Rating without data – how to estimate the loss frequency of loss-free risks

Michael Fackler, Aktuar DAV

freelance actuary
Munich, Germany
michael_fackler@web.de

Abstract

In insurance and even more in reinsurance it occurs that from a risk you only know that it had no losses in the past say seven years. Some of these risks are furthermore such particular that there are no similar risks to infer the loss frequency from.
In this paper we propose a loss frequency estimator which is able to handle such situations reasonably, just relying on the information coming from the risk itself. This estimator can be defined in a mathematically consistent manner and turns out to have desirable properties both in statistical and in business strategic sense.

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 (single risk or 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 in order 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 (apart from random fluctuations), i.e. there are no structural changes in the risk or the environment that make the loss experience a priori unreliable for forecasting.
Now assume that – unfortunately – in the observation period (the period of time for which data about the risk is available) the risk was loss-free. Say in past seven years no losses occurred (and the years before are either unknown, unreliable, or not representative).

This situation may appear theoretical but is not. Certainly it is very rare in the Personal Lines business where we can usually build large collectives of (quite) homogeneous risks providing plenty of loss data (with the important exception of new lines of business which in the first years are so small that they hardly produce losses). But in Commercial and Industrial Lines there are indeed particular risks being such different from all 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 risk portfolios with a very high deductible, long loss-free periods are rather normal than exceptional.

In this paper we propose a loss frequency estimator which 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 describes what many practitioners do in the described situation, leading to several heuristic properties a loss frequency estimator should have. Section 3 is about the volume dependency of loss frequencies. Section 4 develops a new type of loss frequency estimators and finds examples fulfilling the above heuristic requirements in a strictly mathematical sense. Section 5 gives practical applications for their use. Section 6 and 7 calculate bias and mean squared error of the developed estimators and find the optimal one in statistical sense. The numerical evaluation in section 8 completes the paper.

2 What practitioners do

Typically we expect well-established rating methods to work in all situations we come across. But what if the loss statistics consists of – no losses? Then the most common frequency estimator, the sample mean, is equal to zero, as well as the Maximum Likelihood estimators in the classical loss number models Poisson, Binomial, and Negative Binomial. (The first two 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 people in the market 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 being 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 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 has occurred. One could then (correctly in the sense of the sample mean) take 1 loss in eight years but it would appear more logical (and certainly be accepted by the insured) if the premium were somewhat increased after the occurrence of a loss. So one better keeps the additional year, thus assumes 2 losses in 9 years, etc. 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 assumes 0.5 losses in the 7 observed years.) The question is how to proceed 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 other variants but thinking about these three is sufficient to get quite some ideas about what properties we feel a loss frequency estimator should have:

- **One method:** It should work in all situations, not just be a method for loss-free risks. For all possible outcomes it has to be defined in advance what to do. This is not just a management issue: Only thoroughly defined methods can be (mathematically) evaluated.
- **Justifiability:** Many losses should yield a frequency equal or at least very close to the sample mean as this 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 charge more than the sample mean and in situations with many losses we charge about as much as the sample mean we are likely to get a positive bias. This is okay – a negative bias is not acceptable anyway – but we should possibly try to avoid a very high bias.
- **Monotonicity:** Although few losses are likely to be very much a random outcome (which justifies high uncertainty loadings), they must be rated cheaper than many losses, which are less random (lower uncertainty). Lack of losses does be a (statistical) item of information – despite of random errors: If the loss frequency is say one loss per year then seven years without a loss are in theory possible but extremely unlikely. 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 but wants to renew the policy though. He/she is indeed prepared to pay a higher premium next year, but the increase must appear “reasonable”.
- **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 with 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 Risk volume

The above pragmatic considerations will turn out to be translatable into strict mathematics but first we need to introduce an important item: the volume dependency of loss frequencies. Typically risks, especially 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 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. We assume θ 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 loss number (a random variable).

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

Now regard the observation period as a whole. Its volume, frequency, and loss number are the sums of the respective quantities of the single years: $v_+ := v_1 + \ldots + v_k$ and analogously for λ_+, N_+. As above we have $E(N_+) = \lambda_+ = v_+\theta$. N_+ is an unbiased estimator for λ_+.
N_+/v_+ is a further unbiased estimator for θ. This is the already mentioned **sample mean** of the observation period. It is the volume-weighted average of the above estimators N_i/v_i.

If we now want to predict the outcome of a future year with (known) 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 v_eN_i/v_i or any weighted average thereof, in particular the (re-scaled) sample mean $v_eN_+/v_+ = N_+/k_+$.
Here we have introduced the **volume-weighted number of years** $k_+ := v_+/v_e$, 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 are very easy. However, it shall be noted that in practice it may be difficult to determine an adequate volume measure. Many readily available and well-established ways to quantify whether risks are small or large contain both the increase of the loss sizes (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 available one has to factor out the inflation, which can be difficult because inflation may vary a lot from business to business. Note that this uncertainty is not particular to our rating problem, it arises independently of whether there is plenty of loss experience or none at all in any rating situation requiring a separate modelling of frequency and severity.

Although volumes quantifying only the loss frequency – let us call them **frequency volumes** – have lower yearly increases than those containing also the inflation it would be wrong to assume that k and k_+ are anyway such close that it is not worth calculating the latter. They can be surprisingly different, as can be seen from the following example:

Assume a steadily growing portfolio with (frequency) volumes $v_1, \ldots, v_k$ being a geometric sequence: $v_{i+1} = v_i\,(1+s)$. Suppose loss reporting to be 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 $k_+ = \dfrac{1-(1+s)^{-k}}{s}$

E.g. if the yearly increase is 10% then k = 7 years become k_+ = 4.9 years. If the yearly increase is 20% then k = 10 years become k_+ = 4.2 years, 20 years become 4.9 years. In practice all periods of steady and rapid growth will come to an end, but as long as they last k_+ will be much smaller than k. It is even bounded from above – from the formula we see that k_+ must be smaller than 1/s.
What in practice lets further increase the difference between k and k_+ is the fact that there is always a 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 could be used at the earliest to rate the year k+2. Thus in the above situation we would have $v_e = v_{k+2}$, which leads to values for k_+ being lower by the factor 1/(1+s).

Having discussed the denominator k_+ of the sample mean now we regard the numerator N_+.

4 Designing the estimator

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: We call the estimator $g(N_+)/k_+$ the **amended sample mean** (**ASM**) and the rating method applying this estimator **ASM method**. The function g is called **amending function** (or **sequence**).

Now it is possible to translate the above requirements for the frequency estimator into criteria for the amending function. We call an amending function **admissible** if it satisfies:

$g(n)$ is defined for all $n = 0, 1, 2, \dots$	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 condition from section 2 ("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 pretend the desired ratio. Then k_+ equals 5 in the first case and 10 in the second one. The estimator for λ_e is $g(0)/5$ in the first case, which is twice as much as $g(0)/10$.

Excursus to Credibility:

As said at the beginning we cannot rate the risk with the help of data from other risks, therefore it is in particular impossible to apply Bayes/Credibility rating. Well, being this such a strong method for risks having too few data for being rated independently – is it possibly worth trying to find a Credibility-like frequency estimator for our situation? That would be a formula looking like a Bayes premium for one of the risks of a collective, where the parameters that normally are 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 ASM? If not, is it similar? Is it a better approach?

Let us regard the classical Bühlmann-Straub model for (Poisson distributed) loss frequencies (see Bühlmann & Gisler [1, section 4.10]). Translated to our notation in this Credibility model N_+ would be replaced by a linear combination $wN_+ + (1–w)v_+\theta_0$ with θ_0 being the "global" average frequency per volume unit of the collective. The weight w has the structure $w = v_+ / (v_+ + \text{const.})$.

This is indeed a very smooth-looking formula, also monotonic, etc., different from the amending functions described above but not so much different. However, this and other Bayes models are optimised for the collective, not for the single risk. In particular the model is unbiased on collective level but on the individual level of our risk the bias is equal to $(1-w)v_+(\theta_0–\theta)$. In order to be on the safe side we would have to "select" a very high θ_0 but then we are likely to get a bias being too high to be acceptable.
In short, Bayes does not seem to be the best option in our situation where we could "simulate" a surrounding collective but do not have it. We have to optimise the estimator of our individual risk without regarding other risks.

So we go on trying to find 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.
"Additional year" is not a candidate, yielding the formula $(N_++1)/(k_++1)$, which turns out to be Credibility-like as described above with weight $w = v_+ / (v_++v_e)$ and global frequency $v_e\theta_0 = 1$.

As already stated g_1 seems 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 which we have already defined precisely. We will have to quantify the bias but first we should get a clearer concept for the smoothness.

Regard the following typical situation: A risk having suffered n losses in k_+ years was rated $g(n)/k_+$. We write the risk, then a new loss occurs. To rate the treaty for the next renewal of the insurance note that we now have n+1 losses in $(k+1)_+$ years leading to the frequency $g(n+1)/(k+1)_+$. Especially when one has several observation years and/or rapid volume growth k_+ and $(k+1)_+$ will be close, which means that the relative change in the premium will be about $g(n+1)/g(n)$. Hence this ratio can be used as a benchmark for premium increases after one new loss. The following properties come into mind which should ideally be fulfilled to avoid anger among clients:

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

Interpretation:
(1) The premium increases are similar to the loss record increases.
(2) The more losses we have already had the lesser is the 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, rising less steeply.)
(3b) If a loss occurs the premium might double but never 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.)

Recall that for large n we have g(n) = n. We call the integer d wherefrom g equals the identity the **dimension** of g. In dimension d we have to determine the d values g(0), …, g(d–1).
Condition (3a) together with "positive bias" yields the inequality $(n+1)/g(n) \leq g(n+1)/g(n) \leq (n+1)/n$, 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, …, i.e. for all larger integers. Thus for an 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 ones 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) yields $g(0) \leq 1/2$. (3) yields $g(0) \geq 1/2$. Hence we have a unique solution g(0) = 1/2 which leads to the already known function 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 $g(0) \geq g(1)/2$ and $g(1) \geq 1$. The second inequality is already known (positive bias). Both inequalities together yield $g(0) \geq 1/2$. (2) means $3/2 \leq 2/g(1) \leq g(1)/g(0)$. The left inequality yields $g(1) \leq 4/3$ so altogether we have $1 \leq g(1) \leq 4/3$. The right inequality yields $g(0) \leq g(1)^2/2$ so altogether we have $g(1)/2 \leq g(0) \leq g(1)^2/2$. There are (infinitely) many solutions:
- The cheapest choice for g(1) is 1, leading back to the one-dimensional g_2.

- The largest choice for g(1) is 4/3 = 1.33. Then we have $2/3 \le g(0) \le 8/9$ with the most expensive variant g(0) = 8/9 = 0.89, which we call g_3. The first four values of g_3 are a geometric sequence where the ratio of subsequent elements equals 3/2, which means that if 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) be a geometric sequence of second order. From the values g(2), g(3), and g(4) we see that the ratio of second order (g(n+2)/g(n+1))/(g(n+1)/g(n)) of this sequence equals 8/9, which leads to the first two elements g(1) = 32/27 = 1.19 and g(0) = 4096/6561 = 0.62. Here the ratio of subsequent elements decreases slowly (by the factor 8/9). It is easy to verify that all desired conditions are fulfilled. We call this function g_4.

If we regard 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) yields an upper bound which between zero and d+1 is a geometric sequence with ratio (d+1)/d. E.g. for d = 3 the function starts with the values g(0) = 81/64 = 1.27, g(1) = 27/16 = 1.69, g(2) = 9/4 = 2.25, g(3) = 3, g(4) = 4. Here the ratio of subsequent elements is 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 assumes more than one loss. We call it g_5.

The upper bounds of higher dimensions d yield even lower premium increases but the initial values become very high: One calculates easily $g(0) = d\,(1+1/d)^{-d} > d/e$ with the Euler number e = 2.718. For large d these functions will fail condition (1) and anyway have an unacceptably large bias. One will have to look for cheaper variants, maybe geometric sequences of higher order like g_4.
It is obvious that there is a trade-off between bias and smoothness: Cheap amending functions must have low initial values but then increase sharply. Functions increasing smoothly from the beginning need to start at a more expensive level.

See the table displaying 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 “bias” and “justifiability” and is indeed very smooth.

n	**0**	**1**	**2**	**3**	**4**
$g_1(n)$	1	2	3	4	5
$g_2(n)$	0.5	1	2	3	4
$g_3(n)$	0.89	1.33	2	3	4
$g_4(n)$	0.62	1.19	2	3	4
$g_5(n)$	1.27	1.69	2.25	3	4
$g_6(n)$	0.86	1.39	2.11	3	4

We have added one more variant: The first 6 elements of g_6 are a geometric sequence of second order. It is close to g_3, a bit cheaper for loss-free risks but more expensive for 1 and 2 losses. The surcharges after a new loss decrease very smoothly: 62%, 52%, 42%, 33%, …, which could be preferred to 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%).

Before we start with the (quite technical) discussion of the bias let us illustrate the ASM method by calculating some examples. These are inspired from actual cases rated with a variant of the method.

5 Examples

We will apply $g = g_3$. The calculation with other variants is analogous, one just has to replace the values of the function g.

1. Catastrophe reinsurance ("NatCat")

Regard a catastrophe excess of loss reinsurance treaty covering natural disasters with retention 50 and cover 100 (say million Euro).
The treaty protects against accumulation losses (such a loss is the aggregate of all single losses stemming from the same natural catastrophe) that exceed 50 by paying the exceeding part up to a maximum payment of 100. One calls this type of reinsurance a layer (ranging from loss size 50 to loss size 150) and writes shortly 100 xs 50.

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, so 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 portfolio growth). About the time before we either have no data or do not want to use it because the portfolio was substantially different then.

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

We have $k_+ = 10$, hence we get the frequency $g(0)/10 = 0.89/10 = 8.9\%$.

To get the net premium we need an assumption about the average loss of the treaty (loss severity).

- If we have no idea we could go the extremely conservative way and assume that all losses are total losses. Then we get a net premium of 100 x 8.9% = 8.9
- Although external knowledge could not help at all to assess the loss frequency it might help to assess the average loss: World-wide experience with NatCat reinsurance shows that the loss size distributions of such layers are quite heavy-tailed and fairly modelled by (European) Pareto distributions $P(X \leq x) = 1 - (\text{deductible}/x)^{\text{alpha}}$ with a Pareto-alpha close to 1. To be on the safe side we could use alpha = 0.8, which is a very heavy-tailed distribution yielding an average loss of 61.3 in the layer. This leads to a net premium of 5.5.

Now assume a loss occurs and the reinsurance treaty is to be renewed. Then we have 1 loss in $k_+ = 11$ years, therefore we rate the frequency $g(1)/11 = 1.33/11 = 12.1\%$, which is an increase of 36% compared to the year before. This is a notable surcharge but due to the high uncertainty inherent in catastrophe reinsurance this kind of business tends to accept high fluctuations of technical premiums.

The net premium now is 12.1 if we assume total losses only and 7.4 if we use the Pareto model.

This approach might appear inexpensive on the frequency side but it is certainly quite conservative on the severity side. If one talks to practitioners many will confirm that if in the above rating situation one had to come up with a totally judgemental net premium one would probably charge more than what we have calculated here. Thus with the ASM method we are able to be much more technical and at the same time cheaper than with "pragmatic" rating approaches.

2. Fire risk

Regard a per risk excess of loss treaty 4 xs 1 (say million dollar) covering a Fire portfolio. Per risk means that unlike the first example the cover applies to single losses (affecting single objects in the portfolio), not to accumulation losses from catastrophes.

Assume you got information about three years only and the frequency volume grows rapidly. Say $v_1 = 0.8$, $v_2 = 1.0$, $v_3 = 1.2$, and $v_e = 1.35$. Then we have $v_+ = 3.0$ and $k_+ = 2.22$.

Suppose the list of the biggest losses ("as if", having taken account for inflation) in these three years is as follows: 4.5 0.1 0.1

That means that there was one huge loss while all other losses were far below the retention 1. It seems hopeless to quote a treaty with such poor data from such a short observation period. However, if we have knowledge about the loss size distribution of this kind of portfolio we can give it a try: Experience from Property reinsurance tells that Fire per risk layers mostly have somewhat heavy-tailed distributions, however, less heavy tails than NatCat layers. In many markets European Pareto fits yield alphas in the range of 2, maybe about 1.5 for some Industrial Fire portfolios. This applies to layers in the million Euro or US Dollar range, for retentions higher than about half a million.

Assuming a Pareto distribution with alpha = 1.3 should be well on the safe side (unless the portfolio is mainly Industrial Fire). Note we can apply this distribution to any layer with a retention of about 0.5 or more. To be on the safe side we choose the somewhat higher so-called loss threshold 0.6 and proceed as if we had to quote two layers, the really existing 4 xs 1 and an artificial lower layer 0.4 xs 0.6:

- We first assess the frequency of the latter: g(1)/2.22 = 1.33/2.22 = 60% (it suffered one loss).
- We do not need the average loss of this layer but only the frequency of the total losses (all losses exceeding 1), as this 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). We simply have to multiply the loss frequency of the lower layer by the factor $(0.6/1)^{\text{alpha}}$ and get 31%.
- The average loss of the layer 4 xs 1 according to the Pareto model is 1.28.
- Altogether we get a net premium of 0.39, which is about 10% of the layer capacity 4.

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 impose high premiums and to recoup some money quickly.
The assessment of the technical premium should be free from such considerations. Instead one should try to distinguish what in the loss record is rather prone to random errors and what is less. The size of the biggest loss is very random – note we have a heavy-tailed loss size distribution. But the total absence of other losses in the million dollar range, and somewhat below, is unlikely to be just a random fluctuation, which is taken into account by assessing 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 net premium of 0.58 – still less than what quite some practitioners would rate.

Bear in mind that if a new loss of say 0.8 occurs one has to increase the technical premium, as the loss exceeds the 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 necessary to be coherent. If the model parameters are changed (say to a new threshold of 0.81) every time the model leads to a premium increase which is hard to enforce one is most likely to loose money in the long run.

Note that this ASM rating required extremely few information: We did not use the exact loss sizes but just the information that there was a very big one and all others were far from affecting the layer. What we needed precisely instead was the frequency volume of the years of the observation period. In practice 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, which is, however, required for other reinsurance rating methods as well.

Layers like the one described in this example can also be found in Industrial primary insurance. For such risks ASM rating is in principle applicable. An adequate volume measure could be the aggregate sum insured, adjusted for inflation.

Fire is just an example. Of course we can rate other lines of business in the same way, as long as we have good market knowledge about tails of loss size distributions.

3. Liability

The Third Party Liability lines ("long-tail business") pose an additional problem as here late losses occur – losses that are known or reported with a a delay of several months or even years. This is part of the IBNR problem, normally the minor part compared to the run-off of the case reserves. However, if we want to apply the ASM method to Liability risks we have to take the "loss number IBNR" into account. This is in principle easy but requires fair market experience about the run-off of loss numbers which may not readily be available – we need the so-called lag factors: If we regard the losses a risk produces in a certain year the lag factor a_j describes the percentage of losses hat are known j years later. (Lag factors are more frequently calculated for loss amounts but they exist for loss numbers as well.) Of course a_j is an (almost always) increasing sequence converging to 100% (more or less rapidly according to the kind of business, the country, deductibles, etc.).

Assume we have an observation period of k years with frequency volumes $v_1, \ldots, v_k$. Say the rating is done towards the end of the year after year k. Hence year i has a 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 a_{k+1-i} (plus random error).
If the rating is done earlier or later one has to adjust accordingly. If (very) lucky lag factors on a monthly basis are available, otherwise interpolations might be necessary.
It is easy to see that to account correctly for the late losses not yet known at the moment of the rating one simply has to replace the volumes v_i by the values $a_{k+1-i}\, v_i$. , then one proceeds with the rating as if there weren't any reporting lags.

6 Bias

In the following the statistical properties of N_+ and $g(N_+)$ are studied. 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. Again for convenience we will sometimes drop the "$_+$" and write λ for λ_+ and N for N_+.

An ASM of dimension d (or less) can be represented as $g(N) = N + \sum_{j=0}^{d-1} r_j \chi_j(N)$ with $\chi_j(N)$ being the function that equals 1 if N = j and equals 0 else. Then we have $g(n) = n + r_n$ for $n < d$ and $g(n) = n$ for $n \geq d$. The coefficients r_j defining g are nonnegative ("positive bias").

With the probability function $p_j = P(N=j)$ we get: $E(g(N)) = \lambda + \sum_{j=0}^{d-1} r_j p_j$, therefore

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

Proposition 1: The ASM with the lowest bias is $g_2(N)/k_+$.

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, hence we are done.

The fact that g_2 is optimal in terms of the bias does not mean that we should use this amending function anyway. If we regard smoothness as very important we might in return accept a somewhat higher bias. E.g. if we want the maximum increase g(n+1)/g(n) to be less than 2 we cannot use g_2 but have to look for other amending functions with low bias. To be flexible we try to calculate the bias for arbitrary ASMs.

The bias is a (linear) function of the probabilities p_j. To calculate these we need a distribution model for N_+, hence for the N_i. As usual we assume the latter as independent. Now let us apply the three most common models: Poisson, Binomial, and Negative Binomial.

A) Poisson

If the N_i are Poisson distributed with expected value $\lambda_i=v_i\theta$ 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 $\lambda=E(N_+)$, not the single λ_i.

B) 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 have the same value q then N_+ is indeed Binomial with parameters $m := m_+ := m_1 + \ldots + m_k$ and q. This is in fact the most interesting case in actuarial practice. It can be interpreted as follows: In year i the risk consists of m_i insured items suffering each either a loss (with probability q) or no loss. What changes over the years is the number of insured items. Hence m_i is a measure for the volume of the risk in year i. The expected values are $\lambda_i = m_i q$ and $\lambda = mq$. The probability function of N_+ is $p_j = \binom{m}{j} q^j (1-q)^{m-j}$ and in particular we have

$p_0 = (1-q)^m$, $p_1 = mq(1-q)^{m-1}$.

To be easier to compare to the Poisson case we change the parametrisation replacing the parameter q by the expected value λ: $p_j = \binom{m}{j}\frac{\lambda^j (m-\lambda)^{m-j}}{m^m}$, $p_0 = (1-\frac{\lambda}{m})^m$, $p_1 = \lambda(1-\frac{\lambda}{m})^{m-1}$.

As above we get formulae that require (together with the second parameter m) λ but not the single λ_i.

If m is large (like e.g. in group Life insurance) the Binomial distribution is very similar to 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 one then we have m = k and the single years have a Bernoulli distribution (with either one loss or no loss). This model can be applied to Stop Loss reinsurance. A Stop Loss is a layer that applies on the aggregate of all losses occurring in a year in a line of business. I.e., if the total loss amount exceeds the layer retention we have a loss to the layer, otherwise the year is loss-free. In the cases where the loss probability can be assumed to be constant over the observation period ($\lambda_i = q$) the adequate model for N_+ is Binomial(k, λ), $\lambda = kq$.

C) Negative Binomial

Among the many existing parametrisations we choose $p_j = \binom{\alpha+j-1}{j}\left(\frac{\alpha}{\alpha+\lambda}\right)^{\alpha}\left(\frac{\lambda}{\alpha+\lambda}\right)^{j}$

where λ is the expectation and $\alpha>0$ 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 not in all cases again Negative Binomial. This applies unfortunately also to the most interesting case in actuarial practice, the well-established Poisson-Gamma model 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 has a Poisson model as above but the frequency per volume unit is not constant any more, instead it fluctuates according to a Gamma distribution with expected value θ and shape parameter α (see Mack [3, section 1.4.2]). We will regard only this Negative Binomial case with constant α.

Although (unless the v_i are equal) N_+ is not Negative Binomial it is not too difficult to calculate the probability function of N_+. 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. Luckily 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 $p_{i,j} = P(N_i = j)$ and $p_j = P(N_+ = j)$, $i = 1, \ldots, k$. Then we have

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

Proposition 2: For the probability function of N_+ we have

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

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

$$\text{(III)} \quad \frac{\alpha\lambda}{\alpha+\lambda} \le p_1/p_0 = \sum_{i=1}^{k}\frac{\alpha\lambda_i}{\alpha+\lambda_i} \le \frac{k\alpha\lambda}{k\alpha+\lambda}$$

Proof: Note that all factors in these formulae are positive. Let us first prove the equations.

The equation in the middle of (I) means that p_0 is the product of the $p_{i,0}$. As for equation (II), one loss in all years together means that there is one year i suffering a loss and all other years are loss-free, thus $p_1 = \sum_{i=1}^{k}\left(p_{i,1}\prod_{l\neq i}p_{l,0}\right) = \sum_{i=1}^{k}\frac{p_{i,1}}{p_{i,0}}\prod_{i=1}^{k}p_{i,0}$. If we plug in what we already have we are done. The third equation is the quotient of the first two. Let us now regard the inequalities.

(I): If we take the left inequality to the power $-1/k\alpha$ we get $\frac{\alpha+\lambda/k}{\alpha} \ge \left(\prod_{i=1}^{k}\frac{\alpha+\lambda_i}{\alpha}\right)^{1/k}$, 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 $-1/\alpha$ we get

$$\prod_{i=1}^{k}\frac{\alpha+\lambda_i}{\alpha} = \prod_{i=1}^{k}\left(1+\frac{\lambda_i}{\alpha}\right) \ge 1+\sum_{i=1}^{k}\frac{\lambda_i}{\alpha} = \frac{\alpha+\lambda}{\alpha}, \quad \text{which is trivial.}$$

(III): 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 inequality 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, (II) is the product of (I) and (III).

(I) and (III) mean that p_0 and p_1/p_0 lie between the corresponding values of the two Negative Binomial distributions with common expected value λ and shape parameters α and $k\alpha$, respectively. The latter is the distribution of N_+ in case the λ_i (i.e. the v_i) are equal, the former is the limiting case if one λ_i (v_i) is much greater than the others. One could say that in a way N_+ is a blend of these two extreme cases.

Whether the derived inequalities are good, i.e. yield small intervals for p_0 and p_1, depends on the parameters. We will see that in realistic situations this is often the case, but before doing numerical examples we finally come to the classical mathematical criterion for estimators, which we have not regarded yet on our way to the ASM model but are able to quantify now.

7 Mean squared error

The most common way to measure the accuracy of an estimator is the mean squared error (MSE). The results of the last section (together with some more inequalities) will enable us to calculate 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 (at least for the dimensions 1 and 2).

We have MSE(g(N)) = E((g(N)–E(N))^2) = Bias^2(g(N)) + Var(g(N)).

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

Lemma 1: For j = 0, …, d–1 we have

$E(\chi_j) = p_j$, $\quad Var(\chi_j) = p_j(1-p_j)$, $\quad Cov(\chi_j, \chi_i) = -p_j p_i$ for $j \neq i$, $\quad Cov(N, \chi_j) = p_j (j-\lambda)$.

Proof: The (co)variance formulae follow immediately from Cov(A, B) = E(AB) – E(A)E(B).

Lemma 2: The variance of N_+ is equal to

Poisson: λ

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

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

We call κ the **volume homogeneity coefficient**. We have $1 \le \kappa \le k$. $\kappa = k$ iff the v_i are equal. The lower bound 1 is (approximately) taken on if one v_i is much greater than the other ones.

Proof: The Poisson and Binomial formulae are well known. (Recall the N_i are independent.) In the Negative Binomial case for each N_i we have $Var(N_i) = \lambda_i + \frac{\lambda_i^2}{\alpha} = \theta v_i + \frac{(\theta v_i)^2}{\alpha}$. $Var(N_+)$ is the sum of these variances, thus $Var(N_+) = \theta\sum_{i=1}^k v_i + \frac{\theta^2}{\alpha}\sum_{i=1}^k v_i^2 = \theta\sum_{i=1}^k v_i + \frac{\theta^2}{\kappa\alpha}\left(\sum_{i=1}^k v_i\right)^2 = \lambda + \frac{\lambda^2}{\kappa\alpha}$

The stated inequality is equivalent to $\sum_{i=1}^k v_i^2 \le \left(\sum_{i=1}^k v_i\right)^2 \le k\sum_{i=1}^k v_i^2$, where the left inequality is trivial and the right one is an application of the inequality of the arithmetic and the quadratic mean. What was stated about the cases of equality is clear therewith.

The closer the volumes of the single years the larger is κ. Values very close to k do occur in practice. However, the opposite case that a year dominates such that k is close to 1 is rather theoretical. In order to find out what values κ may take on in practice we come for a moment back to the heterogeneous example (in terms of the volume) of a geometrically growing portfolio with volumes $v_1, \ldots, v_k$ fulfilling $v_{i+1} = v_i\,(1+s)$. Some algebra yields $\kappa = \left(1+\frac{2}{s}\right)\frac{(1+s)^k-1}{(1+s)^k+1} < 1+\frac{2}{s}$

Thus here we have yet another upper bound, which is independent of k. E.g. for s = 50% κ cannot exceed 5, whatever large k is. As for lower bounds, there is no easy rule. However, noting that the above formula 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 non-geometrically increasing volumes should be not too different from this result.

For simplicity of notation we unite the three the cases of Lemma 2 in one:

Corollary 1: $\mathrm{Var}(N_+) = \lambda + c\lambda^2$, 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 & Meyers (see [2, section 3]) in order to give an intuitive meaning to the deviations of Negative Binomial and Binomial from the Poisson distribution.

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

Proposition 3: For the ASM $g(N) = N + \sum_{j=0}^{d-1} r_j \chi_j$ we have

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

$$\mathrm{Var}(g(N)) = \lambda + c\lambda^2 + \sum_{j=0}^{d-1}\big(r_j + 2(j-\lambda)\big) r_j p_j - \sum_{j,i=1}^{d-1} r_j r_i p_j p_i$$

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

Proof: Lengthy but straightforward calculation applying Lemma 1 and Corollary 1.

For completeness we state the corresponding formulae for the "quasi"-ASM $g_1(n) = n+1$:
$\mathrm{Bias}(g_1(N)) = 1$, $\mathrm{Var}(g_1(N)) = \mathrm{Var}(N) = \lambda + c\lambda^2$, $\mathrm{MSE}(g_1(N)) = \lambda + c\lambda^2 + 1$

Let us define **Δ(λ)** := MSE(g(N)) – MSE(N) = MSE(g(N)) – Var(N) = $\sum_{j=0}^{d-1}\big(r_j + 2(j-\lambda)\big) r_j p_j$

This is the deviation of the MSE of the ASM from the MSE of the sample mean. If Δ is negative then the ASM is more accurate than the sample mean (in terms of MSE). Δ enables us moreover to compare the ASMs among each other: The lower Δ, the lower is MSE(g(N)), the "better" is g(N).

Δ can be interpreted as a function of λ (and the r_j), indeed a smooth function, as the p_j are differentiable functions of λ.

$\Delta(0)$ is positive. This is not surprising, as for very low λ the (squared) bias must have an enormous impact on the MSE. But we see at a glance that Δ is negative for very large λ: If λ is greater than any of the $j+r_j/2$ we have $r_j+2(j-\lambda)<0$ for $j = 0, \ldots, d-1$. Thus for these λ the amended sample mean (despite of having a positive bias) is more accurate than the sample mean itself.
That means that in the Poisson and Negative Binomial case for any ASM there is an interval $]c; \infty[$ 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 assume arbitrarily large λ. Checking different values for the parameters of the loss number distribution and the ASM one finds (rather unrealistic) parameters where Δ can be proved to be always positive, on the other hand one gets many (rather realistic) examples where Δ takes on negative values. There is no easy rule apparent to distinguish these cases. Anyway, for practical purposes we need to know more about lower values of the frequency λ, be the distribution model Binomial, Poisson, or Negative Binomial.

We will see soon that there is no ASM having globally (for all λ) a lower MSE than the other ones. However, there is another optimality worth being regarded:
Recall that in principle we were not unhappy with the sample mean. Our intention was just to slightly amend it in a way to avoid zeros (and possibly to have a smooth rating scheme), but apart from that we tried to be 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. If we get lower values we are certainly not unhappy 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, no matter how much lower. In other words we could say: We want to "beat" the sample mean in as many situations as possible but do not bother about high "victories".
In this sense we could say than an amending function g is superior to another one if it has a wider range of values of λ where the ASM is better than the sample mean, i.e. $\Delta(\lambda) < 0$.

Proposition 4: The ASM having the widest range of frequencies λ where it is more accurate (in terms of MSE) than the sample mean is $g_2(N)/k_+$. The range of these frequencies is the interval $]0.25; \infty[$.

Proof: For g_2 we have $\Delta(\lambda) = (r_0 - 2\lambda)\, r_0\, p_0$, hence $\Delta(\lambda) < 0$ is equivalent to $\lambda > r_0/2 = g(0)/2 = 0.25$.

We will now show that for all amending functions and all $\lambda \leq 0.25$ we have $\Delta(\lambda) \geq 0$, which proves the proposition:
If we regard the summands of Δ for an arbitrary g we see that $(r_j + 2\,(j - \lambda))\, r_j\, p_j \geq 2\,(j - 0.25)\, r_j\, p_j \geq 0$ for all $j > 0$. Hence Δ is not smaller than the first summand $(r_0 - 2\lambda)\, r_0\, p_0$, which is nonnegative, as $\lambda \leq 0.25$ and for all ASMs $r_0 = g(0) \geq 0.5$.

Note that the proof yields the stronger result that on the interval $[0; 0.25]$ g_2 has the (strictly) lowest MSE among all ASMs: For any other ASM the second summand $(r_1+2(j - \lambda))r_1p_1$ will be positive.

To see that (as claimed above) there is no ASM which has globally (for all λ) a lower MSE than the other ones we just have to find an amending function g and a frequency λ such that $\Delta(\lambda)$ is smaller for g than for g_2: $\Delta(\lambda)(g) < \Delta(\lambda)(g_2)$.
We prove a much stronger result. Let g be any ASM different from g_2 and λ be any value greater than r_0 and any of the $j+r_j/2$. Then $\Delta(\lambda)(g)$ is (strictly) smaller than $\Delta(\lambda)(g_2) = (0.5 - 2\lambda)0.5p_0$: As above all summands of $\Delta(\lambda)(g)$ with $j>0$ are not greater than zero and at least $(r_1+2(j - \lambda))r_1p_1$ must be negative. To compare the first summand $(r_0 - 2\lambda)\, r_0\, p_0$ with $(0.5 - 2\lambda)0.5p_0$ note that $(r - 2\lambda)\, r\, p_0$ is a decreasing function on the interval $[0; \lambda]$. As $0.5 \leq r_0 \leq \lambda$, we are done.

The fact that g_2 is the best 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 1, we could regard 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.

Smoothness is a hybrid issue – half mathematics, half business. The criteria (1), (2), and (3a) reflect consistency in very much the same manner as say monotonicity, though one could possibly weaken them a bit. The last condition (3b) $g(1)/g(0) \leq 2$ is clearly 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 ASMs of dimension 1 (yielding extremely cheap rates for loss-free risks and huge increases after the first loss).
On the contrary if one wants to have as few clients as possible getting angry due to premium rises this could be ensured by an additional condition $g(n+1)/g(n) \leq b$ with a maximum increase $b < 2$, which excludes g_2 and a range of ASMs of higher dimension.

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 could introduce an additional condition $g(n) \geq b$ with a minimum rate $b > 0.5$, which excludes again g_2 and certain ASMs of higher dimension (yet not exactly the same as above).

Business strategies matter so much for the topic of this (in principle mathematical) paper as we deal with risks that are very difficult to assess (high uncertainty). If a risk can be priced 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 gets a large variety of offered premiums due to different decisions made by the offering companies, being all rather political than actuarial. In short, premiums of "difficult" risks, even if they are rated on a case by case basis, depend (implicitly) on the underwriting policy.

Strategies for insurance portfolios are very diverse according to the company's position in the market, the financial strength, etc., however, usually they are 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 high premium increases as soon as losses occur. That is exactly what very cheap (and steep) amending functions like g_2 and g_4 would do. 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 do not want to accept the premium surcharge and will look for a cheaper offer on 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_5 would do. The disadvantage of this strategy is that one will possibly not be able to write a lot of new business.

In spite of the many requirements the amending functions must meet they indeed have quite diverse "behaviour" corresponding to the most different business strategies. (g_3 and g_6 can be seen as somehow intermediate.) Thus one should always be able to find an ASM being suitable for the strategy and the environment. (If not, the aspired strategy might have a negative bias. In particular it is not possible to ensure both very cheap premiums and very moderate increases after a loss.)
In case the business policy 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.
Note that although being able to reflect some commercial aspects ASM rating is a step in the calculation of the **technical** premiums, which should not be mistaken for the **commercial** premiums possibly deviating from the technical ones for a number of (political) reasons.

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

In short, unless an insurer wants to be the overall cheapest (use g_2 then) it would be natural to restrict the range of the admissible ASMs according to the underwriting policy. However, if for such a subclass of ASMs we then want to find the best ones in terms of bias and mean squared error we come to complex optimisation problems not easily accessible for theoretical analysis. At this point numerical evaluation comes into play.

And even if we do not think about introducing additional conditions, numerical results from realistic cases are of interest anyway. Such results shall conclude this paper.

8 Numerical evaluation

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

Corollary 2: For the ASMs of dimension 1 or 2 we have

Bias(g(N)) = $r_0 p_0 + r_1 p_1$

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

MSE(g(N)) = $\lambda + c\lambda^2 + (r_0 - 2\lambda)r_0 p_0 + (r_1 + 2 - 2\lambda)r_1 p_1$

While bias, variance, and MSE are functions of λ only if N_+ is Poisson, respectively functions of λ and the aggregate volume m if N_+ is Binomial, the Negative Binomial case is generally far more complex: A look at the equations of Proposition 2 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 lengthy distinctions of cases having very similar vs. very different λ_i (i.e. v_i): The inequalities of Proposition 2 yield upper and lower bounds for p_0 and p_1 that are functions of α, k, and λ. These enable us now to derive approximations for bias, variance, and MSE of the ASMs of dimensions 1 and 2 being functions of α, k, κ, and λ.
If we regard the formulae in Corollary 2 closely we see that bias and MSE are (affine) linear in p_0 and p_1, thus we can calculate upper and lower bounds summand-wise – we just have to check whether the coefficients 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. Nevertheless we could again proceed summand-wise, though this will possibly not yield the optimal bounds.

What values for the parameters m, α, k, κ, and λ appear in practice?

We have often used the example of seven loss-free years. This is indeed a typical observation period, at least for the rating of layers of reinsurance. (The above Fire example with k=3 was an extreme case to demonstrate the power of the ASM method.) Typically one would have about 5 to 10 observed years, in the case of NatCat rating it happen to be 15 or more.
As explained in section 3, k_+ in practice is often somewhat lower than k, so 5 to 10 years correspond to about 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. We choose m=k which is as different as possible from Poisson and has the practical application Stop Loss reinsurance mentioned earlier.

For the Negative Binomial parameter α a wide range of values is possible. In practice very large values like 100 can be observed, leading again to results not much different from the limiting case Poisson. On the contrary one could 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 already found the realistic range of values: from 3 to k. Note that κ affects the variance and the MSE but not Δ. If we are most interested in the bias and the difference between the MSE of the ASM and that of the sample mean we do not need κ.

To illustrate the calculations some tables are provided in the appendix, displaying bias and MSE of the ASM g_3 together with the variance of the sample mean for a number of parameter constellations. For Negative Binomial we provide in addition the ranges for p_0 and p_1, here for bias and MSE the upper bounds according to the approximation are shown. The Binomial and Negative Binomial parameters are chosen a bit lower than typical in practice in order to be farther away from the Poisson case. Real cases will thus mostly lead to figures somewhere in between those of the displayed scenarios.

Results:

Not surprisingly all ASMs introduced in the paper are far better than g_1 as for bias and MSE. Adding always one loss is just too much, as well as would be (non-smooth) variants adding say always 0.7 or 0.5 losses. Amending functions, being essentially sequences g(n) converging to the sequence n, are indeed the much better solution.

For very low λ the bias is large but it decreases with growing λ. However, Δ decreases much more quickly (albeit not always monotonically) and if we look for the root (i.e. the value λ beyond which Δ 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 (the latter two calculated for Poisson and Binomial only). Let us call this value (the smallest root of Δ) the **critical frequency** of the ASM. Interestingly it hardly depends on the model assumptions. For a wide range of (realistic) parameters all three distribution models yield very similar results deviating by not more than 3% from the stated figures.
It seems that for realistic parameters Δ has only one root, i.e. the range of values λ satisfying $\Delta(\lambda) < 0$ is always an interval $]c; \infty[$ (for Binomial $]c; m[$, respectively). However, at least for the Binomial model this is a bit uncertain, as one can find (somewhat unrealistic) counterexamples where Δ has either no root (very expensive ASMs) or a second root such that for large frequencies Δ turns positive (yet for frequencies such large that they do not matter in practice).

Do the low frequencies where Δ is positive impair the quality of the ASM method right in the range of (small) frequencies we are mainly interested in? We can see easily that this is not the case: The critical frequencies of the analysed amending functions are (mostly much) lower than 100%. Now regard the rating situation of a risk having a frequency λ below the critical frequency of the ASM you have chosen. In particular λ is (mostly very far) less than 100%. Then the probability of no loss in the observation period is (mostly very much) greater than 33%, no matter which of the three distribution models (in case of realistic parameters) be adequate. In this situation the sample mean cannot be used at all for being zero, people will somehow “set” a greater frequency. Even in the case of one ore two losses the sample mean is mostly replaced by a higher figure (recall the workarounds of section 2), which means that in the described rating situation the sample mean is unlikely to be applied.
In short, there is no point in comparing the ASM to the sample mean in situations where the latter is not used: The cases where the ASM is less accurate than the sample mean in practice hardly exist.

We have discussed the frequency λ of the observation period throughout the second half of this paper because this is the parameter the properties of the mathematical model mainly depend on. Now we finally come back to the frequency λ_e of the risk which we wanted to assess. We want to get a feeling for what the results of this paper mean to the problem we started from.

A critical frequency of say λ = 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 the critical frequency corresponds to λ_e = 12%. That means that in this rating situation only if the risk has a loss frequency well below 12% the ASM method could possibly be regarded as inadequate in terms of MSE (and bias) – bearing in mind that the comparison with the cheaper sample mean does not make too much sense for such frequencies yielding so many cases where the latter will not be applicable.

Anyway, 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 amended sample mean method is able to produce from such few information (7 years loss-free) a fair rating for risks having a loss frequency of about 10%, i.e. one loss in 10 years.

If in the case of (re)insurance layers frequency extrapolation from a lower loss threshold is possible (see the Fire example of section 5) and if k_+ is rather large we might even be able to get a fair rate for layers with a frequency of 2%. Being possibly very expensive in the case of frequencies lower than this is, at least in reinsurance practice, not a problem because minimum premium requirements will not allow much cheaper layers anyway.
On the other hand if we think of a primary insurance risk having a loss frequency of 0.01% then even 25 loss-free years will lead to a far too high rating result. Such risks cannot be assessed at all without using "external" data, whatever the rating method. However, we would have a chance with ASM rating if we managed to collect a small portfolio of several hundred similar risks of this kind.

9 Conclusion

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

As for the economical view:

- It is very easy to implement into existing rating tools. Wherever the sample mean is applied (which is very often the case, even if 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 (we then have g(n) = n), which ensures continuity of the rating.
- In situations with poor data where traditional rating tools stop working properly one will always get a loss frequency, even if there are no losses.
- 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 contained.
- Generally by choosing the amending function the ASM method can easily be aligned with the underwriting policy.
- Last but not least: ASM rating saves a lot of time. The automatic assignment of a frequency is certainly much quicker than thinking about each case of very poor data individually.

As for the mathematical view:

- It is a consistent method for all situations, from those with great data to those with no losses.
- It takes correctly account for the volume dependency of loss frequencies.
- 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 rating to situations where underwriters and actuaries usually (have to) abandon their well-established methods.

References

[1] Bühlmann H, Gisler A (2005) A course in Credibility theory and its applications. Springer, Berlin Heidelberg
[2] Heckman PE, Meyers GG (1983) The calculation of aggregate loss distributions from claim severity and claim count distributions. PCAS LXX: 22-61
[3] Mack T (1999) Schadenversicherungsmathematik. Verlag Versicherungswirtschaft, Karlsruhe

Appendix

Table with scenarios for g_3

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

Appendix

Statistical properties of the amended sample mean

lambda	0%	20%	40%	60%	80%	100%	120%	140%	160%	180%	200%	220%	240%	260%	280%	300%
Poisson																
Bias	88,9%	78,2%	68,5%	59,8%	51,9%	45,0%	38,8%	33,4%	28,7%	24,6%	21,0%	18,0%	15,3%	13,0%	11,1%	9,4%
MSE	0,79	0,66	0,59	0,57	0,60	0,68	0,79	0,93	1,09	1,28	1,48	1,69	1,91	2,13	2,36	2,59
Var N+	0,00	0,20	0,40	0,60	0,80	1,00	1,20	1,40	1,60	1,80	2,00	2,20	2,40	2,60	2,80	3,00
DELTA	0,79	0,46	0,19	-0,03	-0,20	-0,32	-0,41	-0,47	-0,51	-0,52	-0,52	-0,51	-0,49	-0,47	-0,44	-0,41
Binomial																
Bias	88,9%	78,1%	68,1%	58,9%	50,4%	42,8%	35,9%	29,7%	24,3%	19,6%	15,5%	12,1%	9,2%	6,9%	5,0%	3,5%
MSE	0,79	0,66	0,57	0,52	0,50	0,52	0,56	0,62	0,69	0,77	0,84	0,91	0,97	1,02	1,05	1,06
Var N+	0,00	0,19	0,37	0,53	0,67	0,80	0,91	1,01	1,09	1,15	1,20	1,23	1,25	1,25	1,23	1,20
DELTA	0,79	0,46	0,20	-0,01	-0,17	-0,28	-0,35	-0,39	-0,40	-0,39	-0,36	-0,32	-0,28	-0,23	-0,18	-0,14
NegBin 1																
p0 min	100,0%	81,9%	67,2%	55,2%	45,4%	37,4%	30,9%	25,5%	21,1%	17,5%	14,5%	12,0%	10,0%	8,3%	6,9%	5,8%
max	100,0%	82,3%	68,3%	57,2%	48,2%	41,0%	35,0%	30,1%	26,0%	22,6%	19,8%	17,3%	15,3%	13,5%	12,0%	10,7%
p1 min	0,0%	15,6%	24,4%	28,8%	30,3%	29,9%	28,5%	26,5%	24,1%	21,7%	19,3%	17,1%	15,0%	13,1%	11,4%	9,9%
max	0,0%	16,3%	26,9%	33,6%	37,5%	39,5%	40,3%	40,1%	39,4%	38,3%	36,9%	35,3%	33,7%	32,1%	30,5%	28,9%
Bias	88,9%	78,6%	69,7%	62,0%	55,4%	49,6%	44,5%	40,1%	36,3%	32,9%	29,8%	27,2%	24,8%	22,7%	20,8%	19,1%
MSE	0,79	0,67	0,60	0,60	0,66	0,76	0,90	1,09	1,31	1,56	1,83	2,11	2,41	2,72	3,04	3,37
Var N+	0,00	0,20	0,41	0,63	0,85	1,08	1,32	1,56	1,81	2,07	2,33	2,60	2,88	3,16	3,45	3,75
DELTA	0,79	0,46	0,19	-0,03	-0,20	-0,33	-0,42	-0,47	-0,50	-0,51	-0,51	-0,49	-0,47	-0,44	-0,41	-0,38
NegBin 2																
p0 min	100,0%	82,3%	68,3%	57,2%	48,2%	41,0%	35,0%	30,1%	26,0%	22,6%	19,8%	17,3%	15,3%	13,5%	12,0%	10,7%
max	100,0%	83,3%	71,4%	62,5%	55,6%	50,0%	45,5%	41,7%	38,5%	35,7%	33,3%	31,3%	29,4%	27,8%	26,3%	25,0%
p1 min	0,0%	13,7%	19,5%	21,4%	21,4%	20,5%	19,1%	17,6%	16,0%	14,5%	13,2%	11,9%	10,8%	9,7%	8,8%	8,0%
max	0,0%	15,9%	26,0%	32,6%	37,0%	40,0%	42,0%	43,2%	44,0%	44,3%	44,4%	44,4%	44,1%	43,8%	43,3%	42,9%
Bias	88,9%	79,4%	72,1%	66,4%	61,7%	57,8%	54,4%	51,4%	48,8%	46,5%	44,4%	42,6%	40,8%	39,3%	37,8%	36,5%
MSE	0,79	0,68	0,64	0,68	0,80	0,97	1,21	1,51	1,87	2,27	2,71	3,19	3,70	4,24	4,82	5,42
Var N+	0,00	0,21	0,45	0,72	1,01	1,33	1,68	2,05	2,45	2,88	3,33	3,81	4,32	4,85	5,41	6,00
DELTA	0,79	0,46	0,19	-0,04	-0,21	-0,36	-0,47	-0,54	-0,58	-0,61	-0,62	-0,62	-0,62	-0,61	-0,60	-0,58