpha_{d^*})_q\frac{ID(\alpha_{d^*})_q}{\widehat{ID(\alpha_{d^*})}_q}\end{aligned}$$

is a very precise approximation. Then we calculate for the corresponding expectations exactly as above

$$\begin{aligned}&\left|\mathrm{E}\left(\bar{N}_q \mid \underline{I}\right) - \tilde{\theta}\mathrm{E}\left(\mathit{\Delta} D(\alpha_{d^*})_q\frac{ID(\alpha_{d^*})_q}{\widehat{ID(\alpha_{d^*})}_q} \mid \underline{I}\right)\right|\\ &\quad\le \mathrm{E}_{\underline{I}}\left(\left|\mathrm{E}\left(\bar{N}_q \mid \mathcal{W}\right) - \tilde{\theta}\mathit{\Delta} D(\alpha_{d^*})_q\frac{ID(\alpha_{d^*})_q}{\widehat{ID(\alpha_{d^*})}_q}\right| \mid |\mathit{\Delta} B_k - 1| \le h\right) + \mathrm{P}\left(\left|\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q - 1\right| > h \mid \underline{I}\right)\\ &\qquad\left\{\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(\bar{N}_q \mid \mathcal{W}\right) \mid \left|\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q - 1\right| > h\right) + \mathrm{E}_{\underline{I}}\left(\tilde{\theta}\mathit{\Delta} D(\alpha_{d^*})_q\frac{ID(\alpha_{d^*})_q}{\widehat{ID(\alpha_{d^*})}_q} \mid \left|\frac{IB_q}{\widehat{IB}_q}\mathit{\Delta} B_q - 1\right| > h\right)\right\}\end{aligned}$$

and get again a small first term in the final expression, being followed by the same probability as before. Analogously to the as-if loss count, the final terms in braces have as upper bound, or equal, respectively, terms $\mathrm{E}_{\underline{I}}\left(\Delta D(a)_q \, \frac{ID(a)_q}{\widehat{ID(a)}_q} \mid \left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| > h\right)$ times a positive constant, where a equals 0 or α_{d^*}.

B.1.3 Second moments

Albeit more complex, the calculation for the *squared quantities* (which yield the variances) is parallel to that for the first moments, using the same conditioning and case differentiation. Let us illustrate this in detail for the as-if loss. We study three approximation variants, starting with the one mentioned in Remark 7.13:

$$\begin{aligned}\mathrm{E}\left(S_k^2 \mid \underline{I}\right) &= \mathrm{E}\left(\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right) \mid \underline{I}\right)\\ &\overset{!}{\approx} \tilde{e}^2 \mathrm{E}\left(\left(1+\beta\right)\Delta D\big(\alpha_{(d^*,u^*)}\big)_k^2 + \frac{1}{IV_k}\max\left[\sum_{i=1}^4 a_i \Delta D(\alpha_i)_k \,, \sum_{i=1}^j a_i^{(j)} \Delta D\Big(\alpha_i^{(j)}\Big)_k\right] \mid \underline{I}\right)\end{aligned}$$

Let now $h > 0$ be such small that for $\Delta B_k \in [1-h, 1+h]$ the pointwise approximation

$$\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right) \approx \tilde{e}^2 \left\{\left(1+\beta\right)\Delta D\big(\alpha_{(d^*,u^*)}\big)_k^2 + \frac{1}{IV_k}\max\left[\sum_{i=1}^4 a_i \Delta D(\alpha_i)_k \,, \sum_{i=1}^j a_i^{(j)} \Delta D\Big(\alpha_i^{(j)}\Big)_k\right]\right\}$$

is very precise. With the abbreviation $\alpha = \alpha_{(d^*,u^*)}$ we get, by splitting the expectation into the cases of ΔB_k being close to / far from 1:

$$\begin{aligned}&\left|\mathrm{E}\left(S_k^2 \mid \underline{I}\right) - \tilde{e}^2 \mathrm{E}\left(\left(1+\beta\right)\Delta D(\alpha)_k^2 + \frac{1}{IV_k}\max\left[\sum_{i=1}^4 a_i \Delta D(\alpha_i)_k \,, \sum_{i=1}^j a_i^{(j)} \Delta D\Big(\alpha_i^{(j)}\Big)_k\right] \mid \underline{I}\right)\right|\\ &= \left|\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right) - \tilde{e}^2\left\{\left(1+\beta\right)\Delta D(\alpha)_k^2 + \frac{1}{IV_k}\max\left[\sum_{i=1}^4 a_i \Delta D(\alpha_i)_k \,, \sum_{i=1}^j a_i^{(j)} \Delta D\Big(\alpha_i^{(j)}\Big)_k\right]\right\}\right)\right|\\ &\le \mathrm{E}_{\underline{I}}\left(\left|\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right) - \tilde{e}^2\left\{(1+\beta)\Delta D(\alpha)_k^2 + \frac{1}{IV_k}\max\left[\sum_{i=1}^4 a_i \Delta D(\alpha_i)_k \,, \sum_{i=1}^j a_i^{(j)} \Delta D\Big(\alpha_i^{(j)}\Big)_k\right]\right\}\right| \mid |\Delta B_k - 1| \le h\right)\\ &\qquad + \mathrm{P}\left(|\Delta B_k - 1| > h \mid \underline{I}\right)\left\{\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right) \mid |\Delta B_k - 1| > h\right)\right.\\ &\left. + \mathrm{E}_{\underline{I}}\left(\tilde{e}^2\left(\left(1+\beta\right)\Delta D(\alpha)_k^2 + \frac{1}{IV_k}\left[\sum_{i=1}^4 |a_i|\, \Delta D(\alpha_i)_k + \sum_{i=1}^j \left|a_i^{(j)}\right| \Delta D\Big(\alpha_i^{(j)}\Big)_k\right]\right) \mid |\Delta B_k - 1| > h\right)\right\}\end{aligned}$$

Again the final expression starts with a small term, being followed by the product of a (already seen) probability and a sum of two terms in braces. These latter are more complex than in the first-moment case, but mathematically closely related:

The last term adds up a number of expectations about terms $\Delta D(a)_k$ of the kind appearing in the first-moment case, however, more values for a come in, namely the α_i and the $\alpha_i^{(j)}$. Its first summand is the analogous conditional expectation of $\Delta D(\alpha)_k^2$ times a positive constant – this is essentially the only novel ingredient.

As for the preceding term, it can be seen easily from the calculations at the beginning of Section 7.2 that

$$\begin{aligned}&\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right)\\&=\widehat{IV_q}^2\widehat{IB_q}^2\theta\left\{(1+\beta)\,\theta\left[\Delta V_k^2+\frac{\gamma}{IV_k}\Delta V_k\right]E^2\left({}_{d^*}^{u^*}\left(\Delta B_k\mathring{Z}\right)\mid\mathcal{W}\right)+\frac{1}{IV_k}\Delta V_k\mathrm{E}\left(\left({}_{d^*}^{u^*}\left(\Delta B_k\mathring{Z}\right)\right)^2\mid\mathcal{W}\right)\right\}\end{aligned}$$

The terms $E^2\left({}_{d^*}^{u^*}\left(\Delta B_k\mathring{Z}\right)\mid\mathcal{W}\right)$ and $\mathrm{E}\left(\left({}_{d^*}^{u^*}\left(\Delta B_k\mathring{Z}\right)\right)^2\mid\mathcal{W}\right)$ have upper bounds according to the second and third formula in Lemma B.1, namely $\max\left(u^{*2},\Delta B_k^2\,\mathrm{E}^2\left(\mathring{Z}\right)\right)$ and $\max\left(u^{*2},\Delta B_k^2\,\mathrm{E}\left(\mathring{Z}^2\right)\right)$. Even for unlimited layers these upper bounds are finite, as long as $\mathring{Z}$ has finite *first* and *second* moment. Resorting terms we see that $\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right)$ has an upper bound being a linear combination of terms $\Delta D(a)_k$ and $\Delta D(a)_k^2$, where a equals 0, 1, or 2.

Thus, we can again assess the final two terms in braces by the same kind of expectations, the only difference being that we have to look at $\Delta D(a)_k$ for many more values of a, most of which stem from the alphas in the variance formulae in Proposition 7.4; further we need to look at $\Delta D(a)_k^2$ for some a. For completeness we list all values of interest for a:
0, 1, 2, $\alpha_{(d^*,u^*)}$, $2\alpha_{(d^*,u^*)}$, $\mathring{\alpha}_{(d^*,u^*)}$, $2\alpha_{(d^*,u^*)}-\alpha_{d^*}$.

In exactly the same way we can assess an alternative approximation:

$$\mathrm{E}\left(S_k^2 \mid \underline{I}\right)\overset{!}{\approx}\tilde{e}^2\mathrm{E}\left((1+\beta)\,\Delta D\big(\alpha_{(d^*,u^*)}\big)_k^2+\frac{1}{IV_k}\sum_{i=1}^{4}a_i\Delta D(\alpha_i)_k\mid\underline{I}\right)$$

To this end, let now $h>0$ be such small that for $\Delta B_k\in[1-h,1+h]$ *both* pointwise approximations

$$\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right)\approx\tilde{e}^2\left\{(1+\beta)\,\Delta D\big(\alpha_{(d^*,u^*)}\big)_k^2+\frac{1}{IV_k}\max\left[\sum_{i=1}^{4}a_i\Delta D(\alpha_i)_k\,,\sum_{i=1}^{j}a_i^{(j)}\Delta D\Big(\alpha_i^{(j)}\Big)_k\right]\right\}$$

$$\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right)\approx\tilde{e}^2\left\{(1+\beta)\,\Delta D\big(\alpha_{(d^*,u^*)}\big)_k^2+\frac{1}{IV_k}\sum_{i=1}^{4}a_i\Delta D(\alpha_i)_k\right\}$$

are very precise. This is possible as these approximations must coincide for all ΔB_k in a very small neighborhood of 1: Recall that ultimately the first sum is part of a first order approximation in ΔB_k, such that very close to 1 the second sum, which yields a (possibly far) lower bound, will be smaller and thus not impact the maximum of the sums.

If we now calculate the basic inequality, deviating from above only by setting the coefficients $a_i^{(j)}$ of the second sum to 0, we get the same final term, save that the summands with the $a_i^{(j)}$ are missing:

$$\begin{aligned}&\left|\mathrm{E}\left(S_k^2 \mid \underline{I}\right)-\tilde{e}^2\mathrm{E}\left((1+\beta)\,\Delta D(\alpha)_k^2+\frac{1}{IV_k}\sum_{i=1}^{4}a_i\Delta D(\alpha_i)_k\mid\underline{I}\right)\right|\\&\quad\le\mathrm{E}_{\underline{I}}\left(\left|\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right)-\tilde{e}^2\left\{(1+\beta)\,\Delta D(\alpha)_k^2+\frac{1}{IV_k}\sum_{i=1}^{4}a_i\Delta D(\alpha_i)_k\right\}\right|\mid|\Delta B_k-1|\le h\right)\\&\qquad+\mathrm{P}\left(|\Delta B_k-1|>h\mid\underline{I}\right)\left\{\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(S_k^2 \mid \mathcal{W}\right)\mid|\Delta B_k-1|>h\right)\right.\\&\qquad\left.+\mathrm{E}_{\underline{I}}\left(\tilde{e}^2\left((1+\beta)\,\Delta D(\alpha)_k^2+\frac{1}{IV_k}\sum_{i=1}^{4}|a_i|\,\Delta D(\alpha_i)_k\right)\mid|\Delta B_k-1|>h\right)\right\}\end{aligned}$$

Thus, the final term here is smaller. This implies that, with both pointwise approximations holding, if we have circumstances making the first approximation in expectation precise, so is the second one.

Now note an obvious inequality for the sums appearing in the two basic inequalities derived here:

$$\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{4} a_i \mathit{\Delta} D(\alpha_i)_k \mid \underline{I}\right) &\leq \max\left[\sum_{i=1}^{4} a_i \mathrm{E}\left(\mathit{\Delta} D(\alpha_i)_k \mid \underline{I}\right), \sum_{i=1}^{j} a_i^{(j)} \mathrm{E}\left(\mathit{\Delta} D\left(\alpha_i^{(j)}\right)_k\right) \mid \underline{I}\right] \\
&\leq \mathrm{E}\left(\max\left[\sum_{i=1}^{4} a_i \mathit{\Delta} D(\alpha_i)_k, \sum_{i=1}^{j} a_i^{(j)} \mathit{\Delta} D\left(\alpha_i^{(j)}\right)_k\right] \mid \underline{I}\right)
\end{aligned}$$

If first and last term lead to fair approximations, so does the term in the middle. It finally yields the preferred approximation variant, as it appears in Proposition 7.12:

$$\mathrm{E}\left(S_k^2 \mid \underline{I}\right) \overset{!}{\approx} \tilde{e}^2\left((1+\beta)\,\mathrm{E}\Big(\mathit{\Delta} D\big(\alpha_{(d^*,u^*)}\big)_k^2 \mid \underline{I}\Big) + \frac{1}{IV_k} \max\left[\sum_{i=1}^{4} a_i \mathrm{E}(\mathit{\Delta} D(\alpha_i)_k \mid \underline{I}), \sum_{i=1}^{j} a_i^{(j)} \mathrm{E}\Big(\mathit{\Delta} D\Big(\alpha_i^{(j)}\Big)_k \mid \underline{I}\Big)\right]\right)$$

Remark B.2. How can there be three useful approximations at the same time? Generally, if one of them is very good, the other ones must be either very close or not precise. However, we have shown here that each variant has the chance to be very precise an approximation and that their precisions depend on *the same* mathematical properties. Thus, given these properties, all three approximations are precise and we can choose among them, without having to worry which one is the best.

Having thoroughly studied the as-if loss, the three remaining squared quantities pose no additional challenges:

- For the squared as-if loss count the calculation is essentially the same, albeit a bit simpler, as instead of the intricate maximum there is only one sum involved, having two summands. The values of interest for a are less and slightly different; ultimately one has to look at $\mathit{\Delta} D(\alpha_{d^*})_k^2$ and at $\mathit{\Delta} D(a)_k$ for a taking the values 0, α_{d^*}, $2\alpha_{d^*}$.
- For the squared future loss the calculation works like that for the as-if loss, with the same values for a. However, like in the first-moment case, the probability of interest is $\mathrm{P}\left(\left|\frac{IB_q}{\widehat{IB_q}} \mathit{\Delta} B_q - 1\right| > h \mid \underline{I}\right)$, while the conditional expectations to be assessed are the $\mathrm{E}_{\underline{I}}\left(\mathit{\Delta} D(a)_q \frac{ID(a)_q}{\widehat{ID(a)}_q} \mid \left|\frac{IB_q}{\widehat{IB_q}} \mathit{\Delta} B_q - 1\right| > h\right)$ and additionally some $\mathrm{E}_{\underline{I}}\left(\mathit{\Delta} D(a)_q^2 \frac{ID(a)_q^2}{\widehat{ID(a)}_q^2} \mid \left|\frac{IB_q}{\widehat{IB_q}} \mathit{\Delta} B_q - 1\right| > h\right)$.
- Finally, the calculation for the squared future loss count is again a simpler copy of the preceding calculation, with the same values a as for the as-if loss count.

For the *mixed products* (which yield the covariances) the calculation is very similar to that for the first moments, apart from a more complex case differentiation. Let us first look at the interaction of two different past years $l < k \leq 0$. We want to study the approximation

$$\mathrm{E}\left(S_k S_l \mid \underline{I}\right) \overset{!}{\approx} \tilde{e}^2 \mathrm{E}\left(\mathit{\Delta} D\big(\alpha_{(d^*,u^*)}\big)_k \mathit{\Delta} D\big(\alpha_{(d^*,u^*)}\big)_l \mid \underline{I}\right)$$

and use again the abbreviation $\alpha = \alpha_{(d^*,u^*)}$. Let now $h > 0$ be such small that for $\mathit{\Delta} B_k, \mathit{\Delta} B_l \in [1-h, 1+h]$ both approximations

$$\mathrm{E}\left(S_k \mid \mathcal{W}\right) \approx \tilde{e} \mathit{\Delta} D(\alpha)_k, \qquad \mathrm{E}\left(S_l \mid \mathcal{W}\right) \approx \tilde{e} \mathit{\Delta} D(\alpha)_l$$

are very precise. Define $h' := \sqrt{1+h} - 1 < h$. We have

$$1 - h \leq (1-h')^2 < 1 < (1+h')^2 \leq 1 + h$$

Suppose that $\mathit{\Delta} B_k$ and $\frac{\mathit{\Delta} B_l}{\mathit{\Delta} B_k}$ lie in the smaller interval $[1-h', 1+h']$. Then both $\mathit{\Delta} B_k$ and $\mathit{\Delta} B_l = \mathit{\Delta} B_k \frac{\mathit{\Delta} B_l}{\mathit{\Delta} B_k}$ lie in the interval $[1-h, 1+h]$, such that the above approximations hold. So we get, by splitting cases for both $\mathit{\Delta} B_k$ and $\frac{\mathit{\Delta} B_l}{\mathit{\Delta} B_k}$,

$$\begin{aligned}
&\left|\mathrm{E}\left(S_k S_l \mid \underline{I}\right) - \tilde{e}^2 \mathrm{E}\left(\mathit{\Delta} D\big(\alpha_{(d^*,u^*)}\big)_k \, \mathit{\Delta} D\big(\alpha_{(d^*,u^*)}\big)_l \mid \underline{I}\right)\right| = \left|\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(S_k S_l \mid \mathcal{W}\right) - \tilde{e}^2 \mathit{\Delta} D(\alpha)_k \, \mathit{\Delta} D(\alpha)_l\right)\right| \\
&\qquad \leq \mathrm{E}_{\underline{I}}\left(\left|\mathrm{E}\left(S_k S_l \mid \mathcal{W}\right) - \tilde{e}\mathit{\Delta} D(\alpha)_k \, \tilde{e}\mathit{\Delta} D(\alpha)_l\right| \mid |\mathit{\Delta} B_k - 1| \leq h', \left|\frac{\mathit{\Delta} B_l}{\mathit{\Delta} B_k} - 1\right| \leq h'\right) \\
&+ \mathrm{P}\left(|\mathit{\Delta} B_k - 1| > h' \vee \left|\frac{\mathit{\Delta} B_l}{\mathit{\Delta} B_k} - 1\right| > h' \mid \underline{I}\right) \Bigg\{\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(S_k S_l \mid \mathcal{W}\right) \mid |\mathit{\Delta} B_k - 1| > h' \vee \left|\frac{\mathit{\Delta} B_l}{\mathit{\Delta} B_k} - 1\right| > h'\right) \\
&\qquad\qquad + \mathrm{E}_{\underline{I}}\left(\tilde{e}^2 \mathit{\Delta} D(\alpha)_k \, \mathit{\Delta} D(\alpha)_l \mid |\mathit{\Delta} B_k - 1| > h' \vee \left|\frac{\mathit{\Delta} B_l}{\mathit{\Delta} B_k} - 1\right| > h'\right)\Bigg\}
\end{aligned}$$

The final expression has a small first summand: In the inflationary world structure S_k and S_l, which are calculated from losses of different years, are uncorrelated conditionally on $\mathcal{W}$. So we can decompose

$$\mathrm{E}\left(S_k S_l \mid \mathcal{W}\right) = \mathrm{E}\left(S_k \mid \mathcal{W}\right) \mathrm{E}\left(S_l \mid \mathcal{W}\right)$$

and apply the precise pointwise approximations. The next term is the probability, which embraces two events and is small, as long as the probability of each of these events is small:

$$\mathrm{P}\left(|\mathit{\Delta} B_k - 1| > h' \vee \left|\frac{\mathit{\Delta} B_l}{\mathit{\Delta} B_k} - 1\right| > h' \mid \underline{I}\right) \leq \mathrm{P}\left(|\mathit{\Delta} B_k - 1| > h' \mid \underline{I}\right) + \mathrm{P}\left(\left|\frac{\mathit{\Delta} B_l}{\mathit{\Delta} B_k} - 1\right| > h' \mid \underline{I}\right)$$

It remains to assess the two summands in braces. Analogously to the first moments, the first summand, which averages over the product of first moments $\mathrm{E}\left(S_k \mid \mathcal{W}\right) \mathrm{E}\left(S_l \mid \mathcal{W}\right)$, has an upper bound looking like the second one, with $\mathit{\Delta} D(a)$ taking the place of $\mathit{\Delta} D(\alpha)$ and a equaling 0 or 1. Thus, ultimately we have to analyze only one kind of term, namely

$$\mathrm{E}_{\underline{I}}\left(\mathit{\Delta} D(a)_k \, \mathit{\Delta} D(a)_l \mid |\mathit{\Delta} B_k - 1| > h' \vee \left|\frac{\mathit{\Delta} B_l}{\mathit{\Delta} B_k} - 1\right| > h'\right)$$

where a equals 0, 1, or $\alpha_{(d^*,u^*)}$.

For the corresponding loss count the calculation is the same, coming up with the same probability to assess and the analogous final term, now with a equaling 0 or α_{d^*}.

For the past-future interaction we can use the same approach, deviating in a few details. Let now $h > 0$ be such small that for $\mathit{\Delta} B_k, \frac{IB_q}{\widehat{IB_q}} \mathit{\Delta} B_q \in [1-h, 1+h]$ both approximations

$$\mathrm{E}\left(S_k \mid \mathcal{W}\right) \approx \tilde{e}\mathit{\Delta} D(\alpha)_k\,, \qquad \mathrm{E}\left(X_q \mid \mathcal{W}\right) \approx \tilde{e}\mathit{\Delta} D(\alpha)_q \, \frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q}$$

with the abbreviation $\alpha = \alpha_{(d^*,u^*)}$ are very precise. Then we get, by splitting cases for both $\mathit{\Delta} B_k$ and $\frac{IB_q}{\widehat{IB_q}} \mathit{\Delta} B_q$,

$$\begin{aligned}
&\left|\mathrm{E}\left(X_q S_k \mid \underline{I}\right) - \tilde{e}^2 \mathrm{E}\left(\Delta D(\alpha)_q \frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q} \Delta D(\alpha)_k \mid \underline{I}\right)\right| \\
&\qquad = \left|\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(X_q S_k \mid \mathcal{W}\right) - \tilde{e}^2 \Delta D(\alpha)_q \frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q} \Delta D(\alpha)_k\right)\right| \\
&\qquad \leq \mathrm{E}_{\underline{I}}\left(\left|\mathrm{E}\left(X_q S_k \mid \mathcal{W}\right) - \tilde{e} \Delta D(\alpha)_q \frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q} \tilde{e} \Delta D(\alpha)_k\right| \mid |\Delta B_k - 1| \leq h, \left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| \leq h\right) \\
&+ \mathrm{P}\left(|\Delta B_k - 1| > h \vee \left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| > h \mid \underline{I}\right) \left\{\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(X_q S_k \mid \mathcal{W}\right) \mid |\Delta B_k - 1| > h \vee \left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| > h\right)\right. \\
&\qquad \left. + \mathrm{E}_{\underline{I}}\left(\tilde{e}^2 \Delta D(\alpha)_q \frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q} \Delta D(\alpha)_k \mid |\Delta B_k - 1| > h \vee \left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| > h\right)\right\}
\end{aligned}$$

The final expression has a small first summand: S_k and X_q are, conditionally given $\mathcal{W}$, uncorrelated, such that we can decompose

$$\mathrm{E}\left(X_q S_k \mid \mathcal{W}\right) = \mathrm{E}\left(X_q \mid \mathcal{W}\right) \mathrm{E}\left(S_k \mid \mathcal{W}\right)$$

and apply the precise pointwise approximations. The next term is the probability, which embraces two events and is small, as long as the probability of each of these events is small:

$$\mathrm{P}\left(|\Delta B_k - 1| > h \vee \left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| > h \mid \underline{I}\right) \leq \mathrm{P}\left(|\Delta B_k - 1| > h \mid \underline{I}\right) + \mathrm{P}\left(\left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| > h \mid \underline{I}\right)$$

It remains to assess the two summands in braces. Analogously to the first moments, the first summand, which averages over the product of first moments $\mathrm{E}\left(X_q \mid \mathcal{W}\right) \mathrm{E}\left(S_k \mid \mathcal{W}\right)$, has an upper bound looking like the second one, with a taking the place of α and equaling 0 or 1. Thus, ultimately we have to analyze only one kind of term, namely

$$\mathrm{E}_{\underline{I}}\left(\Delta D(a)_q \frac{ID(a)_q}{\widehat{ID(a)}_q} \Delta D(a)_k \mid |\Delta B_k - 1| > h' \vee \left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| > h'\right)$$

where a equals 0, 1, or $\alpha_{(d^*,u^*)}$.

For the corresponding loss count the calculation is the same, coming up with the same probability to assess and the analogous final term, now with a equaling 0 or α_{d^*}.

B.2 Comprehensive analysis

Collecting the pieces and summing up, the desired approximations in expectation (conditionally on $\underline{I}$) require that for $l < k \leq 0$ the three probabilities

$$\mathrm{P}\left(|\Delta B_k - 1| > h \mid \underline{I}\right), \qquad \mathrm{P}\left(\left|\frac{\Delta B_l}{\Delta B_k} - 1\right| > h \mid \underline{I}\right), \qquad \mathrm{P}\left(\left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| > h \mid \underline{I}\right)$$

be very small and at the same time the four conditional expectations

$$\mathrm{E}_{\underline{I}}\left(\varDelta D(a)_k^n \mid |\varDelta B_k - 1| > h\right), \qquad \mathrm{E}_{\underline{I}}\left(\varDelta D(a)_k\, \varDelta D(a)_l \mid |\varDelta B_k - 1| > h \vee \left|\frac{\varDelta B_l}{\varDelta B_k} - 1\right| > h\right),$$

$$\mathrm{E}_{\underline{I}}\left(\left(\varDelta D(a)_q \frac{ID(a)_q}{\widehat{ID(a)_q}}\right)^n \mid \left|\frac{IB_q}{\widehat{IB_q}}\varDelta B_q - 1\right| > h\right),$$

$$\mathrm{E}_{\underline{I}}\left(\varDelta D(a)_q \frac{ID(a)_q}{\widehat{ID(a)_q}}\varDelta D(\alpha)_k \mid |\varDelta B_k - 1| > h \vee \left|\frac{IB_q}{\widehat{IB_q}}\varDelta B_q - 1\right| > h\right)$$

be bounded. This must hold for some appropriate, rather small $h > 0$; $n = 1, 2$; and a equaling various values in the interval $[0, a_{max}]$, where $a_{max} = \max\left(2, 2\alpha_{d^*}, 2\alpha_{(d^*,u^*)}, \mathring{\alpha}_{(d^*,u^*)}\right)$.

We emphasize that we do not need to specify precise upper bounds for the four expectations; it is sufficient to show that they cannot become arbitrarily large in case the three probabilities are very small.

So far in this appendix we have formulated the theory for any $\underline{I} \supseteq \sigma\left(\underline{IV}_{(j\le 0)}, \underline{IB}_{(j\le 0)}\right)$. Now we choose

$$\underline{I} = \sigma\left(\underline{M}_{(j\le 0)}, \varUpsilon_{pr}\right) = \sigma\left(\underline{I\bar{M}}_{(j\le 0)}, \varUpsilon_{pr}\right)$$

in order to apply the comprehensive index system, together with certain (both convenient and realistic) additional properties, as we have already used them especially in Chapter 9.

B.2.1 Probabilities

We start with the third probability and again work with the smaller $h' := \sqrt{1+h} - 1 > 0$. One sees quickly that if $\left|\frac{IB_q}{\widehat{IB_q}}\varDelta B_q - 1\right| > h$, we must have $\left|\frac{IB_q}{\widehat{IB_q}} - 1\right| > h'$ or $|\varDelta B_q - 1| > h'$ (or both). Hence,

$$\mathrm{P}\left(\left|\frac{IB_q}{\widehat{IB_q}}\varDelta B_q - 1\right| > h \mid \underline{I}\right) \le \mathrm{P}\left(\left|\frac{IB_q}{\widehat{IB_q}} - 1\right| > h' \mid \underline{I}\right) + \mathrm{P}\left(|\varDelta B_q - 1| > h' \mid \underline{I}\right)$$

which means that overall it is sufficient to know whether and when, for appropriately small $h > 0$, the four probabilities

$$\mathrm{P}\left(|\varDelta B_k - 1| > h \mid \underline{I}\right), \qquad \mathrm{P}\left(\left|\frac{\varDelta B_l}{\varDelta B_k} - 1\right| > h \mid \underline{I}\right), \qquad \mathrm{P}\left(|\varDelta B_q - 1| > h \mid \underline{I}\right), \qquad \mathrm{P}\left(\left|\frac{IB_q}{\widehat{IB_q}} - 1\right| > h \mid \underline{I}\right)$$

are small. The last one is small if, conditionally on $\underline{I}$, $\frac{IB_q}{\widehat{IB_q}}$ is close to 1 in the sense of *convergence in probability.* This is a fair assumption, meaning essentially that (given the index history) IB_q fluctuates few and is very precisely predicted by $\widehat{IB_q}$. However, unlike in Section 7.5, we do not require any more the property "close to 1" to hold almost surely, but admit larger deviations, albeit with low probability.

For the remaining probabilities we could reason the same way, however, these are about gaps. Recall that gaps describe index basis risk, thus reflect the quality of the selection of the picked indices, which should have nothing to do with the specific index history. In other words, instead of looking at $\mathrm{P}(|\varDelta B_k - 1| > h \mid \underline{I})$, it is much more natural to assume that $\mathrm{P}(|\varDelta B_k - 1| > h)$ be very small, which means that $\varDelta B_k$ be *unconditionally* very close to 1 in the sense of convergence in probability.

Looking at the structure of the inflation gap as a time series, it is straightforward to assume that the error term generating the time series $\underline{\varDelta B}$ is close to 1 in the sense of convergence in probability, which means that the picked inflation index reflects the year-to-year inflation very likely very precisely. If so, then all three gaps $\varDelta B_k = \frac{\underline{\varDelta B}_k}{\underline{\varDelta B}_0}$, $\frac{\varDelta B_l}{\varDelta B_k} = \frac{\underline{\varDelta B}_l}{\underline{\varDelta B}_k}$, and $\varDelta B_q = \frac{\underline{\varDelta B}_q}{\underline{\varDelta B}_0}$ fluctuate closely about 1 in the same sense, such that we have very small probabilities

$$\mathrm{P}\left(|\varDelta B_k - 1| > h\right), \qquad \mathrm{P}\left(\left|\frac{\varDelta B_l}{\varDelta B_k} - 1\right| > h\right), \qquad \mathrm{P}\left(|\varDelta B_q - 1| > h\right)$$

How do we come from these to the conditional probabilities? For the third one this is very easy, as long as the comprehensive index system is *strongly* independent: Then $\varDelta B_q$ is independent of all past indices, such that $\mathrm{P}\left(|\varDelta B_q - 1| > h \mid \underline{I}\right) = \mathrm{P}\left(|\varDelta B_q - 1| > h\right)$.

For the first probability we claim

$$\mathrm{P}\left(|\varDelta B_k - 1| > h \mid \underline{I}\right) \approx \mathrm{P}\left(|\varDelta B_k - 1| > h\right)$$

reasoning for $\varDelta B_k = \varDelta A_k^{-\chi\delta} \varDelta G_k^{-\chi} \varDelta H_k \varDelta \bar{B}_k^{\tau-\chi\zeta}$ exactly as done in detail with the structurally analogous $\varDelta D(\alpha)_k = \varDelta A_k^{1-\delta\omega} \varDelta G_k^{-\omega} \varDelta H_k^{\alpha} \varDelta \bar{B}_k^{\tau\alpha-\zeta\omega}$ in Section 9.1: Without further assumptions we get a crude approximation by averaging over $\underline{I}$:

$$\mathrm{P}\left(|\varDelta B_k - 1| > h \mid \underline{I}\right) = \mathrm{E}\left(\chi_{(|\varDelta B_k - 1| > h)} \mid \underline{I}\right) \approx \mathrm{E}\left(\chi_{(|\varDelta B_k - 1| > h)}\right) = \mathrm{P}\left(|\varDelta B_k - 1| > h\right)$$

More on the sophisticated side, let us once more take advantage of the specific structure of the condition $\underline{I} = \sigma\left(\underline{I\bar{M}}_{(j\le 0)}, \varUpsilon_{pr}\right)$. If the comprehensive index system is independent, we can drop the condition on the primordial picked indices $\varUpsilon_{pr}$, which are independent of all gaps. If further the true deflated official volume fluctuates considerably more (as a time series) than the primordial gaps, then the remaining conditions $\underline{I\bar{M}}_j = \underline{\bar{M}}_j \underline{\varDelta A}_j^{\delta} \underline{\varDelta G}_j \underline{\varDelta \bar{B}}_j^{\zeta}$ are dominated by the $\underline{\bar{M}}_j$, which are independent of the gaps too. Thus, heuristically, conditioning on $\underline{I\bar{M}}_{(j\le 0)}$ has only low impact on the above probability, which justifies the approximation by the unconditional probability. For the multivariate lognormal case this heuristics can be made rigorous along the lines of Example 9.4.

The remaining approximation $\mathrm{P}\left(\left|\frac{\varDelta B_l}{\varDelta B_k} - 1\right| > h \mid \underline{I}\right) \approx \mathrm{P}\left(\left|\frac{\varDelta B_l}{\varDelta B_k} - 1\right| > h\right)$ follows analogously. Collecting results found and assumptions used, the required four probabilities are small if

- the comprehensive index system is strongly independent,
- all normalized primordial gaps (some appear in $\varDelta B$, some in $\underline{I\bar{M}}$) fluctuate very closely about 1, meaning here: are very close to 1 in the sense of convergence in probability,
- $\underline{\bar{M}}$ is much more volatile a time series than the primordial gaps (or alternatively we are in the trivial volume case),
- given the index history $\underline{I} = \sigma\left(\underline{M}_{(j\le 0)}, \varUpsilon_{pr}\right)$, IB_q fluctuates few and is quite precisely predicted by $\widehat{IB}_q$, such that $\frac{IB_q}{\widehat{IB}_q}$ is conditionally close to 1 in the sense of convergence in probability.

What we have assembled here are the mathematical properties of the index system, as we have already used them in Chapter 9, plus a formal notion of the (initially heuristic) property of time series to be "close to 1", which essentially means *in probability*. However, this has to be complemented by an additional property relating to a number of expectations, which we shall study now.

B.2.2 Expectations in general

We have to assess conditional expectations of the terms

$$\varDelta D(a)_k^n, \qquad \varDelta D(a)_k \, \varDelta D(a)_l, \qquad \left(\varDelta D(a)_q \frac{ID(a)_q}{\widehat{ID(a)}_q}\right)^n, \qquad \varDelta D(a)_q \frac{ID(a)_q}{\widehat{ID(a)}_q} \varDelta D(a)_k$$

involving some pretty complex conditions. There is, however, a special case where these conditions do not matter at all, as the four terms are bounded pointwise. We could call it the *limited fluctuation case.*

Example B.3. Suppose the random errors generating the time series $\underline{\Delta D(a)}$ have finite and strictly positive support (this shall mean an interval $[min, max]$ with $0 < min \leq max < \infty$), such that both the random error generating the time series and its inverse are bounded. Then $\Delta D(a)_k = \frac{\underline{\Delta D(a)}_k}{\underline{\Delta D(a)}_0}$, $\Delta D(a)_l$, and $\Delta D(a)_q$ are bounded – they can all be written recursively as products, where essentially (positive or negative) powers of the error term appear.

Thus, $\mathrm{E}_{\underline{I}}\left(\Delta D(a)_k^n \mid |\Delta B_k - 1| > h\right)$, the first conditional expectation, is bounded and the same holds for $\mathrm{E}_{\underline{I}}\left(\Delta D(a)_k \, \Delta D(a)_l \mid |\Delta B_k - 1| > h \vee \left|\frac{\Delta B_l}{\Delta B_k} - 1\right| > h\right)$.

If further the random error generating the time series $\underline{ID(a)}$ has finite support too, then $ID(a)_q$ is bounded and so are the remaining expectations $\mathrm{E}_{\underline{I}}\left(\left(\Delta D(a)_q \frac{ID(a)_q}{\widehat{ID(a)}_q}\right)^n \mid \left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| > h\right)$ and $\mathrm{E}_{\underline{I}}\left(\Delta D(a)_q \frac{ID(a)_q}{\widehat{ID(a)}_q}\Delta D(\alpha)_k \mid |\Delta B_k - 1| > h \vee \left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| > h\right)$.

How can such finite supports come about? In terms of the comprehensive index system, recall from Section 8.3 the primordial decompositions

$$\underline{ID(a)} = \underline{\bar{M}}^{\omega'} \, \underline{IA} \, \underline{IH}^a \, \underline{I\bar{B}}^{\tau a} \, \underline{\Delta A}^{\delta\omega'} \, \underline{\Delta G}^{\omega'} \, \underline{\Delta\bar{B}}^{\zeta\omega'}$$

$$\underline{\Delta D(a)} = \underline{\Delta A}^{1-\delta\omega'} \, \underline{\Delta G}^{-\omega'} \, \underline{\Delta H}^a \, \underline{\Delta\bar{B}}^{\tau a - \zeta\omega'}$$

where ω' shall equal a for Cat-loss models else 1. If the random errors generating the appearing primordial time series have finite support, it is ensured that the same holds for $\underline{ID(a)}$. If in addition the random errors generating the primordial gaps have finite and strictly positive support, the error generating $\underline{\Delta D(a)}$ has a finite and strictly positive support.

This special case should have quite some real-world applications, however, time series with lognormal errors are of far higher interest to the practitioner.

B.2.3 Expectations in the lognormal case

If the error terms of a time series are lognormally distributed, they are unbounded, albeit moderately tailed at both ends, such that, in particular, all positive and negative moments of the time series elements are finite. We now treat the comprehensive index system with all time series being generated by lognormal errors, plus additional properties holding as assembled at the end of Section B.2.1.

Let us, however, start with a few general properties of lognormal models. Notice that a lognormal RV is close in probability to the constant RV 1 iff both parameters μ and σ are close to 0.

Proposition B.4. *If Y is lognormal and close to 1 in probability, for $d > 1$ the expectation $E(Y \mid Y > d)$ is close to d.*

Proof. From the formulary in Appendix A of [Klugman et al., 2008] we get the overall lognormal expectation $\mathrm{E}(Y) = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$ and that for the first risk up to d:

$$\mathrm{E}(\min(Y, d)) = \exp\left(\mu + \frac{1}{2}\sigma^2\right)\Phi\left(\frac{\ln d - \mu}{\sigma} - \sigma\right) + d\left(1 - \Phi\left(\frac{\ln d - \mu}{\sigma}\right)\right)$$

The difference of these expectations is that of the second risk:

$$\mathrm{E}\left((Y - d)^+\right) = \exp\left(\mu + \frac{1}{2}\sigma^2\right)\left(1 - \Phi\left(\frac{\ln d - \mu}{\sigma} - \sigma\right)\right) - d\left(1 - \Phi\left(\frac{\ln d - \mu}{\sigma}\right)\right)$$

Using $\mathrm{P}(Y>d)=1-\Phi\left(\frac{\ln d-\mu}{\sigma}\right)$ we get

$$\mathrm{E}(Y \mid Y>d)=d+\mathrm{E}(Y-d \mid Y>d)=d+\frac{\mathrm{E}\left((Y-d)^{+}\right)}{\mathrm{P}(Y>d)}=\exp\left(\mu+\frac{1}{2}\sigma^{2}\right)\frac{1-\Phi\left(\frac{\ln d-\mu}{\sigma}-\sigma\right)}{1-\Phi\left(\frac{\ln d-\mu}{\sigma}\right)}$$

If both parameters μ and σ tend to 0, the first factor in the last expression tends to 1, while in the second factor the arguments in the Gaussian cdf's are ultimately positive ($\ln d>0$). We can apply L'Hôpital's rule here, by letting $x:=\sigma^{-1}$ tend to infinite (and applying the limit as of μ subsequently):

$$\begin{aligned}\lim\frac{1-\Phi\left((\ln d-\mu)x-\frac{1}{x}\right)}{1-\Phi\left((\ln d-\mu)x\right)}&=\lim\left(\frac{\phi\left((\ln d-\mu)x-\frac{1}{x}\right)}{\phi\left((\ln d-\mu)x\right)}\frac{\ln d-\mu+\frac{1}{x^{2}}}{\ln d-\mu}\right)\\&=\lim\exp\left\{-\frac{1}{2}\left[\left((\ln d-\mu)x-\frac{1}{x}\right)^{2}-\left((\ln d-\mu)x\right)^{2}\right]\right\}=\lim\exp\left\{(\ln d-\mu)-\frac{1}{2x^{2}}\right\}=d\end{aligned}$$

□

Lemma B.5. *Let (Y_0,Y_1) be bivariate normal having expectations μ_i, variances σ_i^2, and covariance $\rho\sigma_0\sigma_1$, where $\sigma_1>0$, such that*

$$r:=\varrho\frac{\sigma_0}{\sigma_1}=\frac{Cov(Y_0,Y_1)}{Var(Y_1)}$$

is well defined. Then Y_0 is, conditionally given $Y_1=y$, normally distributed with the same parameters as the normal RV Y_0-rY_1+ry.

Proof. We have

$$\mathrm{E}(Y_0-rY_1+ry)=\mu_0-r\mu_1+ry=\mu_0+\varrho\frac{\sigma_0}{\sigma_1}(y-\mu_1)$$

$$\mathrm{Var}(Y_0-rY_1+ry)=\sigma_0^2+r^2\sigma_1^2-2r\rho\sigma_0\sigma_1=\sigma_0^2-r^2\sigma_1^2=\sigma_0^2\left(1-\rho^2\right)$$

According to standard theory, Y_0 is, conditionally given $Y_1=y$, normally distributed with parameters $\mu_0+\varrho\frac{\sigma_0}{\sigma_1}(y-\mu_1)$ and $\sigma_0^2\left(1-\rho^2\right)$, so the parameters of the two distributions coincide indeed. This notably even holds for the degenerate case $\rho=\pm1$, where Y_0 and Y_1 are linearly dependent. □

Corollary B.6. *Let (U_0,U_1) be bivariate lognormal, $(Y_1,Y_2)=(\ln U_0,\ln U_1)$ have parameters as in the preceding lemma, $\sigma_1>0$ (i.e. U_1 be not a scalar). Then U_0 is, conditionally given $U_1=u$, lognormally distributed with the same parameters as the lognormal RV $u^rU_0U_1^{-r}$. In particular, we have*

$$E(U_0 \mid U_1=u)=u^rE\left(U_0U_1^{-r}\right)$$

or equivalently

$$E(U_0 \mid U_1)=U_1^rE\left(\tilde{U}_0\tilde{U}_1^{-r}\right)$$

where $\left(\tilde{U}_0,\tilde{U}_1\right)$ is a copy of (U_0,U_1) having the same distribution.

Proof. If we apply the lemma to (Y_1,Y_2) and $\ln u$, we get the assertion on the distributions, noting that $\ln\left(u^rU_0U_1^{-r}\right)=Y_0-rY_1+r\ln u$. The expectations follow immediately. □

Proposition B.7. *In the setting of the corollary, let Λ be a Borel set of values that U_1 may take on. Then we have*

$$E(U_0 \mid U_1\in\Lambda)=E(U_1^r \mid U_1\in\Lambda)\,E\left(U_0U_1^{-r}\right),\qquad r=\frac{Cov(\ln U_0,\ln U_1)}{Var(\ln U_1)}$$

Proof. As the binary RV $\{U_1 \in \Lambda\}$ is a measurable function of U_1, we have from above, setting $c := \mathrm{E}\left(U_0 U_1^{-r}\right)$, in terms of abstract conditional expectation

$$\mathrm{E}\left(U_0 \mid U_1, \{U_1 \in \Lambda\}\right) = \mathrm{E}\left(U_0 \mid U_1\right) = U_1^r c$$

$$\mathrm{E}\left(U_0 \mid \{U_1 \in \Lambda\}\right) = \mathrm{E}\left(\mathrm{E}\left(U_0 \mid U_1, \{U_1 \in \Lambda\}\right) \mid \{U_1 \in \Lambda\}\right) = \mathrm{E}\left(U_1^r c \mid \{U_1 \in \Lambda\}\right)$$

which yields

$$\mathrm{E}\left(U_0 \mid U_1 \in \Lambda\right) = \mathrm{E}\left(U_1^r c \mid U_1 \in \Lambda\right) = \mathrm{E}\left(U_1^r \mid U_1 \in \Lambda\right) \mathrm{E}\left(U_0 U_1^{-r}\right)$$

in terms of elementary conditional expectation. □

After these preliminaries we look at the independent comprehensive index system in the lognormal case, where it constitutes a multivariate lognormal distribution. We study the four conditional expectations of interest, starting with $\mathrm{E}_{\underline{I}}\left(\Delta D(a)_k^n \mid |\Delta B_k - 1| > h\right)$. We have already shown (see Example 9.4 and the preceding considerations) that the distribution of $\Delta D(a)_k$, conditionally given $\underline{I}$, has parameters very similar to those of the corresponding unconditional distribution. So we have

$$\mathrm{E}_{\underline{I}}\left(\Delta D(a)_k^n \mid |\Delta B_k - 1| > h\right) \approx \mathrm{E}\left(\Delta D(a)_k^n \mid |\Delta B_k - 1| > h\right)$$

To the RHS we can apply the above results for bivariate lognormal RV's, leaving aside, as throughout this book, the degenerate case of a constant RV ΔB_k, which would mean $\Delta B_k \equiv 1$, i.e. no inflation basis risk.

Proposition B.8. *In the lognormal independent index system with non-degenerate ΔB_k we have*

$$\begin{aligned} E\left(\Delta D(a)_k^n \mid |\Delta B_k - 1| > h\right) &= E\left(\Delta B_k^{nr} \mid |\Delta B_k - 1| > h\right) E\left(\Delta D(a)_k^n \, \Delta B_k^{-nr}\right) \\ &= E\left(\Delta B_k^{nr} \mid |\Delta B_k - 1| > h\right) E\left(\Delta D(s)_k^n\right) \end{aligned}$$

where $s := a - r$, and

$$r = \frac{Cov\left(\ln \Delta D(a)_k, \ln \Delta B_k\right)}{Var\left(\ln \Delta B_k\right)}, \qquad s = -\frac{Cov\left(\ln \Delta V_k, \ln \Delta B_k\right)}{Var\left(\ln \Delta B_k\right)} = \frac{\chi\delta\sigma_1^2 + (1-\chi)\,\tau\zeta\sigma_4^2}{\chi\delta^2\sigma_1^2 + \chi\sigma_2^2 + \sigma_3^2 + (\tau - \chi\zeta)^2\,\sigma_4^2}$$

with the abbreviations

$$\sigma_1^2 = Var\left(\ln \Delta A_k\right), \quad \sigma_2^2 = Var\left(\ln \Delta G_k\right), \quad \sigma_3^2 = Var\left(\ln \Delta H_k\right), \quad \sigma_4^2 = Var\left(\ln \Delta \bar{B}_k\right)$$

If $\delta = 0$, we have $0 \le s \le \frac{\zeta}{\tau}$, else $0 \le s \le \max\left(\frac{\zeta}{\tau}, \frac{1}{\delta}\right)$. Accordingly, r lies between $a - \frac{\zeta}{\tau}$ or $a - \max\left(\frac{\zeta}{\tau}, \frac{1}{\delta}\right)$, respectively, and a, thus may be negative. We have simple bounds for $|r|$ and $|s|$, which depend on the structure (namely the exponents) of the index system and on the regional and local alphas of the normalized severity, but notably not on the parameters of the indices themselves:

$$|s| = s \le \max\left(\frac{\zeta}{\tau}, \frac{1}{\delta}\right), \qquad |r| \le \max\left(a_{max}, \frac{\zeta}{\tau}, \frac{1}{\delta}\right)$$

where in case $\delta = 0$ one drops $\frac{1}{\delta}$.

Proof. The second equality in the first formula follows from

$$\Delta D(a)_k^n \, \Delta B_k^{-nr} = \Delta V_k^n \Delta B_k^{na} \Delta B_k^{-nr} = \Delta V_k^n \Delta B_k^{ns} = \Delta D(s)_k^n$$

The first equality is the preceding proposition applied to the bivariate lognormal RV $\left(\Delta D(a)_k^n, \Delta B_k\right)$ and the Borel set $\Lambda = (0, 1-h) \cup (1+h, \infty)$, where, however, nr takes the place of r. It remains to check the formulae for r and s. The above proposition yields the first of the two following equivalent formulae:

$$nr = \frac{\mathrm{Cov}\left(\ln \Delta D(a)_k^n, \ln \Delta B_k\right)}{\mathrm{Var}\left(\ln \Delta B_k\right)}, \qquad r = \frac{\mathrm{Cov}\left(\ln \Delta D(a)_k, \ln \Delta B_k\right)}{\mathrm{Var}\left(\ln \Delta B_k\right)}$$

The latter together with $\ln \Delta D(a)_k = \ln \Delta V_k + a \ln \Delta B_k$ gives the first equality for s:

$$s = a - r = a - \frac{\mathrm{Cov}\left(\ln \Delta V_k + a \ln \Delta B_k, \ln \Delta B_k\right)}{\mathrm{Var}\left(\ln \Delta B_k\right)} = -\frac{\mathrm{Cov}\left(\ln \Delta V_k, \ln \Delta B_k\right)}{\mathrm{Var}\left(\ln \Delta B_k\right)}$$

Finally, recall the primordial decompositions

$$\Delta V_k = \Delta A_k^{1-(1-\chi)\delta} \Delta G_k^{\chi-1} \Delta \bar{B}_k^{(\chi-1)\zeta}$$

$$\Delta B_k = \Delta A_k^{-\chi\delta} \Delta G_k^{-\chi} \Delta H_k \Delta \bar{B}_k^{\tau-\chi\zeta}$$

The logarithms of both gaps are linear combinations of the same four independent normal RV's. One sees quickly that

$$\mathrm{Cov}\left(\ln \Delta V_k, \ln \Delta B_k\right) = -\chi\delta\sigma_1^2 + (\chi - 1)\,\tau\zeta\sigma_4^2$$

$$0 < \mathrm{Var}\left(\ln \Delta B_k\right) = \chi\delta^2\sigma_1^2 + \chi\sigma_2^2 + \sigma_3^2 + (\tau - \chi\zeta)^2\,\sigma_4^2$$

where we have used that χ is an indicator equaling 0 or 1, such that $\chi^2 = \chi$. So we have

$$s = -\frac{\mathrm{Cov}\left(\ln \Delta V_k, \ln \Delta B_k\right)}{\mathrm{Var}\left(\ln \Delta B_k\right)} = \frac{\chi\delta\sigma_1^2 + (1-\chi)\,\tau\zeta\sigma_4^2}{\chi\delta^2\sigma_1^2 + \chi\sigma_2^2 + \sigma_3^2 + (\tau - \chi\zeta)^2\,\sigma_4^2}$$

In the last term only nonnegative coefficients appear, thus s must be nonnegative. For the upper bound we split the cases for χ. If $\chi = 0$, we have (recall that $\tau > 0$)

$$s = \frac{\tau\zeta\sigma_4^2}{\sigma_3^2 + \tau^2\sigma_4^2} \le \frac{\zeta}{\tau}$$

If instead $\chi = 1$, we have

$$s = \frac{\delta\sigma_1^2}{\delta^2\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + (\tau - \zeta)^2\,\sigma_4^2}$$

which equals 0 if $\delta = 0$, otherwise it may be positive but cannot exceed $\frac{1}{\delta}$. The interval for $r = a - s$ follows immediately. $\square$

To assess the resulting product $\mathrm{E}\left(\Delta B_k^{nr} \mid |\Delta B_k - 1| > h\right) \mathrm{E}\left(\Delta D(s)_k^n\right)$, notice that the first factor is a weighted average of $\mathrm{E}\left(\Delta B_k^{nr} \mid \Delta B_k > 1 + h\right)$ and $\mathrm{E}\left(\Delta B_k^{nr} \mid \Delta B_k < 1 - h\right)$. We show that both terms are bounded. If $r = 0$, this is clear.

Let now be $r > 0$. Then the second expectation cannot exceed 1 (this holds even pointwise), while the first one can be written as $\mathrm{E}\left(\Delta B_k^{nr} \mid \Delta B_k^{nr} > (1+h)^{nr}\right)$. If instead $r < 0$, the first expectation cannot exceed 1, while the second one can be written as $\mathrm{E}\left(\Delta B_k^{nr} \mid \Delta B_k^{nr} > (1-h)^{nr}\right)$.

If ΔB_n is close to 1 in probability, the power ΔB_k^{nr} is close to 1 too, as the (absolute) exponent $|nr|$ cannot be arbitrarily large due to the bound for $|r|$. Thus, we can apply Proposition B.4 and see that the expectations in question are close to $(1+h)^{nr}$ or $(1-h)^{nr}$, respectively. Both terms are greater than but not too far from 1, as $h > 0$ is small.

Summing up, $\mathrm{E}\left(\Delta B_k^{nr} \mid |\Delta B_k - 1| > h\right)$ is the weighted average of a term that is less than 1 and a term that is not much greater than 1. Therefore, $\mathrm{E}\left(\Delta D(a)_k^n \mid |\Delta B_k - 1| > h\right)$ and the close term $\mathrm{E}_{\underline{I}}\left(\Delta D(a)_k^n \mid |\Delta B_k - 1| > h\right)$ are smaller than or in the range of $\mathrm{E}\left(\Delta D(s)_k^n\right)$. This latter term is bounded too, more strongly it is close to 1: $\Delta D(s)_k^n = \Delta V_k^n \Delta B_k^{ns}$ inherits the closeness (in probability) to 1 from ΔV_k and ΔB_k, as its (nonnegative) exponents n and ns cannot be arbitrarily large due to the upper bound for s.

Let us go to the next conditional expectation of interest ($l < k \leq 0$):

$$\mathrm{E}_{\underline{I}}\left(\Delta D(a)_k \, \Delta D(a)_l \mid |\Delta B_k - 1| > h \wedge \left|\frac{\Delta B_l}{\Delta B_k} - 1\right| > h\right)$$

Like for $\Delta D(a)_k$, we know from Section 9.1 that the conditional distribution of $\Delta D(a)_k \, \Delta D(a)_l$, given $\underline{I}$, has parameters very similar to those of the corresponding unconditional distribution. Thus, we have

$$\begin{aligned}\mathrm{E}_{\underline{I}}&\left(\Delta D(a)_k \, \Delta D(a)_l \mid |\Delta B_k - 1| > h \vee \left|\frac{\Delta B_l}{\Delta B_k} - 1\right| > h\right)\\ &\approx \mathrm{E}\left(\Delta D(a)_k \, \Delta D(a)_l \mid |\Delta B_k - 1| > h \vee \left|\frac{\Delta B_l}{\Delta B_k} - 1\right| > h\right)\end{aligned}$$

By splitting into two disjoint cases, we see that the latter expectation is a weighted average of

$$\mathrm{E}\left(\Delta D(a)_k \, \Delta D(a)_l \mid |\Delta B_k - 1| > h\right), \qquad \mathrm{E}\left(\Delta D(a)_k \, \Delta D(a)_l \mid |\Delta B_k - 1| \leq h \wedge \left|\frac{\Delta B_l}{\Delta B_k} - 1\right| > h\right)$$

Now notice that

$$\Delta D(a)_k \, \Delta D(a)_l = \Delta D(a)_k^2 \, \frac{\Delta D(a)_l}{\Delta D(a)_k}$$

where the RHS would be an independent product if $\Delta D(a)_k$ were a (geometric) random walk. Let us assume that, even more strongly, all primordial gaps are geometric random walks. We treat only this special case – it is sufficient for the application of our theory to lognormal models in Chapter 10. Then

$$\left(\Delta D(a)_k \, , \Delta B_k\right), \qquad \left(\frac{\Delta D(a)_l}{\Delta D(a)_k}, \frac{\Delta B_l}{\Delta B_k}\right)$$

are independent, such we can decompose according to Lemma 9.2:

$$\mathrm{E}\left(\Delta D(a)_k^2 \, \frac{\Delta D(a)_l}{\Delta D(a)_k} \mid |\Delta B_k - 1| > h\right) = \mathrm{E}\left(\Delta D(a)_k^2 \mid |\Delta B_k - 1| > h\right) \mathrm{E}\left(\frac{\Delta D(a)_l}{\Delta D(a)_k}\right)$$

$$\begin{aligned}\mathrm{E}&\left(\Delta D(a)_k^2 \, \frac{\Delta D(a)_l}{\Delta D(a)_k} \mid |\Delta B_k - 1| \leq h \wedge \left|\frac{\Delta B_l}{\Delta B_k} - 1\right| > h\right)\\ &= \mathrm{E}\left(\Delta D(a)_k^2 \mid |\Delta B_k - 1| \leq h\right) \mathrm{E}\left(\frac{\Delta D(a)_l}{\Delta D(a)_k} \mid \left|\frac{\Delta B_l}{\Delta B_k} - 1\right| > h\right)\end{aligned}$$

Now we have four factors, which can be seen quickly to be bounded. For $\mathrm{E}\left(\Delta D(a)_k^2 \mid |\Delta B_k - 1| > h\right)$ this was proven above. In a geometric random walk $\left(\frac{\Delta D(a)_l}{\Delta D(a)_k}, \frac{\Delta B_l}{\Delta B_k}\right)$ is distributed like $\left(\Delta D(a)_{l-k} \, , \Delta B_{l-k}\right)$, therefore $\mathrm{E}\left(\frac{\Delta D(a)_l}{\Delta D(a)_k}\right)$ and $\mathrm{E}\left(\frac{\Delta D(a)_l}{\Delta D(a)_k} \mid \left|\frac{\Delta B_l}{\Delta B_k} - 1\right| > h\right)$ are bounded. For the remaining term we apply Proposition B.7 to the bivariate lognormal RV $\left(\Delta D(a)_k^n \, , \Delta B_k\right)$ and the Borel set $\Lambda = [1-h, 1+h]$, which yields

$$\mathrm{E}\left(\Delta D(a)_k^n \mid |\Delta B_k - 1| \leq h\right) = \mathrm{E}\left(\Delta B_k^{nr} \mid 1 - h \leq \Delta B_k \leq 1 + h\right) \mathrm{E}\left(\Delta D(s)_k^n\right)$$

Here the first factor of the RHS is bounded from above either by $(1+h)^{nr}$ or (if $r < 0$) by $(1-h)^{-nr}$, while the second factor is close to 1.

The third expectation is

$$\mathrm{E}_{\underline{I}}\left(\left(\Delta D(a)_q \frac{ID(a)_q}{\widehat{ID(a)}_q}\right)^n \mid \left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| > h\right) = \mathrm{E}_{\underline{I}}\left(\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\right)^n \mid \left|\frac{B_q}{\widehat{IB}_q} - 1\right| > h\right)$$

which can be decomposed, as we can see from the corresponding abstract conditional expectation with the σ-algebra of the appearing binary event.

Proposition B.9. *In the lognormal strongly independent comprehensive index system we have*

$$E\left(\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\right)^n \mid \left\{\left|\frac{B_q}{\widehat{IB}_q} - 1\right| > h\right\}, \underline{I}\right) = E\left(\left(\frac{B_q}{\widehat{IB}_q}\right)^{na} \mid \left\{\left|\frac{B_q}{\widehat{IB}_q} - 1\right| > h\right\}, \underline{I}\right) E\left(\left(\frac{V_q}{\widehat{IV}_q}\right)^n \mid \underline{I}\right)$$

Proof. Recall that $\underline{I}$ equals

$$\sigma\left(\underline{I\bar{M}}_{(j\leq 0)}, \Upsilon_{pr}\right) = \sigma\left(\underline{I\bar{M}}_{(j\leq 0)}, \underline{IA}_{(j\leq 0)}, \underline{IG}_{(j\leq 0)}, \underline{IH}_{(j\leq 0)}, \underline{I\bar{B}}_{(j\leq 0)}\right)$$

In the independent comprehensive index system the products

$$V = \bar{M}^{1-\chi}A, \qquad B = \bar{M}^{\chi}H\bar{B}^{\tau}$$

are independent products and independent of each other: $\bar{M}$ appears only in one of them, for $\chi = 0$ in V, for $\chi = 1$ in B. The analogous result holds for the corresponding picked indices (while the respective gaps can notably be dependent) and thus also for their predictors, which are measurable functions of the past index values. Due to this multiplicative structure we find, by checking the primordial decompositions and noting that past and future gaps are independent (strong independence), that

$$\left(V_q, \widehat{IV}_q, \underline{I\bar{M}}^{1-\chi}_{(j\leq 0)}, \underline{IA}_{(j\leq 0)}, \underline{IG}_{(j\leq 0)}\right), \qquad \left(B_q, \widehat{IB}_q, \underline{I\bar{M}}^{\chi}_{(j\leq 0)}, \underline{IH}_{(j\leq 0)}, \underline{I\bar{B}}_{(j\leq 0)}, \left\{\left|\frac{B_q}{\widehat{IB}_q} - 1\right| > h\right\}\right)$$

are independent. Thus, we can apply Lemma 9.2 and get

$$\begin{aligned}
&\mathrm{E}\left(\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\right)^n \mid \left\{\left|\frac{B_q}{\widehat{IB}_q} - 1\right| > h\right\}, \underline{I}\right) \\
&\quad = \mathrm{E}\left(\left(\frac{V_q}{\widehat{IV}_q}\right)^n \left(\frac{B_q}{\widehat{IB}_q}\right)^{na} \mid \left\{\left|\frac{B_q}{\widehat{IB}_q} - 1\right| > h\right\}, \underline{I\bar{M}}_{(j\leq 0)}, \underline{IA}_{(j\leq 0)}, \underline{IG}_{(j\leq 0)}, \underline{IH}_{(j\leq 0)}, \underline{I\bar{B}}_{(j\leq 0)}\right) \\
&\quad = \mathrm{E}\left(\left(\frac{B_q}{\widehat{IB}_q}\right)^{na} \mid \left\{\left|\frac{B_q}{\widehat{IB}_q} - 1\right| > h\right\}, \underline{I}\right) \mathrm{E}\left(\left(\frac{V_q}{\widehat{IV}_q}\right)^n \mid \left\{\left|\frac{B_q}{\widehat{IB}_q} - 1\right| > h\right\}, \underline{I}\right) \\
&\quad = \mathrm{E}\left(\left(\frac{B_q}{\widehat{IB}_q}\right)^{na} \mid \left\{\left|\frac{B_q}{\widehat{IB}_q} - 1\right| > h\right\}, \underline{I}\right) \mathrm{E}\left(\left(\frac{V_q}{\widehat{IV}_q}\right)^n \mid \underline{I}\right)
\end{aligned}$$

□

In terms of the probability measure $\mathrm{P}_{\underline{I}}$, the same decomposition holds for the elementary conditional expectation with the event $\left|\frac{B_q}{\widehat{IB}_q}-1\right|>h$. The resulting factors of

$$\mathrm{E}_{\underline{I}}\left(\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\right)^n \mid \left|\frac{B_q}{\widehat{IB}_q}-1\right|>h\right)=\mathrm{E}_{\underline{I}}\left(\left(\frac{B_q}{\widehat{IB}_q}\right)^{na} \mid \left|\frac{B_q}{\widehat{IB}_q}-1\right|>h\right)\mathrm{E}_{\underline{I}}\left(\left(\frac{V_q}{\widehat{IV}_q}\right)^n\right)$$

can be evaluated, and proven to be bounded, exactly as done above with $\mathrm{E}\left(\Delta B_k^{nr}\mid|\Delta B_k-1|>h\right)\mathrm{E}\left(\Delta D(s)_k^n\right)$, by using that, conditionally on $\underline{I}$, $\frac{B_q}{\widehat{IB}_q}$ and $\frac{V_q}{\widehat{IV}_q}$ are close to 1 in probability.

It remains to look at

$$\begin{aligned}\mathrm{E}_{\underline{I}}&\left(\Delta D(a)_q\frac{ID(a)_q}{\widehat{ID(a)}_q}\Delta D(a)_k\mid|\Delta B_k-1|>h\vee\left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q-1\right|>h\right)\\&=\mathrm{E}_{\underline{I}}\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\Delta D(a)_k\mid|\Delta B_k-1|>h\vee\left|\frac{B_q}{\widehat{IB}_q}-1\right|>h\right)\end{aligned}$$

which is a weighted average of the expectations

$$\mathrm{E}_{\underline{I}}\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\Delta D(a)_k\mid|\Delta B_k-1|>h\right),\qquad\mathrm{E}_{\underline{I}}\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\Delta D(a)_k\mid|\Delta B_k-1|\le h,\left|\frac{B_q}{\widehat{IB}_q}-1\right|>h\right)$$

We use the same approach as before, now getting approximate decompositions. We will work with approximate independence, which is somewhat rougher an approach than how we have generally proceeded in this book. However, the calculations will yield upper bounds (notably pretty small ones) for a basic inequality, which do not directly enter any formulae for optimal weights etc. Hence, an approximate assessment should suffice, the more so as it is clear how the desired (approximate) decompositions can be verified more rigorously: by exploring the multivariate (log)normal setting via (a lot of) linear algebra, along the lines of Example 9.4.

Recall from Section 9.1 that the condition on $\sigma\left(\underline{I\bar{M}}_{(j\le 0)}\right)\subseteq\underline{I}$ ties past gaps and deflated volume, creating an intricate but typically very weak dependence. So the RV's

$$\left(\Delta D(a)_k,\{|\Delta B_k-1|>h\}\right),\qquad\left(D(a)_q,\widehat{ID(a)}_q,\underline{I},\left\{\left|\frac{B_q}{\widehat{IB}_q}-1\right|>h\right\}\right)$$

are approximately independent, such that Lemma 9.2 gives a fair approximation

$$\begin{aligned}&\mathrm{E}\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\Delta D(a)_k\mid\{|\Delta B_k-1|>h\},\left\{\left|\frac{B_q}{\widehat{IB}_q}-1\right|>h\right\},\underline{I}\right)\approx\\&\mathrm{E}\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\mid\{|\Delta B_k-1|>h\},\left\{\left|\frac{B_q}{\widehat{IB}_q}-1\right|>h\right\},\underline{I}\right)\mathrm{E}\left(\Delta D(a)_k\mid\{|\Delta B_k-1|>h\},\left\{\left|\frac{B_q}{\widehat{IB}_q}-1\right|>h\right\},\underline{I}\right)\\&\qquad\approx\mathrm{E}\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\mid\left\{\left|\frac{B_q}{\widehat{IB}_q}-1\right|>h\right\},\underline{I}\right)\mathrm{E}\left(\Delta D(a)_k\mid\{|\Delta B_k-1|>h\}\right)\end{aligned}$$

In terms of $\mathrm{P}_{\underline{I}}$, the analogous approximate decompositions hold for the elementary conditional expectations, using first $|\Delta B_k - 1| > h$, then $|\Delta B_k - 1| \leq h$ and $\left|\frac{B_q}{\widehat{IB}_q} - 1\right| > h$:

$$\mathrm{E}_{\underline{I}}\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\Delta D(a)_k \mid |\Delta B_k - 1| > h\right) \approx \mathrm{E}_{\underline{I}}\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\right)\mathrm{E}\left(\Delta D(a)_k \mid |\Delta B_k - 1| > h\right)$$

$$\begin{aligned}\mathrm{E}_{\underline{I}}&\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\Delta D(a)_k \mid |\Delta B_k - 1| \leq h, \left|\frac{B_q}{\widehat{IB}_q} - 1\right| > h\right)\\ &\approx \mathrm{E}_{\underline{I}}\left(\frac{D(a)_q}{\widehat{ID(a)}_q} \mid \left|\frac{B_q}{\widehat{IB}_q} - 1\right| > h\right)\mathrm{E}\left(\Delta D(a)_k \mid |\Delta B_k - 1| \leq h\right)\end{aligned}$$

With these decompositions we are done: All factors to the RHS of the last two formulae we have already investigated, and seen to be bounded, except for $\mathrm{E}_{\underline{I}}\left(\frac{D(a)_q}{\widehat{ID(a)}_q}\right)$, which is trivially close to 1 as $\frac{D(a)_q}{\widehat{ID(a)}_q}$ is close to 1 in probability, given $\underline{I}$.

B.3 The Pareto case

Let us close a last gap in our reasoning. Recall that the approximations in this appendix are essentially about the leveraged effect of ground-up inflation on layers. For proportional insurance this effect is linear, the relevant regional alphas equal 1, and the approximations are trivial and exact. For layers covering losses in a Pareto distributed tail, the relevant regional/local alphas equal the Pareto alpha and the approximations are exact for non-random inflation, as long as the Pareto tail starts sufficiently low, see Proposition 7.4. We show that for random inflation having normalized gaps close to 1 (in the sense explored in this appendix) our approximations are not exact for Pareto, but extremely precise, which much less room for deviations than in the general case. To see this, we study a variant of the first basic inequality, which assesses the as-if loss; for the further quantities the calculations are analogous.

Let the normalized severity $\mathring{Z}$ be Pareto distributed beyond a threshold t, $0 < t < d^*$. Then for $\Delta B_k \leq \frac{d^*}{t}$ we have the (exact, not approximative) formula

$$\mathrm{E}\left({}_{d^*}^{u^*}\left(\Delta B_k \mathring{Z} \mid \mathcal{W}\right)\right) = \Delta B_k^{\alpha}\,\mathrm{E}\left({}_{d^*}^{u^*}\mathring{Z} \mid \widehat{IB}_q\right)$$

or equivalently

$$\mathrm{E}\left(S_k \mid \mathcal{W}\right) = \tilde{e}\Delta D(\alpha)_k$$

By splitting the expectation into the cases of ΔB_k being greater than $\frac{d^*}{t}$ or not, we get the following upper bound for the error of the approximation in question:

$$\begin{aligned}\left|\mathrm{E}\left(S_k \mid \underline{I}\right) - \tilde{e}\mathrm{E}\left(\Delta D(\alpha)_k \mid \underline{I}\right)\right| &= \left|\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(S_k \mid \mathcal{W}\right) - \tilde{e}\Delta D(\alpha)_k\right)\right|\\ &\leq \mathrm{P}\left(\Delta B_k \leq \frac{d^*}{t} \mid \underline{I}\right)\left|\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(S_k \mid \mathcal{W}\right) - \tilde{e}\Delta D(\alpha)_k \mid \Delta B_k \leq \frac{d^*}{t}\right)\right|\\ &\quad + \mathrm{P}\left(\Delta B_k > \frac{d^*}{t} \mid \underline{I}\right)\left|\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(S_k \mid \mathcal{W}\right) - \tilde{e}\Delta D(\alpha)_k \mid \Delta B_k > \frac{d^*}{t}\right)\right|\\ &\leq \mathrm{P}\left(\Delta B_k > \frac{d^*}{t} \mid \underline{I}\right)\left\{\mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(S_k \mid \mathcal{W}\right) \mid \Delta B_k > \frac{d^*}{t}\right) + \mathrm{E}_{\underline{I}}\left(\tilde{e}\Delta D(\alpha)_k \mid \Delta B_k > \frac{d^*}{t}\right)\right\}\end{aligned}$$

Here the term in the second row vanishes, such that we get a quite simple final expression. The two terms in braces can be assessed as done in the past section; again they cannot be arbitrarily large. As $\frac{d^*}{t} > 1$, the probability $\mathrm{P}\left(\Delta B_k > \frac{d^*}{t} \mid \underline{I}\right)$ is small if ΔB_k is close to 1 in probability. However, if $\frac{d^*}{t}$ is *much* greater than 1 (which means that the Pareto tail starts far below d^*), $\mathrm{P}\left(\Delta B_k > \frac{d^*}{t} \mid \underline{I}\right)$ is very small even in case ΔB_k is not very close to 1 in probability. Hence, in such situations we have much more *tolerance* about ΔB_k for the desired approximation to hold. If the assumption that ΔB_k be in probability very close to 1, is fulfilled, the approximation will be extremely precise. If the assumption is only fulfilled to some extent, the approximation will still be good.

This point is arguably not essential for the as-if loss – in practice, if indices are selected carefully, we expect gaps to be very likely pretty close to 1. However, such tolerance may help with the future quantities. To illustrate this, let us look at the Pareto variant of the basic inequality for the future loss

$$\left| \mathrm{E}\left(X_q \mid \underline{I}\right) - \tilde{e}\mathrm{E}\left(\Delta D(\alpha)_q \frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q} \mid \underline{I}\right) \right| \leq$$

$$\mathrm{P}\left(\frac{IB_q}{\widehat{IB}_q}\Delta B_q > \frac{d^*}{t} \mid \underline{I}\right) \left\{ \mathrm{E}_{\underline{I}}\left(\mathrm{E}\left(X_q \mid \mathcal{W}\right) \mid \frac{IB_q}{\widehat{IB}_q}\Delta B_q > \frac{d^*}{t}\right) + \mathrm{E}_{\underline{I}}\left(\tilde{e}\Delta D(\alpha)_q \frac{ID(\alpha)_q}{\widehat{ID(\alpha)}_q} \mid \frac{IB_q}{\widehat{IB}_q}\Delta B_q > \frac{d^*}{t}\right) \right\}$$

and, in particular, at the probability $\mathrm{P}\left(\frac{IB_q}{\widehat{IB}_q}\Delta B_q > \frac{d^*}{t} \mid \underline{I}\right)$. It should be small in *many more* real-world situations than the corresponding probability $\mathrm{P}\left(\left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| > h \mid \underline{I}\right)$ of the original basic inequality: The RV ΔB_q is not critical in this context – as a gap it can be expected to be close to 1 in probability (both unconditionally and conditionally; in the strongly independent index system it is anyway independent of $\underline{I}$). But can we also expect $\frac{IB_q}{\widehat{IB}_q}$ to be close to 1 in probability, conditionally given $\underline{I}$? This requires (conditionally) that the prediction of IB_q be quite precise *and* that IB_q fluctuate few. The former seems manageable, but the latter might be harder to believe in some real-world situations, namely in the case of accumulation losses: Here among the primordial components of IB the deflated volume $\bar{M}$ appears, which we notably assume to be considerably more volatile than the gaps. Thus, in practice we will come across situations where $\mathrm{P}\left(\left|\frac{IB_q}{\widehat{IB}_q}\Delta B_q - 1\right| > h \mid \underline{I}\right)$ is not very small, while $\mathrm{P}\left(\frac{IB_q}{\widehat{IB}_q}\Delta B_q > \frac{d^*}{t} \mid \underline{I}\right)$ is.

Analogously one sees that the desired boundedness of the expectations in braces is easier to fulfill for the Pareto variant of the basic inequalities, as long as $\frac{d^*}{t}$ is large.

Luckily, in real-world Cat-loss situations Pareto-like tails are not uncommon, starting often pretty low, such that one can apply the Pareto basic inequalities and take advantage of their higher tolerance with respect to closeness of indices to 1.

B.4 Wrap up

By formally detailing in what way certain indices need to be close to 1, we have finally managed to extend the theory about leveraged layer inflation to stochastic inflation models that are appealing for practical use: firstly the model with limited year-to-year fluctuation, secondly and more importantly the well-established lognormal model. This was possible within the range of (plausible) assumptions used in the main part of the book, as assembled in Theorem 9.18. We only had to add that in the lognormal case the gaps be geometric random walks, but this is no restriction, as random walk gaps are exactly the one lognormal setting we have applied our theory to: first in the basic example, then in the general case treated in Chapter 10.

Appendix C

Overview of predictors and indices

Future loss variable	Description	Predictor	Constant
X_q	aggregate loss	S_k	$\tilde{e} = \widehat{IV}_q \widehat{IB}_q \, e_{(d^*,u^*)}$
Y_q	aggregate loss per official volume unit	R_k	$\mathring{e} = \tilde{e} \, \widehat{M}_q^{-1}$
$\bar{N}_q$	loss count	C_k	$\tilde{\theta} = \widehat{IV}_q \, \theta_{d^*}$
$\bar{N}_q M_q^{-1}$	loss count per official volume unit	$C_k \widehat{M}_q^{-1}$	$\mathring{\theta} = \tilde{\theta} \, \widehat{M}_q^{-1}$

Index	Description	Formula
B	loss inflation	$B = \bar{M}^{\chi} H \bar{B}^{\tau}$
V	frequency volume	$V = \bar{M}^{1-\chi} A$
M	official volume	$M = \bar{M} A^{\delta} \widetilde{B}$
$\widetilde{B}$	official-volume inflation	$\widetilde{B} = G \bar{B}^{\zeta}$
$\bar{M}$	deflated official volume	
A	frequency per deflated official volume unit	
$\bar{B}$	basic inflation	
H	superimposed loss inflation	
G	superimposed official-volume inflation	
D	combined leveraged scale-frequency impact	$D(\alpha) = V B^{\alpha}$
J	combined impact per official volume unit	$J(\alpha) = D(\alpha) \, M^{-1}$
K	combined impact per squared official volume unit	$K(\alpha) = D(\alpha) \, M^{-2}$
$\widetilde{G}$	complement of $\bar{M}$ in M	$M = \bar{M} \widetilde{G}$
P	represents D, J, K	$P(\alpha, z) = D(\alpha) \, M^{-z}$
Q	complement of $\bar{M}$ in P	$P(\alpha, z) = Q(\alpha, z) \, \bar{M}^{\omega - z}$
τ	leverage of basic inflation in loss inflation	
ζ	leverage of basic inflation in official-volume inflation	
δ	indicator for A in M	
χ	indicator for Cat vs non-Cat loss model	
ω	leverage of $\bar{M}$ in D	$\omega = 1 + (\alpha - 1) \, \chi$

Bibliography

[Ansley and Kohn, 1986] Ansley, C. F. and Kohn, R. (1986). Prediction mean squared error for state space models with estimated parameters. *Biometrika*, 73(2):467–473.

[Bohnert et al., 2016] Bohnert, A., Gatzert, N., and Kolb, A. (2016). Assessing inflation risk in non-life insurance. *Insurance: Mathematics and Economics*, 66:86–96.

[Brickmann et al., 2005] Brickmann, S., Forster, W., and Sheaf, S. (2005). Claims inflation – uses and abuses. In *GIRO Convention 2005*.

[Brockwell and Davis, 2006] Brockwell, P. J. and Davis, R. A. (2006). *Introduction to time series and forecasting*. Springer.

[Clark, 2006] Clark, D. R. (2006). Variance and covariance due to inflation. In *CAS Forum, Fall 2006*, pages 61–95.

[Cramer et al., 2005] Cramer, M., Lengfeld, T., Neeb, M., and Voggenauer-von Bothmer, R. (2005). Die sinkende Schadenhäufigkeit in der deutschen Kraftfahrtversicherung. *Versicherungswirtschaft*, 60(19):1460–1465.

[Embrechts et al., 2013] Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events for insurance and finance*. Springer.

[Fackler, 2011] Fackler, M. (2011). Panjer Class United: one formula for the probabilities of the Poisson, Binomial, and Negative Binomial distribution. *Anales del Instituto de Actuarios Españoles*, (17):1–12.

[FINMA, 2006] FINMA (2006). *Technical document on the Swiss Solvency Test*. FINMA.

[Fuller and Hasza, 1981] Fuller, W. A. and Hasza, D. P. (1981). Properties of predictors for autoregressive time series. *Journal of the American Statistical Association*, 76(373):155–161.

[Gerathewohl, 1979] Gerathewohl, K. (1979). *Rückversicherung: Grundlagen und Praxis*. Verlag Versicherungswirtschaft.

[Gerber and Jones, 1975] Gerber, H. U. and Jones, D. A. (1975). Credibility formulas of the updating type. *Transactions of the Society of Actuaries*, 27:31–52.

[Gonçalves Mazzeu et al., 2015] Gonçalves Mazzeu, J. H., Ruiz, E., and Veiga, H. (2015). Model uncertainty and the forecast accuracy of ARMA models: A survey. *UC3M Working Papers*.

[Härdle and Hlávka, 2007] Härdle, W. K. and Hlávka, Z. (2007). *Multivariate Statistics: Exercises and Solutions*. Springer.

[Heckman and Meyers, 1983] Heckman, P. E. and Meyers, G. G. (1983). The calculation of aggregate loss distributions from claim severity and claim count distributions. *Proceedings of the Casualty Actuarial Society*, 70(133-134).

[Heilmann and Schröter, 2013] Heilmann, W.-R. and Schröter, K. J. (2013). *Grundbegriffe der Risikotheorie*. Verlag Versicherungswirtschaft.

[Klugman et al., 2008] Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2008). *Loss models: from data to decisions*. Wiley.

[Kunitomo and Yamamoto, 1985] Kunitomo, N. and Yamamoto, T. (1985). Properties of predictors in misspecified autoregressive time series models. *Journal of the American Statistical Association*, 80(392):941–950.

[Leoni, 2009] Leoni, G. (2009). *A first course in Sobolev spaces*. American Mathematical Society.

[Li and Ferrara, 2015] Li, Z. M. and Ferrara, P. G. (2015). Advances in common shock modeling. *Variance*, 9(1):20–35.

[Lippe et al., 1991] Lippe, S., Sedlmair, H., and Witting, T. (1991). Praxisrelevante Aspekte der Burning-Cost-Kalkulation. *Blätter der DGVFM*, 20(2):97–121.

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

[Mack, 1997] Mack, T. (1997). *Schadenversicherungsmathematik*. Verlag Versicherungswirtschaft.

[Mahler, 1998] Mahler, H. C. (1998). Credibility with shifting risk parameters, risk heterogeneity, and parameter uncertainty. *Proceedings of the Casualty Actuarial Society*, 85:455–653.

[Meyers, 2007] Meyers, G. G. (2007). The common shock model for correlated insurance losses. *Variance*, 1(1):40–52.

[Parodi, 2014] Parodi, P. (2014). *Pricing in General Insurance*. CRC Press.

[Pfeifer, 1997] Pfeifer, D. (1997). A statistical model to analyse natural catastrophe claims by means of record values. In *Proceedings of the 28th International Astin Colloquium*.

[Phillips, 1979] Phillips, P. C. (1979). The sampling distribution of forecasts from a first-order autoregression. *Journal of Econometrics*, 9(3):241–261.

[Rhoades, 1996] Rhoades, D. A. (1996). Estimation of the Gutenberg-Richter relation allowing for individual earthquake magnitude uncertainties. *Tectonophysics*, 258(1):71–83.

[Riegel, 2008] Riegel, U. (2008). Generalizations of common ILF models. *Blätter der DGVFM*, 29(1):45–71.

[Riegel, 2015] Riegel, U. (2015). A quantitative study of chain ladder based pricing approaches for long-tail quota shares. *Astin Bulletin*, 45(02):267–307.

[Scarrott and MacDonald, 2012] Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60.

[Schmutz and Doerr, 1998] Schmutz, M. and Doerr, R. R. (1998). *The Pareto Model in Property Reinsurance: Formulas and Applications*. Swiss Reinsurance Company.

[Scollnik, 2007] Scollnik, D. P. (2007). On composite lognormal-pareto models. *Scandinavian Actuarial Journal*, 2007(1):20–33.

[Stephan, 2011] Stephan, R. (2011). Measuring the impact of inflation on undiscounted loss reserves. In *ASTIN Colloquium Madrid*.

[Stine, 1987] Stine, R. A. (1987). Estimating properties of autoregressive forecasts. *Journal of the American Statistical Association*, 82(400):1072–1078.

[Sundt, 1987] Sundt, B. (1987). Credibility and old estimates. *Insurance: Mathematics and Economics*, 6(3):189–194.

[SwissRe, 2007] SwissRe (2007). *European Motor Markets*. Swiss Reinsurance Company.

[Wilkie, 1995] Wilkie, A. D. (1995). More on a stochastic asset model for actuarial use. *British Actuarial Journal*, 1(05):777–964.