Premium rating without losses – how to estimate the loss frequency of loss-free risks

Michael Fackler*

Abstract

In insurance and even more in reinsurance it occurs that about a risk you only know that it has suffered no losses in the past say seven years. Some of these risks are furthermore such particular or novel that there are no similar risks to infer the loss frequency from.

In this paper we propose a loss frequency estimator that copes with such situations, by just relying on the information coming from the risk itself: the "amended sample mean". It is derived from a number of reasonable mathematical first principles and turns out to have desirable statistical properties.

Some variants are possible, which enables insurers to align the method to their preferred business strategy, by trading off between low initial premiums for new business and moderate premium increases for renewal business after a loss.

We further give examples where it is possible to assess also the average loss, from some market or portfolio information, such that overall one has an estimator of the risk premium.

Keywords: Loss frequency, loss-free, sample mean, mean squared error, reinsurance

1 Introduction

Assume you have to assess the loss frequency of an insured risk (a single risk or a portfolio) that is not comparable to other risks, i.e. you cannot rate it with the help of data from other risks. In short, assume that to assess the loss frequency you have to rely on the loss record of the risk – nothing else is available.

Assume further that – fortunately – past losses predict future losses, i.e. there are no structural changes in the risk or the environment that make the loss history a priori unreliable for forecasting. So, it is in principle adequate to do *experience rating*, which usually means frequency/severity modelling with parameters estimated from the loss history ([Parodi, 2014a]).

Now assume that – unfortunately – in the observation period (the period of time for which relevant data about the risk is available) the risk was loss-free. Say in the past seven years no losses occurred (and data from the years before are unknown, unreliable, or not representative).

This situation may appear remote but is not. Certainly it is quite rare in the *Personal Lines* business where we can usually build large collectives of (quite) homogeneous risks providing plenty of loss data, however, *emerging lines* of business in their first years are usually so small that they hardly produce losses. In *Commercial* and *Industrial Lines* even among well-established insurance covers there are quite a lot of particular risks being so different from other ones that it does not seem adequate to assign them to any collective for rating purposes. When such risks moreover have high deductibles, they are likely to be loss-free for several years in a row. In the case of *non-proportional reinsurance* treaties, which are basically portfolios of risks with a very high first-loss retention, long loss-free periods are rather normal than exceptional.

In this paper we propose a loss frequency estimator that is able to handle such loss-free situations reasonably. It can be defined in a mathematically consistent manner and turns out to have desirable properties both in statistical and in business strategic sense.

Section 2 tells what many practitioners do in the described situation, which leads to several heuristical properties a loss frequency estimator should have. Section 3 is about the volume dependency of loss frequencies. Section 4 develops a novel class of loss frequency estimators and finds examples fulfilling the above heuristical requirements in a strictly mathematical sense. Section 5 gives practical applications for their use, illustrating situations where the average loss too can be can be assessed (from scarce additional information). Sections 6 and 7 calculate bias and mean squared error of the developed frequency estimators and find the optimal one in statistical sense. The numerical examples in Section 8 complete the picture.

*Dr Michael Fackler, DAV / independent consulting actuary / Munich, Germany / michael_fackler@web.de

2 What practitioners do

Typically we expect well-established rating methods to work in all situations we come across. But what if the loss record consists of: no losses? Then the most common frequency estimator, the sample mean, equals zero, as well as the Maximum Likelihood estimators in classical loss number models like Poisson, Binomial, and Negative Binomial (which in the first two models coincide anyway with the sample mean.)

A loss frequency result of zero is unacceptable – we must not provide insurance for free. Thus, we have to "select" a loss frequency greater than zero. Some pragmatic workarounds:

Additional loss: Some practitioners simply add a loss to the loss record. The idea is that this avoids zeros and at the same time in case of many losses the results are close to the sample mean. However, adding always 1 yields an estimator with a considerable positive bias, and maybe it is difficult to explain to an insured that had 7 losses in 7 years (an outcome that will be considered as not random at all) that he has to pay for 8 losses.

Additional year: Others adjoin a year with one loss, i.e. if they had seven loss-free years, they choose a loss frequency of $\frac{1}{8}$. That means assuming a bit less than one loss in the seven observed years. The question is how to rate the risk the year after if in the meantime a loss occurred. One could then (correctly in the sense of the sample mean) take 1 loss in 8 years, i.e. the same premium as before, but it would appear more natural (and would likely be accepted by the insured) if the premium was somewhat increased after the occurrence of a loss. So, one better keeps the additional year, thus assumes 2 losses in 9 years, etc. However, if one never drops the additional loss, this leads to values close to the first approach, hence again to a considerable positive bias.

Additional period: A cheaper variant is to adjoin not just a year, but a whole observation period of seven more years, i.e. instead of 0 losses in 7 years one takes 1 loss in 14 years. (Equivalently one can assume 0.5 losses in the 7 observed years.) The question is how to continue in the subsequent years. It does not seem reasonable to stay with such a long additional period, it should rather be shortened or dropped, but when and how?

There are certainly more variants out there, but thinking about the above three is sufficient to get quite some ideas about what first principles we feel a loss frequency estimator should fulfil:

One method: It should work in all situations, not just be a method for loss-free risks. For all possible outcomes it should be defined *in advance* what to do. This is not just a practical issue: Only thoroughly defined methods can be evaluated as for their statistical properties.

Justifiability: Many losses should yield a frequency equal or at least very close to the sample mean, which is the only result that in case of abundant loss experience (i.e. a low random error) can be explained to the insured.

Non-zero: All data situations must result in a strictly positive loss frequency.

Bias: The sample mean is unbiased. As in loss-free situations we must charge more than the sample mean and in situations with many losses we want to charge about as much as the sample mean, we are likely to get a positive bias. This is okay – the lesser evil than a negative bias – but we should possibly try to avoid a very high bias.

Monotonicity: Although few losses are likely to be a somewhat random outcome (which justifies a substantial uncertainty loading), they must be rated cheaper than many losses, which are less random (lower uncertainty). Lack of losses is a (statistical) item of information, despite of random errors: if the average loss frequency is say one loss per year, then seven years without a loss are possible but extremely unlikely. So, long loss-free periods do indicate a rather low frequency, although it is difficult to say how low.

Smoothness: The steps between the premiums for $0, 1, 2, 3, ...$ losses need to follow a smooth pattern, otherwise conflicts arise in the following classical situation: An insured has just suffered a loss, yet wants to renew the policy. He is indeed prepared to pay a higher premium, but the increase must appear "reasonable", which in this context essentially means: moderate.

Ratio: Everyone would agree that a risk having suffered 20 losses in 5 years (4 per year) should pay twice as much as a risk reporting 20 losses in 10 years (2 per year). Would it not be coherent to charge twice as much for 5 loss-free years than for 10 loss-free years?

3 Volume change

The above pragmatic considerations will turn out to be largely translatable into strict mathematics, but first we need to introduce an important item: the volume dependency of loss frequencies. Many large risks, above all portfolios, change their volume over time. (A volume in this sense could be e.g. the number of insured objects, vehicles, persons, etc.) In this case the loss frequency is not constant any more, but is typically assumed to be proportionate to the volume of the risk: If λ is the loss frequency and v the volume, we have $\lambda = v\theta$, where θ is the *frequency per volume unit*, which is assumed to be constant over time (unless there are structural changes).

Say the observation period consists of k years. For $i = 1, ..., k$ let be

- v_i the volume of the risk in year i,
- λ_i the corresponding loss frequency,
- N_i the corresponding number of losses (a random variable).

Then we have $\mathrm{E}(N_i) = \lambda_i = v_i\theta$. N_i is an unbiased estimator for λ_i, $\frac{N_i}{v_i}$ is an unbiased estimator for θ.

Now consider the observation period as a whole. Its volume, frequency, and loss number are the sums of the respective quantities of the single years: $v_+ = v_1 + ... + v_k$ and analogously for λ_+, N_+. As above we have $\mathrm{E}(N_+) = \lambda_+ = v_+\theta$. N_+ is an unbiased estimator for λ_+.

$\frac{N_+}{v_+}$ is a further unbiased estimator for θ. This is the already mentioned *sample mean* of the observation period and at the same time the volume-weighted average of the above estimators $\frac{N_i}{v_i}$.

If we now want to predict the outcome of a future year with (known or estimated) volume v_e, we need an estimator for its frequency λ_e. From $\lambda_e = v_e\theta$ we see that the product of v_e with any estimator for θ is an estimator for λ_e. To get an unbiased estimator, we can use any $\frac{v_e}{v_i}N_i$ or any weighted average thereof, in particular the (re-scaled) sample mean $\frac{v_e}{v_+}N_+ = \frac{N_+}{k_+}$.

Definition 3.1. Here we have introduced the *volume-weighted number of years* $k_+ := \frac{v_+}{v_e}$, which equals k in case all volumes are equal, including the future year.

One could think of it saying in a generalised sense: *We have N_+ losses in k_+ years.*

This well-known mathematics of volume-dependent frequencies (see e.g. [Mack, 1997], [Riegel, 2015]) is very easy. However, it shall be noted that in practice it may be difficult to determine an adequate volume measure, see Section 8.3 of [Parodi, 2014a]. Many readily available and well-established ways to quantify whether risks are small or large contain both the increase of the loss size (inflationary effects) and the increase of the loss frequency, e.g. the aggregate sum insured of Property accounts or the aggregate payroll in the case of certain Third Party Liability risks. If there is only such a volume measure ("exposure") available, one has to factor out the inflation, which can be difficult because inflation may be different from business to business, see Chapter 2 of [Fackler, 2017]. Note that this uncertainty is not particular to our rating problem, but arises independently of whether there is plenty of loss experience, or none at all, in all rating situations that require a separate modelling of frequency and severity.

Definition 3.2. Let us call exposures reflecting only the loss frequency *frequency volumes.*

Although such volume measures usually have lower yearly increases than those embracing inflation, it would be inadequate to assume that k and k_+ are anyway such close that calculating the latter is not worth the effort. The two can be surprisingly different, as the following example illustrates:

Example 3.3. Consider a steadily growing portfolio whose frequency volumes $v_1, ..., v_k$ constitute a geometric sequence: $v_{i+1} = v_i(1+s)$. Suppose loss reporting is such prompt that at the end of the year k we already know all losses having occurred in that year and can use them to rate the subsequent year. Then we have $v_e = v_{k+1}$ and from this we quickly get

$$k_+ = \frac{1 - (1+s)^{-k}}{s}$$

E.g., if we have a yearly increase of $s = 10\%$, then $k = 7$ years mean $k_+ = 4.9$ years. If the yearly increase is 20%, then $k = 10$ years mean $k_+ = 4.2$ years, while 20 years are 4.9 weighted years. In practice, all periods of steady and rapid growth one day come to an end, but as long as they last, k_+ will be much smaller than k. It is even bounded – from the above formula we see that k_+ must be smaller than $\frac{1}{s}$. What in practice lets further increase the difference between k and k_+ is the usual delay in loss reporting.

It would not be untypical that the data of the year k were not complete until late in the subsequent year, i.e. it can be used at the earliest to rate the year $k+2$, hence in the above situation we would have $v_e = v_{k+2}$, which leads to values for k_+ being lower by the factor $\frac{1}{1+s}$.

Having discussed the denominator k_+ of the sample mean, now we turn to the numerator N_+.

4 Designing the estimator

4.1 Basic structure

The idea is to "amend" the empirical loss number N_+ in a way to fulfil as many as possible of the properties we have collected in Section 2. We try the following ansatz:

Create a new frequency estimator by replacing N_+ in the sample mean formula by $g(N_+)$ with a suitable function $g(n)$ of nonnegative integers.

Definition 4.1. We call the estimator $\frac{g(N_+)}{k_+}$ the *amended sample mean (ASM)* and the rating method applying this estimator the *ASM method*. The function (or sequence) g is called *amending function*.

Now we can translate the requirements for the frequency estimator into criteria for the amending function, at least preliminarily.

Definition 4.2. (Provisional definition:) We call an amending function *admissible* if it satisfies:

$g(n)$ is defined for all $n = 0, 1, 2, ...$	*one method*
$g(n) = n$ for large n	*justifiable*
$g(n) > 0$	*non-zero*
$g(n) \geq n$, but not much greater	positive but moderate *bias*
$g(n+1) > g(n)$	*monotonic*
$g(n+1)/g(n)$ takes on "reasonable" values	*smooth*

The last requirement (ratio) can be dropped as it turns out to be an immediate consequence of the ansatz we have chosen: Suppose we have two loss-free situations with 5 and 10 observed years, respectively. Suppose the volumes are constant (otherwise it does not make sense to require the desired ratio). Then k_+ equals 5 and 10, respectively. The resulting estimate for λ_e is $\frac{g(0)}{5}$ in the first case, which is twice as much as $\frac{g(0)}{10}$.

Note that all conditions but the sixth and partly the fourth are mathematically strict. For a clear definition it remains to get a precise notion of what smoothness and moderate bias mean.

4.2 Credibility-like?

As mentioned at the beginning, we cannot rate the risk with the help of data from other risks, therefore it is in particular impossible to apply Bayesian/Credibility rating. But, as this is such a strong method for risks having not enough data for being rated independently, we want to think a moment about whether we could find a Credibility-like frequency estimator for our situation. That would be a formula looking like a Credibility premium for one of the risks of a collective, where the parameters that ideally would be estimated from the whole collective (which we do not have) are "selected" in a suitable way. Some questions arise: Is such a Credibility-like formula an admissible ASM? If not, is it similar? Is it a better approach?

Let us look at the classical Bühlmann-Straub model for (Poisson distributed) loss frequencies, see Section 4.10 of [Bühlmann and Gisler, 2005]. Translated to our notation, in this Credibility model the empirical loss count N_+, which we replace by $g(N_+)$, would be replaced by a linear combination

$$wN_+ + (1-w)\,v_+\theta_0$$

where $\theta_0 > 0$ is the "global" frequency per volume unit of the collective and the weight w has the structure

$$w = \frac{v_+}{v_+ + const.}$$

This is indeed a very smooth-looking, non-zero and monotonic amending function, which misses justifiability but not by much. However, this and other Bayes methods are optimised for the *collective*, not for the single risk. In particular, the Credibility estimator is unbiased on collective level, but on the individual level of our risk there is a bias, namely $(1-w)\, v_+ \left[\theta_0 - \theta\right]$. In order to be on the safe side for our individual risk, we would have to choose a very high θ_0, but then we are likely to get a bias being too high to be acceptable.

In short, Bayes/Credibility does not seem to be the best option in our situation where we could "imagine" a surrounding collective but do not have it, and have no idea what value a global θ_0 would have. We better try to optimise the estimator of our particular risk without relating it other risks. So, let us continue with our search for good amending functions. We have already found two candidates, see the table:

n	0	1	2	3	4
$g_1(n)$	1	2	3	4	5
$g_2(n)$	0.5	1	2	3	4

g_1 is the workaround "additional loss" from Section 2. g_2 is "additional period" if we adjoin the period only in the loss-free case and use the sample mean in case of one or more losses.

Notice that "additional year" is not a candidate, as it yields the formula

$$\frac{N_+ + 1}{k_+ + 1}$$

which is Credibility-like as described above with weight $w = \frac{v_+}{v_+ + v_e}$ and global frequency $\theta_0 = \frac{1}{v_e}$:

$$\frac{N_+ + 1}{k_+ + 1} = \frac{k_+}{k_+ + 1}\frac{N_+}{k_+} + \frac{1}{k_+ + 1}1 = \frac{1}{k_+}\left(\frac{v_+}{v_+ + v_e}N_+ + \left(1 - \frac{v_+}{v_+ + v_e}\right)v_+\frac{1}{v_e}\right)$$

As stated earlier, g_1 is too expensive for large n and fails the criterion justifiability. g_2 might intuitively appear very cheap, but we see at a glance that it meets all those requirements that are already defined precisely. It remains to quantify its bias, but let us first get a clearer concept of smoothness.

4.3 Smoothness

Consider the following typical situation: A risk that suffered n losses in k_+ years was rated $\frac{g(n)}{k_+}$. We write the risk, then a new loss occurs. To rate the risk for the next renewal of the insurance cover, note that we now have $n+1$ losses in $(k+1)_+$ years, which leads to the frequency

$$\frac{g(n+1)}{(k+1)_+}$$

Especially when one has many observation years and/or rapid volume growth, k_+ and $(k+1)_+$ are close, such that the relative change in the premium is close to $\frac{g(n+1)}{g(n)}$. This latter ratio can thus be used as a benchmark for relative premium increases after one new loss. The following properties come into mind that should ideally be fulfilled to avoid anger among the insureds:

(1)	$\frac{g(n+1)}{g(n)} \approx \frac{n+1}{n}$	for $n > 0$
(2)	$\frac{g(n+2)}{g(n+1)} \leq \frac{g(n+1)}{g(n)}$	
(3a)	$\frac{g(n+1)}{g(n)} \leq \frac{n+1}{n}$	for $n > 0$
(3b)	$\frac{g(1)}{g(0)} \leq 2$	

Interpretation:

(1) The premium increases are similar to the loss record increases.

(2) The more losses we have already had, the lesser the relative impact of a new loss on the premium.

(3a) The premium increases are not greater than the loss record increases. We could say that in a way $g(n)$ is even smoother than the sample mean, raising less steeply.

(3b) If a loss occurs, the premium might double but should not increase more. (For $n > 0$ this is ensured by (3a), where the right hand side cannot exceed 2; for $n = 0$ we need the extra condition.)

4.4 Wrap up

Recall that for an amending function to be admissible, we require $g(n) = n$ for large n.

Definition 4.3. We call the integer d where an amending function g starts to equal the identity the *dimension* of g.

In this sense g_1 has infinite dimension, while admissible amending functions have a finite dimension d, thus are determined by the d values $g(0), ..., g(d-1)$. The smoothness condition (3a) together with "positive bias" yield the inequality $\frac{n+1}{g(n)} \leq \frac{g(n+1)}{g(n)} \leq \frac{n+1}{n}$, $n > 0$, which shows that as soon as we have $g(n) = n$ for an integer $n > 0$, the same equation holds for $n+1$, hence for $n+2, \ldots$, i.e. for all larger integers. Thus, for an admissible amending function with dimension d we have $g(n) = n$ for all $n \geq d$ and $g(n) > n$ for all $n < d$.

To see whether there is a chance to fulfil the properties (1), (2), and (3), together with the requirements developed earlier, let us check the lowest dimensions.

Dimension 1: $g(n) = n$ for all $n \geq 1$.

We only have to find $g(0)$. (2) means $g(0) \leq 0.5$, while (3b) means $g(0) \geq 0.5$. Thus, we have a unique solution $g(0) = 0.5$ yielding the already known g_2. All conditions are fulfilled (bearing in mind that we should do a closer analysis of the bias).

Dimension 2: $g(n) = n$ for all $n \geq 2$.

We have to find $g(0)$ and $g(1)$. (3) yields $g(0) \geq \frac{g(1)}{2}$ and $g(1) \geq 1$, but the second inequality is already known (positive bias). Both inequalities together yield $g(0) \geq \frac{1}{2}$. (2) applied twice yields $\frac{3}{2} \leq \frac{2}{g(1)} \leq \frac{g(1)}{g(0)}$. The left inequality yields $g(1) \leq \frac{4}{3}$, so altogether we have $1 \leq g(1) \leq \frac{4}{3}$. The right inequality yields $g(0) \leq \frac{g(1)^2}{2}$, so altogether we have $\frac{g(1)}{2} \leq g(0) \leq \frac{g(1)^2}{2}$. There are (infinitely) many solutions:

- The cheapest choice for $g(1)$ is 1, which leads back to the one-dimensional g_2.
- The largest choice for $g(1)$ is $\frac{4}{3} = 1.333$. Then we have $\frac{2}{3} \leq g(0) \leq \frac{8}{9}$ with the most expensive variant $g(0) = \frac{8}{9} = 0.889$, which we call g_3. The first four values of g_3 constitute a *geometric sequence* where the ratio of subsequent elements equals $\frac{3}{2}$, which means that when one has 0, 1, or 2 losses, then a new loss triggers a premium increase of about 50%.
- We get an intermediate and in a way very smooth variant if we let the first five values of $g(n)$ constitute a *geometric sequence of second order*. From the values $g(2)$, $g(3)$, and $g(4)$ we see that the ratio of second order $\frac{g(n+2)}{g(n+1)} / \frac{g(n+1)}{g(n)}$ of this sequence equals $8/9$, which leads to the first two elements $g(1) = \frac{32}{27} = 1.185$ and $g(0) = \frac{4096}{6561} = 0.624$. Here the ratio of subsequent elements decreases slowly (by the factor $\frac{8}{9}$). It is easy to verify that all conditions are fulfilled. We call this function g_4.

If we look at *higher* dimensions d, we can see easily that in all cases (3) determines the same *lower bound* for the values of g: the one-dimensional function g_2.

(2) instead yields an *upper bound* which between 0 and d constitutes a geometric sequence with ratio $\frac{d+1}{d}$. E.g., for $d = 3$ the function starts with the values $g(0) = \frac{81}{64} = 1.266$, $g(1) = \frac{27}{16} = 1.688$, $g(2) = \frac{9}{4} = 2.25$, $g(3) = 3$. Here the ratio of subsequent elements is $\frac{4}{3}$, which means that a new loss leads to a premium increase of about 33%. Note that $g(0) > 1$, i.e. in case of zero losses this variant sets more than one loss. We call it g_5.

The analogous upper bounds of higher dimensions d yield even lower premium increases, but the initial values become very high: one calculates easily $g(0) = d\left(1 + \frac{1}{d}\right)^{-d} > \frac{d}{e}$ with the Euler number $e = 2.718$. For large d these functions will fail condition (1) and anyway have an unacceptably large bias. To obtain admissible amending functions of higher dimensions, one must look for cheaper variants, maybe geometric sequences of higher order like g_4.

It is clear that there is a trade-off between *bias* and *smoothness*: cheap amending functions must have low initial values but then increase sharply. Functions increasing very moderately from the beginning must in turn start at a more expensive level.

The following table unites the candidates discussed so far. For a better orientation we leave $g_1(n) = n + 1$ in the overview although it does not meet all criteria. Note that it only fails "moderate bias" and "justifiability" and is indeed very smooth.

n		0	1	2	3	4
$g_1(n)$	$g_{+1}(n)$	1	2	3	4	5
$g_2(n)$	$g_{min}(n)$	0.5	1	2	3	4
$g_3(n)$	$g_{max2}(n)$	0.889	1.333	2	3	4
$g_4(n)$	$g_{so2}(n)$	0.624	1.185	2	3	4
$g_5(n)$	$g_{max3}(n)$	1.266	1.688	2.25	3	4
$g_6(n)$	$g_{so3}(n)$	0.859	1.390	2.109	3	4

We have added a further variant: the first 6 elements of g_6 constitute a geometric sequence of second order. It is close to g_3, a bit cheaper for loss-free risks, but more expensive in case of 1 or 2 losses. The surcharges after a new loss decrease very smoothly: 62%, 52%, 42%, 33%, ..., which could be preferred over constant increases at the beginning of the sequence, as is the case with g_2 (2x 100%), g_3 (3x 50%), and g_5 (4x 33%).

Further we have (in the second column) given the six amending functions alternative labels that compactly describe their properties. The first label obviously means "plus one loss", the second "minimum g". g_3 and g_5 are renamed "maximum g" of dimension 2 and 3, respectively, while g_4 and g_6 are the amending functions starting as a geometric sequence of "second order" and having dimension 2 and 3, respectively.

4.5 IBNR

In some business lines, most markedly in any kind of Third Party Liability ("long-tail") business, "late damage" occurs – losses that are reported with a a delay of several months or even years. This is part of the run-off problem, albeit often the minor part compared to the run-off of the case reserves. But, if we want to apply the ASM method to such risks, we must take the run-off of the loss numbers into account, which is often called *(true) IBNR*: incurred but not (yet) reported. This is in principle straightforward reserving calculus, but as we cannot estimate any run-off from our data, we have to work with market experience about the run-off of loss numbers in the respective line of business, which may not easily be available: We need the so-called *lag factors*, the inverses of the so-called *age-to-ultimate factors*. (These factors are popular for loss amounts, but work analogously for loss numbers.) If we consider the losses that a risk produces in a certain year, the lag factor a_j describes the percentage thereof that is reported (on average) until j years later. The a_j constitute an increasing sequence converging to 100% (more or less rapidly according to the kind of business, the country, deductibles, etc.), they are discrete points on the so-called *delay distribution* (see [Parodi, 2014b] for how it can be estimated and applied).

Suppose we have an observation period of k years with frequency volumes $v_1, ..., v_k$. Say the rating is done late in the year after the year v_k. Hence year i has a time-lag of $k + 1 - i$, i.e. the percentage of losses produced by year i which are known at the moment of the rating equals (on average) a_{k+1-i}. (If the rating is done some months earlier or later than assumed above, one must choose slightly different lag factors to reflect exactly the resulting time-lag. If (very) lucky, lag factors on a monthly basis are available or even perhaps a continuous parametric delay distribution, otherwise interpolations might be necessary.)

When calculating empirical frequencies (per volume unit) on good data, to account for the losses incurred in the year i but not yet reported at the moment of the rating, one would gross up the empirical loss count N_i by the age-to-ultimate factor, or equivalently multiply the volume by the lag-factor:

$$\frac{N_i\,(1/a_{k+1-i})}{v_i} = \frac{N_i}{a_{k+1-i}v_i}$$

The IBNR-adjusted sample mean is the weighted average of these estimators

$$\frac{N_+}{\sum_{i=1}^k a_{k+1-i}v_i} = \frac{\sum_{i=1}^k N_i}{\sum_{i=1}^k a_{k+1-i}v_i} = \sum_{i=1}^k \left(\frac{a_{k+1-i}v_i}{\sum_{l=1}^k a_{k+1-l}v_l}\right)\frac{N_i}{a_{k+1-i}v_i}$$

where the weights are notably not proportional to the v_i, but to the $a_{k+1-i}v_i$, which can interpreted as reflecting the part of the risk in the year i for which the losses are known at the moment when the rating is done. This weighting of the yearly empirical frequencies is not just heuristics, but optimises the variance in the most common model where variance is proportional to volume: If U_i counts the ultimate losses of the year i, the losses out of these being known at the moment of the rating equal

$$N_i = \sum_{j=1}^{U_i} B_{ij}$$

where $B_{ij} \sim \mathrm{Bin}\,(1, a_{k+1-i})$ is a Bernoulli distributed random variable describing whether or not the j-th loss occurring in the year i has already been reported at the moment of the rating. So, if U_i is Poisson distributed with mean and variance $v_i\theta$, one sees quickly that $N_i \sim \mathrm{Poi}\,(a_{k+1-i}v_i\theta)$.

The IBNR-adjusted sample mean can be adapted to scarce-data situations in a straightforward manner, by using $g\,(N_+)$ instead of N_+. The resulting

$$\frac{g\,(N_+)}{\sum_{i=1}^{k} a_{k+1-i}v_i}$$

is the amended sample mean analogously to the case without run-off issues. Thus, ASM rating with IBNR works just the same, by simply replacing the volumes v_i by the reduced volumes $a_{k+1-i}v_i$.

5 Premium rating

Before the (quite technical) discussion of the bias, let us illustrate the ASM method by calculating some examples, which are inspired from real-life pricing cases where ASM variants were successfully used.

5.1 Assessing the average loss

The second ingredient of premium rating via frequency/severity modelling is the average loss, which, of course, cannot be assessed from a loss-free history. However, a few observed losses would not change this: loss severities are typically skewed, such that estimates for the average loss, whether direct (empirical mean) or indirect (via a parametric fit of the severity) are volatile unless you have more than say fifty losses, see [Brazauskas and Kleefeld, 2009] and related papers of the first author, exploring old and novel (more robust) methods to fit the GPD and other models in case of scarce data. So, no matter whether one has a few observed losses or none at all, one needs additional knowledge to assess the average loss. Let us present three cases where this can be done.

- For a very *restricted insurance capacity*, say a first-loss cover with a rather low limit per loss, one can use this limit as upper bound for the average loss. This is of course a very pessimistic estimate, but in case of a low limit it can work well, yielding premiums that don't appear excessively conservative. Note that here no additional information is required.

- For *layers* protecting losses *in the million Euro range*, as they are common in reinsurance and industrial insurance, the industry has gathered decades of world-wide market experience, at least in large lines of business. The observed loss size distributions of such layers are typically heavy-tailed (having in particular a slowly falling density) and can often be modelled fairly by single-parameter Pareto distributions
$$\mathrm{P}\,(X \le x) = 1 - \left(\frac{x_0}{x}\right)^{\alpha}$$
where the observed Pareto alphas vary across the business lines, but much less so within the lines. So practitioners know typical "market" alphas for certain covers, which can complement or replace a parameter estimate on the specific loss history ([Schmutz and Doerr, 1998], [FINMA, 2006]). The examples in the following subsection will illustrate this approach.

- Assessing the average loss is much harder for *ground-up business*, i.e. business covering all losses from the very first Dollar (or after some low deductible) up to a maximum capacity. Apart from the case of very limited capacity treated above, the typical situation is that losses can be as high as several millions of Euros and for the average loss anything between say some hundred Euro and half a million Euro is thinkable. This is because for the loss severity distribution of ground-up business very different shapes can be observed: from similar to a Gaussian (albeit slightly skewed) to heavy-tailed from the first Dollar, see the examples assembled in [Fackler, 2013]. So, there is no handy market experience like the Pareto-like geometries with typical alphas as one finds them in reinsurance layers.

However, there is a special case that can be assessed with not too many calculations and not too many assumptions, namely portfolios covering a large number of *units* (meaning insured objects / persons / activities) which are similar in coverage but quite heterogeneous in size, such that the maximum loss potential stems from a few very large units, while the vast majority of units is much smaller in size, say fifty times smaller. The heuristics about this situation is: if only a few units are large, unless these produce the majority of losses, the average loss must be very small compared to the maximum loss, say in the range of thirty times smaller. We will illustrate this approach, which requires quite detailed portfolio information, in the subsection after next.

In all examples we will use the ASM $g = g_3 = g_{max2}$; the calculation with other variants is analogous.

Let us further emphasise that, just like in traditional experience rating on better data (see Chapter 10 of [Parodi, 2014a]), the volumes and losses (if any) used in ASM rating must be *as if*, i.e. adjusted to the terms of the future year by taking into account inflation and other things changing over time.

5.2 Layers of (re)insurance

We assemble some layer terminology, borrowing notation and basic results from [Riegel, 2018].

Definition 5.1. A (re)insurance layer $c \,\mathrm{xs}\, a$ (c *in excess of* a) pays the part of each loss X that exceeds the *attachment point* (also called *retention*, *deductible*) $a \geq 0$, up to a maximum (*cover*) of $c > 0$, which mathematically means paying: $\min\left((X-a)^+, c\right)$.

If e is the risk premium of the layer, we call $r := e/c$ the *risk rate on line* (*RRoL*).

If f is the expected frequency of losses exceeding a (*layer losses*), and h the expected frequency of losses reaching or exceeding the layer capacity $a + c$, we must have $f \geq r \geq h$.

The risk rate on line can be seen as a kind of average loss frequency across the layer area $[a, a+c]$. For reinsurance practitioners it is the key figure showing at a glance how frequently a layer is affected by losses. It usually receives much more attention than the frequencies f and h, firstly because the RRoL immediately yields the risk premium, secondly because severity distributions in reinsurance layers are typically heavy-tailed, such that, unless a layer is very long, the quantities f, r, h are anyway of similar range.

5.2.1 Catastrophe reinsurance ("NatCat")

Consider a catastrophe excess of loss reinsurance treaty $100 \,\mathrm{xs}\, 50$ (say million US Dollar) covering accumulation losses from natural disasters. An accumulation loss is the aggregate of all single losses caused by the same natural event.

There are very sophisticated models for the rating of NatCat reinsurance business, but they do not cover all natural perils everywhere in the world. Assume we reinsure say earthquake, flood, windstorm, and hail in an exotic country for which such models are not available, such that we can only rely on the loss experience of the portfolio itself.

Say the portfolio has been loss-free in the past 10 years (to be precise: loss-free "as if", having taken into account inflation and portfolio growth). About the time before we either have no data or do not want to use them because the portfolio was substantially different then.

Accumulation losses are particular in that the loss frequency is not affected by changes in size of the portfolio. If a portfolio grows, it does not suffer more natural disasters: it suffers bigger accumulation losses but not more. Here the adequate frequency volume is a constant, hence $k_+ = k$.

We have $k_+ = 10$, so we get the frequency $\frac{g(0)}{10} = \frac{0.889}{10} = 8.89\%$.

To get the risk premium, we need an assumption about the average loss of the layer.

- We could go the most prudent way and assume that all losses are total losses. Then we would get a risk premium of $100 \cdot 8.89\% = 8.89$.

- Alternatively, we can use internationally established market Pareto alphas for NatCat reinsurance, which are close to 1. To be on the safe side, we choose $\alpha = 0.8$, a very heavy-tailed distribution yielding an average loss of 61.43 in the layer. This leads to a risk premium of 5.46.

Now assume that a loss occurs and the treaty must be rated for renewal. Then we have one loss in $k_+ = 11$ years, so for the frequency we rate $\frac{g(1)}{11} = \frac{1.333}{11} = 12.12\%$, which is an increase of 36% compared to the year before. This is a notable surcharge, but for the high uncertainty inherent in catastrophe reinsurance it is often possible to enforce such increases after a loss.

The risk premium now would be 12.12 if we assumed total losses only, and is 7.45 if we use the above Pareto model.

This is not a quick and cheap rating. If one talks to practitioners, many would confirm that if they, in the above rating situation, had to come up with a completely judgemental risk premium, they would probably charge more than what we have calculated here, maybe something in the range of 10 (which means 10% RoL) before the loss and substantially more after. Here with the ASM method we are able to be more technical and at the same time cheaper.

5.2.2 Fire per risk cover

Consider a *per risk* excess of loss reinsurance treaty 4 xs 1 (say million Euro) covering a Fire portfolio. Per risk means that, unlike the above example, this cover applies to single losses (affecting single objects in the portfolio), not to accumulation losses from catastrophes. Here volume change does affect the loss frequency.

You got information about three years only and the frequency volume grows rapidly: $v_1 = 0.8$, $v_2 = 1.0$, $v_3 = 1.2$, and $v_e = 1.35$. Then we have $v_+ = 3.0$ and $k_+ = 2.222$.

The list of the largest losses ("as if", having taken inflation into account) in these three years is as follows: 4.5; 0.1; 0.1.

We see that there was one huge loss, while all other losses were far below the retention. It seems hopeless having to rate a treaty with such poor data from such a short observation period. But, if we have knowledge about the loss size distribution of this kind of portfolio, we can give it a try: Market Pareto alphas for Fire per risk layers are in the range of 2, maybe rather of 1.5 for Industrial Fire portfolios. This market experience is based on layers in the million Euro or US Dollar range, starting at say half a million.

Unless the portfolio is mainly Industrial Fire, assuming a Pareto distribution with $\alpha = 1.3$ should be well on the safe side. We could apply this distribution to any layer having a retention of about 0.5 or higher. To be even more on the safe side, we choose the somewhat higher *model threshold* 0.6 and proceed as if we had to rate two layers, the actual 4 xs 1 and an artificial lower layer 0.4 xs 0.6:

- We first assess the frequency of the latter, which suffered one loss: $\frac{g(1)}{2.222} = \frac{1.333}{2.222} = 60.0\%$.
- We do not need the average loss of this layer, but the frequency of total losses exceeding 1, which is at the same time the loss frequency of the layer 4 xs 1. In the Pareto model this is a very easy calculation called Pareto extrapolation ([Schmutz and Doerr, 1998]). We simply have to multiply the loss frequency of the lower layer by the factor $\left(\frac{0.6}{1}\right)^{\alpha}$ and get 30.9%.
- The average loss of the layer 4 xs 1 according to the chosen Pareto model is 1.277.
- Altogether we get a risk premium of 0.394, which means about 10% RoL.

Probably many practitioners would charge much more – there was a recent huge loss and we have extremely few years of data.

However, one should bear in mind that a high premium after a bad year is often a rather *commercial* than *technical* premium: one charges more because in this situation it is possible to enforce high premiums and recoup some money quickly. The assessment of the technical premium should be free from such considerations. Instead, one should try to distinguish which elements of the loss record are much affected by random effects and which not.

Here the exact size of the biggest loss is clearly very random – recall we have a heavy-tailed loss size distribution. But the total absence of other losses in the million Euro range, and far below, is unlikely to be just a random error. We took this into account by modelling the frequency at the threshold 0.6. The rest of the calculation depends (in an admittedly sensitive manner) on the correctness of the selected loss size distribution, but here it was definitely tried to be on the safe side. For comparison: had we applied the very pessimistic $\alpha = 1$, we would have got a risk premium of 0.579, i.e. about 14% RoL – still less than what quite some practitioners would rate.

Be aware that if a new loss of say 0.8 occurs, one has to increase the technical premium at renewal, as the loss exceeds the chosen threshold 0.6 (i.e. affects the artificial layer). The surcharge might be difficult to explain to the reinsured who will argue that the loss did not hit the layer $4\,xs\,1$. But it is imperative to be coherent. If the model specifications are changed (say to a new, loss-free, model threshold of 0.85) every time the model leads to an increase being hard to enforce, one is most likely to loose money in the long run.

Note how extremely scarce the information was that this ASM rating required: We did not use the exact loss sizes, but just the information that there was a very big loss and all others were far from hitting the layer. What we instead needed precisely was the frequency volume of the years of the observation period. In reinsurance pricing this is mostly calculated from the aggregate primary insurance premium of the reinsured portfolio by factoring out inflation and premium cycles – a sometimes difficult task (explored in [Riegel, 2015] and Chapters 2 and 8 of [Fackler, 2017]), which is, however, required for other reinsurance premium rating methods as well.

Note that layers like the one described here are common in Industrial direct insurance too (for a multitude of examples see [Parodi, 2014a]), so for such risks ASM rating is in principle applicable. An adequate volume measure could be the aggregate sum insured, adjusted for inflation.

Lines of business other than Fire can be rated in the same way, as long as we have good market knowledge about tails of loss size distributions.

5.3 Ground-up business, or: How to use a wrong tariff

Consider a portfolio or a large complex commercial/industrial risk that is (re)insured from the ground up, i.e. with no or only small deductibles. Large portfolios usually produce many enough losses for classical experience rating, but single risks and small portfolios (e.g. Special Lines, emerging or newly started business) can be loss-free over some years, which makes them candidates for ASM rating. As stated above, ground-up business appears much more heterogeneous than layer business, such that we cannot expect easy rules like market Pareto alphas. On the other hand, for such risks the (re)insurer may get quite granular information about how the portfolio/risk is composed, namely whether the single insured *units* (objects, persons, vehicles, ...) are large or small with respect to a measure of size indicating the maximum loss potential, e.g.: sum insured (Property lines, Personal Accident, some Third Party Liability business), PML/EML/MPL (large Property business), insured value (Ocean Hull), etc. This often comes with some qualitative info about relevant characteristics of the units, which may be clustered into groups according to size and characteristics (risk profile), or reported each separately (bordero).

Let us explain how such information can help to roughly assess the average loss. As an example see the following bordero, where each row represents a risk or a number of risks having equal characteristics. (Suppose the bordero has been adjusted for inflation etc., representing as if the future year.)

No.	size in Euro	rate ‰
1	1’000’000	0.70
1	800’000	1.00
2	500’000	0.80
3	200’000	0.90
5	100’000	1.50
75	50’000	1.12
165	30’000	1.20
490	20’000	1.25

The first and second column show that this portfolio is very heterogeneous as for size, having only a few units in the million Euro range, while the majority is fifty times smaller. If we knew that the large units don’t produce more losses on average than the small ones, we could infer that the average loss is dominated by the many small units and must be in the range of some ten thousand Euro, if not lower.

How could we manage to verify this? Look at the third column, which displays the (gross) premium rate in per mil of the size. (Note that in case of Property PMLs this rate deviates from the usual premium rate, which is always related to the sum insured.)

If we have been provided with such premium rates (or equivalently the premiums) and if we believe that the given premiums are reasonable, the rating is essentially done: one can infer the risk premium from the given gross premium with low uncertainty. But, this is not the situation we aim to address here. Instead, we want to study cases where we have “found” or been given premium rates, but don’t trust them too much, thus want to use them only as a vague indication, e.g. in one of the following situations:

- The reinsurer receives from the insurer a bordero with the premiums actually charged by the latter, but suspects the overall premium level of the insurer to be inadequate, i.e. heavily underpriced or overpriced.

- The (re)insurer receives from the (re)insured a bordero without premiums, but with other information enabling us to assign premium rates to the units, by using a tariff from “similar” business.

The bottom line in both cases is the following: We have premium rates from a tariff (or alike) and feel that this tariff discerns fairly well between good and bad risks (assigning accordingly low/high premium rates), but we have doubts about the overall level of the tariff. Our goal is to use this tariff only *up to a factor*, to be precise:

- use the tariff only to assess the average loss,
- use the empirical loss count (via ASM rating) to assess the frequency.

Notice that this is not a Credibility approach. It seems straightforward to apply Credibility by using the tariff premium as the global premium, but this is only adequate if one believes that the tariff overall yields a reasonable premium level. Instead it will turn out that we are able to work with much weaker (albeit a bit unusual) assumptions, which would lead to the same result if we replaced our tariff by one having say three times higher or lower premium rates.

We need some notation. For a single unit j we consider the following quantities:

S_j	size	
G_j	gross premium	
g_j	gross premium rate	$g_j = G_j/S_j$
R_j	risk premium	
f_j	loss frequency	
L_j	average loss	$R_j = f_j L_j$
l_j	average loss degree	$l_j = L_j/S_j$
q_j	loss ratio	$q_j = R_j/G_j$
w_j	auxiliary quantity	$w_j := f_j/g_j$

In business lines insuring units of variable size it is usually possible to split the losses into a (usually major) part depending strongly (albeit not always proportionally) on unit size and a rather independent part (e.g. certain legal expenses). Accordingly, one gets a split of the average loss into a "constant" and a "variable" component:

$$L_j = {}^cL_j + {}^vL_j = {}^cL_j + {}^vl_j S_j$$

The idea behind this split is that, while the sizes S_j may vary a lot across units, the average constant loss ${}^cL_j$ and the average variable loss degree ${}^vl_j = {}^vL_j/S_j$ should usually vary much less and thus be easier to assess.

Definition 5.2. For a finite set of real figures u_j und corresponding weights $a_j \geq 0$, we write $\overline{u}$ for the ordinary arithmetic mean, while for the weighted average we write

$${}^a\overline{u} := \frac{\sum a_j u_j}{\sum a_j}$$

Let us calculate an upper bound for the risk premium, which can be used as a prudent estimate. In the following the sums run over the units j and are sometimes indicated briefly by the subscript Σ.

$$R_\Sigma = \sum R_j = \sum f_j L_j = \sum f_j {}^cL_j + \sum f_j {}^vl_j S_j \leq$$
$$\leq {}^cL_{max} \sum f_j + {}^vl_{max} \sum f_j S_j = f_\Sigma \left({}^cL_{max} + {}^vl_{max} \frac{\sum f_j S_j}{\sum f_j} \right) = f_\Sigma \left({}^cL_{max} + {}^vl_{max} {}^f\overline{S} \right)$$

The final term is the product of the loss frequency of the whole portfolio/risk and an upper bound for the average loss. As for the pieces of this term,

- f_Σ can be estimated (via the ASM method),
- ${}^cL_{max}$ and ${}^vl_{max}$ are not known, but in many situations reasonable prudent estimates should be possible, say Euro 4′000 and 25%, respectively,
- ${}^f\overline{S}$ is completely unknown. This is the *frequency-weighted* average of the insured values, but its weights, the single frequencies, are not known.

Notice that, however, another weighted average of the insured values can be calculated from the given data, the *gross-premium-rate-weighted* average:

$$^{g}\overline{S} = \frac{\sum g_j S_j}{\sum g_j} = \frac{\sum G_j}{\sum g_j}$$

How are these averages related? Noting $f_j S_j = w_j G_j$ we calculate

$$\frac{^{f}\overline{S}}{^{g}\overline{S}} = \frac{\sum f_j S_j}{\sum f_j} \frac{\sum g_j}{\sum g_j S_j} = \frac{\sum w_j G_j}{\sum w_j g_j} \frac{\sum g_i}{\sum G_i} = \frac{\sum G_j w_j}{\sum G_j} \frac{\sum g_j}{\sum g_j w_j} = \frac{^{G}\overline{w}}{^{g}\overline{w}}$$

The final term looks promising, although the auxiliary quantities w_j are unknown. They are averaged in two different ways. Can the results be very different?

To see first how much the w_j themselves may vary across the units, note that $w_j = f_j/g_j = q_j/l_j$. The loss ratios q_j should hardly vary, provided that the tariff discerns fairly well between good and bad risks: then the loss ratios of large units can be expected to be somewhat higher (due to a lower percentage of administration expenses and more power to negotiate low premiums), but differences should be rather small. The $l_j = {}^{c}L_j/S_j + {}^{v}l_j$ certainly vary more, yielding possibly smaller values for larger units: this is plausible for both summands, although huge variation shouldn't be the norm for the second one. As this latter will mostly dominate, we can expect the l_j to vary rather moderately, such that overall the w_j shouldn't vary too much, certainly much less than the unit's sizes S_j. However, overall it is plausible that larger units have somewhat larger w_j.

Let us look now at the weights of the two averages. The G_j are usually much larger than average for large units, while the g_j will mostly be more balanced. Thus, $^{G}\overline{w}$ should usually be the larger average. But, with the w_i being fairly homogeneous, it is hard to imagine that the two averages be extremely far from each other. Of course, assuming $^{G}\overline{w} \approx {}^{g}\overline{w}$ would carry approximations too far, but in many situations it should be fair to assume

$$\frac{^{f}\overline{S}}{^{g}\overline{S}} = \frac{^{G}\overline{w}}{^{g}\overline{w}} \leq C$$

with a prudently chosen constant, say $C = 5$. Overall we get as upper bound for the risk premium

$$R_\Sigma \leq f_\Sigma \left({}^{c}L_{max} + {}^{v}l_{max} C \, {}^{g}\overline{S}\right)$$

where f_Σ is estimated from the loss history, $^{g}\overline{S}$ is calculated from the bordero, and the three terms in between are assessed by expert judgment.

To finalise the premium rating of this example, say we have observed $k_+ = 5.8$ loss-free years. Then ASM rating yields as estimated frequency $\frac{g(0)}{5.8} = \frac{0.889}{5.8} = 15.3\%$.

To approximate the average loss, we complete the above table. The three last columns show the *subtotals* per row.

No.	S_j	G_j	g_j	stS_j	stG_j	stg_j
1	1'000'000	700	0.70	1'000'000	700	0.7
1	800'000	800	1.00	800'000	800	1.0
2	500'000	400	0.80	1'000'000	800	1.6
3	200'000	180	0.90	600'000	540	2.7
5	100'000	150	1.50	500'000	750	7.5
75	50'000	56	1.12	3'750'000	4'206	84.0
165	30'000	36	1.20	4'950'000	5'940	198.0
490	20'000	27	1.25	9'800'000	13'230	661.5

Calculating now the sums over all units and collecting quantities estimated earlier, we get

$$\sum G_j = 26'960, \quad \sum g_j = 957.0‰, \quad {}^{g}\overline{S} = 28.171, \quad {}^{c}L_{max} = 4'000, \quad {}^{v}l_{max} = 25\%, \quad C = 5$$

which yields

$$R_+ \leq f_+ \left({}^{c}L_{max} + {}^{v}l_{max} C \, {}^{g}\overline{S}\right) = 15.3\% \, (4'000 + 35'214) = 6'011$$

If the tariff premium is the real one, this means a loss ratio of 22.3%.

Notice that the only input taken from the tariff is the ratio $\sum G_i / \sum g_j$, such that, as anticipated, a three times higher/lower tariff would yield the same final result. We have in fact used only mild assumptions and in particular not the overall level of the tariff.

Nevertheless this approach should be *handled with care*, it is not a panacea for scarce-data situations of any kind. In particular the assumptions about the homogeneity of the ${}^{v}l_j$ and the w_j can break down easily if among the units there are few very unusual ones, e.g. large units having an extremely low premium rate being due to a specific high deductible. Most Third Party Liability business is problematic too. In some cases the sums insured are closely tied to the loss potential (e.g. insolvency insurance), but for many TPL covers the insureds are rather free to choose their sum insured, or more precisely their first-loss policy limit, according to their budget and risk aversion. Then very similar risks can have sums insured between e.g. 1 and 20 million Euro, such that the latter hardly tell anything about the average loss, which will mostly be far below the policy limit and not very sensitively affected by it.

6 Bias

Having seen premium rating examples, let us now study the statistical properties of N_+ and $g(N_+)$. We will sometimes a bit loosely call these the sample mean and the ASM, always keeping in mind that to estimate the frequency in the year v_e, we have to divide by k_+. The mathematical properties of N_+ and $g(N_+)$ translate to the sample mean and the ASM in an obvious way; it is just more convenient not to carry the factor k_+ along the way through all formulae. For further ease of reading we will mostly write λ for λ_+ and N for N_+.

Now we need to finally make our definition of admissible amending functions strict.

Definition 6.1. We call a real-valued sequence or function on (all) nonnegative integers g an *admissible amending function* if it is finite-dimensional, strictly positive, strictly increasing, and furthermore fulfils $g(n) \geq n$ and the above smoothness criteria (2) and (3).

Compared to the provisional definition, we have dropped the first smoothness condition and "moderate bias", as they are hard to make precise via e.g. inequalities, the more so as there is the bias/smoothness trade-off. It goes without saying that we still want to meet these criteria to some extent, however, it is more convenient to have them not formally included – the remaining conditions are many enough. Recall that our five candidates g_2 to g_6 meet them all, thus are admissible in a strict sense.

An amending function of dimension d (or less) can be represented as

$$g(n) = n + \sum_{j=0}^{d-1} r_j \chi_j(n)$$

with coefficients r_j, where $\chi_j(n)$ is the function that equals 1 if $n = j$ and else equals 0. Then we have $g(n) = n + r_n$ for $n < d$ and $g(n) = n$ for $n \geq d$. For admissible amending functions the r_j are nonnegative. Notice that this representation can be extended to the case $d = \infty$, yielding an infinite, but well-defined sum.

Lemma 6.2. *With the probability function $p_j = P(N = j)$ we have*

$$E(g(N)) = \lambda + \sum_{j=0}^{d-1} r_j p_j, \qquad \mathit{Bias}(g(N)) = E(g(N)) - E(N) = \sum_{j=0}^{d-1} r_j p_j$$

provided $g(N)$ has finite expectation, as is in particular the case if g has finite dimension.

The Lemma applies e.g. to the infinite-dimensional g_1, where all r_j equal 1.

Proposition 6.3. *For any given loss count distribution, the admissible amending function having the lowest bias is $g_2(N)$.*

Proof. The bias is monotonic in the coefficients $r_j \geq 0$. As stated in Section 4, g_2 is a lower bound to all admissible amending functions, having the lowest coefficients r_i. $\square$

The fact that g_2 is optimal in terms of bias does not mean that we should use this amending function anyway. If we consider smoothness as very important, we might in return accept a somewhat higher bias. Say we want the maximum increase $\frac{g(n+1)}{g(n)}$ to be less than 2. Then we cannot use g_2, but have to look for other amending functions having a low bias.

Let us try to calculate the bias for admissible ASMs in general. It is a (linear) function of the probabilities p_j. To calculate the latter, we need a distribution model for N_+ and ultimately for the N_i. As usual, we assume the latter as independent (independence of the years).

We treat the three classical loss count models: Poisson, Binomial, and Negative Binomial (see e.g. [Klugman et al., 2008]).

6.1 Poisson

If the N_i are Poisson distributed with expected value λ_i, then N_+ is again Poisson with expected value $\lambda = v_+\theta$ and we have $p_j = \frac{\lambda^j}{j!}e^{-\lambda}$ and in particular $p_0 = e^{-\lambda}$, $p_1 = \lambda e^{-\lambda}$. Note that these formulae just use $\lambda = \mathrm{E}(N_+)$, not the single λ_i.

6.2 Binomial

If the N_i are Binomial with parameters m_i (number of trials) and q_i (probability of success), then N_+ is possibly not Binomial any more. However, if all q_i equal the same value q, then N_+ is indeed Binomial having parameters $m := m_+ := m_1 + \ldots + m_k$ and q. This is indeed the most interesting case in the actuarial practice and can be interpreted as follows: In the year i the risk consists of m_i insured units suffering each either a loss (with probability q) or no loss. What changes over the years is the number of insured units. So m_i is a measure for the frequency volume of the risk in the year i. The expectations are $\lambda_i = m_i q$ and $\lambda = mq$. The probability function of N_+ is $p_j = \binom{m}{j}q^j(1-q)^{m-j}$, in particular we have $p_0 = (1-q)^m$, $p_1 = mq(1-q)^{m-1}$.

To make this model easier comparable to the Poisson case, we change the parametrisation by replacing the parameter q by the expected value λ:

$$p_j = \binom{m}{j}\frac{\lambda^j(m-\lambda)^{m-j}}{m^m}, \qquad p_0 = \left(1-\frac{\lambda}{m}\right)^m, \qquad p_1 = \lambda\left(1-\frac{\lambda}{m}\right)^{m-1}$$

As for Poisson, we get formulae that require (together with the second parameter m) λ, but not the single λ_i. If m is large, the Binomial distribution is very similar to the Poisson (which is in fact the limiting case for $m \to \infty$). Thus, Binomial is probably not worth being studied for large m.

The opposite case instead is very interesting, as it has an important practical application: If all m_i equal 1, we have $m = k$ and the single years have a Bernoulli distribution producing either one loss or no loss. This reflects the payments of *Stop Loss reinsurance* layers, which are triggered by the *aggregate* of all losses occurring in a year in a line of business: if the aggregate loss amount exceeds the attachment point, we have a loss to the layer, otherwise the year is loss-free. In situations where the loss probability can be assumed to be constant over the observation period ($\lambda_i = q$), the adequate model for N_+ is indeed $\mathrm{Bin}(k, \lambda)$, where $\lambda = kq$.

6.3 Negative Binomial

Among the many existing parameterisations (see [Fackler, 2011] for an overview) we choose

$$p_j = \binom{\alpha+j-1}{j}\left(\frac{\lambda}{\alpha+\lambda}\right)^j\left(\frac{\alpha}{\alpha+\lambda}\right)^\alpha$$

where as in the preceding models λ is the expectation, while $\alpha > 0$ the so-called shape parameter (not to be mistaken for the Pareto alpha used earlier).

If the N_i are Negative Binomial with parameters λ_i and α_i, then N_+ is possibly not NegBin any more. This applies unfortunately also to the most interesting case in actuarial practice, the well-established Poisson-Gamma model (see e.g. [Bühlmann and Gisler, 2005], [Klugman et al., 2008]), where the years have (possibly different) expectations $\lambda_i = v_i\theta$, but the same shape parameter α. We get this model if we assume that every year is (conditionally) Poisson distributed as above, but the frequency per volume unit is not a constant, instead fluctuates according to a Gamma distribution with expected value θ and shape parameter α. Let us look at this case of varying λ_i and invariable α more closely.

Although (unless the v_i are equal) N_+ is not Negative Binomial, it is not too difficult to calculate its probability function. However, unlike in the above Poisson and Binomial cases, the resulting probabilities are functions of all λ_i, not just of their sum λ. This is inconvenient for further mathematical analysis as one would (for fixed λ) have to distinguish an infinity of different cases. Fortunately, it turns out that it is possible to calculate, at least for the first values of the probability function, upper and lower bounds depending on λ only.

Let be $p_{i,j} = \mathrm{P}(N_i = j)$ and $p_j = \mathrm{P}(N_+ = j)$, $i = 1, \ldots, k$. Then we have

$$p_{i,0} = \left(\frac{\alpha}{\alpha+\lambda_i}\right)^\alpha, \qquad p_{i,1} = \frac{\alpha\lambda_i}{\alpha+\lambda_i}\left(\frac{\alpha}{\alpha+\lambda_i}\right)^\alpha, \qquad \frac{p_{i,1}}{p_{i,0}} = \frac{\alpha\lambda_i}{\alpha+\lambda_i}$$

from which we get:

Proposition 6.4. *If N_+ is the independent sum of NegBin(α, λ_i)-distributed loss counts, for its probability function we have*

$$\left(\frac{k\alpha}{k\alpha+\lambda}\right)^{k\alpha} \le p_0 = \prod_{i=1}^{k}\left(\frac{\alpha}{\alpha+\lambda_i}\right)^{\alpha} \le \left(\frac{\alpha}{\alpha+\lambda}\right)^{\alpha}$$

$$\frac{\alpha\lambda}{\alpha+\lambda}\left(\frac{k\alpha}{k\alpha+\lambda}\right)^{k\alpha} \le p_1 = \left\{\sum_{i=1}^{k}\frac{\alpha\lambda_i}{\alpha+\lambda_i}\right\}\prod_{i=1}^{k}\left(\frac{\alpha}{\alpha+\lambda_i}\right)^{\alpha} \le \frac{k\alpha\lambda}{k\alpha+\lambda}\left(\frac{\alpha}{\alpha+\lambda}\right)^{\alpha}$$

$$\frac{\alpha\lambda}{\alpha+\lambda} \le \frac{p_1}{p_0} = \sum_{i=1}^{k}\frac{\alpha\lambda_i}{\alpha+\lambda_i} \le \frac{k\alpha\lambda}{k\alpha+\lambda}$$

Proof. Note that all factors and summands in these formulae are positive. Let us first prove the equivalences.

The equation in the first row means that p_0 is the product of the $p_{i,0}$. As for the next equation, one loss in the multi-year period means that there is one year i suffering a loss and all other years are loss-free, thus

$$p_1 = \sum_{i=1}^{k}\left(p_{i,1}\prod_{l\neq i}p_{l,0}\right) = \sum_{i=1}^{k}\left(\frac{p_{i,1}}{p_{i,0}}\prod_{l=1}^{k}p_{l,0}\right) = \left\{\sum_{i=1}^{k}\frac{p_{i,1}}{p_{i,0}}\right\}\prod_{i=1}^{k}p_{i,0}$$

If we plug in what we already have, we get the second equation. The third one is the quotient of the first two. Let us now look at the inequalities.

First row: If we take the left inequality to the power $\frac{-1}{k\alpha}$, we get equivalently

$$\frac{\alpha+\lambda/k}{\alpha} \ge \left(\prod_{i=1}^{k}\frac{\alpha+\lambda_i}{\alpha}\right)^{1/k}$$

which is an application of the inequality of the arithmetic and the geometric mean. (Note that λ/k is the arithmetic mean of the λ_i.)

If we take the right inequality to the power $\frac{-1}{\alpha}$, we get the equivalent and trivial inequality

$$\prod_{i=1}^{k}\frac{\alpha+\lambda_i}{\alpha} = \prod_{i=1}^{k}\left(1+\frac{\lambda_i}{\alpha}\right) \ge 1+\sum_{i=1}^{k}\frac{\lambda_i}{\alpha} = \frac{\alpha+\lambda}{\alpha}$$

Third row: Here the left inequality is trivial:

$$\frac{\alpha\lambda}{\alpha+\lambda} = \sum_{i=1}^{k}\frac{\alpha\lambda_i}{\alpha+\lambda} \le \sum_{i=1}^{k}\frac{\alpha\lambda_i}{\alpha+\lambda_i}$$

The right one is equivalent to

$$\frac{1}{k}\sum_{i=1}^{k}\frac{\alpha\lambda_i}{\alpha+\lambda_i} \le \frac{\alpha\lambda/k}{\alpha+\lambda/k}$$

which is an application of Jensen's inequality using the concave function $\psi(x) = \frac{\alpha x}{\alpha+x}$.

Finally, the second row is the product of the first and third one. □

The latter illustrate that p_0 and $\frac{p_1}{p_0}$ lie between the corresponding values of two Negative Binomial distributions, namely those having expectation λ and shape parameters α and $k\alpha$, respectively. The latter is the distribution of N_+ in case all λ_i (i.e. all v_i) are equal, the former is the limiting case if one of the λ_i (i.e. v_i) is much greater than the others. One could say that in a way the distribution of N_+, albeit not being NegBin, is a blend of these two NegBin cases.

A look into the proof yields further insight. We have used two inequalities for means, arithmetic-geometric and Jensen, obtaining the *lower* bound in the *first* row and the *upper* bound in the *third* row. These bounds are taken on if the λ_i, or equivalently the v_i, are equal. So, if the frequency volumes are pretty close, a case fairly common in practice, these bounds are approximately taken on. As for the remaining bounds in the first and third row, one sees quickly that they are approximately taken on if one of the volumes is much greater than the others. This case is in a way the opposite of the first one and rather remote in the real world. So, these latter bounds will hardly be good approximations. Overall we can say that in practice p_0 will often be close to its lower bound, while $\frac{p_1}{p_0}$ will often be close to its upper bound. Unfortunately, this does not lead to a similar result for p_1. The second row is the product of the first and third one, thus combines two lower bounds which cannot be good at the same time, and analogously for the upper bounds. So, p_1 cannot be very close to any of its bounds.

Whether the derived inequalities are useful, yielding sufficiently narrow intervals for p_0 and p_1, depends on the parameters. Scenarios with realistic parameter constellations will illustrate below that we often get fairly narrow intervals for p_0, while intervals are larger for p_1. But, as r_1 is usually small, the latter probability has a rather low numerical impact on the bias and other quantities of interest. So, we overall get good-enough approximations. Before looking at those numerical examples, we finally treat the classical mathematical criterion for estimators.

7 Mean squared error

The usual way to measure the accuracy of an estimator is the mean squared error (MSE), which in this paper shall always mean SPEE (squared parameter estimation error), i.e. we look at the expected squared deviation of the ASM from the parameter $\mathrm{E}(N_+) = \lambda_+ = v_+\theta$. The results of the last section (together with some more inequalities) will enable us to calculate (approximately) the MSE of the ASM as a function of λ_+ for the above three models for N_+, with gives us a complete picture of the quality of the ASM method for the dimensions 1 and 2.

7.1 Basic results

We have $\mathrm{MSE}\left(g\left(N\right)\right) = \mathrm{E}\left(\left(g\left(N\right) - \mathrm{E}\left(N\right)\right)^2\right) = \mathrm{Bias}^2\left(g\left(N\right)\right) + \mathrm{Var}\left(g\left(N\right)\right)$.

In order to calculate the variance of $g\left(N\right) = N + \sum_{j=0}^{d-1} r_j \chi_j\left(N\right)$, we need some formulae:

Lemma 7.1. *For $j = 0, \ldots, d$–1 we have*

$$E(\chi_j) = p_j; \quad Var(\chi_j) = p_j\left(1–p_j\right); \quad Cov(\chi_j, \chi_l) = -p_j p_l,\ j \neq l; \quad Cov(N, \chi_j) = p_j(j - \lambda)$$

Proof. The (co)variance formulae follow immediately from $\mathrm{Cov}(A, B) = \mathrm{E}(AB) - \mathrm{E}(A)\mathrm{E}(B)$. □

Lemma 7.2. *The variance of N_+ is equal to*

Poisson: λ

Binomial: $\lambda\left(1 - \frac{\lambda}{m}\right)$

Negative Binomial: $\lambda\left(1 + \frac{\lambda}{\kappa\alpha}\right)$, *where* $\kappa = \left(\sum_{i=1}^{k} v_i\right)^2 / \sum_{i=1}^{k} v_i^2$.

We have $1 \leq \kappa \leq k$, where $\kappa = k$ iff the v_i are equal, while the lower bound is (approximately) reached if one of the v_i is much greater than the others.

Definition 7.3. We call κ as set in the preceding lemma the *volume homogeneity coefficient.*

Proof. The Poisson and Binomial formulae are well known. In the Negative Binomial case for each N_i we have $Var\left(N_i\right) = \lambda_i + \frac{\lambda_i^2}{\alpha} = v_i\theta + \frac{(v_i\theta)^2}{\alpha}$. $Var\left(N_+\right)$ is the sum of these variances, thus

$$Var\left(N_+\right) = \theta\sum_{i=1}^{k} v_i + \frac{\theta^2}{\alpha}\sum_{i=1}^{k} v_i^2 = \theta\sum_{i=1}^{k} v_i + \frac{\theta^2}{\kappa\alpha}\left(\sum_{i=1}^{k} v_i\right)^2 = \lambda + \frac{\lambda^2}{\kappa\alpha}$$

The stated inequality for κ is equivalent to $\sum_{i=1}^{k} v_i^2 \leq \left(\sum_{i=1}^{k} v_i\right)^2 \leq k\sum_{i=1}^{k} v_i^2$, where the left inequality is trivial, while the right one is an application of the inequality of the arithmetic and the quadratic mean. The cases of equivalence follow immediately. □

The closer to each other the volumes of the single years, the larger κ. Values very close to k do occur in practice. Instead, the opposite case that a year dominates such that k is close to 1, is rather remote. In order to find out what values κ may take on in practice, we go for a moment back to Example 3.3 having a geometrically growing portfolio with volumes $v_1, \ldots, v_k$ such that $v_{i+1} = v_i\left(1 + s\right)$. Some algebra yields

$$\kappa = \left(1 + \frac{2}{s}\right) \frac{(1+s)^k - 1}{(1+s)^k + 1} < 1 + \frac{2}{s}$$

Now we have a second upper bound, which is notably independent of k. E.g., for $s = 50\%$ (extreme growth) the coefficient κ cannot exceed 5, mo matter how large k is. As for lower bounds, there is no easy rule. However, noting that the above formula for κ is an increasing function in k and a decreasing function in s, we can see quickly that if we have a minimum of 4 years and a yearly increase of not more than 70%, then κ is greater than 3.

Results for volumes increasing (as is common) in a similar way, but not exactly geometrically, should be close. Thus, for real-world constellations the assumption $\kappa \geq 3$ is reasonable, apart from rare particular constellations.

For simplicity of notation we unite the three cases of Lemma 7.2 in one:

Corollary 7.4. *$Var(N_+) = \lambda + c\lambda^2$, where according to the distribution we have:*
Poisson: $c = 0$, Binomial: $c = -\frac{1}{m}$, Negative Binomial: $c = \frac{1}{\kappa\alpha}$.

The parameter c is a slight generalisation of the c named *contagion* by [Heckman and Meyers, 1983] in order to give an intuitive meaning to the deviations of Binomial and NegBin from the Poisson distribution. At the same time it is a special case of the contagion as defined generally for loss count distributions in Chapter 5 of [Fackler, 2017].

Now we have all the ingredients at hand to determine variance and MSE of the ASM.

Proposition 7.5. *For the amending function $g(N) = N + \sum_{j=0}^{d-1} r_j \chi_j(N)$ we have*

$$Bias(g(N)) = \sum_{j=0}^{d-1} r_j p_j$$

$$Var(g(N)) = \lambda + c\lambda^2 + \sum_{j=0}^{d-1} (r_j + 2(j - \lambda)) r_j p_j - \sum_{j,l=0}^{d-1} r_j r_l p_j p_l$$

$$MSE(g(N)) = \lambda + c\lambda^2 + \sum_{j=0}^{d-1} (r_j + 2(j - \lambda)) r_j p_j$$

The formulae extend to the infinite-dimensional case provided the appearing moments are finite.

Proof. Straightforward calculation applying the preceding corollary and Lemma 7.1. □

For the non-admissible $g_1(n) = n + 1$ we have in particular:

$$\text{Bias}(g_1(N)) = 1, \quad \text{Var}(g_1(N)) = \lambda + c\lambda^2, \quad \text{MSE}(g_1(N)) = \lambda + c\lambda^2 + 1$$

7.2 Comparison with sample mean

Definition 7.6. For any loss count distribution N having probabilities p_j and expectation λ, and any ASM $g(N)$ having finite first and second moment, we define the *MSE delta*

$$\varDelta g := \text{MSE}(g(N)) - \text{MSE}(N) = \text{MSE}(g(N)) - \text{Var}(N) = \sum_{j=0}^{d-1} (r_j + 2(j - \lambda)) r_j p_j$$

This is the deviation of the MSE of the ASM from the MSE of the sample mean. If the MSE delta is negative, then the ASM is more accurate than the sample mean (in terms of MSE). $\varDelta g$ enables us also to compare different ASMs: The lower $\varDelta g$, the lower $\text{MSE}(g(N))$, the "better" g.

When writing $\varDelta g(\lambda)$, we interpret $\varDelta g$ as a function of λ. This is indeed a smooth function, as the p_j are differentiable functions of the expected loss count λ for all three distributions we discuss here (as well as for many other parametric loss count models).

Proposition 7.7. *For any nontrivial loss count distribution and any admissible ASM given, we have that for very small λ the MSE delta is positive, while for very large λ (provided this is at all possible) it is negative.*

Proof. For $\lambda \searrow 0$ the probability function $(p_j)_j$ tends to $(1, 0, 0, ...)$, such that $\Delta g\,(0+) = r_0^2 > 0$. Conversely, if λ is greater than any of the d values $j + \frac{r_j}{2}$, we have $r_j + 2\,(j - \lambda) < 0$ for $j = 0, \ldots, d$–1, such that $\Delta g\,(\lambda) < 0$. □

Intuitively, for very low λ the (squared) bias must have an enormous impact on the MSE of g, hence the positive delta. For very large λ instead, the amended sample mean (despite of having a positive bias) is more accurate than the ordinary sample mean. Thus, in the Poisson and Negative Binomial case for any admissible ASM there is an interval $]s; \infty[$ of frequencies λ where the ASM is more accurate than the sample mean.

In the Binomial case λ is bounded from above by m, hence we cannot have arbitrarily large λ. Checking different cases of Binomial parameters and ASMs, one sees that for very expensive (and thus not practicable) ASMs it can occur that the MSE delta is always positive, however, one gets mostly situations where Δg takes on negative values for larger λ. There is no easy rule apparent to distinguish these cases. At any rate, for real-world constellations the Binomial case seems to behave very much like the two other loss count models.

We will see soon that there is no best ASM in the sense of having globally (for all λ) a lower MSE than the other ASMs. However, another optimality is worth being considered:

Recall that in principle we were not unhappy with the sample mean. Our intention was to just slightly amend it in a way to avoid zeros (and possibly to have a smooth rating scheme), but apart from that we tried to keep ASMs as close to the sample mean as possible. Thus it would be coherent to accept a mean squared error not too different from that of the sample mean. Lower MSEs are welcome, but perhaps it is less important to get far lower. Instead, we could try to find ASMs that have a lower MSE than the sample mean for as many frequencies λ as possible, but no matter how low we get. In other words: we want to beat the sample mean in as many situations as possible, but don't bother about high victories. In this sense we can say:

Definition 7.8. An amending function is *superior* to another one if it has a wider range of loss frequencies λ where the ASM is better than the sample mean, i.e. $\Delta g\,(\lambda) < 0$.

Proposition 7.9. *For any given nontrivial loss count distribution, the admissible ASM having the widest range of loss frequencies λ where it is more accurate (in terms of MSE) than the sample mean, is $g_2\,(N)$. The range of these frequencies is the interval $]0.25; \infty[$.*

Proof. We have $\Delta g_2\,(\lambda) = (0.5 - 2\lambda)\,p_0$, so $\Delta g_2\,(\lambda) < 0$ is equivalent to $\lambda > 0.25$. To prove the proposition, we show that for any admissible amending function g and any $\lambda \leq 0.25$ we have $\Delta g\,(\lambda) \geq 0$:

$$\Delta g\,(\lambda) = (r_0 - 2\lambda)\,r_0 p_0 + \sum_{j=1}^{d-1} (r_j + 2\,(j - \lambda))\,r_j p_j \geq (r_0 - 0.5)\,r_0 p_0 + \sum_{j=1}^{d-1} 2\,(j - 0.25)\,r_j p_j \geq (r_0 - 0.5)\,r_0 p_0$$

The final term is nonnegative as for all admissible amending functions $r_0 = g\,(0) \geq 0.5$. □

Note that the proof yields the stronger result that if $\lambda \in [0; 0.25]$, g_2 has the (strictly) lowest MSE among all admissible ASMs: For any of them the second summand $(r_1 + 2\,(j - \lambda))\,r_1 p_1$ is positive, so the inequality is strict. However, as anticipated above, this does not hold for large λ, such that neither g_2 nor any other admissible ASM can be "best" for all λ. Stated precisely:

Proposition 7.10. *For parametric loss count distributions whose expectation λ is not bounded from above, there is no admissible ASM having globally (for all λ) a lower MSE than the other ones.*

Proof. We just have to find an admissible amending function g and a frequency λ such that $\Delta g\,(\lambda)$ is smaller than $\Delta g_2\,(\lambda)$. We prove a stronger result.

Let g be any admissible amending function different from g_2 and λ be any value greater than r_0 and any of the $j + \frac{r_j}{2}$. Then $\Delta g\,(\lambda)$ is (strictly) smaller than $\Delta g_2\,(\lambda) = (0.5 – 2\lambda)\,0.5 p_0$:

$$\Delta g\,(\lambda) = (r_0 - 2\lambda)\,r_0 p_0 + \sum_{j=1}^{d-1} 2\left(\frac{r_j}{2} + j - \lambda\right) r_j p_j < (r_0 - 2\lambda)\,r_0 p_0 \leq (0.5 – 2\lambda)\,0.5 p_0$$

The first inequality is strict as the sum runs from 1 to at least 1, thus is not empty. For the second inequality note that $(r – 2\lambda)\,r p_0$ is a decreasing function in r on the interval $[0; \lambda]$. As $0.5 \leq r_0 \leq \lambda$, we are done. □

For "bounded" distributions like Binomial with fixed m (here $\lambda < m$), it could be that g_2 has globally the lowest MSE and MSE delta. However, independently of whether and when this is possible, we feel that the alternative criterion for "good" ASMs introduced above (wide range of λ where the MSE delta is negative) is anyway of more interest for the insurance practice.

7.3 Reflecting business strategy

The fact that g_2 is the best admissible ASM in terms of MSE (in the way we interpret optimality here), and in addition has the lowest bias, does not mean that it is the best amending function for all purposes. As already stated after Proposition 6.3, we could consider smoothness as very important. Then we would accept a somewhat higher bias and a reduced range of frequencies where the ASM is more accurate than the sample mean.

ASM smoothness is a hybrid issue – half mathematics, half business. The above criteria (1), (2), and (3a) reflect consistency in very much the same manner as e.g. monotonicity, though one could possibly weaken them a bit. The last condition (3b) $g(1)/g(0) \leq 2$ is instead a "political" decision, one could in principle have allowed say triplication or quadruplication of the premium after a loss, which would have lead to more admissible ASMs of dimension 1 having $0 < g(0) < 0.5$ and thus yielding extremely cheap rates for loss-free risks and huge increases ($g(1) \geq 1$) after the first loss.

On the contrary, if one wants to have as few angry clients as possible due to drastic premium raises, this could be ensured by a stronger condition $g(n+1)/g(n) \leq h$ with a *maximum increase factor* $h < 2$, which would exclude g_2 and a range of ASMs of higher dimension.

Such a maximum could also be derived from a reasoning on *insurance demand* (see in the following [Hao et al., 2019]), which in basic models is assumed to be proportional to $\pi^{-\epsilon}$, where π is the premium and ϵ is the *demand elasticity.* For the latter in insurance partly quite low values are observed, sometimes lower than 0.5. However, even a very low elasticity of 0.3 would mean loosing $1-2^{-0.3} = 18.8\%$ of the insureds after doubling of their premium. According to such a demand model, to retain more clients one would have to set h lower than 2.

Another business issue is the general level of the rating. If it is felt that the value $g(0) = 0.5$ is anyway too low for loss-free situations (recall it means adjoining an observation period as long as the really existing one), one may introduce a stronger condition $g(0) \geq b$ with a *minimum level* $b > 0.5$, which would exclude again g_2 and certain ASMs of higher dimension (yet not exactly the same as above). Other market players might rather think about a maximum level $g(0) \leq b'$ or a combination of both, some players might consider similar restrictions for $g(1)$ and possibly a few subsequent values of the amending function.

Business strategies matter so much for the topic of this (in principle mathematical) paper as we deal with risks that are very difficult to rate (high model/parameter uncertainty). If a risk can be rated on tons of good data, normally all (re)insurers will calculate about the same rate and the offered premiums will usually be very close. In poor-data situations instead, one typically observes a large variety of offered premiums due to different decisions made by the offering companies, being all partly political rather than purely actuarial. In short, premiums of "difficult" risks, even when they are rated judgmentally on a case by case basis, depend somehow (implicitly) on the underwriting policy.

Strategies for insurance portfolios can be very diverse according to one's position in the market, financial strength, etc. However, a bit simplifying they find themselves somewhere in between the following two:

Growth at any cost: If a company aims for a lot of new business, it needs to be among the cheapest offers in the market for many risks, but has to recoup quickly via sharp premium increases after losses. That is exactly what very cheap (and steep) amending functions like g_{min} and g_{so2} do. (For clarity here we use the descriptive names introduced in the last table in Chapter 4.) The disadvantage of this strategy is that one has to put up with loosing a lot of (opportunistic) clients just after having paid their first loss, as they will not accept the premium surcharge and will look for a cheaper offer in the market.

Portfolio must be stable: If a company rather refrains from writing new risks that are likely to be lost very soon (in the moment of necessary premium increases), it will set a generally higher premium level but will try to avoid drastic changes. That is exactly what e.g. g_{max3} does. The disadvantage of this strategy is that one might possibly not be able to write a lot of new business.

In spite of the many requirements admissible amending functions must meet, they can indeed have quite diverse "behaviour" corresponding to very different business strategies. (g_{max2} and g_{so3} can be seen as somehow intermediate.) Thus, an insurer should always be able to find an ASM being suitable for the given strategy and environment. (If not, the aspired strategy might imply a negative bias. In particular it is not possible to get both very cheap initial premiums and very moderate increases.) In case the business strategy and/or the acceptance of premium increases are not the same in all lines of business, it might be a good idea to work with different ASMs.

The great advantage of the ASM approach over case-by-case dealing with poor-data situations is that one has to decide only *once* how to proceed in such cases and that the decision is made *in advance* and *explicitly*.

Recall that while being able to reflect some commercial aspects, ASM rating is a step in the calculation of the *technical* premium, which should not be mistaken for the offered *commercial* premium possibly deviating from the former for a number of (political) reasons. However, when ASM rating is in line with the underwriting strategy, it might be easier to stay with technical premiums, instead of frequently having to adjust them (case by case, time-consuming) for a commercial offer.

If an insurer wants to adopt ASM rating, unless they don't want to be the overall cheapest (use g_2 then), it would be natural to consider ASMs in a certain range and make a choice among them according to criteria like bias and mean squared error. To support such choices and generally to gather more intuition about how ASMs "behave", we now study a number of realistic scenarios.

8 Numerical illustration

Before calculating examples, let us explain shortly how we can proceed in the Negative Binomial case in order to get results depending from as few parameters as possible. The approximations derived in Proposition 6.4 enable us to study efficiently the amending functions of dimension 1 and 2, which can be represented as $g(N) = N + r_0\chi_0 + r_1\chi_1$, such that Proposition 7.5 can be rearranged (via some algebra) as:

Corollary 8.1. *For the ASMs of dimension 1 or 2 we have*

$$Bias(g(N)) = r_0p_0 + r_1p_1$$

$$Var(g(N)) = \lambda + c\lambda^2 + 2(-\lambda)r_0p_0 + 2(1-\lambda)r_1p_1 + r_0^2p_0(1-p_0) + r_1^2p_1(1-p_1) - 2r_0r_1p_0p_1$$

$$MSE(g(N)) = \lambda + c\lambda^2 + (r_0 - 2\lambda)r_0p_0 + (r_1 + 2 - 2\lambda)r_1p_1$$

$$\Delta g = (r_0 - 2\lambda)r_0p_0 + (r_1 + 2 - 2\lambda)r_1p_1$$

These four quantities are functions of λ if N_+ is Poisson, respectively functions of λ and the aggregate volume m if N_+ is Binomial. Instead, the Negative Binomial case is far more intricate: A look at the *equations* of Proposition 6.4 makes clear that for an exact calculation we need α, k, κ (to determine c), and the single λ_i. However, if we accept approximate results, we can avoid tedious distinctions of cases having very similar vs. very different λ_i: The *inequalities* of Proposition 6.4 yield upper and lower bounds for p_0 and p_1 which are functions of α, k, and λ, enabling us to derive approximations for bias, variance, MSE, and MSE delta of the ASMs of dimensions 1 and 2 which are functions of α, k, κ, and λ.

If we have a close look at the formulae in the above corollary, we see that all but the second are linear functions in p_0 and p_1, such that we can calculate upper and lower bounds summand-wise, by simply checking whether the factors the two probabilities are multiplied with are positive or negative. The variance instead is a quadratic form in p_0 and p_1. The derivation of the minimum and maximum of such a function on a two-dimensional interval is a tedious analysis exercise. Yet, one could again proceed summand-wise (with terms collected for p_0, p_1, p_0^2, p_1^2, p_0p_1), though this will possibly not yield the optimal bounds.

8.1 Real-world parameters

What values for the parameters m, α, k, κ, and λ can we expect to appear in the real world?

We have often used the example of seven loss-free years. This is indeed a typical observation period, at least for the rating of reinsurance layers. Typically one would have about 5 to 10 observed years, in the case of NatCat rating it happen to be 15 or more. (The Fire example in Section 5, where $k = 3$, is an extreme case, which was presented to demonstrate the power of the ASM method.)

As explained in Section 3, k_+ in practice is often somewhat lower than k, so 5 to 10 years correspond to say 4 to 10 volume-weighted years.

As for the Binomial parameter m, it was already mentioned that large values m yield almost the same results as Poisson. Most interesting is the case $m = k$, which is as different as possible from Poisson and embraces Stop Loss reinsurance as we have mentioned earlier.

For the Negative Binomial parameter α a wide range of values is observed, see e.g. the examples in [Mack, 1997]. Very large values like 100 occur, leading again to results not much different from the limiting case Poisson. On the contrary, one can come across values in the range of 1. It makes sense to check a variety of (rather low) values for α.

For κ we have discussed the realistic range of values in the preceding chapter: from 3 to k. Notice that κ affects variance and MSE, but not bias and MSE delta. If we are most interested in the latter, we do not need to think about κ.

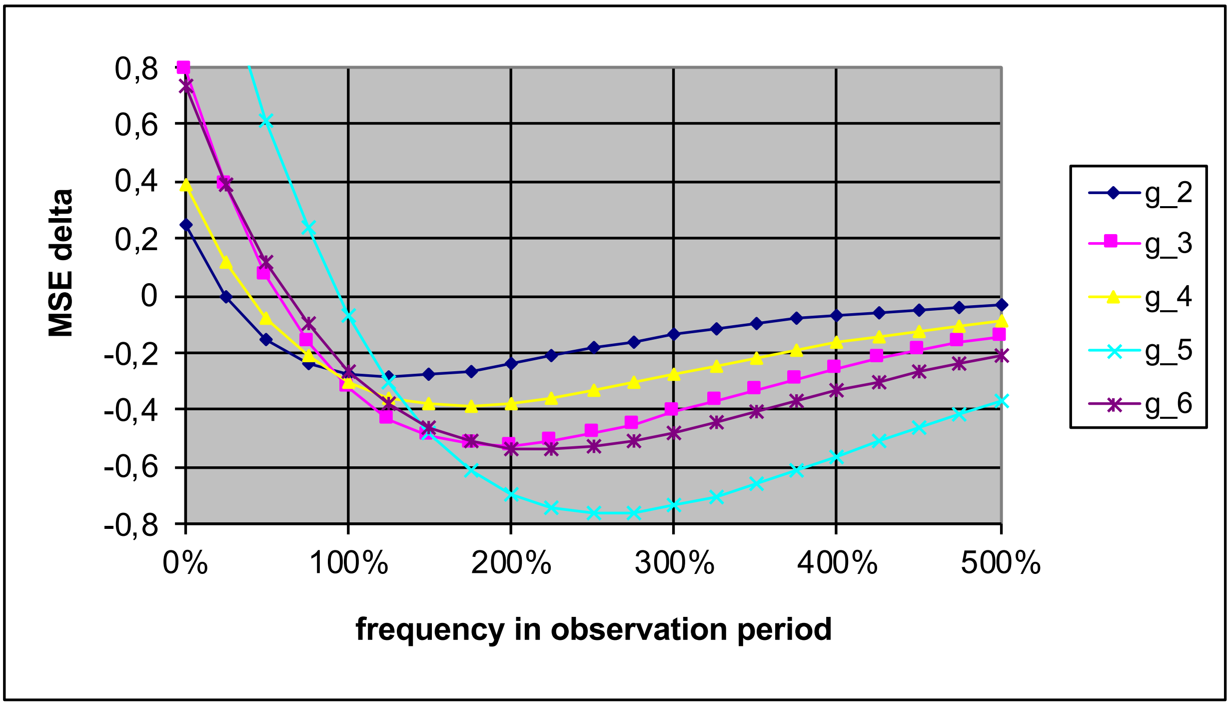
8.2 Results

To illustrate the calculations, two tables are provided in the appendix.

The *first* one displays bias, MSE, and MSE delta of the ASM $g_3 = g_{max2}$ together with the variance of the sample mean $\mathrm{Var}(N) = \lambda + c\lambda^2$, for four scenarios: a Poisson, a Binomial and two NegBin variants. For the latter we provide in addition the ranges for p_0 and p_1, while for bias, MSE, and MSE delta we show the upper bounds, all calculated according to the inequalities derived above. The second parameters for Binomial and Negative Binomial are chosen significantly lower than what would be mostly seen in practice, in order to be farther away from the Poisson case. Real-world cases will thus usually lead to figures somewhere in between those of the displayed scenarios and the Poisson case.

We can see that the intervals containing the NegBin probabilities are not really narrow, especially for p_1 in case of low α, but by running calculations with both interval endpoints one can check that the impact of this uncertainty on the MSE is moderate. The table shows that all four scenarios yield quite similar results for the MSE delta, with the NegBin scenarios being largely lowest, which is remarkable as here we display the upper bounds. The similarity of results across scenarios indicates that the MSE delta is (for realistic parameters) not much affected by model uncertainty. In all examples the bias is (as expected) large for very low λ and decreases with growing λ, while the MSE delta decreases too, but notably much more quickly.

The *second* table and the corresponding chart here below compare, for the (in a way intermediate) Poisson scenario, the MSE delta for the ASMs g_2 to g_6. The latter turn out to be (in terms of MSE) far better than g_1 (where $\Delta g(\lambda) \equiv 1$), which confirms that adding always one loss to the loss record is too expensive, as well as would be (non-smooth) variants adding say always 0.7 or 0.5 losses. Amending functions, being sequences $(g(n))_n$ converging quickly to the sequence $(n)_n$, are the far better option.



If we look for the root (i.e. the value λ beyond which $\Delta g(\lambda)$ becomes negative), we find 25% for g_2 (already known, holds for any distribution model) and further about 57% for g_3, 39% for g_4, 97% for g_5, and 64% for g_6. This value deserves a name:

Definition 8.2. Let us call the smallest root of $\Delta g(\lambda)$ the *critical frequency* of the ASM g.

The critical frequency too depends only weakly on the loss count model. By checking a range of (realistic) parameter constellations one can see that all three models yield very similar results, deviating by not more than 3% from the stated figures. (g_5 and g_6 were not calculated for NegBin because we lack a handy approximation for the probability p_2.) It seems that for realistic parameters λ has only one root, i.e. the range of values λ satisfying $\Delta g(\lambda) < 0$ is always an interval $]s;\infty[$ (for Binomial $]s;m[$, respectively). However, for the Binomial model one can find (unrealistic, very expensive) ASMs where $\Delta g(\lambda)$ has no root.

Does the interval of very low frequencies where $\Delta g(\lambda)$ is positive spoil the quality of the ASM method in the range of (small) frequencies we are mainly interested in? The answer is: maybe in theory, but in practice no. To see this, notice that the critical frequencies of the analysed amending functions are (mostly much) lower than 100%. Now consider the rating situation of a risk having a frequency λ below the critical frequency of the ASM you have chosen. Then in particular λ is (mostly very much) smaller than 100%. Then the probability of no loss in the observation period is (mostly very much) greater than 33% – this holds for all three studied distribution models (in case of realistic parameters). But, in loss-free situations the sample mean is not used: it equals zero, practitioners will somehow "choose" a frequency > 0. Even in case of one or two losses practitioners often replace the empirical loss count by a higher figure (recall the workarounds of Section 2), which means doing something similar to ASM rating. Overall, for the low frequencies λ considered here, the sample mean is in practice unlikely to be applied, thus it doesn't matter that here it is (would be) more accurate than the ASM.

In other words: there is no point in comparing the ASM to the sample mean for situations where the latter is not used. The cases where the ASM is less accurate than the sample mean are largely theoretical, hardly reflecting real-world pricing situations.

8.3 One-year perspective

We have discussed the frequency $\lambda = \lambda_+$ of the *observation period* throughout most of this paper because this is the parameter the properties of the ASM mainly depend on. Let us finally come back to the frequency λ_e of the risk in the *future year,* which we wanted to assess. We want to get a feeling for what the results of this paper mean to this risk in the one-year perspective relevant for pricing.

A critical frequency of say $\lambda_+ = 60\%$ corresponds to a frequency λ_e between 6% and 15% if we assume a number of volume weighted years between 4 and 10. To get an idea, let us take once again the average case of 7 years, say $k_+ = 5$. Then $\lambda_+ = 60\%$ corresponds to $\lambda_e = 12\%$. That means that only if the risk has a yearly loss frequency well below 12%, the ASM method could in theory be regarded as suboptimal in terms of MSE (and bias), yet the comparison with the sample mean has not much practical relevance for such low frequencies leading mostly to rating situations where people don't apply the sample mean.

Overall we can say that for frequencies larger than about 10% the statistical properties of the ASM are similar to those of the sample mean or even better. It is indeed remarkable that the ASM method is able to produce from such few information (7 years loss-free) a fair rating for risks producing about one loss in 10 years.

If in the case of layer business frequency extrapolation from a lower model threshold is possible (see the Fire example in Section 5) and if k_+ is rather large, we are even able to get a fair rating for layers with a frequency as low as say $\lambda_e = 2\%$. Only for extremely low frequencies (towards the per mil range) the ASM method will have a huge MSE and bias, but doing a possibly very expensive rating in case of such frequencies is not a problem, at least in the reinsurance practice where common minimum RRoL requirements in the range of 1% will not allow much cheaper ratings anyway.

If we, as an extreme case, imagine a risk having a yearly loss frequency of 0.01%, it is clear that even 25 observed loss-free years will lead to a far too high ASM rating result. Such risks cannot be assessed at all without using any "external" data, whatever the rating method. Experience rating of such risks can only work if one manages to collect a small portfolio of several hundred similar ones, and rates the resulting pool together.

9 Conclusion

The ASM method for the rating of loss frequencies has a lot of desirable properties both from a statistical and a business driven standpoint. As for the business view:

- It is very easy to implement into existing rating tools. Wherever the sample mean is applied to estimate the loss frequency (and this is very often the case, even when tools use distributions like Negative Binomial where it does not coincide with the Maximum Likelihood estimator), one simply corrects the empirical loss count via the amending function. Nothing else has to be changed.
- In situations with several losses the results will not change (then we have $g(n) = n$), which ensures continuity of the rating methodology for such situations.
- In situations with poor data, where traditional rating tools stop working, now one will always get a loss frequency, even when the loss record is empty.
- The rating results of subsequent years may be volatile, but indeed *less volatile than the data itself* (smoothness). If one takes an amending function with low increases, the premium jumps after a new loss can be kept low.
- Generally, by choosing the amending function appropriately, the ASM method can be aligned with the underwriting policy.
- Last but not least: ASM rating safes a lot of time. The automatic assignment of a frequency is certainly much quicker than thinking individually about each case of very poor data.

As for the mathematical/statistical view:

- ASM rating is a consistent method for all situations, from those with abundant loss history to those with no losses.
- It takes the volume dependency of loss frequencies consistently into account and is altogether in line with traditional frequency/severity modelling, as it is used when more loss data is available.
- On average the method is on the safe side (positive bias).
- It is more accurate than the sample mean in terms of MSE. The only exception are extremely low frequencies where, however, in practice the sample mean would hardly ever be used.

In short, the ASM method is a coherent and very efficient extension of traditional experience rating to scarce-data situations where underwriters and actuaries usually (have to) abandon their well-established methods.

Appendix: Tables

All quantities are displayed as functions of $\lambda = \lambda_+$.

Table 1:

Four scenarios for g_3:

- Poisson
- Binomial: $m = 5$
- Negative Binomial 1: $\alpha = 4, k = 7, \kappa = 3$
- Negative Binomial 2: $\alpha = 1, k = 4, \kappa = 3$

λ [%]	0	25	50	75	100	125	150	175	200	250	300	400	500
Poisson													
Bias [%]	88.9	75.7	64.0	53.8	45.0	37.4	31.0	25.6	21.0	14.1	9.4	4.1	1.7
MSE	.79	.64	.57	.59	.68	.82	1.01	1.23	1.48	2.02	2.59	3.75	4.86
Var (N_+)	.00	.25	.50	.75	1.00	1.25	1.50	1.75	2.00	2.50	3.00	4.00	5.00
$\Delta g_3(\lambda)$	.79	.39	.07	-.16	-.32	-.43	-.49	-.52	-.52	-.48	-.41	-.25	-.14
Binomial													
Bias [%]	88.9	75.6	63.4	52.5	42.8	34.3	26.9	20.7	15.5	8.0	3.5	0.2	0
MSE	.79	.63	.54	.51	.52	.58	.65	.75	.84	1.00	1.06	.79	0
Var (N_+)	.00	.24	.45	.64	.80	.94	1.05	1.14	1.20	1.25	1.20	.80	0
$\Delta g_3(\lambda)$	.79	.39	.09	-.13	-.28	-.36	-.40	-.39	-.36	-.25	-.14	-.01	0
NegBin 1													
p_0 min [%]	100	78.0	60.9	47.7	37.4	29.4	23.2	18.3	14.5	9.1	5.8	2.4	1.0
p_o max [%]	100	78.5	62.4	50.3	41.0	33.7	28.0	23.4	19.8	14.3	10.7	6.3	3.9
p_1 min [%]	0	18.3	27.1	30.1	29.9	28.0	25.3	22.3	19.3	14.0	9.9	4.8	2.2
p_1 max [%]	0	19.4	30.7	36.7	39.5	40.3	39.8	38.6	36.9	32.9	28.9	21.9	16.6
Bias [%]	88.9	76.2	65.7	56.9	49.6	43.4	38.1	33.7	29.8	23.7	19.1	12.8	9.0
MSE	.79	.65	.60	.64	.76	.94	1.20	1.49	1.83	2.56	3.37	5.09	6.94
Var (N_+)	.00	.26	.52	.80	1.08	1.38	1.69	2.01	2.33	3.02	3.75	5.33	7.08
$\Delta g_3(\lambda)$	.79	.39	.08	-.16	-.33	-.44	-.49	-.51	-.51	-.46	-.38	-.24	-.14
NegBin 2													
p_0 min [%]	100	78.5	62.4	50.3	41.0	33.7	28.0	23.4	19.8	14.3	10.7	6.3	3.9
p_o max [%]	100	80.0	66.7	57.1	50.0	44.4	40.0	36.4	33.3	26.8	25.0	20.0	16.7
p_1 min [%]	0	15.7	20.8	21.6	20.5	18.7	16.8	14.9	13.2	10.2	8.0	5.0	3.3
p_1 max [%]	0	18.8	29.6	36.1	40.0	42.3	43.6	44.3	44.4	44.0	42.9	40.0	37.0
Bias [%]	88.9	77.4	69.1	62.8	57.8	53.6	50.1	47.1	44.4	40.0	36.5	31.1	27.2
MSE	.79	.66	.65	.76	.97	1.28	1.69	2.17	2.71	3.97	5.42	8.84	12.93
Var (N_+)	.00	.27	.58	.94	1.33	1.77	2.25	2.77	3.33	4.58	6.00	9.33	13.33
$\Delta g_3(\lambda)$	.79	.39	.07	-.17	-.36	-.49	-.56	-.60	-.62	-.62	-.58	-.49	-.40

Table 2:

Comparison of the MSE delta of five admissible amending functions, Poisson model:

λ [%]	0	25	50	75	100	125	150	175	200	250	300	400	500
$g_2 = g_{min}$	.25	.00	-.15	-.24	-.28	-.29	-.28	-.26	-.24	-.18	-.14	-.07	-.03
$g_3 = g_{max2}$	.79	.39	.07	-.16	-.32	-.43	-.49	-.52	-.52	-.48	-.41	-.25	-.14
$g_4 = g_{so2}$	.39	.12	-.08	-.21	-.30	-.36	-.38	-.39	-.38	-.33	-.27	-.16	-.09
$g_4 = g_{max3}$	1.60	1.07	.62	.24	-.06	-.30	-.48	-.61	-.70	-.76	-.74	-.56	-.37
$g_5 = g_{so3}$	.74	.39	.12	-.10	-.26	-.38	-.46	-.51	-.53	-.53	-.48	-.33	-.21

References

[Brazauskas and Kleefeld, 2009] Brazauskas, V. and Kleefeld, A. (2009). Robust and efficient fitting of the generalized Pareto distribution with actuarial applications in view. *Insurance: Mathematics and Economics*, 45(3):424–435.

[Bühlmann and Gisler, 2005] Bühlmann, H. and Gisler, A. (2005). *A course in credibility theory and its applications*. 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.

[Fackler, 2013] Fackler, M. (2013). Reinventing Pareto: Fits for both small and large losses. In *ASTIN Colloquium 2013*.

[Fackler, 2017] Fackler, M. (2017). *Experience rating of (re)insurance premiums under uncertainty about past inflation*. PhD thesis, Universität Oldenburg, urn:nbn:de:gbv:715-oops-33087.

[FINMA, 2006] FINMA (2006). *Technical document on the Swiss Solvency Test*. FINMA.

[Hao et al., 2019] Hao, M., Macdonald, A. S., Tapadar, P., and Thomas, R. G. (2019). Insurance loss coverage and social welfare. *Scandinavian Actuarial Journal*, 2019(2):113–128.

[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).

[Klugman et al., 2008] Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2008). *Loss models: from data to decisions*. Wiley.

[Mack, 1997] Mack, T. (1997). *Schadenversicherungsmathematik*. Verlag Versicherungswirtschaft.

[Parodi, 2014a] Parodi, P. (2014a). *Pricing in General Insurance*. CRC Press.

[Parodi, 2014b] Parodi, P. (2014b). Triangle-free reserving: a non-traditional framework for estimating reserves and reserve uncertainty. *British Actuarial Journal*, 19(1):168–218.

[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.

[Riegel, 2018] Riegel, U. (2018). Matching tower information with piecewise Pareto. *European Actuarial Journal*, 8(2):437–460.

[Schmutz and Doerr, 1998] Schmutz, M. and Doerr, R. R. (1998). *The Pareto Model in Property Reinsurance: Formulas and Applications*. Swiss Reinsurance Company.