Thinning of loss counts and the Mixed Contagion model

Michael Fackler*

Abstract

We investigate representations of counting distributions that are largely invariant to thinning: where most parameters do not change when we, e.g., go from ground-up losses to the subset of losses exceeding a certain threshold. An important thinning-invariant parameter is the contagion, which was introduced for the Panjer (a,b,0) class of distributions, but can be defined in general.

We show that the finite moments of loss count distributions can be written as polynomials of the expectation with thinning-invariant coefficients. This yields representations where all key figures but one are unaffected by thinning. This helps compare heterogeneous data sets and respective models, like loss records having different reporting thresholds.

Further, we study the Mixed Contagion loss count model for portfolios of variable size, which strongly generalizes the common mixed Poisson-Gamma model, by allowing both for global and local fluctuations. We show how the three model parameters change when the losses are thinned, or restricted to a sub-portfolio, or both.

Finally, we illustrate how the parameters can be estimated by combining two data sources. This enables, e.g., inference by combining portfolio and whole-market data, even when the two data sets are thinned in a different way.

Keywords:
Discrete loss distribution, thinning, contagion

1 Introduction

1.1 Motivation

Statistical inference works well when the empirical data are abundant. We want to explore ways how to reasonably make inference of loss counts from scarce data, namely by comparing data sets that are different but related. We address questions like the following.

Problem 1.1. Can one infer loss distribution parameters from some subset? E.g.,

- market pool vs single portfolio,
- all losses vs subclass.

Problem 1.2. Can one combine data sets of somewhat different kind for inference? E.g.,

- all losses of a small portfolio,
- the million dollar losses of a large portfolio.

1.2 Objective

To this end, we study how loss count distributions transform when one compares the loss data produced by a portfolio to some subset, which is obtained by one of the following two options, or by a combination thereof:

- *thinning*,
- by considering a *sub-portfolio*.

Our aim is a realistic model that coherently describes both ways of selecting loss subsets, as a function of the (sub-)portfolio volume and of the probability underlying the thinning process.

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

1.3 Outline

This paper is organized as follows.

Section 2 treats *thinning*, slightly generalizing the common approach.

Section 3 explores the *contagion* of counting distributions, which is related to the variance, but more practical for many purposes and, in particular, thinning-invariant.

Section 4 develops representations of moments and probabilities that use mostly *thinning-invariant quantities*, which generalizes the representation of the second moment as a polynomial of the first.

Section 5 *relates* the loss counts of a portfolio to those of a single risk, which is not trivial.

Section 6 treats the *Mixed Contagion model*, a flexible and intuitive generalization of the mixed Poisson-Gamma model for the loss counts of portfolios of variable size.

Section 7 develops a procedure for the *inference* of the Mixed Contagion parameters from *two data sources*, which are related by similar overall characteristics, but difficult to compare as their losses are selected by different thinnings. A numerical example is given.

2 Thinning

Intuitively, thinning relates the losses produced by an insured portfolio to a subset, according to some *selection criterion*, e.g.,

- loss size (severity),
- loss cause,
- duration of run-off,
- claim components, e.g., the heads of damage in injury cases to be compensated by (M)TPL policies.

Thinning according to loss size is most common. If the losses produced by an insured portfolio constitute a *collective model* (iid loss severities independent of loss count), the subset of losses exceeding a chosen threshold constitutes a random draw, which makes the set/subset relation accessible for mathematical analysis. Other selection criteria can likewise yield such random draws.

Let us formalize thinning, generalizing a bit beyond the traditional setting, in order to embrace various ways to thin losses.

Definition 2.1. Consider the insurance losses produced by a certain portfolio of risks, in a certain period. The *loss count* is denoted by N.

For each loss we call the set of all relevant information *loss record*, symbolized by $\overrightarrow{X}$, while for its (ultimate) *size* we write X. The loss record reflects how the losses are administered and reported. Technically, it is an array of numerical and categorical items, hence the notation. It typically contains things like: policy information, loss cause, occurrence date, dates of payments and case reserves, split of the final amount paid (or of each payment) into heads of damage, etc. The ultimate loss size X is formally a function of $\overrightarrow{X}$, namely the sum of all payments made until final settlement. It may include certain allocated expenses according to the reporting standard.

A *selection criterion* is a binary function b of the loss records. If applied to the i-th loss, we get the Bernoulli distributed random variable $B_i := b\left(\overrightarrow{X_i}\right)$.

The *thinned loss count* (according to the selection criterion) is

$$\bar{N} := \sum_{i=1}^{N} B_i = \sum_{i=1}^{N} b\left(\overrightarrow{X_i}\right)$$

The loss records $\overrightarrow{X_i}$ constitute a *generalized collective model* (according to the selection criterion b) if the B_i are iid and independent of N. We then call $q := \mathrm{E}(B)$ the *selection probability.*

Example 2.2. If the loss sizes X_i constitute a classical collective model, all criteria $b\left(\overrightarrow{X}\right) = \widetilde{b}(X)$ referring only to loss size yield generalized collective models, in particular $b\left(\overrightarrow{X}\right) = \chi_{\{X>d\}}$ selecting the losses that exceed a certain threshold d.

One can, more in general, consider counting distributions without relating them to respective losses.

Definition 2.3. For a discrete random variable N we call any $\bar{N} = \sum_{i=1}^{N} B_i$, where the B_i are iid Bernoulli RVs independent of N, a *(proper) thinning* of N.

If the meaning is clear, we may drop "proper" and just speak of thinning, as is common. We want to emphasize, however, that there might be intricate real-world situations where the selecting B_i don't constitute a collective model, due to heterogeneity or dependencies. In a wider sense this is still thinning, but much harder to treat mathematically.

Before exploring in general how moments and parameters of original and (properly) thinned counting distribution are related, we study a particular representation.

3 Contagion

Definition 3.1. Let N be a counting RV having positive finite first and second moment. We define

Dispersion:

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

Overdispersion:

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

Contagion:

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

While (over)dispersion is a very common concept, contagion is less so. In generality it is defined and explored in Chapter 5 of Fackler (2017), however, at least its empirical counterpart is much older, appearing in some literature as *Poisson Index* (Ross and Preece, 1985). For the distributions of the Panjer $(a, b, 0)$ class, i.e., the Poisson, Binomial, and Negative Binomial model (Klugman et al., 2008), the contagion indicates whether losses (in an ongoing claims production process) become more or less likely after the occurrence of a loss, hence the name (Heckman and Meyers, 1983). Later the contagion was also used for mixed Poisson distributions (Meyers, 2007), where it equals the variance of the mixing distribution. In both cases it proved to be not only intuitive, but also convenient for notation and computation, helping treat a number of distributions in parallel.

As a multiple of the overdispersion, the contagion equals 0 in the Poisson case and is observed to be positive for a lot of real-world insurance data. Beyond that, it leads to a number of compact and intuitive representations. One immediately gets:

Proposition 3.2. *The first two moments of N can be written in terms of mean and variance or equivalently in terms of mean and contagion. If we call the mean λ and the contagion c, we have*

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

Example 3.3. This representation is known for the Panjer $(a, b, 0)$ class, where c equals 0 for Poisson, is the inverse of the shape parameter α for NegBin, and for Binomial $c = -\frac{1}{n}$. It can be used for a unified parameterization of the three models (Fackler, 2011). The *probability function* (pf) reads

$$p_k = (1 + c\lambda)^{-\frac{1}{c}} \frac{\lambda^k}{k!} \prod_{i=0}^{k-1} \frac{1 + ci}{1 + c\lambda}$$

which is not more complex than the traditional NegBin representations. The parameter space is intricate due to Binomial, namely $c \in \left(\left(-\frac{1}{\lambda}; 0\right) \cap \left\{\frac{1}{x} \mid x \in \mathbb{Z}^*\right\}\right) \cup [0; \infty)$, such that for treating Binomial alone the traditional parameterizations (n with p or $\lambda = np$) are preferable. Yet, if one wants to treat the whole $(a, b, 0)$ class, the parameters λ and c are at least as practical as other parameterizations of the class, see Table 1 in Fackler (2021) for an overview.

By replacing variance by contagion in Wald's equation $\mathrm{Var}(S) = \mathrm{Var}(N)\, E^2(X) + \mathrm{E}(N)\, \mathrm{Var}(X)$ and normalizing, one gets:

Proposition 3.4. *In any collective model* $S = \sum_{i=1}^{N} X_i$ *where loss severity* X *and loss count* N *have positive finite first and second moment, we have*

$$CV^2(S) = Ct(N) + \frac{1}{E(N)} \frac{E(X^2)}{E^2(X)} = c + \frac{1 + CV^2(X)}{\lambda}$$

This formula makes clear immediately that if the contagion is positive, the volatility of S cannot approach zero, however high the loss frequency is. If the latter is very large, the CV of S depends mainly on the loss count contagion, while the severity plays a minor role.

The most interesting property in our context is the following.

Proposition 3.5. *The contagion is invariant to proper thinning. With* $\bar{\lambda} = E(\bar{N})$ *we have*

$$E(N) = \lambda, \qquad Var(N) = \lambda + c\lambda^2$$

$$E(\bar{N}) = \bar{\lambda}, \qquad Var(\bar{N}) = \bar{\lambda} + c\bar{\lambda}^2$$

Proof. With $\bar{N} := \sum_{i=1}^{N} B_i$ and $\mathrm{E}(\bar{N}) = q\lambda$ we get

$$\mathrm{Ct}(\bar{N}) = \mathrm{CV}^2(\bar{N}) - \frac{1}{\mathrm{E}(\bar{N})} = \mathrm{Ct}(N) + \frac{1 + \mathrm{CV}^2(B)}{\mathrm{E}(N)} - \frac{1}{q\lambda} = c + \frac{1/q}{\lambda} - \frac{1}{q\lambda} = c$$

by using the preceding proposition. □

In particular, the contagion is equal for the respective loss count distributions of all layers of a (re)insurance program (provided the loss severities constitute a collective model, as is mostly assumed).

4 Polynomials

The above representations for the variances of N and $\bar{N}$ are *polynomials* of the respective expectations, notably having the *same* coefficients (1 and c). We now show that such *thinning-invariant polynomial representations* exist for higher moments, too.

Definition 4.1. For the *notation of moments* of N we use, for $k \geq 0$,

raw:

$$m_k = \mathrm{E}(N^k), \qquad \lambda = m_1 = \mathrm{E}(N^k)$$

central:

$$\mu_k = \mathrm{E}\left((N - \lambda)^k\right)$$

factorial:

$$\mu_{(k)} = \mathrm{E}\left(N_{(k)}\right) = \mathrm{E}\left(N(N-1)\ldots(N-k+1)\right)$$

and analogously $\bar{\lambda}$, $\bar{m}_k$, $\bar{\mu}_k$, $\bar{\mu}_{(k)}$ for thinned $\bar{N}$.

Moments of order 0 equal 1 and are sometimes used for notational convenience.

Lemma 4.2. *For any proper thinning* $\bar{N}$ *of* N*, we have* $\bar{\mu}_{(k)} = q^k \mu_{(k)}$ *with the selection probability* q*, as long as the moments are finite. Thus, the* normalized factorial moments *are proper-thinning-invariant:*

$$\frac{\bar{\mu}_{(k)}}{\bar{\lambda}^k} = \frac{\mu_{(k)}}{\lambda^k} =: \varphi_k$$

Proof. With the *probability generating function* (pgf) $P_N(z) = \mathrm{E}(z^N)$ of N, we have $\mu_{(k)} = P_N^{(k)}(1_-)$ and analogously for $\bar{N}$. The collective model $\bar{N} := \sum_{i=1}^{N} B_i$ yields the pgf formula $P_{\bar{N}}(z) = P_N(1 + q(z-1))$, where $q = \mathrm{E}(B)$. Taking the derivative (repeatedly), we get $P_{\bar{N}}^{(k)}(z) = q^k P_N^{(k)}(1 + q(z-1))$, hence the first formula, which also holds for $k = 0$. With

$$\frac{\bar{\mu}_{(k)}}{\bar{\lambda}^k} = \frac{q^k \bar{\mu}_{(k)}}{(q\lambda)^k} = \frac{\mu_{(k)}}{\lambda^k}$$

we are done. □

Proposition 4.3. *Let N represent a discrete loss distribution. Any finite k-th (raw, central, factorial) moment of N can be written as a polynomial of the expectation λ, having order k or less, whose coefficients are thinning-invariant. This means that for a proper thinning $\bar{N}$ of N, the corresponding moment has a polynomial representation with the same coefficients, just with $\lambda = E(N)$ replaced by $\bar{\lambda} = E(\bar{N})$. The coefficients are as follows.*

$$\mu_{(k)} = \varphi_k \lambda^k, \qquad \varphi_k = \frac{P_N^{(k)}(1_-)}{\left(P_N'(1_-)\right)^k}$$

$$m_k = \sum_{i=0}^{k} \gamma_{ki} \lambda^i, \qquad \gamma_{ki} = \left\{ {k \atop i} \right\} \varphi_i$$

$$\mu_k = \sum_{i=0}^{k} \varepsilon_{ki} \lambda^i, \qquad k \geq 2$$

with the Stirling numbers of second kind $\left\{ {k \atop i} \right\}$ *and with linear combinations* ε_{ki} *of the* φ_j*, namely, for* $0 \leq i \leq k$*,*

$$\varepsilon_{ki} = \sum_{j=0}^{i} (-1)^j \binom{k}{j} \left\{ {k-j \atop i-j} \right\} \varphi_{i-j}$$

Proof. The first formula is clear from the preceding lemma. The raw moment of order $k \geq 1$ can be written as a linear combination

$$m_k = \sum_{i=0}^{k} \left\{ {k \atop i} \right\} \mu_{(i)}$$

of the factorial moments of order up to k (Johnson et al., 2005), hence the second formula, which holds for $k = 0$ because $\left\{ {0 \atop 0} \right\} \varphi_0 = 1 \cdot 1 = 1$. For $k \geq 1$ one can, with $\left\{ {k \atop 0} \right\} = 0$, equivalently let the sum start from $i = 1$.

Now we express μ_k, $k \geq 2$, in terms of the raw moments and order for powers of λ.

$$\mu_k = \sum_{j=0}^{k} \binom{k}{j} m_{k-j} (-\lambda)^j = \sum_{j=0}^{k} (-1)^j \binom{k}{j} \sum_{l=0}^{k-j} \gamma_{k-j,l} \lambda^{j+l} = \sum_{i=0}^{k} \varepsilon_{ki} \lambda^i$$

where for $0 \leq i \leq k$ the coefficients equal

$$\varepsilon_{ki} = \sum_{j=0}^{i} (-1)^j \binom{k}{j} \gamma_{k-j,i-j} = \sum_{j=0}^{i} (-1)^j \binom{k}{j} \left\{ {k-j \atop i-j} \right\} \varphi_{i-j}$$

Finally, for the thinned moments the lemma yields $\bar{\mu}_{(k)} = \varphi_k \bar{\lambda}^k$, leading to the same polynomials for $\bar{m}_k$ and $\bar{\mu}_k$ as above, just with λ replaced by $\bar{\lambda}$. □

Corollary 4.4. *The above polynomials can be simplified as follows, for $k \geq 2$.*

$$m_k = \lambda + \left(\sum_{i=2}^{k-1} \gamma_{ki} \lambda^i \right) + \varphi_k \lambda^k$$

$$\mu_k = \lambda + \sum_{i=2}^{k} \varepsilon_{ki} \lambda^i, \qquad \varepsilon_{kk} = \sum_{j=0}^{k-2} (-1)^j \binom{k}{j} \psi_{k-j}, \qquad \psi_j = \varphi_j - 1$$

Proof. For $k \geq 2$ we have $\gamma_{k0} = 0$ as stated above, $\gamma_{k1} = \left\{ {k \atop 1} \right\} \varphi_1 = 1 \cdot 1 = 1$, and $\gamma_{kk} = \left\{ {k \atop k} \right\} \varphi_k = \varphi_k$, by using that $\left\{ {i \atop i} \right\} = 1$ for $i \geq 0$. Further we have

$$\varepsilon_{k0} = \binom{k}{0} \left\{ {k \atop 0} \right\} \varphi_0 = 0, \qquad \varepsilon_{k1} = \binom{k}{0} \left\{ {k \atop 1} \right\} \varphi_1 - \binom{k}{1} \left\{ {k-1 \atop 0} \right\} \varphi_0 = 1$$

Finally, we get

$$\varepsilon_{kk} = \sum_{j=0}^{k} (-1)^j \binom{k}{j} \left\{ \begin{matrix} k-j \\ k-j \end{matrix} \right\} \varphi_{k-j} = \sum_{j=0}^{k} (-1)^j \binom{k}{j} \psi_{k-j} = \sum_{j=0}^{k-2} (-1)^j \binom{k}{j} \psi_{k-j}$$

by using $\sum_{j=0}^{k} (-1)^j \binom{k}{j} = 0$ and $\psi_0 = 0 = \psi_1$. □

Let us for illustration give the polynomials for the raw and central moments, up to order five. Conversion of the lower raw / central / factorial moments into each other is rather easy and given, e.g., in Chapter 2 of Panjer and Willmot (1992); one does not have to use the complex general formulae derived above. After (quite) some algebra we get the following expressions, where we write the central moments in terms of the ψ_j, which yield somewhat simpler coefficients than the φ_j. The resulting formula for μ_2 gives indeed the variance expressed via the contagion $c = \psi_2$.

Corollary 4.5. *For counting distributions we have, as long as moments are finite,*

$$m_1 = \lambda$$

$$\begin{aligned} m_2 &= \lambda + \varphi_2 \lambda^2 \\ m_3 &= \lambda + 3\varphi_2 \lambda^2 + \varphi_3 \lambda^3 \\ m_4 &= \lambda + 7\varphi_2 \lambda^2 + 6\varphi_3 \lambda^3 + \varphi_4 \lambda^4 \\ m_5 &= \lambda + 15\varphi_2 \lambda^2 + 25\varphi_3 \lambda^3 + 10\varphi_4 \lambda^4 + \varphi_5 \lambda^5 \end{aligned}$$

$$\begin{aligned} \mu_2 &= \lambda + \psi_2 \lambda^2 \\ \mu_3 &= \lambda + 3\psi_2 \lambda^2 + (\psi_3 - 3\psi_2)\, \lambda^3 \\ \mu_4 &= \lambda + (7\psi_2 + 3)\, \lambda^2 + (6\psi_3 - 12\psi_2)\, \lambda^3 + (\psi_4 - 4\psi_3 + 6\psi_2)\, \lambda^4 \\ \mu_5 &= \lambda + (15\psi_2 + 10)\, \lambda^2 + (25\psi_3 - 35\psi_2)\, \lambda^3 + (10\psi_4 - 30\psi_3 + 30\psi_2)\, \lambda^4 + (\psi_5 - 5\psi_4 + 10\psi_3 - 10\psi_2)\, \lambda^5 \end{aligned}$$

Example 4.6. Two common distributions are examples where some moments don't have the maximum possible polynomial degree. For the *Poisson* distribution with pgf $P_N(z) = e^{\lambda(z-1)}$ one sees quickly that all φ_k equal 1. With all $\psi_k = 0$, one gets immediately

$$m_1 = \mu_2 = \mu_3 = \lambda, \qquad \mu_4 = \lambda + 3\lambda^2, \qquad \mu_5 = \lambda + 10\lambda^2$$

Distributions with maximum possible count n, like the *Binomial* distribution with (traditional) parameter n, have $\mu_{(k)} = 0$ for all $k > n$. Thus, all raw moments m_k can be written as polynomials of order n or less.

If a counting distribution has l parameters, it may be possible to parameterize it by the expectation λ together with the subsequent raw, or equivalently central, moments of order 2 to l. This leads to alternative parameterizations where all parameters but the expectation are invariant to proper thinning.

Corollary 4.7. *Suppose a counting distribution model can be parameterized by its first l (finite) raw moments $m_1, m_2, \ldots, m_l$, or equivalently by using some central moments μ_k instead of m_k, $k \geq 2$. Then the model can be parameterized by the expectation $m_1 = \lambda$ and the proper-thinning-invariant normalized factorial moments $\varphi_2, \ldots, \varphi_l$, or equivalently by using some $\psi_k = \varphi_k - 1$ instead of φ_k.*

For $l > 2$ there could be distributions where the first n moments don't constitute parameters, and even where they do, they may not be handy, leading to overly complex representations. For such cases the representations via the φ_k are instructive, but for practical use one will prefer other parameters. However, many two-parameter distributions can be parameterized compactly by expectation and variance and thus also by expectation and contagion, as was done above for the Panjer $(a, b, 0)$ class and occurs for mixed Poisson models.

Example 4.8. The simplest example where other parameters are preferable, is the Logarithmic distribution, where in a common parameterization (Klugman et al., 2008) with $\beta > 0$ we have the pf

$$p_k = \frac{1}{\ln(1+\beta)} \left(\frac{\beta}{1+\beta}\right)^k \frac{1}{k}, \quad k \geq 1; \qquad p_0 = 0$$

which could in principle be reparameterized by the expectation

$$\lambda = \frac{\beta}{\ln(1+\beta)}$$

but this would ease neither theoretical understanding nor practical use of the distribution.

Another case are *zero-modified* distributions (Klugman et al., 2008), where the *no-loss probability* p_0 is altered and the subsequent probabilities are rescaled accordingly. The *chosen* no-loss probability $\tilde{p}_0$ comes in as a new parameter; together with the parameters of the original distribution it constitutes a *natural* parameterization. For rather simple examples like Poisson and NegBin one can verify (with quite some effort) that one can reparameterize their zero-modifications via respective expectation and first normalized factorial moments, but the parameter conversion formulae are such intricate that the resulting pf's are over-complex. In zero-modified (and *zero-truncated*, where $\tilde{p}_0 = 0$) models one better works with the natural parameterization.

Remark 4.9. Beyond the φ_k, distributions can have other representations having mostly thinning-invariant parameters. Klugman et al. (2013) give (in Section 8.1) a special pgf representation that embraces a wide range of models, namely compound Poisson distributions and all distributions of the Panjer $(a, b, 1)$ class in the wide sense of Fackler (2021), i.e. Poisson, Bin, NegBin, Log, ETNB, and zero-modifications thereof:

$$P_N(z) = \alpha + (1-\alpha)\frac{\Phi(\beta(1-z)) - \Phi(\beta)}{1-\Phi(\beta)}$$

has parameters $0 \leq \alpha < 1$ (the no-loss probability), $0 < \beta < \beta_{\text{sup}} \leq \infty$, and possibly further parameters of the function Φ. β looks like a scale and may, in the special *non-zero-modified* case $\alpha = \Phi(\beta)$, be the expectation, but in many cases it is not.

If N is properly thinned with selection probability q, $P_{\bar{N}}(z)$ has the same structure with the same parameters of Φ, while $\bar{\beta} = q\beta$ and $\bar{\alpha} = P_N(1-q)$, see Lemma 8.1 of Klugman et al. (2013).

One could, however, replace the parameter α by

$$\eta = \frac{1-\alpha}{1-\Phi(\beta)}, \qquad 0 < \eta \leq \frac{1}{1-\Phi(\beta)}$$

which is, as one sees quickly, thinning-invariant. This yields a parameterization where all parameters but β are invariant to proper thinning. The pgf simplifies to

$$P_N(z) = 1 - \eta + \eta\,\Phi(\beta(1-z))$$

while the parameter space becomes a bit intricate with the range of η depending on β.

For orientation: $\eta = 1$ is the non-zero-modified case $\alpha = \Phi(\beta)$, where $P_N(z) = \Phi(\beta(1-z))$, while the maximum η corresponds to the zero-truncated case $\alpha = 0$.

Example 4.10. The *zero-modified Binomial* model with the three natural parameters n, p, and chosen no-loss probability $\alpha = \tilde{p}_0$, can accordingly be represented as

$$P_N(z) = 1 - \eta + \eta\,\{1 - p(1-z)\}^n$$

with $\beta = p$, $\Phi_n(s) = \{1-s\}^n$, and

$$\eta = \frac{1-\tilde{p}_0}{1-p_0}$$

with the no-loss probability $p_0 = (1-p)^n$ of the original Binomial model. The resulting pf has

$$\tilde{p}_k = \eta \binom{n}{k} p^k (1-p)^{n-k}, \qquad 1 \leq k \leq n$$

which is a lot simpler a representation than via the normalized factorial moments. The parameters n and η are thinning-invariant, while $\bar{p} = qp$.

5 Sub-portfolios

Let us now investigate a loss count reduction that is a bit like thinning, but not the same. How do the loss count moments / parameters change if we consider the losses produced by a sub-portfolio?

We focus on *representative* sub-portfolios. E.g., in a fairly homogeneous MTPL market with insurers having quite similar portfolios (and many insureds regularly changing their insurer), each such insurer has a representative share of the market, such that its loss production, and the underlying distributional properties, should be closely related to the overall ones. On the contrary, if an insurer were specialized on trucks or had a far higher or lower share of motorcycles than the market average, we could not consider it representative.

The easiest case is a homogeneous portfolio of v risks having equal characteristics, which mathematically means that they have identical loss distributions (not the same losses, but the same distributional laws). We call each such risk *unit*; if we understand how the loss count parameters are related between unit and portfolio, we can derive them for any sub-portfolio.

Let us for the moment assume that each unit loss count U_i have a Negative Binomial distribution with expectation θ. This (overdispersed) model is common, plausible, and has a very intuitive interpretation as a mixed Poisson-Gamma model (Klugman et al., 2008), i.e., the Poisson parameter is interpreted as a RV being Gamma distributed with expectation 1.

The portfolio loss count $N_v := \sum_{i=1}^{v} U_i$ has expectation $\lambda = v\theta$. What about the variance? There are two straightforward possibilities, being both plausible, but surprisingly different.

Example 5.1. $U_i \sim \mathit{Poi}\left(\theta Q\right)$ with a *common* (positive) mixing RV Q having mean 1 and (finite) variance β, reflecting *global* fluctuation. This can be interpreted as market-wide environmental changes from year to year, see Section 1.4.2 of Mack (1997), or as common shocks (Meyers, 2007). The units are independent *conditionally* on the common *market factor* Q, which implies $N_v \sim \mathit{Poi}\left(v\theta Q\right)$ and

$$\mathrm{Var}\left(U_i\right) = \theta + \beta\theta^2, \qquad \mathrm{Ct}\left(U_i\right) = \beta, \qquad \mathrm{Var}\left(N_v\right) = \lambda + \beta\lambda^2, \qquad \mathrm{Ct}\left(N_v\right) = \beta$$

Here the contagion does not depend on the volume – it is an invariant just as in the case of proper thinning studied above.

By changing just a detail we get:

Example 5.2. $U_i \sim \mathit{Poi}\left(\theta R_i\right)$ with *individual* iid mixing RVs R_i having each mean 1 and variance κ, reflecting *local* fluctuations. The units are (unconditionally) independent, variances are additive, such that

$$\mathrm{Var}\left(U_i\right) = \theta + \kappa\theta^2, \qquad \mathrm{Ct}\left(U_i\right) = \kappa, \qquad \mathrm{Var}\left(N_v\right) = \lambda + \frac{\kappa}{v}\lambda^2, \qquad \mathrm{Ct}\left(N_v\right) = \frac{\kappa}{v}$$

Here the contagion shrinks with the volume and is, for very large volumes, close to the Poisson case $\kappa = 0$.

The fundamental difference between the two cases is reflected in the variances, but it is more clearly revealed by the contagion, which proves to be a most intuitive quantity.

6 Mixed Contagion model

In the insurance practice it is felt that independence of units is often too optimistic an assumption, while a quite volatile common market factor seems to exaggerate dependency – something intermediate would be more plausible. We now present such a model, see in the following Section 5.4 of Fackler (2017). It embraces both above examples and generalizes them in many ways. Again, there is a global fluctuation, but there is a further source of volatility (and dispersion), such that one can have a considerable overall volatility (which is often observed) even when the market factor varies few.

Definition 6.1. Consider a family $\left(N_v\right)_v$ of discrete loss distributions for risks having volume v, where the volumes may vary across a certain range of admissible values. We call this a *Mixed Contagion model* (MCM) if there is a nonnegative RV Q having mean 1 and (finite) variance β, such that for a certain $\theta > 0$ and a real γ, for all possible values of v and conditionally on Q, the distribution of N_v has the following (finite) moments:

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

such that, in particular, $\mathrm{Ct}\left(N_v \mid Q\right) = \frac{\gamma}{v}$ does not depend on Q.

Proposition 6.2. *In any Mixed Contagion model we have*

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

$$D(N_v) = 1 + \theta\gamma(1+\beta) + v\theta\beta, \qquad OD(N_v) = \theta(\gamma(1+\beta) + v\beta)$$

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

Note that $\mathrm{Var}(N_v \mid Q)$ grows linear with v, which corresponds to the idea that, given Q, N_v (for integer v) behaves like a sum of v iid units. The latter implies that all sub-portfolios have the same characteristics and MC parameters. After averaging over Q, the variance has instead a component in v and one in v^2, and the respective contagion is, with $\kappa = \gamma(1+\beta)$, a generalization and "mixture" of the two above examples, hence the name Mixed Contagion. The two components of the contagion together with the mean yield a model in three parameters θ, β, γ (or equivalently κ). More precisely, with only the first two moments specified, the MCMs constitute a class of models, leaving room for a number of variants.

Example 6.3. The natural and very intuitive MCM example is $U_i \sim Poi(\theta R_i Q)$, combining local and global fluctuations of the Poisson parameter, with independent mixing RVs R_i and Q, which are as above, but now $\mathrm{Var}(R_i) = \gamma$. One sees quickly that this model meets the MCM properties. If the R_i are, in particular, Gamma distributed, one has, given Q, $N_v \sim NegBin\left(v\theta Q, \frac{\gamma}{v}\right)$ with the contagion as second parameter (and inverse of the common shape parameter). This model can be extended to real v. Unless Q is trivial ($\beta = 0$), the unconditional model is not NegBin any more, but geometrically quite similar as long as the market factor varies few.

A possibly interesting variant emerges when the R_i and Q are Lognormal RVs. Then N_v has the same distribution type both conditionally on Q and unconditionally, namely mixed Poisson-Lognormal. This model, while geometrically similar to NegBin, has no closed-form formulae for the probabilities. Yet, it has been popular in Biostatistics long since, see Izsák (2008), Bulmer (1974).

For further details see Fackler (2017), where one also finds the interesting mixed-Binomial MCM $N_v \sim Bin(nv, pQ)$, which has $\gamma = \frac{-1}{n} < 0$ and $\mathrm{Ct}(N_v) = \beta - \frac{1+\beta}{nv}$, such that N_v, with growing v, shifts from underdispersion to overdispersion.

One can slightly generalize MCMs via a simpler definition without mixing RV Q. Yet, the intuitively appealing MCMs are those having an explicit market factor, which models correlation in a "causal" manner, being much easier to interpret as when inferred from a copula model.

Definition 6.4. Consider a family $(N_v)_v$ of discrete loss distributions for risks having volume v, where the volumes may vary across a certain range of admissible values. We call this a *generalized Mixed Contagion model* (GMCM) if for some $\theta > 0$, $\beta \geq 0$, and real γ, for all possible values of v, N_v has the following (finite) moments:

$$\mathrm{E}(N_v) = v\theta, \qquad \mathrm{Var}(N_v) = v\left(\theta + \theta^2\kappa\right) + v^2\theta^2\beta = v\theta + (v\theta)^2\left(\beta + \frac{\kappa}{v}\right)$$

Remark 6.5. Discrete real-world volumes (like the number of insured objects/people) are usually rather large, such that they can mostly be approximated well by continuous ones, and vice versa. For the insurance practice it is thus secondary whether a GMCM admits real volumes or only integers.

Note that, like θ, the parameters γ and κ are related to the volume unit, and must be adapted if the scale of the volume measure is changed, as it occurs with real volumes. To be precise, the products $\theta\gamma$ and $\theta\kappa$ are dimensionless, as is β.

MCMs are arguably too complex for most cases where one just wants to analyze the loss count history of one portfolio over a number of years. The portfolio volume varies from year to year, but in practice the shifts are mostly moderate, such that the contagion does not vary much over the years and the parameters β and κ have similar influence on the model – thus, their estimators may strongly interact. For such data a mixed Poisson model will mostly be just as good.

Yet, the situation is totally different when one has data from *two portfolios* that *differ in size* considerably, but are *qualitatively similar*, such that one can imagine them as being parts of the same macro-portfolio. Think of two insurers in a quite homogeneous market, or of a typical portfolio vs the whole market. The similarity of the two portfolios means mathematically that we assume that both consist of units having equal characteristics, namely distributions being iid conditionally on the market factor, and that this market factor is *the same* for the two portfolios. When the two data sets have quite different volumes (per year), Mixed Contagion modeling will accordingly reveal marked differences in contagion (unless $\kappa = 0$).

The MCM class seems to be designed for such data situations. To cater for possibly thinned data, we need a last ingredient, which combines the results of this section with those of Section 3.

Proposition 6.6. *If N_v is GMC distributed with expectation $v\theta$ and contagion $\beta + \frac{\kappa}{v}$, each proper thinning $\bar{N}_v$ thereof has, with selection probability q and $\bar{\theta} = q\theta$, the moments*

$$E\left(\bar{N}_v\right) = v\bar{\theta}, \qquad Var\left(\bar{N}_v\right) = v\left(\bar{\theta} + \bar{\theta}^2\kappa\right) + v^2\bar{\theta}^2\beta, \qquad CV^2\left(\bar{N}_v\right) = \beta + \frac{\kappa + 1/\bar{\theta}}{v}$$

Proof. The first formula is clear; the second is the thinning-invariant polynomial representation of the variance, with $\bar{\lambda} = v\bar{\theta}$ and $c = \beta + \frac{\kappa}{v}$. □

Now we have a coherent rule for how the loss count parameters change when we consider a subset of losses: for sub-portfolios (via a reduced volume v), for thinning (via a selection probability $q < 1$), and for combinations of both.

7 Parameter inference

Let us illustrate a parameter-free inference procedure for GMCMs. It is inspired from a method-of-moments estimator given in Section 1.4.2 of Mack (1997) for mixed-Poisson loss counts, where the volume varies over the years. Our approach is analogous, but it combines data from two sources.

7.1 Available empirical data

Consider the following situation.

Portfolio 1: For an observation period we have been given, for $i \in I$, yearly volumes v_i and respective loss counts $\bar{n}_i$, which are possibly (properly) thinned, by some criterion with (unknown) selection probability q.

Portfolio 2: For a (possibly different) observation period we have been given, for $j \in J$, yearly volumes v'_j and respective loss counts $\dot{n}'_j$, which are likewise possibly thinned, by some (possibly different) criterion with selection probability q'.

There could be missing years in between (lack of reporting), but for each portfolio we suppose a rather large number of observed years, well over 10, as are necessary for reliable use of the empirical variance.

Example 7.1. Two insurers report losses to their (same) reinsurer, using different reporting thresholds. Typically large insurers have higher reinsurance retentions and accordingly higher reporting thresholds, so it could be that the larger insurer reports (on average) less losses than the smaller one.

Example 7.2. The first data are the million dollar losses from a MTPL insurer, while the second are nationwide and ground-up (no reporting threshold), but embracing only MTPL losses involving severe personal injury (thinning by claim components involving long-term care, death indemnity, etc.). The two thinnings are possibly related, but not in a hierarchical order.

At first glance, this data situation seems hopeless. If a data set is thinned, we cannot infer the underlying overall frequency. Nevertheless, we can do quite some inference.

7.2 First data set

Let us find out what the first data set alone reveals. To begin with, each yearly frequency $\bar{N}_i/v_i$ is an unbiased estimator for $\bar{\theta} = q\theta$, where θ is the overall frequency per volume unit and $\bar{\theta}$ the respective "thinned" frequency.

Definition 7.3. Consider data sets v_i, $\bar{n}_i$ (realization of $\bar{N}_i$), $i \in I$. For each vector of nonnegative weights $\overrightarrow{w} = (w_i)_{i \in I}$, $\sum_{i \in I} w_i = 1$, we call

$$M_{\overrightarrow{w}} := \sum_{i \in I} w_i \frac{\bar{N}_i}{v_i}, \qquad S_{\overrightarrow{w}} := \sum_{i \in I} w_i \left(\frac{\bar{N}_i}{v_i} - M_{\overrightarrow{w}}\right)^2 = \sum_{i \in I} w_i \left(\frac{\bar{N}_i}{v_i}\right)^2 - M_{\overrightarrow{w}}^2$$

the $\vec{w}$*-weighted sample mean* and respective *sum of squares, and call*

$$a_{\vec{w}} := \sum_{i \in I} \left(w_i - w_i^2\right), \qquad b_{\vec{w}} := \sum_{i \in I} \frac{w_i - w_i^2}{v_i}$$

the variance coefficients.

For any weights, $M_{\vec{w}}$ is an unbiased estimator of $\bar{\theta}$. For further results we must assume some structure.

Proposition 7.4. *If the yearly loss counts $N_i := N_{i;v_i}$ are distributed according to a generalized Mixed Contagion model with parameters θ, β, κ, we have*

$$E\left(\frac{\bar{N}_i}{v_i}\right) = \bar{\theta}, \qquad Var\left(\frac{\bar{N}_i}{v_i}\right) = \frac{\bar{\theta} + \bar{\theta}^2\kappa}{v_i} + \bar{\theta}^2\beta$$

For any given weights $\vec{w}$, $E(M_{\vec{w}}) = \bar{\theta}$. If further the N_i are independent, we have

$$E(S_{\vec{w}}) = a_{\vec{w}}\bar{\theta}^2\beta + b_{\vec{w}}\left(\bar{\theta} + \bar{\theta}^2\kappa\right)$$

Then the optimal weights minimizing the variance of the weighted sample mean are inversely proportional to

$$CV^2\left(\frac{\bar{N}_i}{v_i}\right) = \beta + \frac{\kappa + 1/\bar{\theta}}{v_i}$$

Proof. From the preceding the first three formulae are clear, while standard sum-of-squares algebra yields

$$\begin{aligned}
\mathrm{E}\left(S_{\vec{w}}\right) &= \sum_{i \in I} w_i \mathrm{E}\left(\frac{\bar{N}_i}{v_i}\right)^2 - \mathrm{E}\left(\left(\sum_{i \in I} w_i \frac{\bar{N}_i}{v_i}\right)^2\right) = \sum_{i \in I} w_i \mathrm{Var}\left(\frac{\bar{N}_i}{v_i}\right) - \mathrm{Var}\left(\sum_{i \in I} w_i \frac{\bar{N}_i}{v_i}\right) \\
&= \sum_{i \in I} \left(w_i - w_i^2\right) \mathrm{Var}\left(\frac{\bar{N}_i}{v_i}\right) = \sum_{i \in I} \left(w_i - w_i^2\right)\left[\frac{\bar{\theta} + \bar{\theta}^2\kappa}{v_i} + \bar{\theta}^2\beta\right] = b_{\vec{w}}\left(\bar{\theta} + \bar{\theta}^2\kappa\right) + a_{\vec{w}}\bar{\theta}^2\beta
\end{aligned}$$

The $\vec{w}$-weighted sample mean is a linear combination of the unbiased estimators $\bar{N}_i/v_i$ of $\bar{\theta}$. If the $\bar{N}_i/v_i$ are independent, $M_{\vec{w}}$ has minimum variance if the w_i are inversely proportional to the variances $\mathrm{Var}\left(\bar{N}_i/v_i\right)$, or equivalently to the respective squared CVs. □

Remark 7.5. The assumptions of the proposition are not restrictive at all. Supposing a (G)MC distribution for each year is far more general than assuming Poisson or Negative Binomial, as is common. The independence of the years in the sample mean is a standard assumption, for an exception (and the resulting huge complexity) see Fackler (2017). Note that for a MCM this independence means, in particular, that the yearly market factors Q_i are independent. This excludes structures like random walks or autoregressive time series, which could be plausible in some cases, e.g., when insurance losses are closely tied to economic cycles or long-term weather oscillations like El Niño.

With $M_{\vec{w}}$ and $S_{\vec{w}}$ being unbiased estimators of $\bar{\theta}$ and $a_{\vec{w}}\bar{\theta}^2\beta + b_{\vec{w}}\left(\bar{\theta} + \bar{\theta}^2\kappa\right)$, respectively, a natural estimator for

$$a_{\vec{w}}\beta + b_{\vec{w}}\kappa$$

emerges, in the same way as in standard moment matching:

$$D_{\vec{w}} := \frac{S_{\vec{w}}}{M_{\vec{w}}^2} - \frac{b_{\vec{w}}}{M_{\vec{w}}}$$

We cannot expect this estimator to be exactly unbiased, but it is fair to assume that it is often close. Without making further specific distributional assumptions for the N_i, it is arguably hard to improve. Note that $D_{\vec{w}}$ estimates a linear combination of the contagion parameters β and τ, which are the same for all thinnings of the N_i and not tied to the specific selection of the first data set and the respective thinned frequency $\bar{\theta}$.

The above results hold for any given weights $\vec{w}$. The *optimum weights* for the estimation of $\bar{\theta}$ are a priori unknown. Qualitatively one can say that if β dominates the terms $\beta + \left(\kappa + 1/\bar{\theta}\right)/v_i$, the optimum weights are approximately equal. If the second summands dominate, the optimum weights are approximately proportional to the volumes v_i, just as in the Poisson model. Intermediate cases have optimum weights somewhere in between. In practice, the v_i typically vary only moderately, such that all these options lead to quite similar weights being not far from the optimum. Thus, any of them yields a fair estimate of $\bar{\theta}$, at least a good initial estimate to be subsequently improved, as we will do below.

7.3 Second data set

As the portfolios are qualitatively related, we assume that the second one has a GMC distributed loss count with the same parameters θ, β, κ, but a (possibly) different thinned frequency $\dot{\theta}$.

Remark 7.6. For a MCM, one would typically assume that the underlying market factors Q and Q' of the two portfolios be *the same*. Although this is plausible, we want to emphasize that our inference method uses only the weaker assumption that Q and Q' have the same variance, which means same MC parameter β. The far stronger assumption $Q = Q'$ does not alter our estimators, but affects their properties: it implies a dependency between the two portfolios even when they are disjoint. (How inference can be improved by using $Q = Q'$, is a potential question for further research.)

We can apply the above procedure analogously to the second data set, i.e., choose weights u_j, $j \in J$, and calculate the respective variance coefficients $a'_{\vec{u}}$ and $b'_{\vec{u}}$, the unbiased estimator $M'_{\vec{u}}$ of $\dot{\theta}$, and $S'_{\vec{u}}$ estimating $a'_{\vec{u}}\dot{\theta}^2\beta + b'_{\vec{u}}\left(\dot{\theta} + \dot{\theta}^2\kappa\right)$. Then

$$D'_{\vec{u}} := \frac{S'_{\vec{u}}}{M'^2_{\vec{u}}} - \frac{b'_{\vec{u}}}{M'_{\vec{u}}}$$

is a natural (and likely not too biased) estimator for

$$a'_{\vec{u}}\beta + b'_{\vec{u}}\kappa$$

which is again a linear combination of β and κ.

7.4 Joining sources

By combining the estimates from both data sets, we get a linear equation system for β and κ, which can be solved unless the vectors

$$(a_{\vec{w}}, b_{\vec{w}}) = \left(\sum_{i\in I}\left(w_i - w_i^2\right), \sum_{i\in I}\frac{w_i - w_i^2}{v_i}\right), \qquad (a'_{\vec{u}}, b'_{\vec{u}}) = \left(\sum_{j\in J}\left(u_j - u_j^2\right), \sum_{j\in J}\frac{u_j - u_j^2}{v'_j}\right)$$

are parallel. This is not the case when the two underlying portfolios are quite different in size: $b_{\vec{w}}/a_{\vec{w}}$ is a $\left(w_i - w_i^2\right)$-weighted average of the $1/v_i$, while $b'_{\vec{u}}/a'_{\vec{u}}$ is a $\left(u_j - u_j^2\right)$-weighted average of the $1/v'_j$. Thus, if the v_i are (largely) considerably greater, or smaller, than the v'_j, the vectors are far from parallel.

So, apart from rare unfortunate cases, the equation system

$$a_{\vec{w}}\widehat{\beta} + b_{\vec{w}}\widehat{\kappa} = D_{\vec{w}}$$
$$a'_{\vec{u}}\widehat{\beta} + b'_{\vec{u}}\widehat{\kappa} = D'_{\vec{u}}$$

can be solved. However, as is common in moment matching, we must make sure that the resulting estimates be in the admissible range. β is nonnegative, so we must have $\widehat{\beta} \geq 0$. κ may be negative, but $\mathrm{CV}^2\left(\bar{N}_v\right) = \beta + \left(\kappa + 1/\bar{\theta}\right)/v \geq 0$ for all admissible volumes v. We can ensure consistency by the constraint $\widehat{\kappa} \geq -1/\widehat{\bar{\theta}}$ together with the analogous $\widehat{\kappa} \geq -1/\widehat{\dot{\theta}}$. Wrapping up we get:

Definition 7.7. With data situation and notation as above, we define the *two-data GMCM estimators*. The *provisional* estimators, *for given weights* $\vec{w}$ and $\vec{u}$, are

$$\widehat{\bar{\theta}}_{\vec{w}} = M_{\vec{w}}, \qquad \widehat{\dot{\theta}}_{\vec{u}} = M'_{\vec{u}}, \qquad \widehat{\beta}_{\vec{w},\vec{u}} = \max\left(\frac{b'_{\vec{u}}D_{\vec{w}} - b_{\vec{w}}D'_{\vec{u}}}{a_{\vec{w}}b'_{\vec{u}} - a'_{\vec{u}}b_{\vec{w}}}, 0\right)$$

$$\widehat{\kappa}_{\vec{w},\vec{u}} = \max\left(\frac{a_{\vec{w}}D'_{\vec{u}} - a'_{\vec{u}}D_{\vec{w}}}{a_{\vec{w}}b'_{\vec{u}} - a'_{\vec{u}}b_{\vec{w}}}, \frac{-1}{\widehat{\bar{\theta}}_{\vec{w}}}, \frac{-1}{\widehat{\dot{\theta}}_{\vec{u}}}\right)$$

The *overall* estimators result by iteration:

Start with *chosen weights* that are rather balanced, like: equal, volume-proportional, or in between.

Estimate the four parameters, calculate the respective (estimated) squared CVs of the loss count per volume, by plugging the parameter estimates in the formulae

$$\mathrm{CV}^2\left(\frac{\bar{N}_i}{v_i}\right) = \beta + \frac{\kappa + 1/\bar{\theta}}{v_i} \qquad \mathrm{CV}^2\left(\frac{\dot{N}'_j}{v'_j}\right) = \beta + \frac{\kappa + 1/\dot{\theta}}{v'_j}$$

Set new weights for both data sets, being inversely proportional to the respective squared CVs.
Repeat the *estimation.*
Reiterate until convergence.

The definition adds the common iteration for inferences where the *optimum weights* are not a priori known, but result as a fix point of the estimating function that one iterates. Testing with realistic (not too weird) data input indicates that the procedure typically converges and a few steps are sufficient.

Some comments on potential (and possibly unwelcome) dependencies in the inference procedure are adequate:

- If more than one parameter is estimated, one generally prefers procedures where the estimations of the single parameters interact as few as possible, such that, loosely speaking, a potential misestimate of one parameter does not spoil the estimation of the others. E.g., for ML estimation of NegBin, unlike other common parameterizations (see Appendix B of Fackler (2021) for an overview), the pair of parameters λ and c (or shape parameter $\frac{1}{c}$) has *orthogonality*: the respective MLE estimators are asymptotically independent (Panjer and Willmot, 1992). Our four-parameter procedure is quite intertwined, such that interaction of (mis)estimates cannot be ruled out. However, given the challenging data input with two (possibly very differently) thinned data sets, it is difficult to imagine a perfectly orthogonal inference method.

- The stochastic processes that produced the two input data sets are not independent, which links the estimates of the four parameters. This dependency is in so far unavoidable as the (usually assumed) common market factor ties the two portfolios. Apart from that, the portfolios may not be disjoint, such that some losses appear on both sides, which creates a dependency that can often, and should possibly, be avoided. The typical case is that the larger portfolio is the whole market, containing the other one. Unless the latter has only a small market share, this is a strong interaction; if possible, one should split the market data (losses and volumes) and use portfolio vs rest of market.

7.5 Variants

Let us now treat the (seemingly simpler) case that the two thinnings are the same, such that $\bar{\theta} = \dot{\theta}$ and only three parameters have to be estimated. There are two straightforward options to deal with that (finding out which one works better is a potential question for future research):

1. Run the inference procedure as if $\bar{\theta} \neq \dot{\theta}$. Take the resulting estimates for β and κ, but reestimate $\bar{\theta}$ from the *merged* data set.

2. Estimate $\bar{\theta}$ from the *merged* data at every step.

Merging the data means here: for all years where you have only one observation available, use it; for years covered by both data sets, merge losses and add volumes if you can rule out double counting, otherwise use the data having the larger volume.

Whichever variant one adopts, one may come across cases where the separate estimates of $\bar{\theta}$ from the two data sets differ considerably. There are two obvious options how this could come about:

1. The thinnings seem to be the same, but are not. E.g., thinning by claim components could be different when the latter are defined (or administered) in not exactly the same way. Then one has a $\bar{\theta} \neq \dot{\theta}$ case with four parameters to estimate.

2. A portfolio has significantly lower (or higher) frequencies than the other one.

The latter case leads to a variant of the four-parameter estimate. We choose a portfolio, typically the one having larger volumes, as the reference, say this is the first data set. Then the second portfolio is considered still similar, but having overall a higher (or lower) loss frequency per volume unit. A simple (but not implausible) assumption is that these frequencies differ by the same factor across all years. To make the two portfolios comparable and to describe them via the same GMCM model, we must replace the given volumes of the second portfolio by *rescaled* ones, being multiplied by a *conversion factor* ϱ, which has to be estimated from the data and is thus the fourth parameter.

Example 7.8. An exotic variant would be: two portfolios using different volume measures, like two commercial MTPL fleets having different premium rules, such that one reports driven miles, the other driven hours. Yet, if both volumes are adequate for the modeling of the frequency, they can be integrated in one model by a conversion factor.

Table 1: Data, empirical frequencies, final weights; final estimators for key figures

year	v_i	$\bar{n}_i$	$\frac{\bar{n}_i}{v_i}$	v'_j	$\dot{N}'_j$	$\frac{\bar{N}'_j}{v'_j}$	w_i	u_j
1	4.108	38.043	9.3				7.7%	
2	4.403	42.179	9.6				7.7%	
3	4.502	46.214	10.3	200	177	0.89	7.7%	5.9%
4	4.604	43.268	9.4	200	180	0.90	7.7%	5.9%
5	4.708	43.037	9.1	200	201	1.01	7.7%	5.9%
6	4.770	56.226	11.8	200	188	0.94	7.7%	5.9%
7	4.793	48.703	10.2	200	170	0.85	7.7%	5.9%
8	4.857	52.052	10.7	200	223	1.12	7.7%	5.9%
9	4.869	46.654	9.6	200	199	1.00	7.7%	5.9%
10	5.017	47.757	9.5	200	202	1.01	7.7%	5.9%
11	5.082	43.344	8.5	200	186	0.93	7.7%	5.9%
12	5.342	47.848	9.0	200	185	0.93	7.7%	5.9%
13	5.663	60.063	10.6	200	211	1.06	7.7%	5.9%
14				200	221	1.11		5.9%
15				200	190	0.95		5.9%
16				200	160	0.80		5.9%
17				200	254	1.27		5.9%
18								
19				100	88	0.88		3.9%
20				100	82	0.82		3.9%
21				100	121	1.21		3.9%

$\hat{\hat{\theta}} = M_{\vec{w}}$	9.81
$a_{\vec{w}}$	0.923
$b_{\vec{w}}$	0.000193
$S_{\vec{w}}$	0.706
$D_{\vec{w}}$	0.00732
$\hat{\dot{\theta}} = m'_{\vec{u}}$	0.981
$a'_{\vec{u}}$	0.943
$b'_{\vec{u}}$	0.00528
$S'_{\vec{u}}$	0.0151
$D'_{\vec{u}}$	0.01033
$a_{\vec{w}}b'_{\vec{u}} - a'_{\vec{u}}b_{\vec{w}}$	0.00469
$b'_{\vec{u}}D_{\vec{w}} - b_{\vec{w}}D'_{\vec{u}}$	0.0000367
$a_{\vec{w}}D'_{\vec{u}} - a'_{\vec{u}}D_{\vec{w}}$	0.00263
$\hat{\beta}$	0.00781
$\hat{\kappa}$	0.560
$\hat{\gamma}$	0.556

Technically, we have $v'_j = \varrho v^*_j$, where the v^*_j are given and the v'_j go into the model, reflecting yearly frequencies $\bar{\theta} v'_j$ being comparable to those of the reference portfolio. For ease of orientation we don't alter the notation, but emphasize that the $v'_j = \varrho v^*_j$ now are RVs via the factor ϱ, as are the second variance coefficients

$$b'_{\vec{u}} = \sum_{j \in J} \frac{u_j - u_j^2}{v'_j} = \frac{1}{\varrho} \sum_{j \in J} \frac{u_j - u_j^2}{v^*_j}$$

Each estimation step starts as above with the setting of the weights and $\hat{\hat{\theta}}_{\vec{w}} = M_{\vec{w}}$, then calculates

$$M^*_{\vec{u}} = \sum_{j \in J} u_j \frac{\bar{N}'_j}{v^*_j}, \qquad \hat{\varrho} = \frac{M^*_{\vec{u}}}{M_{\vec{w}}}$$

and finally estimates β and τ as above. The choice of $\hat{\varrho}$ ensures indeed that

$$M'_{\vec{u}} = \sum_{j \in J} u_j \frac{\bar{N}'_j}{v'_j} = \frac{M^*_{\vec{u}}}{\hat{\varrho}} = \hat{\hat{\theta}}_{\vec{w}}$$

reflecting the same frequency level of the two portfolios after rescaling volumes.

This procedure is still more intertwined than the standard two-data GMCM estimator. However, essentially both are straightforward moment matching.

7.6 Numerical example

Let us illustrate the two-data GMCM parameter estimation by a numerical example. Figures are assembled in Table 1, with volumes being rescaled (divided by 1.000) for ease of reading.

The first data set was used in Section 1.4.2 of Mack (1997) for the analogous moment-matching (one-data) mixed-Poisson parameter inference: a large set of fire losses covering 13 years reporting on average 47.000 losses, having a slowly increasing volume. The mixing distribution or market factor $\tilde{Q}$ is quite volatile (estimates for the variance $\tilde{\beta}$ are in the range of 0.008, which means a CV of about 9%); the Poisson hypothesis is clearly rejected. What is more, together with the high loss count it results that the yearly volatilities $\mathrm{CV}^2\left(\frac{\bar{N}_i}{v_i}\right) = \tilde{\beta} + \frac{1/\bar{\theta}}{v_i}$, which give the optimal weights for the inference of $\bar{\theta}$, are almost equal; the volume-dependent summand is negligible. For this data alone, mixed-Poisson modeling does fine; the respective $\mathrm{CV}^2\left(\frac{\bar{N}_i}{v_i}\right) = \beta + \frac{\kappa+1/\bar{\theta}}{v_i}$ of a GMCM fit will likewise be dominated by β.

GMC modeling does make a difference, however, if we embrace the second data set, which is a sampled one reflecting strongly thinned losses from a portfolio similar to the first one, but having much lower volume (with a recent sharp shift). It covers 18 years; the loss count is some hundred per year.

The GMCM parameter estimates result after four iteration steps, irrespective of whether the initial weights are equal (which is close to optimum for the first data) or volume-proportional. We have $\widehat{\bar{\theta}} = 9.8$, $\widehat{\dot{\theta}} = 0.98$, $\widehat{\beta} = 0.0078$, $\widehat{\kappa} = 0.56$ and see (note the resulting non-equal weighting) that for the second data the volatilities $\mathrm{CV}^2\left(\frac{\dot{N}'_j}{v'_j}\right) = \beta + \frac{\kappa+1/\dot{\theta}}{v'_j}$ have a considerable volume-dependent component, despite the large β, which in MCMs indicates (analogously to mixed Poisson above) a rather volatile underlying market factor Q. Each model parameter plays a specific role, with $\kappa \neq 0$ distinguishing the model from mixed Poisson.

8 Conclusion

We have addressed the two questions posed at the beginning of this paper.

The *contagion* describes, in a compact and intuitive way, the effects of thinning on the volatility (second moment) of the loss count.

It is also the key ingredient of the definition of the flexible, realistic, and intuitive *Mixed Contagion model*, which describes how the loss count parameters are related between (homogeneous) portfolio and single risk, and between portfolio and (representative) sub-portfolio.

Altogether we have a method to construct a common loss count model out of two quite different sets of empirical data, like a portfolio and a market statistics being thinned in different ways.

References

Bulmer, M. (1974). On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics*, 30:101–110.

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

Fackler, M. (2021). Panjer class revisited: One formula for the distributions of the Panjer (a,b,n) class. In *ASTIN Colloquium 2021.*

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).

Izsák, R. (2008). Maximum likelihood fitting of the Poisson lognormal distribution. *Environmental and Ecological Statistics*, 15(2):143–156.

Johnson, N. L., Kotz, S., and Kemp, A. W. (2005). *Univariate discrete distributions.* John Wiley & Sons.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2008). *Loss models: from data to decisions.* John Wiley & Sons.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2013). *Loss models: Further topics.* John Wiley & Sons.

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

Meyers, G. G. (2007). The common shock model for correlated insurance losses. *Variance*, 1(1):40–52.

Panjer, H. H. and Willmot, G. E. (1992). *Insurance risk models.* Society of Actuaries.

Ross, G. and Preece, D. (1985). The negative binomial distribution. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 34(3):323–335.