Three-layer problems and the Generalized Pareto Distribution

Michael Fackler*

Abstract

The classical way to get an analytical model for the (supposedly heavy) tail of a loss severity distribution is via parameter inference from empirical large losses. However, in the insurance practice it occurs that one has much less information, but nevertheless needs such a model, say for reinsurance pricing or capital modeling.

We use the Generalized Pareto distribution to build consistent underlying models from very scarce data like: the frequencies at three thresholds, the risk premiums of three layers, or a mixture of both. It turns out that for typical real-world data situations such GPD "fits" exist and are unique.

We also provide a scheme enabling practitioners to construct reasonable models in situations where one has even less, or somewhat more, than three such bits of information.

Finally, we have a look at model risk, by applying some parameter-free inequalities for distribution tails and a particular representation for loss count distributions. It turns out that, in the data situation given above, the uncertainty about the severity can be surprisingly low, such that the overall uncertainty is driven by the loss count.

Keywords: Generalized Pareto, Heavy tail, Scarce data, Reinsurance, Premium rating

1 Introduction

1.1 Motivation

In theory, loss modeling in insurance should work like this: You have abundant loss data available, find a parametric model that fits this data well, and do all the calculations you need with this model, using the parameters estimated from the data. So, you get all output you need from the parametric distribution model: moments, quantiles, TVaR, etc.

In practice instead, it occurs that the loss data are too scarce for this procedure, but nevertheless the actuary is expected to produce the same outputs as if he/she had abundant data available. Again a parametric model is needed, but its selection and parameter estimation are much harder now, as standard procedures (best fit, parameter inference) give very volatile results or are not applicable at all. This occurs e.g. when large losses have to be modeled for solvency matters or the premium rating of (re)insurance layers, in particular when data from smaller losses are either not available or considered inadequate for extrapolation into the large-loss area.

An illustrative example: Assume you have to model losses in a certain line of business, in the range between 1 and 20 (say million USD), from a loss history of say 12 losses, which do fine for a *Burning Cost* (BC, mathematically the sample mean) premium rating of a 2 xs 1 layer, but are insufficient to assess higher layers. Further, for the layer 5 xs 5 there is e.g. a market benchmark BC rate (as is common for Motor Liability in several countries) or a reliable risk premium estimate from a geophysical simulation model for natural disasters (as is available for most Property lines in many countries). For higher loss sizes you don't have such benchmarks or don't find them reliable, however, you may use a (possibly a bit "political", see Section 6.3.6.4 of Schwepcke (2004)) payback approach, i.e., you assign a return period of say 200 years to the loss size 20 (million USD).

This is the challenge that we are going to explore in this paper: how to assign a consistent and plausible model to (about) 3 bits of information, namely risk premiums of layers or loss frequencies at thresholds. We will use the collective model with the Generalized Pareto distribution (GPD) as severity model. That the GPD is a very reasonable model for losses exceeding a large threshold, is supported both by theory (namely Extreme Value Theory and in particular the famous asymptotic properties found by Pickands-Balkema-De Haan, see e.g. Embrechts et al. (2013)) and by widespread practical experience.

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

1.2 Research context

There is not too much literature about how to build models from scarce data in a way other than conventional parameter inference. Old and novel methods to fit the GPD in case of scarce data are explored in Brazauskas and Kleefeld (2009) and related papers of the first author, however, they require some dozen losses at least. Two recent papers are closely related to our work; both were notably written by practitioners, both notably appeared in leading actuarial journals. Aviv (2018) proves that if for a (limited) layer we have the risk premium, the loss frequency, and the frequency of total losses, then there is one and only one Generalized Pareto model consistent with this information. This is a special case of our main result below. Riegel (2018) constructs, for the same data input, a piecewise Pareto model (two pieces) and develops procedures how to build an overall consistent model for a tower of n layers where the risk premiums and (possibly) the frequencies at the attachment points are given, yielding a spliced distribution consisting of about $2n$ Pareto pieces.

1.3 Outline

Section 2 explores the GPD, from memoryless properties over special cases and parameterizations to an instructive map of the parameter space. Section 3 introduces compact layer notation, gives our main result (the three-layer problem and its solution with the GPD), and discusses a number of variants and applications, including proposals about what to do when one has a bit more, or less, than 3 data inputs. Section 4 gives some non-parametric inequalities helping assess the model risk of layer losses. We will see that the uncertainty about the severity distribution has in many cases a surprisingly low impact, possibly less than the uncertainty about the loss count model. This justifies our choice to work with just one severity distribution model instead of comparing a number of parametric models, as is usually done in case of abundant loss data.

2 Generalized Pareto

Definition 2.1. The *Generalized Pareto (distribution)* = GP(D) is a model for severity distribution tails, *starting at* a known threshold $s \geq 0$. Note that $s = 0$ is possible and yields a ground-up model as we may need it in some situations. The *survival function* (1 minus the cdf) in its common form (see e.g. Embrechts et al. (2013)), reads

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

where the shape parameter ξ may be any real number, while $\sigma > 0$.

For negative ξ the distribution has a finite *supremum loss* $s + \frac{\sigma}{-\xi}$. The (Exponential) case $\xi = 0$ results from taking the (well-defined) limit. For $\xi \in (-\infty, 1)$ we have finite expectation.

An important property of the GPD is that it is closed to (upward) threshold change: any tail distribution $\bar{F}_Z\left(x \mid Z > t\right)$ starting from a higher tail $t > s$ is again GP (for $\xi < 0$: as long as t is lower than the supremum loss), notably having the same exponent ξ. The second parameter shifts, instead: from σ to $\sigma + \xi\left(t - s\right) > 0$. This is sometimes confusing, but the standard GP parameterization is convenient for parameter inference and for the mathematics below; we will use it throughout most of this paper.

However, in order to get some intuition about the GPD, let us look shortly at alternative parameterizations. Some are indeed tail-shifting-invariant in *both* parameters and thus easier to interpret. For details see Section 6.5 in Fackler (2017) and Sections 2-3 in Fackler (2013); here we give a brief account.

A variant of the classical GP parameterization replaces σ by the so-called *modified scale* (see e.g. Scarrott and MacDonald (2012)) $\sigma^* = \sigma – \xi s > –\xi s$. This yields

$$\bar{F}_Z\left(x \mid Z > s\right) = \left(\frac{\xi s + \sigma^*}{\left(\xi x + \sigma^*\right)^+}\right)^{\frac{1}{\xi}}, \qquad x \geq s$$

and one sees quickly that the survival function of a higher tail looks the same; one just has to replace s by the new threshold:

$$\bar{F}_Z\left(x \mid Z > t\right) = \left(\frac{\xi t + \sigma^*}{\left(\xi x + \sigma^*\right)^+}\right)^{\frac{1}{\xi}}, \qquad x \geq t \geq s$$

So, the model "forgets" the original threshold. We can interpret the *tail-invariant* parameters ξ and σ^* as geometric properties of the tail, no matter where this tail starts. They can help compare loss data and possibly identify typical parameters for certain business. Say if we fit various data sets of large fire losses with the GPD, no matter whether or not these sets have the same large-loss threshold, we can judge how similar their tails are, by just comparing the respective inferred parameters ξ and σ^*. For this ease of interpretation one has to pay with the somewhat complicate parameter space of the modified scale, which extends a bit into the negative real numbers.

For further useful representations we treat three cases separately, according to the sign of the GP exponent ξ.

2.1 Proper GPD

The case $\xi > 0$, which we call *proper* GPD, is largely considered the most interesting case for the insurance practice. Here a parameterization proposed by Scollnik (2007) is most intuitive.

Set $\alpha := \frac{1}{\xi} > 0$, $\lambda := \alpha\sigma^* = \alpha\sigma - s > -s$. Now we have

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

This representation is also tail-invariant and further extremely handy, revealing in particular a lot of analogies to the *single-parameter Pareto* model, which is indeed the case $\lambda = 0$ or equivalently $\sigma^* = 0$ (and requires a threshold $s > 0$). Pareto is the border of two proper-GP subclasses differing in terms of the so-called *local Pareto alpha* (Riegel, 2008), which is the parameter of local approximations by Pareto curves.

Definition 2.2. At any point $d > 0$ where a survival function $\bar{F}(x)$ is positive and differentiable, such that locally the pdf $f\left(x\right) = -\bar{F}'\left(x\right)$ exists, we call

$$\alpha_d := -\left.\frac{d}{dt}\right|_{t=\ln(d)} \ln\left(\bar{F}\left(e^t\right)\right) = d \cdot \frac{f\left(d\right)}{\bar{F}\left(d\right)}$$

the *local Pareto alpha at d*.

If α_d is (about) constant on an interval, the distribution is (close to) Pareto on this interval. For insurance losses one often observes that for very large d (say in the million Euro range) α_d is a (slowly) increasing function of d: in a way the tail, while being roughly similar to Pareto, gradually becomes somewhat less heavy. As one sees quickly, in general in GP tails we have for $d > s$

$$\alpha_d = \frac{d}{\sigma + \xi\left(d - s\right)} = \frac{d}{\sigma^* + \xi d}$$

For the proper GPD $\alpha_d = \frac{d}{d+\lambda}\alpha$, which tends to α as d becomes large. The specific behavior of α_d as a function of d is as follows:

- $\lambda > 0$: α_d increases (as is often observed for large insurance losses)
- $\lambda = 0$: Pareto $(s > 0)$
- $\lambda < 0$: α_d decreases $(s > -\lambda > 0)$

2.2 Power curve

For $\xi < 0$ the negatives of the above parameters α and λ are very intuitive, besides being tail-invariant. Indeed we get with $\beta := -\frac{1}{\xi} > 0$, $\nu := \beta\sigma + s > s$ the formula

$$\bar{F}_Z\left(x \mid Z > s\right) = \left(\frac{\left(\nu - x\right)^+}{\nu - s}\right)^{\beta}, \qquad x \geq s$$

which shows at a glance that this GP case is a piece of a shifted power curve having β as (positive) exponent and ν as supremum loss (and center of the curve).

Values ξ far below 0 should hardly appear in fits to insurance loss data: $\xi = -1$ yields the uniform distribution between threshold and supremum, while for $\xi < -1$ the pdf increases, i.e., larger losses are overall more likely than smaller losses, a rather unrealistic case. Yet not impossible: Section 4.1 of Aviv (2018) gives examples of earthquake loss distributions from a common geophysical model, including one having a rising pdf for a certain range of loss sizes, yielding a GPD fit with ξ well below -1.

The local Pareto alpha here always increases, rather quickly as is typical for not too heavy-tailed distributions: we have $\alpha_d = \frac{d}{\nu - d}\beta$, which is an increasing and diverging (as $d \nearrow \nu$) function in d.

2.3 Exponential

For $\xi = 0$ the parameters σ and σ^* coincide, yielding the traditional and tail-invariant representation of the Exponential distribution

$$\bar{F}_Z(x \mid Z > s) = \exp\left(-\frac{x-s}{\sigma}\right), \qquad x \geq s$$

Here the local Pareto alpha is an increasing linear function: $\alpha_d = \frac{d}{\sigma}$

2.4 GPD map

To conclude, we illustrate the variety of properties the GPD can have, using the classical parameters ξ and $\sigma > 0$. They span an open half-plane, which can be split in two parts by a half-line in four different ways, see Figure 1:

- $\xi = -1$ (uniform): rising vs falling density
- $\xi = 0$ (Exponential): finite vs infinite support
- $\xi = +1$: finite vs infinite expectation
- $\xi s = \sigma$ (Pareto): rising vs falling local Pareto alpha

If $s = 0$, the last half-line falls out of the parameter space and coincides with the right half of the ξ axis, such that there is no sector between the two where the local Pareto alpha would decrease.

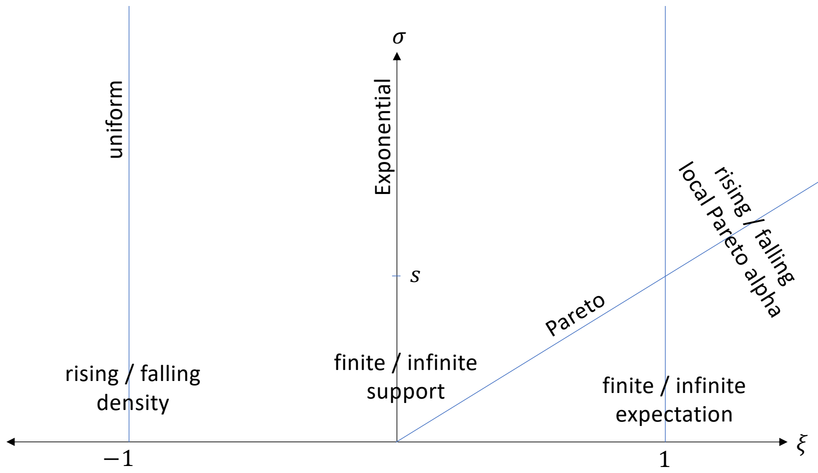

Figure 1: GPD areas

3 Three-layer problems

3.1 Layers

We borrow notation and basic results for layers from Riegel (2018), generalizing a bit and introducing particular sequences of layers.

Definition 3.1. A (re)insurance layer $c \,\mathrm{xs}\, a$ (*c in excess of a*) pays the part of each loss Z that exceeds the *attachment point* $a \geq 0$, up to a maximum $c > 0$ (*cover*, often called *line* or *liability*), which mathematically means paying $\min\left((Z-a)^+, c\right)$.

We set $b := a + c$ (*detachment point*) and often identify the layer with the corresponding interval $[a, b]$, $0 \leq a < b$.

If e is the risk premium of the layer, we call $r := e/c$ the *risk rate on line* (*RRoL*). If f is the expected frequency of losses exceeding a (*layer losses*), and g the expected frequency of losses exceeding b, we must have $f \geq r \geq g$, where equivalence is only possible in the (rather unrealistic) case that all ground-up losses Z hitting the layer ($Z > a$) are total layer losses reaching even beyond the layer ($Z > b$). The RRoL can be interpreted as the average loss frequency across the interval defining the layer.

We include two extremes. In case $\mathrm{E}(Z) < \infty$, we allow c to be infinite (*unlimited layer*). Here we have $r = 0$. On the opposite side, we admit $b = a$, i.e., $c = 0$ (!). Such a *degenerate* layer, shrunk to a point, in practice does not make sense, but as a bit of information it does: Imagine an extremely thin layer with b tending to a. Then e tends to 0, but r does not: it tends to f just as g does, yielding finally the *expected frequency at (the threshold) a*.

Generally, a *layer as an item of information* is an interval $0 \leq a \leq b$ together with the nonnegative figures e and r, which in case $0 < c < \infty$ are tied by the formula $r = e/c$. If $f > 0$, e or r is positive (if not both). The respective information provided by them is either the risk premium (for non-degenerate = *proper* layers) or the expected loss frequency (for *thresholds*).

Hierarchy (order) of layers: We say that a layer $[a_1, b_1]$ is *lower* than the layer $[a_2, b_2]$ if $a_1 \leq a_2$, $b_1 \leq b_2$, and one of the inequalities is strict. We then write $[a_1, b_1] < [a_2, b_2]$. Note that the lower and the *higher* layer may *overlap*, which shall mean that they have a proper intersection ($a_2 < b_1$), not just a common endpoint. Non-overlapping situations are e.g. $[a_1, b_1] < [b_1, b_2]$, $[a, b] < [b, b]$.

We call two ordered layers *strongly ordered* if they either do not overlap or have $a_1 < a_2 < b_1 < b_2$. This just excludes the case that two proper layers have a common attachment or detachment point, such that one would be an initial or final piece of the other one. Instead, $[a, a] < [a, b]$ is a strong ordering. In particular, in a sequence of strongly ordered layers only the highest one can be unlimited.

We finally introduce some particular sequences of ordered layers $[a_i, b_i]$: A sequence of layers is *weakly overlapping* if the layers are ordered and overlap, if at all, only with adjacent layers, i.e., $b_i \leq a_{i+2}$. A set of non-overlapping layers can obviously always be arranged as a sequence of strongly ordered layers. In particular, a *tower of layers* is an ordered sequence of non-overlapping (usually proper) layers attaching exactly at each other, i.e., $b_i = a_{i+1}$.

Ordered layers and towers of layers are classics in the reinsurance practice and do appear in the literature. For orientation we give a scheme of the intermediate kinds of ordering introduced here:

$$\text{tower} \Rightarrow \text{non-overlapping} \Rightarrow \left\{ \begin{array}{c} \text{weakly overlapping} \Rightarrow \\ \text{strongly ordered} \Rightarrow \end{array} \right\} \text{ordered}$$

Coming back to how layers are affected by losses, in the situation of two ordered layers we have $f_1 \geq f_2$, $g_1 \geq g_2$, and notably $r_1 \geq r_2$ (Riegel, 2018). In the tower case we have more strongly a decreasing sequence of altering frequencies and RRoL's:

$$f_1 \geq r_1 \geq g_1 = f_2 \geq r_2 \geq g_2 = f_3 \geq \ldots$$

Remark 3.2. One could alternatively define g as reflecting just the total layer losses ($Z \geq b$) instead of requiring that they properly exceed the layer detachment point ($Z > b$). For this variant the property $f \geq r \geq g$ holds as well, but g is greater than in the above definition in case the severity has a mass point at b. Thus, for a tower of layers one would have

$$f_1 \geq r_1 \geq g_1 \geq f_2 \geq r_2 \geq g_2 \geq f_3 \geq \ldots$$

We will use this variant in Section 4, but for now stay with the original definition. It does fine for our purposes and is easier to deal with, having in particular the convenient property mentioned above: if b tends to a, g tends to f.

3.2 Main result

Consider a loss severity having a GP tail starting at the threshold $s \geq 0$, and a layer $[a, b]$ being located in the area of this tail, i.e., $a \geq s$. If the expected loss frequency at s equals ϑ, we can express the risk premium of the layer with the tail severity as follows (Riegel, 2018):

$$\vartheta \int_a^b \bar{F}_Z(x \mid Z > s)\, dx = \vartheta \int_a^b \left(\left(1+\xi\frac{x-s}{\sigma}\right)^+\right)^{-\frac{1}{\xi}} dx = \vartheta\, E(s, a, b, \xi, \sigma)$$

Here we have used a compact notation for the integral term:

Definition 3.3. For $0 \leq s \leq a < b = a + c$ we write shortly

$$E(s, a, b, \xi, \sigma) := \int_a^b \left(\left(1+\xi\frac{x-s}{\sigma}\right)^+\right)^{-\frac{1}{\xi}} dx$$

and analogously for the corresponding quantity per line

$$R(s, a, b, \xi, \sigma) := \frac{1}{c}\int_a^b \left(\left(1+\xi\frac{x-s}{\sigma}\right)^+\right)^{-\frac{1}{\xi}} dx$$

The latter formula (which emphasizes that the RRoL is the average of the loss frequencies $\vartheta\, \bar{F}_Z(x \mid Z > s)$ over the layer interval) can be extended to the case $a = b$ in a natural way, via

$$\lim_{b \searrow a} R(s, a, b, \xi, \sigma) = \left(\left(1+\xi\frac{a-s}{\sigma}\right)^+\right)^{-\frac{1}{\xi}}$$

If we relate the risk premiums of two layers to each other, the threshold frequency ϑ drops out; the layer premium ratio depends only on integrals over the survival function. This is not specific to the GPD, but holds generally for loss severities in the collective model. A great advantage of the GPD (besides its theoretical and intuitive properties assembled above) is that it is essentially a power function, which makes the calculation of integrals, limits, etc. very easy, yielding analytical formulae whose mathematical properties are much easier to understand than when one has to involve numerical integration of complex functions like Gamma, Beta, Gaussian, as they appear in other common parametric distributions (for an overview see Appendix A of Klugman et al. (2008)).

In the following we need some quantities about a function that is formally similar to the survival function of the GPD, but starts at infinite.

Definition 3.4. For $0 \leq s \leq a < b = a + c$ and $\xi > 0$ we define $\tilde{F}(x) := (x-s)^{-\frac{1}{\xi}}$, $x \geq s$ and write shortly

$$\tilde{E}(s, a, b, \xi) := \int_a^b \tilde{F}(x)\, dx = \int_a^b (x-s)^{-\frac{1}{\xi}}\, dx$$

and analogously for the corresponding quantity per line

$$\tilde{R}(s, a, b, \xi) := \frac{1}{c}\int_a^b \tilde{F}(x)\, dx$$

The latter formula, which gives the average of $\tilde{F}(x)$ over the layer, can be extended to the case $a = b$ in a natural way, via

$$\lim_{b \searrow a} \tilde{R}(s, a, b, \xi) = \tilde{F}(x)$$

We notably admit infinite values here; indeed $\tilde{F}(s) = \infty$, while the integrals are infinite if $a = s$ and $\xi \leq 1$, as well as (like for the GPD) if $b = \infty$ and $\xi \geq 1$.

Remark 3.5. If $0 = s < a$, $\tilde{E}$ gives, up to a factor, the risk premium of the layer $[a, b]$ for the single-parameter Pareto model with $\alpha = \frac{1}{\xi}$ (Riegel, 2008). We can thus interpret $\tilde{E}$ and $\tilde{R}$ as reflecting risk premium and RRoL of the shifted layer $[a-s, b-s]$ for the Pareto model, plus a formal extension to the case where the attachment point $a - s$ equals 0.

An analogous formal extension was introduced by Riegel (2010): *quasi exposure curves*, a useful generalization of exposure curves. The classical example is based on the function $x^{-\alpha}$ being like Pareto but starting at $x = 0$, which is a special case of the "quasi" survival function $\tilde{F}$ introduced above. Likewise, the integrals $\tilde{E}$ and $\tilde{R}$ based on $\tilde{F}$ will turn out to be very helpful for our topic.

Now we can formulate the task this paper is about.

Definition 3.6. The *general three-layer problem* is as follows: If we have three layers $[a_i, b_i]$ with respective risk premiums e_i and RRoL's r_i, is there a severity distribution consistent with this information?

Usually one deals with ordered layers $[a_1, b_1] < [a_2, b_2] < [a_3, b_3]$. Here the inequality $r_1 \geq r_2 \geq r_3$ is a mathematical necessity and in practice mostly strict. Overlapping layers are a possibility, but most real-world three-layer problems are about layers that do not overlap. This case can be largely solved with the GPD, the results hold even for certain layer overlaps.

Theorem 3.7. *Suppose you have got three strongly ordered and weakly overlapping layers $[a_i, b_i]$ with respective risk premiums e_i and strictly decreasing RRoL's $r_1 > r_2 > r_3 \geq 0$, such that for each layer r_i and/or e_i is positive, namely the former for limited (possibly degenerate) layers, the latter for proper layers (the top layer being possibly unlimited).*

If there is a GP tail model starting at a_1 that (together with a loss frequency at a_1) yields the given risk premiums and RRoL's, this model is unique.

Now suppose more strongly that the given top and middle layer do not overlap. If the bottom layer is a threshold and/or the top layer is unlimited, a matching GP tail model always exists. In the remaining case (proper first layer, finite third layer) the existence depends on a technical condition *that essentially means that r_2/r_1 is not extremely small compared to r_3/r_2, as follows. Here for $i = 1, 2$ the functions*

$$\varrho_i : (0, \infty) \ni \xi \mapsto \frac{\tilde{R}\,(a_1, a_{i+1}, b_{i+1}, \xi)}{\tilde{R}\,(a_1, a_i, b_i, \xi)} \in [0, 1)$$

are well defined and increasing; ϱ_2 is invertible with image $(0, 1)$. The set of RRoL's having an underlying GPD is

$$\left\{ r_1 > r_2 > r_3 > 0 \,\middle|\, \frac{r_2}{r_1} > \varrho_1 \left(\varrho_2^{-1} \left(\frac{r_3}{r_2} \right) \right) \right\}$$

Proof. We give a very short sketch of proof here, for the (many) technical details see Appendix B.

We only treat limited layers, the case of an unlimited third layer can (with some effort) be proved analogously. Any GPD solving the problem must allocate some probability of loss to every layer, which means that in case $\xi < 0$ the supremum loss must be greater than a_3, or equivalently $\sigma > -\xi\,(a_3 - s)$. We now work with $s = a_1$.

Some (intricate but quite elementary) calculus shows that for fixed $a_{i,}, b_i$ the mapping

$$\left\{ (\xi, \sigma) \in \mathbb{R}^2 \,\middle|\, \sigma > (-\xi\,(a_3 - a_1))^+ \right\} \ni (\xi, \sigma) \mapsto \left(\frac{R\,(a_1, a_2, b_2, \xi, \sigma)}{R\,(a_1, a_1, b_1, \xi, \sigma)}, \frac{R\,(a_1, a_3, b_3, \xi, \sigma)}{R\,(a_1, a_2, b_2, \xi, \sigma)} \right) \in (0, 1) \times (0, 1)$$

is well-defined, locally invertible and continuously differentiable in both directions, and (globally) injective. Thus we have, if any, unique parameters ξ, σ such that the corresponding GP tail model starting at a_1 yields the given ratios r_2/r_1, r_3/r_2. Setting $f_1 = r_1/R\,(a_1, a_1, b_1, \xi, \sigma)$, we match the three given RRoL's.

More (intricate but quite elementary) calculus shows that for a non-overlapping top layer the mapping is surjective if $a_1 = b_1$, whereas for $a_1 < b_1$ the image of the mapping is somewhat reduced by the above technical condition.

As for the latter, a bit more calculus shows that the function

$$\varrho_1 \circ \varrho_2^{-1} : (0, 1) \to [0, 1)$$

equals 0 on the interval $(0, \varrho_2\,(1)]$, then increases strictly towards the limit 1. Thus, for $\frac{r_3}{r_2} \leq \varrho_2\,(1)$ the technical condition is fulfilled irrespective of $\frac{r_2}{r_1}$; this yields a sufficient condition being easy to check. For greater $\frac{r_3}{r_2}$ things are more complicated as one needs to invert ϱ_2 to verify the technical condition. How restrictive the latter is, depends on how long, and far from each other, the three layers are. However, generally one can say that the technical condition is violated in situations where r_2 is very close to r_3 and at the same time much smaller than r_1. □

Translated into tail geometry, $r_1 \gg r_2 \gtrapprox r_3$ means that the loss severity is extremely heavy-tailed (in the sense of very slow reduction of the loss probability for rising thresholds) in the area extending from the second layer to the third, but much less so below. The GPD can cater for such situations to some extent (via a decreasing local Pareto alpha, see Section 2.1), even much better than many other common two-parameter loss models, it just misses certain extreme cases. Three-layer situations and RRoL inputs as one typically sees them in practice, are mostly far from violating the technical condition.

Note that GP tail models starting at some lower threshold $t < a_1$ cannot yield further solutions of the three-layer problem – their upper tail starting at a_1 would again be the unique GP tail model from the theorem.

Let us illustrate some special three-layer problems, where some layers are thresholds, i.e., shrunk to a point.

- If the top layer is a threshold, we have a situation like in the example given in the introduction.
- If all three layers are thresholds, we get the *three-threshold problem*. Here three given frequencies $f_1 > f_2 > f_3 > 0$ at thresholds $0 \leq a_1 < a_2 < a_3$ are matched by a unique GPD, which always exists as the first layer is a threshold.
- If *bottom* and *top* layer are thresholds that equal attachment and detachment point, respectively, of the *middle* layer, we get the very particular tower of layers $[a, a] < [a, b] < [b, b]$, which always yields a solution. Here the risk premium and the frequencies at the endpoints of the layer are given and fulfill $f > r > g > 0$. Let us call this the *layer-endpoints problem*. A direct proof of existence and uniqueness of the matching GP model can be found in Aviv (2018), together with an example (frequencies and RRoL's from geophysical modeling software) illustrating how this can be used to construct an analytical model for a tower of n layers: simply solve the layer-endpoints problem for each of the layers according to the given input $f_i > r_i > g_i$. This yields a piecewise GPD (n pieces).

Note that the theorem cannot be easily extended to straightforward bordering cases. In particular, it is not difficult to see that the variant with a loss-free top layer/threshold ($e_3 = 0$, $r_3 = 0$) can lead to GP solutions, but these are not unique. Solutions must have a supremum loss that be finite and not greater than a_3, i.e., $\xi < 0$ and $a_1 + \frac{\sigma}{-\xi} \leq a_3$. To get uniqueness, the supremum loss has to be specified.

The uniqueness assertion of the theorem requires strongly ordered and weakly overlapping layers, but this is not very restrictive as many three-layer problems for ordered layers can be equivalently reformulated with such layers, by reducing overlaps like this: Say for the layers $[a, b_1] < [a, b_2]$ with respective risk premiums e_1 and e_2 we have $e_2 > e_1$ (this is necessary to have consistent input). Then we can replace the larger layer by the "difference" of the two and work with the non-overlapping layers $[a, b_1] < [b_1, b_2]$ and the respective risk premiums e_1 and $e_2 - e_1$. Any GP solution to the modified problem solves the original problem, and vice versa. Analogously we can modify similar three-layer problems where two or three layers have a common attachment or detachment point, even situations where the middle layer shares the attachment point with the bottom layer and the detachment point with the top layer. However, for a few situations this procedure does not help, in particular the following:

- In the situation $a_1 = a_2 < a_3 = b_1 < b_2 = b_3$ the middle layer is the non-overlapping union of bottom and top layer, such that its risk premium must equal that of the other layers added up. So, if the given risk premiums fulfill $e_2 = e_1 + e_3$, the problem is consistently posed but insufficiently specified (only two data points given), if $e_2 \neq e_1 + e_3$, the problem has inconsistent input and no solution.
- If $a_1 < a_2 < a_3 < b_1 < b_2 = b_3$, the layers are ordered, but all three contain the proper interval $[a_3, b_1]$. This overlap can by eliminated via replacing the second layer by $[a_2, a_3]$. But, now the first and second layer are not ordered any more; this situation may have a (possibly even unique) solution, but is not covered by the theorem. The same holds for the analogous situation $a_1 = a_2 < a_3 < b_1 < b_2 < b_3$.

3.3 How to use it

Now we turn to real-world applications, embracing cases where one has a bit more or less than three data inputs of the kind discussed here: an ordered sequence of layers with risk premiums, or of thresholds with frequencies, or a mixture of both, usually having few or no overlaps.

Let us start with the scarcest data. If one has less than three inputs, one better works with a simpler tail model than the GPD. At least for reinsurers and their layer business having attachment points in the million dollar range, the (single-parameter) Pareto distribution would usually be the preferred choice. In many lines of business benchmarks for the parameter α are known, see e.g. Schmutz and Doerr (1998) or Section 4.4.8 of FINMA (2006).

More importantly, Pareto solves various *two-layer problems* (defined analogously to the three-layer problems studied above) for layers $[a_1, b_1] < [a_2, b_2]$, $a_i > 0$, yielding unique solutions. For two thresholds this is a simple and well-known one-liner (see e.g. Riegel (2018)) yielding

$$\alpha = \frac{\ln(r_1/r_2)}{\ln(a_2/a_1)}$$

For two proper limited layers it is a classic in the reinsurance practice, for a strict proof see Riegel (2008) – here layers may overlap, but must be ordered. The mixed situation (a proper layer and a threshold in hierarchical order) can be proved analogously. For a detailed proof of the case $[a_1, a_1] < [a_2, b_2]$ and for cases with unlimited layers see Riegel (2018).

Now we give procedures for constructing a severity, according to the number of given data inputs. Finding subsequently the correct frequency works as in the above sketch of proof of the theorem. We focus on the case of proper layers; if some layers are thresholds, the procedures are similar or far simpler.

1: Use Pareto, choose α from market experience.

2: Use Pareto, calculate α numerically (two-layer problem).

3, tentative: Use the GPD, try to solve the three-layer problem numerically.
Instead of proceeding as in the proof of the theorem, one can attempt to find f_1, ξ, σ in one step, by solving the system of three equations $r_i = f_1 R(a_1, a_i, b_i, \xi, \sigma)$ (for an unlimited top layer use instead $e_3 = f_1 E(a_1, a_3, \infty, \xi, \sigma)$ and restrict to $\xi < 1$). The uniqueness of the solution for the layer triples treated in the theorem (and arguably some more), as well as practical experience, suggest that this is not a hard numerical problem; choice of algorithm and start values are apparently not very critical. As for the latter, for $a_1 > 0$ one could start with say $\hat{f}_1 = r_1$ and GP parameters $\hat{\xi} = 1/\hat{\alpha}$, $\hat{\sigma} = \hat{\xi} a_1$ yielding a Pareto model, where $\hat{\alpha}$ is calculated from thresholds a_1, a_2 and frequencies r_1, r_2.
Given the rather benign numerical setting, even in the case where a matching GPD does not always exist, it is mostly quickest to let first a numerical algorithm look for a solution. Only if some attempts fail, it makes sense to take the time and check the above technical condition for existence.

≥3, exact: Use a piecewise-GP model. Try first to solve the three-layer problem for the three highest layers. (GP geometry is most plausible for high tails.) If this works, it yields a frequency at the attachment point of the third highest layer. If that is smaller than the RRoL of the next layer below, attach layer per layer downwards by solving the layer-endpoints problem for each of them. If not, you have inconsistency, thus cannot use a common GPD for the three top layers. In this case and generally if the upper three layers don't yield a GP model, solve the three-layer problem for the top layer, the second highest layer and the attachment point of the latter, choosing for the attachment point a frequency between the RRoL's of the adjacent layers; then use the layer-endpoints approach for each of the layers below the upper two.
To apply the layer-endpoints procedure to a number of layers, one must split overlapping layers and possibly add artificial layers (to close gaps between the given layers), choosing missing parameters appropriately, in order to have a consistent "tower" of parameters

$$f_1 > r_1 > g_1 = f_2 > r_2 > g_2 = f_3 > \dots$$

where each layer can be treated separately, but overall one gets a consistent piecewise-GP model.

≥3, approximate: Use the GPD and solve the system of equations $r_i = f_1 R(a_1, a_i, b_i, \xi, \sigma)$ approximately, according to some deviation metric. In real-world situations one often gets surprisingly close. To get start values, one could first solve the three-layer problem for a low threshold, the highest and a further layer.

This set of concepts should enable practitioners to build severity models in a large variety of situations where the data are too scarce (or otherwise not suitable) for traditional parameter inference.

3.4 Numerical example

Consider the tower of layers (given in say million USD) and respective RRoL's, as displayed in Table 1.

Table 1: Tower of layers

No.	1	2	3	4	(5)
Layering	2 xs 1	2 xs 3	5 xs 5	10 xs 10	20
RRoL [%]	52	13	4.8	1.3	0.5

The fifth layer is not a proper one, but a threshold with an assigned frequency. Let us go through diverse situations where some of these five bits of information are given:

1. *Layer 3 only.* If we know from market experience that for the respective loss range between 5 and 10, in the respective business line in the respective country, distributions are typically similar to Pareto with the parameter α in the range of 2, we can use that (vague) info, which specifies a model.
2. *Layers 1, 3.* This is a variant of the example given in the introduction. If we want to rely on the two Layer premiums only, we calculate, as practitioners name it, the Pareto alpha *between them* (Riegel (2008), Riegel (2018)), which specifies a model.
3. *Layers 1, 3, 5.* This is the example from the introduction. If we combine the two layer premiums and the (possibly a bit politically set) large-loss frequency, we have specified a three-layer problem and can look for a matching GP model, which indeed exists.
For comparison: If we, instead of 0.5%, chose a frequency of 2% at 20 (payback 50 years), there would be no matching GPD. Indeed one would have
$$\frac{r_3}{r_2} = \frac{2.0}{4.8} = 41.67\%, \qquad \varrho_1 \circ \varrho_2^{-1}\left(\frac{r_3}{r_2}\right) = \varrho_1\,(1.3077) = 9.90\% > \frac{r_2}{r_1} = \frac{4.8}{52} = 9.23\%$$
which violates the technical condition, albeit not by much. At first glance, here r_2 does not seem close to r_3, being more than twice as large. However, it must be noted that the large-loss threshold 20 is far higher than the middle layer 5 xs 5, such that a drop of the RRoL to a bit less than half is a slow decrease, in particular compared to the drop between bottom and middle layer, which is down to less than a tenth. Such a tail geometry is (slightly) beyond the flexibility of the GPD.
4. *Layers 1 to 4.* If we have the premiums of layers up to quite large loss size, it is usually neither necessary nor adequate to judgmentally set a large-loss frequency. We have two options:
   (a) Layers 2 to 4 are matched a GP model, which yields a loss frequency of 19% at 3. This figure is smaller than the RRoL of Layer 1 and thus consistent with the latter, such that we can attach a model for Layer 1. For a GPD as last input the loss frequency at 1 is needed, which can typically be inferred from empirical data, otherwise we set a plausible value. The resulting model is built out of two GP pieces, one for Layer 1, one above.
   (b) If some deviations of the given RRoL's are accepted, we can "fit" one GPD to all four RRoL's, according to some deviation metric. An option is the sum of squared distances of given from fitted RRoL's. (To ensure a close fitting of the upper layers, one may give the respective distances a higher weight.) The resulting GP model fits fairly well here (relative premium deviations below 5%), happening to be pretty close to Pareto.

We assemble key figures of the resulting models in the comprehensive Table 2. To be comparable, all appearing (Generalized) Pareto models start at the threshold $s = 1$. We display the (proper) GP parameters and the frequencies/RRoL's across the tower of layers.

3.5 Discussion

Is it disappointing that, due to the technical condition, some three-layer problems don't have a Generalized Pareto solution? As for mathematical beauty, maybe yes. As for practical relevance, not much. While

Table 2: GPD fits

	Layers used	3	1, 3	1, 3, 5	2 to 4	1; 2-4	$\approx$ 1-4
Parameters							
ξ		0.5	0.59	0.41	0.44	0.20	0.56
σ		0.5	0.59	0.96	1.31	0.90	0.60
σ^*		0	0	0.55	0.87	0.70	0.04
α		2	1.70	2.44	2.28	5.04	1.77
λ		0	0	1.35	1.98	3.54	0.06
frequencies / RRoL's [%]							
1		240.0	136.0	108.4		120.0	135.2
2 xs 1	52.0	80.0	52.0	52.0		52.0	52.0
3		26.7	20.9	24.1	19.0	19.0	20.7
2 xs 3	13.0	16.0	13.5	15.3	13.0		13.2
5		9.6	8.8	9.6	8.8		8.5
5 xs 5	4.8	4.8	4.8	4.8	4.8		4.6
10		2.4	2.7	2.3	2.6		2.5
10 xs 10	1.3	1.2	1.5	1.1	1.3		1.4
20	0.5	0.6	0.8	0.5	0.6		0.7

the problematic case (proper first layer, finite top layer) is relevant for practical use (where one often wants to match premiums of proper limited layers with a severity model), the problematic RRoL situation $r_1 \gg r_2 \gtrapprox r_3$ is not common. As mentioned after the theorem, it means that most losses affecting the first layer don't hit the higher layers, while most losses affecting the second layer are total losses both to the second and third layer. Apart from being arguably a rather remote case in the real world, a severity having such a rough tail geometry (within the first layer totally different from above) is hard to match anyway with a two-parameter model; it calls for more complex models, maybe piecewise defined ones as discussed e.g. in Fackler (2013).

So, it is not a big problem that the GPD fails to model some $r_1 \gg r_2 \gtrapprox r_3$ situations. On the other hand, it is remarkable that it deals well with the analogous "opposite" tail geometry $r_1 \gtrapprox r_2 \gg r_3$. Here most losses affecting the first layer are total losses both to the first and second layer, while they mostly don't hit the third layer, or if at all, only a tiny part of it. This (equally remote) tail geometry is matched by the GPD, however, one could find it disappointing that the resulting distribution typically has $\xi < 0$ and a very low supremum loss being not much higher than a_3, which leaves much of the third layer loss free (and premium free). As in practice insurance covers are not given for free, this situation requires manual correction or alternative models. An option is to use the GPD with a more restricted parameter space such that all layers have a positive probability of a total loss. This is easy, but somewhat reduces the range of three-layer problems that can be solved.

Practical experience suggests that for realistic layers and not-too-weird given premiums/RRoL's, the GPD usually solves the three-layer problem, yielding an exact match or at least a close approximation. If the inferred ξ is negative, the resulting supremum loss is mostly such large that no practical issues arise.

Despite the flexibility of the GPD (and its many other advantages mentioned earlier), it could be interesting to investigate whether other two-parameter models can solve the three-layer problem similarly – or even better. The detailed proof of the theorem given in Appendix B, which is subdivided into many steps, could lead the way for such research. While some steps are quite specific about the GPD, most of them ultimately address surjectivity, which will anyway be specific per distribution, requiring rather specific calculus. Instead, the overall setting seems rather general and easy to adapt to other models. Key is Proposition B.1 about some local properties of the logarithm of the survival function, which are necessary (and not far from sufficient) for some essential properties of the mapping given above in the sketch of proof of the theorem, e.g. local invertibility and strict monotonicity in the single variables. From the latter the way to global injectivity is not far, which is not a must for the practitioner but greatly eases interpretation of results and numerical search for solutions. So, Proposition B.1 could be used as a criterion to identify other promising models for the three-layer problem.

4 Where the model risk (not) is

Despite its flexibility, it is clear that the GPD is not always the true tail model. However, with very scarce data, finding the best fit is too ambitious an objective, one should be content with a model that is *reasonable* and *fairly close* to the unknown (to be fair: unknowable) true model. In the words of the literature on *risk vs uncertainty* (see Neth and Gigerenzer (2015) for an overview): in case of uncertainty you cannot *optimize* (best solution), instead *satisfice* (good-enough solution).

To emphasize that with scarce data input the GPD is often *good enough* a tail model, we look at an aspect of model risk (more exactly: model uncertainty) that is easily accessible and can be treated with generality. We will see that, for a not-too-long layer covering a part of the (heavy-tailed) area of large insurance losses, if we have the risk premium (first moment) and a bit more knowledge, there is not much further uncertainty about the second moment, such that different parametric severity models must lead to about the same results. For higher moments the situation is similar, but more complex.

To this end, we use the collective model of risk theory and a particular representation for the loss count. Let $[a, b]$ be a (proper) layer, N be the number of layer losses, and X_k be the severity of the k-th layer loss. The loss sizes are independent of the loss number and iid, represented by X. We don't specify the distributions of N and X further, just assume finite first and second moments. Then Wald's equations for the aggregate layer loss $S = \sum_{k=1}^{N} X_k$ yield

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

4.1 New parameters

Let us now use an alternative parameter which is introduced and explored in Chapter 5 of Fackler (2017), however, it may be much older: its empirical counterpart appears in some literature as *Poisson Index*, see e.g. Ross and Preece (1985).

Definition 4.1. The *contagion* of a counting random variable N having finite first and second moment is

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

Note that the last factor is the *over-dispersion* of N, which equals 0 in the Poisson case and is observed to be positive for a lot of real-world insurance data. Contagion and over-dispersion have the same sign. Another interesting property is that the loss counts of all layers of a (re)insurance program have the same contagion.

With $\kappa := \mathrm{Ct}(N)$ and $f = \mathrm{E}(N)$, Wald's second equation can be rewritten compactly as

$$\mathrm{CV}^2(S) = \mathrm{Ct}(N) + \frac{1}{\mathrm{E}(N)}\frac{\mathrm{E}\left(X^2\right)}{\mathrm{E}^2(X)} = \kappa + \frac{1 + \mathrm{CV}^2(X)}{f}$$

from which we see immediately that if the expected layer loss frequency is very large, the relative second moment (CV) of S depends mainly on the contagion of N, while the severity (and the uncertainty about it) plays a minor role. The assessment of the contagion is instead affected by another model risk: that of the loss count, another hard (and arguably largely underestimated) problem.

However, real-world (re)insurance layers often have very low loss frequencies. Then in the above formula the last term cannot be neglected, so we must try to assess it. For limited layers ($c = b - a < \infty$) we can get more insight by using a quantity that somehow reflects the geometry of the cdf of X.

Definition 4.2. For integer $k \geq 1$ we write

$${}_k\tau_X := \frac{\mathrm{E}\left(X^k\right)}{c^{k-1}\mathrm{E}(X)}, \qquad \tau_X := {}_2\tau_X = \frac{\mathrm{E}\left(X^2\right)}{c\,\mathrm{E}(X)}$$

${}_1\tau_X = 1$, but ${}_2\tau_X$ is most interesting, yielding extremely compact formulae for the second moment using the RRoL $r = f\,\mathrm{E}(X)/c$:

$$\mathrm{CV}^2(S) = \kappa + \frac{\tau_X}{r}, \qquad \mathrm{Var}(S) = \mathrm{E}^2(S)\left(\kappa + \frac{\tau_X}{r}\right) \qquad \mathrm{E}\left(S^2\right) = \mathrm{E}^2(S)\left(1 + \kappa + \frac{\tau_X}{r}\right)$$

For higher moments analogous, albeit much more complex, formulae hold; we will briefly look at them later.

4.2 Bounds for severity moments

Can we “estimate” ${}_k\tau_X$ without knowing much about the distribution of X? Only ${}_k\tau_X \leq 1$ is obvious as $X \leq c$ (equivalence means that all layer losses are total losses). However, in some situations it is possible to narrow down ${}_k\tau_X$ surprisingly well, in particular for heavy severity tails. The heuristics behind the following mathematics is that for such tails the obvious inequality $X^k \leq c^{k-1}X$, which is anything but sharp, is surprisingly so in expectation.

Proposition 4.3. *Suppose we have a proper limited layer $[a, b]$ being affected by layer losses represented by X, which leads to figures $f > r \geq g \geq 0$, where $r > 0$ is the RRoL, $f = E(N)$ is the frequency of layer losses and g is the frequency of total layer losses as defined in Remark 3.2. Then for integer $k \geq 1$ we have*

$$1 - \left(1 - \left(\frac{r-g}{f-g}\right)^{k-1}\right)\frac{r-g}{r} \leq {}_k\tau_X \leq 1$$

and in particular for $k = 2$

$$1 - \frac{f-r}{f-g}\frac{r-g}{r} \leq \tau_X \leq 1$$

If the cdf of X is concave on $[0, c)$ (as is often observed for large insurance losses), we have

$$1 - \left(\frac{1}{3} + \frac{2}{3}\frac{(f-r)-(r-g)}{f-g}\right)\frac{r-g}{r} \leq \tau_X \leq 1 - \frac{1}{3}\frac{r-g}{r}$$

In this case (which implies $r - g \leq f - r$) the resulting interval for τ_X is contained in the preceding one, being smaller by a third at least. The inequalities are sharp, i.e., cannot be amended by narrower bounds. The first and second one also hold if f is replaced by an upper bound and/or g by a lower bound.

The second formula is essentially Formula (1)/(2) in Aviv (2018), slightly amended and generalized – and profoundly rearranged in order to make it more easily interpretable.

Proof. Recall that X represents the *excess* loss to the layer (not the ground-up loss), thus takes values between 0 and $c = b-a$. Its survival function $\bar{F}_X(x)$ decreases on $[0, c)$ from $\bar{F}_X(0) = 1$ to the percentage of total layer losses $\bar{F}_X(c-) = \mathrm{P}(X = c) = \frac{g}{f}$, then at c jumps down to 0 (mass point of the excess loss). The integral over $[0, c]$ yields

$$\int_0^c \bar{F}_X(x)\,dx = \int_0^\infty \bar{F}_X(x)\,dx = \mathrm{E}(X) = \frac{cr}{f}$$

i.e., the average of $\bar{F}_X$ on $[0, c]$ equals $\frac{r}{f}$. Now, with values at interval endpoints and average fixed, how can the decreasing function $\bar{F}_X$ look like? In particular, what values are possible for the integral

$$\int_0^c kx^{k-1}\bar{F}_X(x)\,dx = \int_0^\infty kx^{k-1}\bar{F}_X(x)\,dx = \mathrm{E}\left(X^k\right)$$

giving the k-th moment, and consequently for ${}_k\tau_X$? This is a variational problem, the details of which are assembled in Appendix A. The first inequality of Proposition A.3 reads

$$\left[\frac{g}{f} + \frac{r-g}{f}\left(\frac{r-g}{f-g}\right)^{k-1}\right]c^k \leq \int_0^c kx^{k-1}\bar{F}_X(x)\,dx \leq \frac{r}{f}c^k$$

If we divide by $c^{k-1}\mathrm{E}(X) = \frac{r}{f}c^k$ and rearrange, we get the first and the second inequality above.

The third inequality results from the analogous variational problem having the convexity of $\bar{F}_X$ (equivalent to the concavity of the cdf) on $[0, c)$ as additional constraint. It emerges if we divide the second inequality of Proposition A.3, which reads

$$\left[\frac{g}{f} + \frac{4}{3}\frac{r-g}{f}\frac{r-g}{f-g}\right]c^2 \leq \int_0^c kx^{k-1}\bar{F}_X(x)\,dx \leq \left[\frac{r}{f} - \frac{1}{3}\frac{r-g}{f}\right]c^2$$

by $c\,\mathrm{E}\,(X)$, after rearranging terms. That the resulting interval is contained in the preceding one, can be verified by some algebra, however, it is clear as the second variational problem is a "restriction" of the first one. If we compare the lengths of the two intervals, we get

$$\left(\frac{2}{3}\frac{(f-r)-(r-g)}{f-g}\frac{r-g}{r}\right) \leq \frac{2}{3}\left(\frac{f-r}{f-g}\frac{r-g}{r}\right)$$

As for uncertainty about frequencies, we show that for $k \geq 2$ ($k=1$ is trivial) the lower bound of the first/second inequality

$$1-\left(1-\left(\frac{r-g}{f-g}\right)^{k-1}\right)\frac{r-g}{r} = 1-\left\{1+\left(\frac{r-g}{f-g}\right)+...+\left(\frac{r-g}{f-g}\right)^{k-2}\right\}\left[\left(1-\left(\frac{r-g}{f-g}\right)\right)\frac{r-g}{r}\right]$$

is a decreasing function in f and an increasing one in g. The first assertion is obvious from the LHS. The second assertion can be seen from the RHS, where in the braces we have a sum of powers of $\frac{r-g}{f-g} = 1-\frac{f-r}{f-g}$, which decreases in g, while the term in the squared brackets can be written as $\frac{f-r}{r}\frac{r-g}{f-g}$, which decreases in g as well. Thus, if f is set too high and/or g is set too low, the inequality still holds, it just yields a larger interval due to a smaller lower bound. □

Whether or not the inequalities stated in the Proposition yield narrow bounds, depends essentially on where r is located in the interval $[g, f]$. For orientation: a uniform distribution would yield $r = \frac{f+g}{2}$, while a concave cdf implies $r \leq \frac{f+g}{2}$. The inequalities are very good in two obvious cases:

- r is very close to f. This means that the average layer loss is close to c, which can only occur if the partial layer losses are very few or almost total, i.e., concentrated in the area just below c. This is a rather unrealistic situation, however, as stated earlier, the GPD embraces such tails as well: for $\xi < -1$ it has convex cdf, the density rises until the supremum loss is reached.
- r is very close to g and $g > 0$. By splitting the risk premium $cr = cg + c\,(r-g)$ into the parts stemming from total and partial layer losses, respectively, we see that in this case, where $r - g \ll g$, the partial losses account just for a small part of the risk premium – they must be few and/or small, i.e., rather concentrated near the low end of the layer. This case is easy to find in the real world: If the loss severity distribution has a quite heavy tail and the layer is not too long, then the loss frequency does not drop quickly along the layer, such that g is not far closer to 0 than f is. If the cdf is further concave in the layer area (falling density), the partial losses have a (possibly much) higher probability to be small than to be large.

If we compare the intervals for ${}_k\tau_X$ given by the first inequality of the proposition, we see that their length

$$\left(1-\left(\frac{r-g}{f-g}\right)^{k-1}\right)\frac{r-g}{r}$$

increases with k, but is bounded by $\frac{r-g}{r}$, which is small for heavy-tailed severities.

Summing up, if one wants to fairly assess ${}_k\tau_X$, one needs r and the frequencies f, g, but no more details about the layer loss distribution. For layers in the middle of a tower this works even when the frequencies are not available, provided you have the RRoL's of the adjacent layers: According to the proposition, in the first/second inequality f can be replaced by an upper bound, so one can use the RRoL of the layer below. Analogously, instead of g one can use the RRoL of the layer above. Note that for the narrower interval of the convex case such replacement of f and g does not work. So, this variant is instructive theory, but less applicable in the insurance practice, where parameters are mostly somewhat uncertain.

Example 4.4. To illustrate the above inequalities, let us look at one of the situations discussed in Section 3.4, namely 4a and here the first layer 2 xs 1, to which a GPD was assigned by solving the layer-endpoints problem. The relevant data input (see Table 2) reads in the terminology of the present section: $f = 120\%$, $r = 52\%$, $g = 19\%$. With f being more than six times larger than g, and r being not close to either of them, this is not an ideal data input for the above inequalities – one would expect to see a considerable model uncertainty.

The resulting interval τ_X from the second inequality is $\tau_X \in [57\%, 100\%]$. With the maximum being 75% higher than the minimum, this interval is not narrow, but for the given data it seems more than acceptable – recall it involves no model assumption (apart from the overall applicability of the collective model). If we apply the third inequality (believing strongly in a concave cdf), we get $\tau_X \in [64\%, 79\%]$, which is a remarkably narrow interval.

For comparison we calculate the output from the corresponding GP model. Generally for the proper GPD (case $\xi > 0$, see Section 2.1) starting at $s = a$, the moments of the severity X of the (proper limited) layer $[a, b]$ can be written very compactly with the abbreviation $\gamma := \frac{b+\lambda}{a+\lambda} = 1 + \frac{c}{a+\lambda}$. One sees quickly that

$$\mathrm{E}(X) = (a+\lambda)\,\frac{1-\gamma^{1-\alpha}}{\alpha-1}, \qquad \mathrm{E}\left(X^2\right) = 2\,(a+\lambda)^2 \left(\frac{1-\gamma^{2-\alpha}}{\alpha-2} - \frac{1-\gamma^{1-\alpha}}{\alpha-1}\right)$$

which yields

$$\tau_X = \frac{2}{\gamma-1}\left(\frac{\alpha-1}{\alpha-2}\frac{1-\gamma^{2-\alpha}}{1-\gamma^{1-\alpha}} - 1\right)$$

If we plug in the inferred proper-GP parameters $\alpha = 5.04$ and $\lambda = 3.54$, we get $\tau_X = 70\%$.

4.3 Skewness

We have closely looked at the second moment of S, where the formulae look very elegant. As for higher moments, it is well known that $\mathrm{E}\left(S^k\right)$ can be expressed as a sum of products of moments of N and of X of order up to k. However such a (intricate) formula in detail looks, if we express the appearing $\mathrm{E}\left(X^i\right)$ by the respective ${}_i\tau_X$ and are in a situation where we have narrow intervals for the latter, we overall have a formula for $\mathrm{E}\left(S^k\right)$ depending only weakly on the distribution of X, thus depending mainly on the moments of N.

Let us illustrate this for the third moment. If we denote the third central moment of a r.v. by μ_3 and use the well-known formulae

$$\mu_3(S) = \mu_3(N)\,\mathrm{E}^3(X) + 3\,\mathrm{Var}(N)\,\mathrm{E}(X)\,\mathrm{Var}(X) + \mathrm{E}(N)\,\mu_3(X)$$

$$\mu_3(X) = \mathrm{E}\left(X^3\right) - 3\,\mathrm{Var}(X)\,\mathrm{E}(X) - \mathrm{E}^3(X)$$

we get with

$$f = \mathrm{E}(N), \qquad \kappa = \mathrm{Ct}(N), \qquad cr = f\,\mathrm{E}(X) \;\Rightarrow\; \frac{\mathrm{E}\left(X^k\right)}{\mathrm{E}^k(X)} = {}_k\tau_X\,\frac{f^{k-1}}{r^{k-1}}$$

after (quite) some algebra

$$\mu_3(S) = \mathrm{E}^3(S)\left[\frac{\mu_3(N) - f}{f^3} + 3\frac{\kappa}{f}\left(\frac{\tau_X}{r} - \frac{1}{f}\right) + \frac{{}_3\tau_X}{r^2}\right]$$

which involves the first three moments of N, not just the contagion as the corresponding formula for the second moment. So, we have a pretty complex formula, apart from the Poisson case where the first two summands in the squared brackets disappear.

Overall we can say that layers covering heavy-tailed losses often bear surprisingly low model risk from the severity side. More precisely (and modestly): once we have assessed the risk premium of the layer (which can be a pretty uncertain enterprise), there is not much further uncertainty due to severity model risk. In this case we may choose to work with the GPD simply for *practical* reasons: it solves the three-layer problem and is, as was illustrated in Section 2, easy and intuitive to work with in many ways. Why look for alternative (possibly less handy) parametric models when they will anyway yield similar statistical output?

5 Wrap up

We can conclude that the building of models from very scarce data input, via the GPD (or in extreme cases the single-parameter Pareto model) as described in this paper, is not just powerful in the sense of getting things done (which is what is expected from practitioners in the first place). It it even very satisfi(c)ing in terms of statistical modeling.

Acknowledgments

The author thanks Franz Merkl, Matthias Paulsen, and Ulrich Riegel for elegant ideas that helped keep the calculus in this paper reasonably short.

A Calculus about model risk

This appendix gives not only technical details, but also (towards the end) some useful intuition.

Lemma A.1. *Let ω be a nonnegative real-valued* increasing *(in the sense of: nondecreasing) function on a real interval having finite endpoints $a < b$. Let $\varphi : [a,b] \to [m,M]$ be a* decreasing *function to a real interval with endpoints $m < M < \infty$, having the average value*

$$\overline{\varphi}^{[a,b]} := \frac{1}{b-a}\int_a^b \varphi(x)\,dx = u$$

which in particular implies $m \le u \le M$. Then for the integral of $\omega\varphi$ over $[a,b]$ we have the inequality

$$m\int_a^b \omega(x)\,dx + (M-m)\int_a^z \omega(x)\,dx \le \int_a^b \omega(x)\,\varphi(x)\,dx \le u\int_a^b \omega(x)\,dx$$

where

$$z = a + (b-a)\,\frac{u-m}{M-m}$$

All appearing integrals are well defined; in particular $z \in [a,b]$.

The bounds in the above inequality cannot be improved and are taken on, namely for piecewise constant *functions fulfilling the above conditions: the maximum for $\varphi_{max} \equiv u$ on $[a,b)$; the minimum for*

$$\varphi_{min}(x) = \begin{cases} M, & a \le x < z \\ m, & z \le x < b \end{cases}$$

The two boundary functions are not exactly specified at b as their value at this point does not matter, we just require $m \le \varphi(b) \le \varphi(b-)$.

The results remain the same if we add the additional constraint that φ be right-continuous.

Proof. $z \in [a,b]$ is clear; ω and φ are monotonic, thus (Riemann) integrable. So, all integrals are well defined. φ_{max} and φ_{min} obviously belong to the class of functions φ in question. One sees quickly that the integral of $\omega\varphi_{max}$ ($\omega\varphi_{min}$) over $[a,b]$ equals the upper (lower) bound stated in the above inequality. It remains to show that the integral of $\omega\varphi$ is within these bounds for any φ fulfilling the above conditions.

Consider such a φ. Note that, while φ may be imagined to decrease from M to m, this Lemma is about the wider class of functions having more weakly $\varphi(a) \le M$ and $\varphi(b) \ge m$.

Due to the definition of φ_{min}, $\varphi_{min} \ge \varphi$ on $(0,z)$ and $\varphi_{min} \le \varphi$ on (z,b), such that we have

$$\int_a^b \omega(x)\left[\varphi(x) - \varphi_{min}(x)\right]dx = \int_a^z \omega(x)\left[\varphi(x) - \varphi_{min}(x)\right]dx + \int_z^b \omega(x)\left[\varphi(x) - \varphi_{min}(x)\right]dx$$
$$\ge \int_a^z \omega(z)\left[\varphi(x) - \varphi_{min}(x)\right]dx + \int_z^b \omega(z)\left[\varphi(x) - \varphi_{min}(x)\right]dx = \omega(z)\int_a^b \left[\varphi(x) - \varphi_{min}(x)\right]dx = 0$$

As φ is decreasing having average u, there must be an $y \in [a,b]$ such that either $\varphi(y) = u$ or φ jumps at y from above u to below u. In any case, $u = \varphi_{max} \le \varphi$ on (a,y) and $u = \varphi_{max} \ge \varphi$ on (y,b), such that we have

$$\int_a^b \omega(x)\left[\varphi(x) - \varphi_{max}(x)\right]dx = \int_a^y \omega(x)\left[\varphi(x) - \varphi_{max}(x)\right]dx + \int_y^b \omega(x)\left[\varphi(x) - \varphi_{max}(x)\right]dx$$
$$\le \int_a^y \omega(y)\left[\varphi(x) - \varphi_{max}(x)\right]dx + \int_y^b \omega(y)\left[\varphi(x) - \varphi_{max}(x)\right]dx = \omega(y)\int_a^b \left[\varphi(x) - \varphi_{max}(x)\right]dx = 0$$

As φ_{max} and φ_{min} are right-continuous, the proof is identical for right-continuous functions φ. □

The heuristics behind the proof is as follows: ω can be seen as an increasing weight. To make the integral of $\omega\varphi$ large (small), we must find a φ that be rather small (large) at the beginning of the interval, then decrease few (much), in order to be rather large (small) at the end of the interval.

The same heuristics can be applied to a smaller class of functions, where the resulting boundary functions look quite different.

Lemma A.2. *In the situation of the preceding Lemma we consider the subclass of functions φ decreasing to (exactly) $\varphi(b) = m$ and being convex. This implies $m \leq u \leq \frac{m+M}{2}$. Again there are non-improvable bounds for the integral of $\omega\varphi$ over $[a, b]$, which are taken on for* piecewise linear *functions fulfilling the required conditions: the maximum for*

$$\varphi_{cxmax}(x) = 2u - m - 2(u - m)\frac{x - a}{b - a} = m + 2(u - m)\frac{b - x}{b - a}$$

on $[a, b]$; the minimum for

$$\varphi_{cxmin}(x) = \begin{cases} M - (M - m)\frac{x-a}{t-a}, & a \leq x < t \\ m, & t \leq x \leq b \end{cases}, \qquad t = a + 2(b - a)\frac{u - m}{M - m}$$

Again the results remain the same if we add the additional constraint that φ be right-continuous.

Proof. Any convex function starting at $\varphi(a) \leq M$ and ending at $\varphi(b) = m$ lies below the line connecting the two points $(a, \varphi(a))$ and $(b, \varphi(b))$, thus its average cannot exceed $\frac{M+m}{2}$, i.e., $u \leq \frac{m+M}{2}$.

As the two defined boundary functions are right-continuous, the proof is the same whether or not we add the constraint right-continuity.

One sees quickly that the boundary functions belong to the described class, in particular $t \in [a, b]$. φ_{csmax} is linear, decreasing from $2u - m \leq M$ to m, while φ_{csmin} has two linear pieces, decreasing between a and t from M to m, then being constant.

Recall that convex functions are continuous in open intervals, such that here jumps in (a, b) are impossible. Recall further that convex functions can cross a linear function at most twice.

Now consider a φ fulfilling the above conditions and being different from the two defined boundary functions. As φ has the same average as the boundary functions, it must cross each of them at least once on (a, b). We show that it crosses exactly once – the second crossing is "outside" at the endpoint b.

We must have $\varphi(a) > \varphi_{cxmax}(a)$, otherwise φ would lie below the straight line φ_{cxmax} all the time. To have the same average, somewhere in (a, b) φ must be below φ_{cxmax}, thus the two functions cross in a point $y^* \in (a, b)$. They further cross again at b, so we have $\varphi_{cxmax} \leq \varphi$ on $(0, y^*)$ and $\varphi_{cxmax} \geq \varphi$ on (y^*, b). The rest of the calculation is as in the preceding Lemma.

Due to the definition of φ_{cxmin}, $\varphi_{cxmin}(t) \leq \varphi(t)$. However, this inequality must be strict, otherwise φ would equal φ_{cxmin} on (t, b) and lie below φ_{cxmin} on (a, t). Thus, to have the same average, φ must be lower than φ_{cxmin} somewhere before t, which means that the two functions must cross on (a, t) in a point z^* and due to the linearity of φ_{cxmin} on (a, t) and the convexity of φ there can be no further crossing. So we have $\varphi_{cxmin} \geq \varphi$ on (a, z^*) and $\varphi_{min} \leq \varphi$ on (z^*, b). The rest of the calculation is as in the preceding Lemma. $\square$

Proposition A.3. *With $0 \leq g \leq r < f$, $r > 0$ let $\varphi : [0, c] \to \left[\frac{g}{f}, 1\right]$ be a decreasing right-continuous function having the average $\overline{\varphi}^{[0,c]} = \frac{r}{f}$. Then for integer $k \geq 1$ we have*

$$\left[\frac{g}{f} + \frac{r - g}{f}\left(\frac{r - g}{f - g}\right)^{k-1}\right]c^k \leq \int_0^c kx^{k-1}\varphi(x)\,dx \leq \frac{r}{f}c^k$$

If φ is convex and $\varphi(c) = \frac{g}{f}$, which implies $g \leq r \leq \frac{f+g}{2}$, we have a narrower interval, for $k = 2$ namely

$$\left[\frac{g}{f} + \frac{4}{3}\frac{r - g}{f}\frac{r - g}{f - g}\right]c^2 \leq \int_0^c 2x\varphi(x)\,dx \leq \left[\frac{r}{f} - \frac{1}{3}\frac{r - g}{f}\right]c^2$$

Proof. We are in the situation of the two preceding Lemmas with $a = 0$, $b = b - a = c$, $M = 1$, $m = \frac{g}{f}$, $u = \frac{r}{f}$, $\omega(x) = kx^{k-1}$. It remains to calculate some intermediate points and integrals. The ingredients for the first lemma are

$$z = \frac{r - g}{f - g}c, \qquad \int_0^c kx^{k-1}dx = c^k, \qquad \int_0^z kx^{k-1}dx = \left(\frac{r - g}{f - g}\right)^k c^k$$

and putting pieces together we get (after some algebra) the first formula stated above. For the convex case we have

$$\varphi_{cxmax}(x) = \frac{2r-g}{f} - 2\frac{r-g}{f}\frac{x}{c}$$

$$\varphi_{cxmin}(x) = \begin{cases} 1 - \frac{f-g}{f}\frac{x}{t}, & 0 \le x < t, \\ \frac{g}{f}, & t \le x \le c, \end{cases} \qquad t = 2c\frac{r-g}{f-g}$$

The calculation of the integrals of $2x\varphi_{cxmax}$ and $2x\varphi_{cxmin}$ over $[0, c]$ is straightforward. $\square$

The bounds in the two inequalities can be written in different ways. We have chosen a variant that can be arrived at by not too much algebra and that emphasizes the shrinking from the first to the second interval.

The setting of the proposition has a probabilistic interpretation, namely the situation of Proposition 4.3. There is a one-to-one correspondence between right-continuous functions decreasing on $[0, c]$ from 1 to $\frac{g}{f}$ and survival functions of layer severities, where c is the liability (cover) of the layer and $\frac{g}{f}$ is the total-loss / excess-loss frequency ratio. The only deviation is that the survival functions jump to 0 at c, but the value of the function at the interval endpoint does not matter.

A technicality shall be mentioned: Functions φ may take on a value lower than 1 at 0. (Indeed φ_{max} and φ_{cxmax} do.) In the probabilistic interpretation this means a mass point of layer losses equaling 0. These are no proper excess losses, we could call them "zero losses". When they are counted as layer losses, the true layer loss frequency is lower than f. On the other hand, functions may possibly (but not in the convex case) take on a value greater than $\frac{g}{f}$ at $c-$. (Indeed φ_{max} does.) This means that the mass point of total layer losses has a probability greater than $\frac{g}{f}$, such that the true frequency of total losses is possibly greater than g. We could have excluded these (a bit confusing) cases by requiring $\varphi(0+) = 1$, $\varphi(c-) = \frac{g}{f}$, but this would not have affected the derived inequalities, as the boundary functions we have used can be approximated by similar ones decreasing from exactly 1 to exactly $\frac{g}{f}$.

Let us finally explain the layer loss distributions underlying the four boundary functions. If these functions are layer severity survival functions, they describe:

φ_{max}: only two (layer) loss sizes occur, constituting mass points, namely total layer losses (of size c) with probability $u = \frac{r}{f}$, zero losses with probability $1 - u$.

φ_{min}: only two loss sizes occur, total layer losses with probability $m = \frac{g}{f}$, partial layer losses of size $z = c\frac{r-g}{f-g}$ with probability $1 - m$.

φ_{cxmax}: uniform distribution between the loss sizes 0 and c, plus two mass points, namely zero losses with probability $1 + \frac{g}{f} - 2\frac{r}{f}$, total losses with probability $\frac{g}{f}$.

φ_{cxmin}: uniform distribution between the loss sizes 0 and $t = 2c\frac{r-g}{f-g}$, then a gap, finally the mass point of total losses having probability $\frac{g}{f}$.

B Proof of theorem

We subdivide the proof into many steps: firstly because it is lengthy, secondly to make the steps more transparent and intuitive, thirdly because some intermediate results are interesting in their own right and/or can possibly be generalized or adapted for other distribution models.

B.1 Preliminaries

The core of the proof is a number of mathematical properties of the logarithm of the survival function of the GPD.

Proposition B.1. *The function*

$$\psi(x,s,\xi,\sigma) := -\frac{\ln\left(1+\xi\frac{x-s}{\sigma}\right)}{\xi} = \ln\left(\left(1+\xi\frac{x-s}{\sigma}\right)^{-\frac{1}{\xi}}\right), \qquad x > s \geq 0,\ \xi \in \mathbb{R},\ \sigma > (-\xi(x-s))^{+}$$

is well defined and continuously differentiable (C1) in all variables (for s=0 we mean right-differentiable). The partial derivatives ψ_ξ and ψ_σ are (strictly) positive. Interpreted as functions of x, they are strictly increasing and $\psi_\xi \circ \psi_\sigma^{-1}$ is strictly convex.

Proof. The constraints of the domain (in particular on σ) ensure that the term under the logarithm is positive. With the auxiliary variables $v := \frac{x-s}{\sigma} > 0$, $z := \xi v > -1$, we have

$$\psi = -\frac{\ln(1+\xi v)}{\xi} = -v\frac{\ln(1+z)}{z}$$

where the last factor can be written as a well-defined Taylor series about $z=0$:

$$\rho(z) := \frac{\ln(1+z)}{z} = \sum_{k=0}^{\infty}\frac{(-z)^k}{(k+1)!}$$

This yields in particular $\psi(x,s,0,\sigma) = -v = -\frac{x-s}{\sigma}$. ψ is C1 in z, v, and thus in the original variables. The partial derivatives can be calculated easily, always looking separately at the case $\xi = 0$. We have (for $\xi \neq 0$)

$$\psi_\sigma = \frac{\partial\psi}{\partial v}\frac{\partial v}{\partial\sigma} = \frac{-1}{1+\xi v}\frac{-v}{\sigma} = \frac{v}{(1+\xi v)\sigma} = \frac{1}{\left(\frac{1}{v}+\xi\right)\sigma}$$

The final representation holds, as can be checked quickly, for $\xi = 0$ as well. One sees at a glance that ψ_σ is positive and strictly increasing in v and thus in x. Further we have (for $\xi \neq 0$)

$$\psi_\xi = -v\frac{\partial\rho}{\partial z}\frac{\partial z}{\partial\xi} = -\frac{v^2}{z^2}\left[\frac{z}{1+z} - \ln(1+z)\right] = \frac{1}{\xi^2}\left[\ln(1+z) - \frac{z}{1+z}\right]$$

which is positive as the terms in the last squared brackets can be written as

$$\int_0^z \left(\frac{1}{1+t} - \frac{1}{(1+t)^2}\right)dt = \int_0^z \frac{t}{(1+t)^2}dt$$

which is positive for $z \neq 0$ or equivalently $\xi \neq 0$. As for $\xi = 0$, we get analogously

$$\psi_\xi = -v^2\left.\frac{\partial\rho}{\partial z}\right|_{z=0} = \frac{v^2}{2} > 0$$

If we interpret $\psi_\xi = \frac{1}{\xi^2}\left[\ln(1+\xi v) - \frac{\xi v}{1+\xi v}\right]$ as a function of v and take the derivative, we get $\frac{v}{(1+\xi v)^2} > 0$; the same result holds for $\xi = 0$. So, ψ_ξ is strictly increasing in v and in x.

Finally, as ψ_ξ and ψ_σ are defined on the same domain (that of ψ), $\psi_\xi \circ \psi_\sigma^{-1}$ is well defined. Setting (for $\xi \neq 0$)

$$y = \psi_\sigma = \frac{v}{(1+\xi v)\sigma} = \frac{1}{\xi\sigma}\frac{z}{1+z}$$

we get the equivalent formulae

$$\xi\sigma y = \frac{z}{1+z}, \qquad \frac{1}{1+z} = 1 - \xi\sigma y$$

which can be plugged in the above formula for ψ_ξ. This yields

$$\psi_\xi\left(\psi_\sigma^{-1}(y)\right) = -\frac{1}{\xi^2}\left[\ln(1-\xi\sigma y)+\xi\sigma y\right]$$

which is twice differentiable in y and we have

$$\left(\psi_\xi \circ \psi_\sigma^{-1}\right)''(y) = \frac{\sigma^2}{(1-\xi\sigma y)^2} > 0$$

One checks quickly that the last formula holds for $\xi = 0$ too, such that we have shown the stated strict convexity. □

Definition B.2. If a nonnegative (weighting) function w and its product with another real-valued function g are both integrable on an interval $[a,b]$, $-\infty < a < b \leq \infty$, we call

$$^{w}\overline{g}^{[a,b]} := \frac{\int_a^b g(x)\,w(x)\,dx}{\int_a^b w(x)\,dx}$$

the *w-weighted average of g on* $[a,b]$. If w and g are right-continuous, this definition extends (via $b \searrow a$) in a natural way to the case $a = b$, yielding the "average" $g(a)$.

Where $w(x) = 0$, $g(x)$ has no impact. This definition shall embrace functions g that are defined only on $[a,b] \setminus w^{-1}(0)$.

Proposition B.3. *For given $0 \leq s \leq a < b = a + c$ and with the constraint $\xi < 1$ for $b = \infty$; with $\bar{F}(x) = \left(\left(1+\xi\frac{x-s}{\sigma}\right)^+\right)^{-\frac{1}{\xi}}$, $x \geq s$; the function*

$$E : \left\{(\xi,\sigma) \in \mathbb{R}^2 \mid \sigma > (-\xi(a-s))^+\right\} \ni (\xi,\sigma) \mapsto E(s,a,b,\xi,\sigma) = \int_a^b \bar{F}(x)\,dx \in (0,\infty)$$

is C1 and one can get its partial derivatives by differentiating under the integral. For the logarithmic partial derivatives we have

$$\frac{E_\xi}{E} = {}^{\bar{F}}\overline{\psi_\xi}^{[a,b]} = {}^{\bar{F}}\overline{\psi_\xi}^{[a,b^*]}, \qquad \frac{E_\sigma}{E} = {}^{\bar{F}}\overline{\psi_\sigma}^{[a,b]} = {}^{\bar{F}}\overline{\psi_\sigma}^{[a,b^*]}$$

where $\psi = \ln\left(\bar{F}\right)$ being undefined where $\bar{F} = 0$, and b^* is the upper interval endpoint capped by the supremum loss:

$$b^* := \begin{cases} \min\left(b, a_1 + \frac{\sigma}{-\xi}\right), & \xi < 0 \\ b & \xi \geq 0 \end{cases}$$

For $b < \infty$ the same results hold for the function $R := E/c$, yielding the same logarithmic derivatives, and can be meaningfully extended to the case $b = a$.

Proof. The domain of (ξ,σ) is such that E is positive, which makes logarithmic derivatives in principle possible. As for partial derivatives, in most situations we can apply the Leibniz integral rule in its basic form. The integrand $\bar{F}$ is indeed C1 in the parameters, apart from the case $\xi < 0$, where we have continuity, but possibly a kink at $-\xi(x-s) = \sigma$ or equivalently $x = s + \frac{\sigma}{-\xi}$. If $b < s + \frac{\sigma}{-\xi}$, the kink is beyond the interval, while for $b \geq s + \frac{\sigma}{-\xi}$ we can rewrite the integral as

$$E = \int_a^{s+\frac{\sigma}{-\xi}} \left(1+\xi\frac{x-s}{\sigma}\right)^{-\frac{1}{\xi}} dx$$

with a C1 integrand and a variable C1 upper bound. In both cases we can differentiate under the integral according to the Leibniz rule, and the extra terms coming in for variable interval endpoints equal 0. For $\xi \geq 0$ the integrand is C1 anyway, so applying the Leibniz rule is straightforward if b is finite. In the remaining case $0 \leq \xi < 1$, $b = \infty$ we can see the differentiability (under the integral) e.g. by verifying that the integral with finite b and its partial derivatives converge for $b \to \infty$ to the infinite integral and its derivatives, respectively.

As for the logarithmic derivatives, if we denote ξ or σ by t, we get

$$\frac{E_t}{E} = \frac{\partial_t \int_a^b \bar{F}(x)\,dx}{\int_a^b \bar{F}(x)\,dx} = \frac{\int_a^b \partial_t \bar{F}(x)\,dx}{\int_a^b \bar{F}(x)\,dx} = \frac{\int_a^{b^*} \frac{\partial_t \bar{F}(x)}{\bar{F}(x)} \bar{F}(x)\,dx}{\int_a^{b^*} \bar{F}(x)\,dx} = \frac{\int_a^{b^*} \psi_t \bar{F}(x)\,dx}{\int_a^{b^*} \bar{F}(x)\,dx} = {}^{\bar{F}}\overline{\psi_t}^{[a,b^*]} = {}^{\bar{F}}\,\overline{\psi_t}^{[a,b]}$$

with the function $\psi = \ln\left(\bar{F}\right)$ studied above, which is defined where $\bar{F} > 0$. The last equivalence holds as the weight w equals 0 between b^* and b.

If $a < b$, R equals E up to a constant, such that the properties of the partial derivatives are the same and the logarithmic derivatives coincide. For $b = a$ we have (in the given domain)

$$R(s, a, a, \xi, \sigma) = \left(1 + \xi \frac{a - s}{\sigma}\right)^{-\frac{1}{\xi}}$$

which is C1 in ξ and σ and the partial logarithmic derivatives are the partial derivatives of ψ. Thus,

$$\frac{\partial}{\partial t}\left(\ln\left(R(s, a, a, \xi, \sigma)\right)\right) = \psi_t(a) = {}^{\bar{F}}\overline{\psi_t}^{[a,a]}$$

where the last term is a degenerate, but well defined, weighted average. $\square$

Remark B.4. For $a > s$ the domain of E and R is bounded from below by the half-line $-\xi(a - s) = \sigma > 0$, which separates the parameter situations where the supremum loss is below the layer attachment point from those where it is above, such that the layer has a positive probability of loss. See Figure 2 showing this fifth half-line, which unlike the other four indicates no mathematical property, but applicability for the modeling of a given layer/threshold.

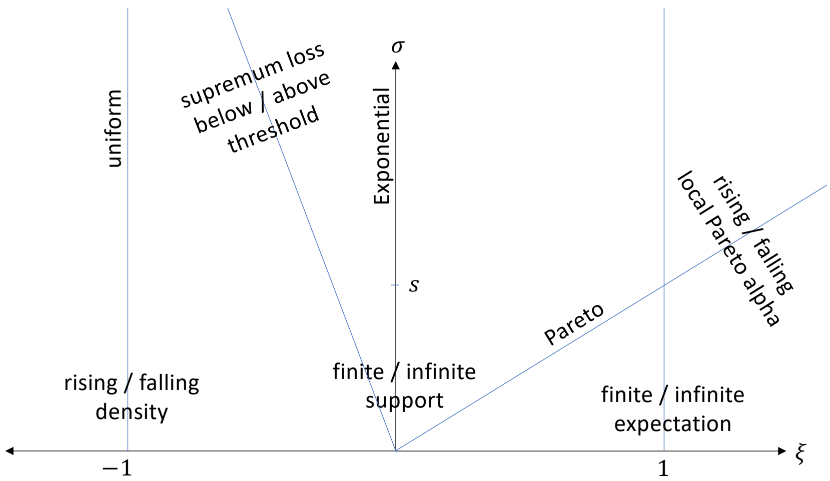

Figure 2: Admissible GPD areas

As a last ingredient we need a Lemma about convex functions on three intervals.

Lemma B.5. *Let f, g, w be real-valued functions on an interval I such that w, fw, gw are integrable; w is positive; f is strictly increasing; and $h := g \circ f^{-1}$ is strictly convex. Let further $[a_1, b_1] < [a_2, b_2] < [a_3, b_3]$ be three weakly overlapping (possibly degenerate) subintervals of I such that $[a_1, b_1] \cup [a_3, b_3] \neq [a_2, b_2]$. Then for the respective weighted averages*

$$\overline{g}^i := {}^{w}\overline{g}^{[a_i, b_i]}, \qquad \overline{f}^i := {}^{w}\overline{f}^{[a_i, b_i]}$$

the following formula holds:

$$\left(\overline{g}^2 - \overline{g}^1\right)\left(\overline{f}^3 - \overline{f}^2\right) - \left(\overline{g}^3 - \overline{g}^2\right)\left(\overline{f}^2 - \overline{f}^1\right) < 0$$

Proof. Let g_k be any real numbers and $f_1 < f_2 < f_3$. Then we have

$$\begin{aligned} D\left(\left(f_k, g_k\right)_k\right) &:= (g_2 - g_1)(f_3 - f_2) - (g_3 - g_2)(f_2 - f_1) \\ &= g_1 f_2 + g_2 f_3 + g_3 f_1 - g_1 f_3 - g_2 f_1 - g_3 f_2 = (f_3 - f_2)(f_2 - f_1)\left\{\frac{g_2 - g_1}{f_2 - f_1} - \frac{g_3 - g_2}{f_3 - f_2}\right\} \end{aligned}$$

where the first representation is as above, the second cyclically symmetric, and the third a product of two positive terms with a difference (in braces) indicating convexity. Indeed, if we choose any values $x_1 < x_2 < x_3$ in I and set $f_k = f(x_k)$, $g_k = g(x_k)$, we have $f_1 < f_2 < f_3$ and $h(f_k) = g(x_k)$, which yields

$$\frac{g_2 - g_1}{f_2 - f_1} - \frac{g_3 - g_2}{f_3 - f_2} = \frac{h(f_2) - h(f_1)}{f_2 - f_1} - \frac{h(f_3) - h(f_2)}{f_3 - f_2} < 0$$

due to the strict convexity of h.

Now assume that the three given subintervals do not overlap and choose $x_k \in [a_k, b_k]$, which implies $x_1 < x_2 < x_3$. First interpret $D\left((f(x_k), g(x_k))_k\right)$ as a function of x_1, which we let vary in $[a_1, b_1]$, while keeping x_2 and x_3 fixed. Note that

$$D(f(x_1), f(x_2), f(x_3), g(x_1), g(x_2), g(x_3)) = g(x_1) f_2 + g_2 f_3 + g_3 f(x_1) - g(x_1) f_3 - g_2 f(x_1) - g_3 f_2 < 0$$

is a linear combination of the functions $f(x_1)$ and $g(x_1)$, which we can integrate over $[a_1, b_1]$ using the positive weight w. Dividing the result by $\int_{a_1}^{b_1} w(x)\, dx > 0$ yields

$$D\left(\overline{f}^1, f(x_2), f(x_3), \overline{g}^1, g(x_2), g(x_3)\right) = \overline{g}^1 f_2 + g_2 f_3 + g_3 \overline{f}^1 - \overline{g}^1 f_3 - g_2 \overline{f}^1 - g_3 f_2 < 0$$

(In the degenerate case $a_1 = b_1$ we get the same inequality by doing nothing, as here $f(x_1) = \overline{f}^1$, $g(x_1) = \overline{g}^1$.) The last inequality holds for any $x_2 \in [a_2, b_2]$ and $x_3 \in [a_3, b_3]$. If we still keep x_3 fixed and integrate by x_2, we get analogously

$$D\left(\overline{f}^1, \overline{f}^2, f(x_3), \overline{g}^1, \overline{g}^2, g(x_3)\right) = \overline{g}^1 \overline{f}^2 + \overline{g}^2 f_3 + g_3 \overline{f}^1 - \overline{g}^1 f_3 - \overline{g}^2 \overline{f}^1 - g_3 \overline{f}^2 < 0$$

and finally by a third integration $D\left(\overline{f}^1, \overline{f}^2, \overline{f}^3, \overline{g}^1, \overline{g}^2, \overline{g}^3\right) < 0$.

Having proved the lemma for non-overlapping intervals, let us now look at interval triples that are just ordered, i.e., $[a_1, b_1] < [a_2, b_2] < [a_3, b_3]$. If all six layer endpoints are different, ordering them yields a partition of the interval $[a_1, b_3]$ into a "tower" of five proper, non-overlapping, *basic subintervals.* If some of the layer endpoints coincide, we can order them nevertheless and interpret any interval between two coinciding points as a degenerate interval of length zero, such that we have again a tower of five subintervals, now with one or more of them being degenerate.

For ease of reading we use a compact (and somewhat loose) notation, naming the basic subintervals (in rising order) u, v, x, y, z and writing for unions of adjacent subintervals vx, xyz, etc. Each interval $[a_i, b_i]$ is a union of some of the basic subintervals. Investigating how the former can overlap (or not), one sees quickly that if one rewrites $[a_1, b_1] < [a_2, b_2] < [a_3, b_3]$ in terms of the basic subintervals, there are five possibilities:

$$u < x < z, \quad u < xy < yz, \quad uv < vx < z, \quad uv < vxy < yz, \quad uvx < vxy < xyz$$

The *first* case is non-overlapping and already done. In the second and third case two intervals overlap (and are thus proper). In the fourth case the middle interval overlaps with each of the others, but the latter do not overlap, which means weakly overlapping. In the last case all three layers share the subinterval x, which cannot be degenerate, otherwise ($uvx = uv$, $xyz = yz$) one would be in the preceding case. The fifth case is not weakly overlapping and we don't have to treat it. (In fact here the stated inequality does not generally hold, as can be seen easily from calculating examples of three largely overlapping layers.)

Now consider the *second* case. For orientation note that in the situation $u < xy < yz$ all basic subintervals may be degenerate apart from y, which must be proper, otherwise we would have $xy = x$, $yz = z$ and be back in the first case. If x is degenerate, we have $xy = y \subset yz$. z can be degenerate too (yielding an analogous situation), but not at the same time as x, otherwise we would have $xy = yz$. Let us introduce some more compact notation. For any two basic subintervals $r \neq s$ we write

$$|r| := \int_r w(x)\, dx, \qquad \overline{g}^r := {}^w\overline{g}^r, \qquad \overline{f}^r := {}^w\overline{f}^r, \qquad \Upsilon_{r,s} := \frac{\overline{g}^s - \overline{g}^r}{\overline{f}^s - \overline{f}^r}$$

where $\overline{f}^r \neq \overline{f}^s$ as f is strictly increasing. Some algebra shows that for proper y we have

$$\bar{g}^{xy} = \frac{|x|}{|x| + |y|} \bar{g}^x + \frac{|y|}{|x| + |y|} \bar{g}^y, \qquad \bar{g}^{xy} - \bar{g}^u = \frac{|x|}{|x| + |y|} (\bar{g}^x - \bar{g}^u) + \frac{|y|}{|x| + |y|} (\bar{g}^y - \bar{g}^u)$$

$$\bar{g}^{yz} = \frac{|y|}{|y|+|z|}\bar{g}^{y} + \frac{|z|}{|y|+|z|}\bar{g}^{z}, \qquad \bar{g}^{yz} - \bar{g}^{xy} = \frac{|y|}{|y|+|z|}\left(\bar{g}^{z} - \bar{g}^{y}\right) + \frac{|y|}{|x|+|y|}\left(\bar{g}^{y} - \bar{g}^{x}\right)$$

which together with analogous formulae for f yields $\Upsilon_{xy,y} = \Upsilon_{x,y}$, $\Upsilon_{yz,z} = \Upsilon_{y,z}$ and (after more algebra)

$$\frac{\overline{g}^2 - \overline{g}^1}{\overline{f}^2 - \overline{f}^1} = \Upsilon_{u,xy} = p_1\Upsilon_{u,x} + (1-p_1)\,\Upsilon_{u,y}, \quad p_1 = \frac{|x|\left(\overline{f}^x - \overline{f}^u\right)}{|x|\left(\overline{f}^x - \overline{f}^u\right) + |y|\left(\overline{f}^y - \overline{f}^u\right)}$$

$$\frac{\overline{g}^3 - \overline{g}^2}{\overline{f}^3 - \overline{f}^2} = \Upsilon_{xy,yz} = p_2\Upsilon_{x,y} + (1-p_2)\,\Upsilon_{y,z}, \quad p_2 = \frac{\overline{f}^y - \overline{f}^{xy}}{\overline{f}^{yz} - \overline{f}^{xy}}$$

As y is a proper interval and f a strictly increasing function, all appearing denominators are positive and $p_1, p_2 \in [0,1]$. Thus, the last two formulae constitute convex combinations and quickly lead to the desired inequality as follows. Each triple of basic subintervals $r < s < t$ is non-overlapping, hence $\Upsilon_{r,s} < \Upsilon_{s,t}$ which, as one sees quickly, is equivalent to both $\Upsilon_{r,s} < \Upsilon_{r,t}$ and $\Upsilon_{r,t} < \Upsilon_{s,t}$. Thus, we have $\Upsilon_{u,x} < \Upsilon_{u,y} < \Upsilon_{x,y} < \Upsilon_{y,z}$, which immediately implies

$$\frac{\overline{g}^2 - \overline{g}^1}{\overline{f}^2 - \overline{f}^1} < \frac{\overline{g}^3 - \overline{g}^2}{\overline{f}^3 - \overline{f}^2}$$

The *third* case $uv < vx < z$ has the same proof as the second, just swap u and z, v and y.

Finally, in the *fourth* case $uv < vxy < yz$ both v and y must be proper, otherwise we would have less than two overlaps and be back in one of the preceding cases. By adapting formulae from the second case, we get

$$\frac{\overline{g}^2 - \overline{g}^1}{\overline{f}^2 - \overline{f}^1} = \Upsilon_{uv,vxy} = p_3\Upsilon_{u,v} + (1-p_3)\,\Upsilon_{v,xy} = p_3\Upsilon_{u,v} + (1-p_3)\,p_4\Upsilon_{v,x} + (1-p_3)\,(1-p_4)\,\Upsilon_{v,y}$$

$$\frac{\overline{g}^3 - \overline{g}^2}{\overline{f}^3 - \overline{f}^2} = \Upsilon_{vxy,yz} = p_5\Upsilon_{vx,y} + (1-p_5)\,\Upsilon_{y,z} = p_5 p_6\Upsilon_{v,y} + p_5\,(1-p_6)\,\Upsilon_{x,y} + (1-p_5)\,\Upsilon_{y,z}$$

which are again convex combinations. As $\Upsilon_{u,v} < \Upsilon_{v,x} < \Upsilon_{v,y} < \Upsilon_{x,y} < \Upsilon_{y,z}$, the first equation yields a smaller value than the second whatever the weights, apart from one case of equivalence which requires $p_3 = 0 = p_4$, $p_5 = 1 = p_6$, such that both equations yield $\Upsilon_{v,y}$. Inspecting the weights, we see that

$$0 = p_3 = \frac{\overline{f}^v - \overline{f}^{uv}}{\overline{f}^{vxy} - \overline{f}^{uv}}, \qquad 1 = p_5 = \frac{\overline{f}^y - \overline{f}^{vxy}}{\overline{f}^{yz} - \overline{f}^{vxy}}$$

imply $v = uv$ and $y = yz$ or equivalently that u and z are degenerate, which is possible. Further we see that

$$0 = p_4 = \frac{|x|\left(\overline{f}^x - \overline{f}^v\right)}{|x|\left(\overline{f}^x - \overline{f}^v\right) + |y|\left(\overline{f}^y - \overline{f}^v\right)}$$

implies $|x| = 0$ or equivalently that x is degenerate, which is possible. But, u, x, z cannot be degenerate intervals at the same time: this situation means $v < vy < y$, where the middle interval is the non-overlapping union of the other two, which is exactly the one weakly overlapping case that the Lemma excludes. Here indeed the inequality turns into an equation. We are done. $\square$

If the three given intervals correspond to layers, the last and forbidden situation means that the middle layer alone provides exactly the same insurance cover as top and bottom layer together. As explained after the theorem, this is no meaningful input for the three-layer problem.

B.2 Main part

Consider the three-layer problem as posed in the theorem, i.e., we have been given

- three strongly ordered and weakly overlapping layers and/or thresholds $[a_i, b_i]$ of finite length c_i,
- respective (finite) risk premiums e_i and corresponding, strictly decreasing RRoL's $r_1 > r_2 > r_3 > 0$.
- As a particular case, $b_3 = \infty$ is admitted; here $r_3 = 0$ and one requires $e_3 > 0$.

The constraints ensure that all layers/thresholds are exposed to losses (i.e., have a positive loss frequency), even the third and highest one. Generalized Pareto models fulfill this property if one restricts the parameter space accordingly. If we choose $s = a_1$ (we don't need to model losses below the lowest layer), the adequate domain is given in the following definition, where finite expectation for unlimited layers is ensured by constraining ξ accordingly.

Definition B.6. For a given, strongly ordered, layer triple $[a_i, b_i]$ we set

$$\varDelta := \{(\xi,\sigma) \in (-\infty, \xi_{\text{sup}}) \times (0,\infty) \,|\, \sigma > -\xi\,(a_3 - a_1)\}\,, \qquad \xi_{\text{sup}} := \begin{cases} \infty, & b_3 < \infty \\ 1, & b_3 = \infty \end{cases}$$

$$\varDelta^* := \{(\xi,\sigma) \in (-\infty, \infty) \times (0,\infty) \,|\, \sigma > -\xi\,(a_2 - a_1)\}$$

Now consider three closely related functions:

Definition B.7. For a given, strongly ordered, layer triple $[a_i, b_i]$ we write $\bar{F}(x) = \left(\left(1 + \xi \frac{x-a_1}{\sigma}\right)^+\right)^{-\frac{1}{\xi}}$, $x \geq a_1$, and set: for proper layers $E_i := E(a_1, a_i, b_i, \xi, \sigma)$, for limited layers $R_i := R(a_1, a_i, b_i, \xi, \sigma)$. We can interpret

$$R_i = \frac{1}{c_i}\int_{a_i}^{b_i} \bar{F}(x)\,dx = \overline{\bar{F}}^{[a_i,b_i]}$$

as a (non-weighted) average of $\bar{F}$ across the layer, which (by taking the limit) embraces the case $a_i = b_i$, where we get $R_i = \bar{F}(a_i)$.

We define three *layer ratio functions* as follows: for proper layers $(a_i < b_i)$

$$\varPhi : \varDelta \to (0,\infty)^2\,, (\xi,\sigma) \mapsto \left(\frac{E_2}{E_1}, \frac{E_3}{E_2}\right)$$

for limited layers $(a_i \leq b_i < \infty)$

$$\varPhi^* : \varDelta \to (0,\infty)^2\,, (\xi,\sigma) \mapsto \left(\frac{R_2}{R_1}, \frac{R_3}{R_2}\right)$$

for an unlimited top layer $(a_i \leq b_i,\, b_3 = \infty)$

$$\varPhi^{**} : \varDelta \to (0,\infty)^2\,, (\xi,\sigma) \mapsto \left(\frac{R_2}{R_1}, \frac{E_3}{R_2}\right)$$

The first component of these functions shall be defined on the extended domain $\varDelta^*$.

Finally, for $\xi > 0$ we set with $\tilde{F}(x) := (x - a_1)^{-\frac{1}{\xi}}$: for proper layers

$$\tilde{E}_i := \tilde{E}(a_1, a_i, b_i, \xi) = \int_{a_i}^{b_i} \tilde{F}(x)\,dx$$

for limited layers $\tilde{R}_i := \tilde{R}(a_1, a_i, b_i, \xi)$, which embraces degenerate layers.

$\varDelta$ was chosen such that all E_i are positive and finite for proper layers, while all R_i are positive and finite for limited layers. For E_1, E_2, R_1, R_2 this holds on the domain $\varDelta^* \supseteq \varDelta$, which is larger if $a_2 < a_3$ or $b_3 = \infty$. So, the three functions are well defined. $\varPhi$ is maybe the straightforward variant, but either $\varPhi^*$ or $\varPhi^{**}$ describes the three-layer problem for proper layers equivalently to $\varPhi$ (being component-wise equal up to a factor), while more generally embracing thresholds. We can in the following largely treat the three functions in parallel (writing sometimes $\varPhi^{(**)}$), keeping in mind that the latter two (shortly written $\varPhi^{*(*)}$) cover all three-layer problems treated in the theorem. Let us first look at the borders of the domain $\varDelta$.

Lemma B.8. *For any given, strongly ordered, layer triple $[a_i, b_i]$, the layer ratio functions are continuous and can be extended continuously to the border of Δ, apart from the half-line $\xi = 1, \sigma \geq 0$ in case of infinite b_3. On the half-line $-\xi(a_3 - a_1) = \sigma > 0$ we get*

$$\Phi(\xi, \sigma) = \left(\frac{E_2}{E_1}, 0\right), \qquad \Phi^{*(*)}(\xi, \sigma) = \left(\frac{R_2}{R_1}, 0\right)$$

while on the half-line $\sigma = 0, \xi_{\sup} > \xi \geq 0$ we get

$$\Phi(\xi, 0) = \left(\frac{\tilde{E}_2}{\tilde{E}_1}, \frac{\tilde{E}_3}{\tilde{E}_2}\right), \qquad \Phi^{*}(\xi, 0) = \left(\frac{\tilde{R}_2}{\tilde{R}_1}, \frac{\tilde{R}_3}{\tilde{R}_2}\right), \qquad \Phi^{**}(\xi, 0) = \left(\frac{\tilde{R}_2}{\tilde{R}_1}, \frac{\tilde{E}_3}{\tilde{R}_2}\right)$$

or more precisely a meaningful extension thereof by zeros, which maps $(0,0)$ to $(0,0)$ and sets $\tilde{E}_2/\tilde{E}_1$ and $\tilde{R}_2/\tilde{R}_1$ equal to 0 in the case $a_1 = b_1 = a_2 < b_2, \xi \leq 1$.

The first component $\Phi_1^{(**)}$ can even be extended continuously to the border of Δ^*, which contains the whole non-negative ξ-axis. Assembling where on this axis the components of the extended layer ratio functions equal zero, we get:

$\Phi_1^{(**)}(\xi, 0) = 0$: for $\xi \in [0,1]$ anyway; for $\xi \in [0, \infty)$ if $b_1 = a_1$.

$\Phi_2^{(**)}(\xi, 0) = 0$: for $\xi = 0$ anyway; for $\xi \in [0,1]$ if $a_2 = a_1$.

For the case of an unlimited top layer this means that on the border of Δ, for $\xi < \xi_{\sup} = 1$ we have $\Phi_1^{**}(\xi, 0) \equiv 0$, while $\Phi_2^{**}(\xi, 0) \equiv 0$ too if $a_2 = a_1$.

Proof. The formulae for Φ follow immediately from those for $\Phi^{*(*)}$, so it suffices to treat the latter.

Recall first how three strongly ordered layers $[a_i, b_i]$ can relate to each other. $a_1 = a_2$ is possible, but implies $a_1 = b_1 = a_2 < b_2$. $a_2 = a_3$ is possible, but implies $a_2 = b_2 = a_3 < b_3$. In any case we must have $a_1 < a_3$, such that $-\xi(a_3 - a_1) = \sigma > 0$ defines a half-line in the quadrant $\xi < 0, \sigma > 0$, for orientation see Figure 2. For a point (ξ, σ) in this quadrant, the corresponding GPD has the supremum loss $a_1 + \frac{\sigma}{-\xi}$, such that a layer with attachment point a has zero probability of loss iff $-\xi(a - a_1) \geq \sigma > 0$.

The R_i and E_3 are well-defined and continuous functions of (ξ, σ) on the whole half-plane $\mathbb{R} \times (0, \infty)$. This domain embraces the half-line $-\xi(a_3 - a_1) = \sigma > 0$, on which R_3 and E_3 equal 0. As $R_1 > 0$ on the whole half-plane, $\frac{R_2}{R_1}$ is well defined and continuous on the whole half-plane (and positive on Δ^*).

R_2 equals 0 outside of Δ^*, i.e., for all points fulfilling $-\xi(a_2 - a_1) \geq \sigma > 0$; these constitute an area not bordering Δ if $a_2 < a_3$. Then $\frac{R_3}{R_2}$ and $\frac{E_3}{R_2}$ too are continuous about the half-line $-\xi(a_3 - a_1) = \sigma > 0$, on which they equal 0.

It remains to look at the case $a_2 = b_2 = a_3 < b_3$, where we have $R_2 = \bar{F}(a_3)$. Note that generally

$$E_i = \int_{a_i}^{b_i} \bar{F}(x)\, dx = \int_{a_i}^{b_i^*} \bar{F}(x)\, dx, \qquad b_i^* := \begin{cases} \min\left(b_i, a_1 + \frac{\sigma}{-\xi}\right), & \xi < 0 \\ b_i, & \xi \geq 0 \end{cases}$$

such that in our situation (degenerate second layer, being the attachment point of the third, which is proper) we have on Δ the inequality

$$0 < \frac{E_3}{R_2} = \frac{\int_{a_3}^{b_3^*} \bar{F}(x)\, dx}{\bar{F}(a_3)} = \int_{a_3}^{b_3^*} \frac{\bar{F}(x)}{\bar{F}(a_3)} dx \leq b_3^* - a_3$$

Thus, for any sequence (ξ_k, σ_k) converging to some point (ξ_0, σ_0) on the half-line $0 < \sigma_0 = \xi_0 (a_3 - a_1)$, we ultimately have $\xi_k < 0$ and with

$$0 < \frac{E_3}{R_2} \leq b_3^* - a_3 \leq \frac{\sigma_k}{-\xi_k} - (a_3 - a_1) \longrightarrow \frac{\sigma_0}{-\xi_0} - (a_3 - a_1) = 0$$

it is clear that $\frac{E_3}{R_2}$ has limit 0 at any point on the half-line.

Let us now investigate the other (open) half-line $\sigma = 0, \xi > 0$, for which it is sufficient to consider the layer ratio functions on the adjacent quadrant $\xi > 0, \sigma > 0$. Here we have for finite b_3

$$\frac{R_3}{R_2} = \frac{\overline{\left(\left(1 + \xi \frac{x - a_1}{\sigma}\right)^+\right)^{-\frac{1}{\xi}}}^{[a_3, b_3]}}{\overline{\left(\left(1 + \xi \frac{x - a_1}{\sigma}\right)^+\right)^{-\frac{1}{\xi}}}^{[a_2, b_2]}} = \frac{\overline{\left(\frac{\sigma}{\xi} + x - a_1\right)^{-\frac{1}{\xi}}}^{[a_3, b_3]}}{\overline{\left(\frac{\sigma}{\xi} + x - a_1\right)^{-\frac{1}{\xi}}}^{[a_2, b_2]}}$$

For any sequence (ξ_k, σ_k) converging to some point $(\xi_0, 0)$ on the positive ξ-axis, the numerator of the last representation tends to

$$\overline{(x-a_1)^{-\frac{1}{\xi_0}}}^{[a_3,b_3]} = \tilde{R}_3(\xi_0)$$

which is positive and finite as $a_3 > a_1$. Likewise, the numerator tends to $\tilde{R}_2(\xi_0)$, which, however, can be infinite, namely if $a_2 = a_1$ and $\xi_0 \leq 1$. Nevertheless we have overall for $\xi_0 > 0$

$$\lim_{k\to\infty} \frac{R_3}{R_2}(\xi_k, \sigma_k) = \frac{\tilde{R}_3}{\tilde{R}_2}(\xi_0)$$

where $\tilde{R}_2(\xi_0) = \infty$ simply means that the limit equals 0.

For infinite b_3 and $\xi \in (0,1)$ the reasoning is analogous with

$$\frac{E_3}{R_2} = \frac{\frac{\sigma}{1-\xi}\left(1+\xi\frac{a_3-a_1}{\sigma}\right)^{1-\frac{1}{\xi}}}{\overline{\left(\left(1+\xi\frac{x-a_1}{\sigma}\right)^+\right)^{-\frac{1}{\xi}}}^{[a_2,b_2]}} = \frac{\frac{\xi}{1-\xi}\left(\frac{\sigma}{\xi}+a_3-a_1\right)^{1-\frac{1}{\xi}}}{\overline{\left(\frac{\sigma}{\xi}+x-a_1\right)^{-\frac{1}{\xi}}}^{[a_2,b_2]}}$$

where, as one sees quickly, the denominator to the right converges to $\tilde{E}_3(\xi_0)$.

For the first component we have as above, in the whole quadrant $\xi > 0, \sigma > 0$,

$$\frac{R_2}{R_1} = \frac{\overline{\left(\frac{\sigma}{\xi}+x-a_1\right)^{-\frac{1}{\xi}}}^{[a_2,b_2]}}{\overline{\left(\frac{\sigma}{\xi}+x-a_1\right)^{-\frac{1}{\xi}}}^{[a_1,b_1]}}$$

If $a_1 < a_2$, we can proceed exactly as above, calculating the limits of denominator and numerator separately. The former is positive and finite, the latter is infinite if $\xi_0 \leq 1$; then the limit of $\frac{R_2}{R_1}$ equals 0.

The case $a_1 = b_1 = a_2 < b_2$ requires some extra reasoning. Here $R_1 = \bar{F}(a_1) \equiv 1$ and

$$0 < \frac{R_2}{R_1} = R_2 = \overline{\bar{F}(x)}^{[a_2,b_2]} \leq \bar{F}(b_2) = \left(1+\xi\frac{b_2-a_1}{\sigma}\right)^{-\frac{1}{\xi}}$$

As $b_2 > a_1$, the last term tends to 0 for any sequence (ξ_k, σ_k) converging to some point $(\xi_0, 0)$ on the positive ξ-axis. Thus, the limit of $\frac{R_2}{R_1}$ equals 0 and it remains to show that $\tilde{R}_2/\tilde{R}_1 = 0$ too for $a_1 = b_1$. As $\tilde{R}_1 = \tilde{F}(a_1) = \infty$, we are done if $\tilde{R}_2$ is finite, i.e., if $\xi > 1$. In the remaining case $\xi \leq 1, a_1 = b_1 = a_2 < b_2$ both $\tilde{R}_1$ and $\tilde{R}_2$ are infinite, but we can meaningfully interpret their ratio as equaling 0 via

$$\frac{\tilde{R}_2}{\tilde{R}_1} := \lim_{\varepsilon\to 0} \frac{\tilde{R}(a_1, a_1+\varepsilon, b_2, \xi)}{\tilde{R}(a_1, a_1+\varepsilon, a_1+\varepsilon, \xi)} = \lim_{\varepsilon\to 0} \frac{\overline{\tilde{F}(x)}^{[a_1+\varepsilon,b_2]}}{\tilde{F}(a_1+\varepsilon)} = 0$$

Finally, similar straightforward calculus shows that for sequences $\Delta \ni (\xi_k, \sigma_k) \to (0,0)$, the limit of all layer ratio functions equals $(0,0)$, while $\Phi_1^{(**)}(\xi_k, \sigma_k)$ converges to 0 even for sequences in the larger Δ^*. Overall we have a continuous extension of the $\Phi_1^{(**)}$ and $\Phi_2^{(**)}$ to all half-lines in question, including the point $(0,0)$ where they meet.

Collecting where on the border $\Phi_1^{(**)}$ and $\Phi_2^{(**)}$ equal 0, we get the concluding assertions of the lemma. □

Proposition B.9. *For any given, strongly ordered, layer triple $[a_i, b_i]$, the components of the layer ratio functions have the following properties at the extremes of the domain Δ:*

For fixed $\sigma > 0$, the corresponding domain of ξ such that $(\xi, \sigma) \in \Delta$, is $(\xi_{\inf}, \xi_{\sup})$ having the lower endpoint $\xi_{\inf} = \frac{-\sigma}{a_3-a_1}$. For $\xi \searrow \xi_{\inf}$ the second components of all layer ratio functions tend to 0, while for $\xi \nearrow \xi_{\sup}$ we have

$$\lim_{\xi\nearrow\xi_{\sup}} \Phi_2 = \frac{c_3}{c_2}, \qquad \lim_{\xi\nearrow\infty} \Phi_2^* = 1, \qquad \lim_{\xi\nearrow 1} \Phi_2^{**} = \infty, \qquad \lim_{\xi\nearrow\infty} \Phi_1^* = 1$$

For fixed ξ, the corresponding domain of σ such that $(\xi, \sigma) \in \Delta$, is $(\sigma_{\inf}, \infty)$, where $\sigma_{\inf} = (-\xi(a_3-a_1))^+$. For $\sigma \searrow \sigma_{\inf}$, $\Phi_2^{(**)}$ tends to 0 if $\xi \leq 0$, while for any ξ we have

$$\lim_{\sigma\nearrow\infty} \Phi_2 = \frac{c_3}{c_2}, \qquad \lim_{\sigma\nearrow\infty} \Phi_2^* = 1, \qquad \lim_{\sigma\nearrow\infty} \Phi_2^{**} = \infty, \qquad \lim_{\sigma\nearrow\infty} \Phi_1 = \frac{c_2}{c_1}, \qquad \lim_{\sigma\nearrow\infty} \Phi_1^{*(*)} = 1$$

Proof. The domains for the one-parameter functions result immediately from the definition of $\varDelta$, for orientation see Figure 2. The limits at the lower endpoints result from the preceding lemma, such that it remains to look at the upper endpoints.

Let us fix $\sigma > 0$ and first look at the case $b_3 < \infty$. For $\xi > 0$ we have

$$\frac{R_3}{R_2} = \frac{\overline{\left(\left(1+\xi\frac{x-a_1}{\sigma}\right)^+\right)^{-\frac{1}{\xi}}}^{[a_3,b_3]}}{\overline{\left(\left(1+\xi\frac{x-a_1}{\sigma}\right)^+\right)^{-\frac{1}{\xi}}}^{[a_2,b_2]}} = \frac{\overline{\left(\frac{1}{\xi}+\frac{1}{\sigma}\left(x-a_1\right)\right)^{-\frac{1}{\xi}}}^{[a_3,b_3]}}{\overline{\left(\frac{1}{\xi}+\frac{1}{\sigma}\left(x-a_1\right)\right)^{-\frac{1}{\xi}}}^{[a_2,b_2]}}$$

where in the last representation both numerator and denominator tend to 1 as $\xi \nearrow \infty$. So, we have $\varPhi_2^* = \frac{R_3}{R_2} \to 1$ and for proper layers $\varPhi_2 = \frac{c_3 R_3}{c_2 R_2} \to \frac{c_3}{c_2}$. The reasoning for $\varPhi_1^* = \frac{R_2}{R_1}$ is the same.

In the case $b_3 = \infty$ we have for $\xi > 0$

$$\frac{E_3}{R_2} = \frac{\frac{\sigma}{1-\xi}\left(1+\xi\frac{a_3-a_1}{\sigma}\right)^{1-\frac{1}{\xi}}}{\overline{\left(1+\xi\frac{x-a_1}{\sigma}\right)^{-\frac{1}{\xi}}}^{[a_2,b_2]}}$$

where for $\xi \nearrow 1$ all factors but $\frac{1}{1-\xi}$ converge to positive real numbers, such that overall we have $\varPhi_2^{**} = \frac{E_3}{R_2} \to \infty$ and for a proper second layer

$$\varPhi_2 = \frac{E_3}{c_2 R_2} \to \infty = \frac{c_3}{c_2}$$

Let us now fix $\xi \neq 0$. The R_i for limited layers and E_3 for an unlimited top layer (here $\xi < 1$), respectively, equal

$$R_i = \overline{\left(\left(1+\xi\frac{x-a_1}{\sigma}\right)^+\right)^{-\frac{1}{\xi}}}^{[a_i,b_i]}, \qquad E_3 = \frac{\sigma}{1-\xi}\left(1+\xi\frac{a_3-a_1}{\sigma}\right)^{1-\frac{1}{\xi}}$$

For $\sigma \nearrow \infty$ the former tend to 1 and the latter tends to ∞. One sees quickly that this holds for $\xi = 0$ as well. The limits for $\varPhi_2^*$, $\varPhi_2^{**}$, $\varPhi_2$, $\varPhi_1$ and $\varPhi_1^* = \varPhi_1^{**}$ follow immediately. ☐

Albeit their extensions will be useful below, many properties of the layer ratio functions hold only on the original (open) domains $\varDelta$ and $\varDelta^*$. In the following, unless specified otherwise, we treat the original functions.

Proposition B.10. *For any given, strongly ordered and weakly overlapping, layer triple, the components of the layer ratio functions have continuous and positive partial derivatives, such that the components are strictly increasing in both variables. Moreover, the layer ratio functions locally have C1 inverses and are thus, in particular, continuous and open functions. As for their range, we have*

$$\varPhi(\varDelta) \subseteq \left(0, \frac{c_2}{c_1}\right) \times \left(0, \frac{c_3}{c_2}\right), \qquad \varPhi^*(\varDelta) \subseteq (0,1)^2, \qquad \varPhi^{**}(\varDelta) \subseteq (0,1) \times (0,\infty)$$

and if we fix any $\sigma > 0$, the second component of the resulting functions in ξ are bijective in the sense that they map bijectively on the respective interval from above:

$\varPhi_2(\_,\sigma)$ *on* $\left(0, \frac{c_3}{c_2}\right)$, $\varPhi_2^*(\_,\sigma)$ *on* $(0,1)$, $\varPhi_2^{**}(\_,\sigma)$ *on* $(0,\infty)$.

Proof. In any distribution model, for limited layers, a higher layer has a smaller or equal RRoL than a lower one, and equivalence occurs iff the survival function is constant across both layers and in between, i.e., from the lower attachment point to the higher detachment point, which means that all losses affecting the lower layer are total losses for both layers. According to the choice of $\varDelta$, the survival function of the GPD is strictly decreasing from a_1 until beyond a_3, i.e., in at least part of each of the three layers. Thus, we have for finite layers $0 < \frac{R_2}{R_1} < 1$, $0 < \frac{R_3}{R_2} < 1$ and for proper layers equivalently $0 < \frac{E_2}{E_1} < \frac{c_2}{c_1}$, $0 < \frac{E_3}{E_2} < \frac{c_3}{c_2}$. The last formula notably embraces the case of an unlimited top layer, where $E_3 < \infty$ and $c_3 = \infty$. So, the ranges of the layer ratio functions lie in the two-dimensional intervals given above.

The choice of $\varDelta$ and $\varDelta^*$ enables us further to apply Proposition B.3 to all three layers. Thus, the E_i and the R_i, as well as their logarithms, are C1 in both ξ and σ; this holds on $\varDelta$ for E_3 and R_3, on $\varDelta^*$ for the remaining quantities.

Let now Λ be the (component-wise) logarithm of any of the layer ratio functions. The three variants of Λ are component-wise equal up to constants; their partial derivatives coincide and follow immediately from Proposition B.3. If we denote ξ or σ by t, we have on Δ

$$\Lambda_t = \left({}^{\bar{F}}\overline{\psi_t}^{[a_2,b_2^*]} - {}^{\bar{F}}\overline{\psi_t}^{[a_1,b_1^*]}, {}^{\bar{F}}\overline{\psi_t}^{[a_3,b_3^*]} - {}^{\bar{F}}\overline{\psi_t}^{[a_2,b_2^*]} \right)$$

where b_i^* is b_i capped by the supremum loss as introduced in the proof of Lemma B.8. The appearing terms are $\bar{F}$-weighted averages of the function $\psi_t(x)$ over the three layer intervals (with layers possibly shortened, which does not change the average). Recall from Proposition B.1 that $\psi_t(x)$ is strictly increasing. The three layers are, despite their eventual capping, still ordered (albeit the second and third capped layer may have the supremum loss as common detachment point, such that their ordering is not strong any more). As the weighting function $\bar{F}$ is positive in the area of the capped layers, we must have

$$\overline{\psi_t}^3 > \overline{\psi_t}^2 > \overline{\psi_t}^1$$

Here for ease of reading we have introduced an abbreviated notation for the averages, which we will use whenever it is clear which weighting function and intervals are being considered.

We see that the two components of Λ_t are (strictly) positive, which carries over to the components of $\Phi_t^{(**)}$. Note that for the first components the properties found extend to the larger domain Δ^*.

It remains to show that Λ (and thus $\Phi^{(**)}$) is locally a C1 diffeomorphism on Δ, for which it is sufficient to show that the Jacobian matrix

$$J := \begin{pmatrix} \Lambda_{1;\xi} & \Lambda_{1;\sigma} \\ \Lambda_{2;\xi} & \Lambda_{2;\sigma} \end{pmatrix} = \begin{pmatrix} \overline{\psi_\xi}^2 - \overline{\psi_\xi}^1 & \overline{\psi_\sigma}^2 - \overline{\psi_\sigma}^1 \\ \overline{\psi_\xi}^3 - \overline{\psi_\xi}^2 & \overline{\psi_\sigma}^3 - \overline{\psi_\sigma}^2 \end{pmatrix}$$

(where we use a compact notation for partial derivatives) is invertible. We show that its determinant is negative for any $(\xi,\sigma) \in \Delta$. We have

$$\det(J) = \Lambda_{1;\xi}\Lambda_{2;\sigma} - \Lambda_{2;\xi}\Lambda_{1;\sigma} = \left(\overline{\psi_\xi}^2 - \overline{\psi_\xi}^1\right)\left(\overline{\psi_\sigma}^3 - \overline{\psi_\sigma}^2\right) - \left(\overline{\psi_\xi}^3 - \overline{\psi_\xi}^2\right)\left(\overline{\psi_\sigma}^2 - \overline{\psi_\sigma}^1\right)$$

The last expression is exactly the formula Lemma B.5 is about. Let us check the conditions of the lemma. The intervals corresponding to the – possibly capped – layers are weakly overlapping like the original intervals. After the capping the strong ordering (of second and third layer) might be spoilt, but the condition $[a_1, b_1^*] \cup [a_3, b_3^*] \neq [a_2, b_2^*]$ is fulfilled. Due to Proposition B.1, ψ_σ is strictly increasing and $\psi_\xi \circ \psi_\sigma^{-1}$ is strictly convex in the area of the capped layers. Hence, the lemma applies and $\det(J)$ is negative.

Finally, let us fix a $\sigma > 0$ and interpret $\Phi^{(**)}$ as a function in ξ, which is defined on the domain $(\xi_{\inf}, \xi_{\sup})$ given in Proposition B.9. From above we know that $\Phi_2^{(**)}(\_,\sigma)$ has a positive derivative, thus is strictly increasing and in particular injective. To determine its range (image), we just need the limits at the endpoints of the domain, which are assembled in Proposition B.9: at $\xi_{\inf}$ the limit equals 0, at $\xi_{\sup}$ we have, according to the layer ratio function: $\frac{c_3}{c_2}$, 1, or ∞. Thus, the respective images of $\Phi_2(\_,\sigma)$ and its variants are $\left(0, \frac{c_3}{c_2}\right)$, $(0,1)$, and $(0,\infty)$. □

Proposition B.11. *For any given, strongly ordered and weakly overlapping, layer triple, the layer ratio functions are injective. Thus, the Generalized Pareto solution of the three-layer problem, if any, is unique and depends on the given input (r_i, e_i) in a continuously differentiable manner.*

Proof. Let again Λ be the component-wise logarithm of $\Phi^{(**)}$. It is sufficient to show the injectivity of Λ. Choose any point $\left(\mathring{\xi}, \mathring{\sigma}\right) \in \Delta$ and set $(p,q) = \Lambda\left(\mathring{\xi}, \mathring{\sigma}\right)$.

For any fixed $\sigma > 0$ we know from the preceding proposition that the function

$$k_\sigma(\xi) := \Lambda_2(\xi,\sigma) = \ln\left(\Phi_2^{(**)}(\xi,\sigma)\right)$$

is invertible and that there is a unique ξ such that $(\xi,\sigma) \in \Delta$ and $q = \Lambda_2(\xi,\sigma) = k_\sigma(\xi)$ (or equivalently $\Phi_2^{(**)}(\xi,\sigma) = e^q$). This means that the function

$$g_q(\sigma) := k_\sigma^{-1}(q)$$

is well defined for all $\sigma > 0$ and by construction all $(g_q(\sigma), \sigma)$ lie in Δ. Further, $g_q(\sigma)$ is C1 in σ as an implicit function solving the equation

$$\Lambda_2(g_q(\sigma), \sigma) = k_\sigma(g_q(\sigma)) = k_\sigma\left(k_\sigma^{-1}(q)\right) = q$$

We can thus take the derivative and get

$$0 = \frac{d}{d\sigma}\Lambda_2(g_q(\sigma), \sigma) = \frac{\partial}{\partial\xi}\Lambda_2(g_q(\sigma), \sigma)\, g_q'(\sigma) + \frac{\partial}{\partial\sigma}\Lambda_2(g_q(\sigma), \sigma)$$

From the preceding proof we know that the partial derivatives of the components of Λ are positive, thus we have (using again a compact notation for partial derivatives)

$$g_q'(\sigma) = -\frac{\Lambda_{2;\sigma}(g_q(\sigma), \sigma)}{\Lambda_{2;\xi}(g_q(\sigma), \sigma)} < 0$$

which means in particular that g_q is strictly decreasing.

g_q is the function which, for any σ, yields (the unique) ξ such that $\Lambda_2(\xi, \sigma) = q$. Now consider

$$h_q(\sigma) := \Lambda_1(g_q(\sigma), \sigma)$$

which is another C1 function of σ. If we can show that it is injective, this implies that there is at most one $\sigma > 0$ such that $\Lambda_1(g_q(\sigma), \sigma) = p$. As (p, q) was chosen arbitrarily in the image of Λ (through choice of $\left(\mathring{\xi}, \mathring{\sigma}\right)$), this would proof the proposition. We get

$$\begin{aligned} h_q'(\sigma) &= \frac{d}{d\sigma}\Lambda_1(g_q(\sigma), \sigma) = \frac{\partial}{\partial\xi}\Lambda_1(g_q(\sigma), \sigma)\, g_q'(\sigma) + \frac{\partial}{\partial\sigma}\Lambda_1(g_q(\sigma), \sigma) \\ &= -\Lambda_{1;\xi}(g_q(\sigma), \sigma)\frac{\Lambda_{2;\sigma}(g_q(\sigma), \sigma)}{\Lambda_{2;\xi}(g_q(\sigma), \sigma)} + \Lambda_{1;\sigma}(g_q(\sigma), \sigma) = \left(-\frac{1}{\Lambda_{2;\xi}}\{\Lambda_{1;\xi}\Lambda_{2;\sigma} - \Lambda_{2;\xi}\Lambda_{1;\sigma}\}\right)\Bigg|_{\xi = g_q(\sigma)} \end{aligned}$$

The last term must be positive: the factor in braces is the determinant of the Jacobian studied in the preceding proof, which is negative for any $(\xi, \sigma) \in \Delta$, while the denominator $\Lambda_{2;\xi}$ is positive. Thus, $h_q(\sigma)$ is a strictly increasing function, thus is injective and so are Λ and $\Phi^{(**)}$.

From the preceding proposition we know that $\Phi^{(**)}$ is locally C1 invertible, so it is a C1 diffeomorphism between Δ and $\Phi^{(**)}(\Delta)$. Thus, its inverse is a C1 function mapping the given premium/RRoL input (provided it is contained in $\Phi^{(**)}(\Delta)$) to the GP parameters solving the three-layer problem. □

B.3 Final step

Lemma B.12. *For any given, strongly ordered, layer triple and the continuous extensions of its layer ratio functions to the non-negative ξ-axis, the following holds: If a component of $\Phi^{(**)}(\xi, 0)$ does not equal 0 for all admissible $\xi \in [0, \xi_{\sup})$, it equals initially (for small ξ) 0, then increases strictly. For $\xi \nearrow \xi_{\sup}$, $\Phi_2^{(**)}(\xi, 0)$ tends to the same supremum as $\Phi_2^{(**)}(\xi, \sigma)$ does (for fixed σ).*

Proof. For each $\tilde{E}_i$ or $\tilde{R}_i$, we call the set of numbers $\xi \in [0, \xi_{\sup})$ where it takes positive finite values: the *inner domain*. From Definition 3.4 we know that if the latter is not empty, it is an open interval, namely either $(0, \xi_{\sup})$ or $(1, \xi_{\sup})$. Thus, the same holds for the ratios $\tilde{E}_{i+1}/\tilde{E}_i$, $\tilde{R}_{i+1}/\tilde{R}_i$, $\tilde{E}_3/\tilde{R}_2$ constituting the components of $\Phi^{(**)}(\xi, 0)$ as specified in Lemma B.8. As $\tilde{\psi} := \ln\left(\tilde{F}\right) = -\frac{1}{\xi}\ln(x - a_1)$ has the partial derivative $\tilde{\psi}_\xi = \frac{1}{\xi^2}\ln(x - a_1)$, which is a strictly increasing function in x, we can work with logarithmic derivatives as in the proofs of Propositions B.3 and B.10. For a proper i-th layer we get for $\tilde{E}_i$ on its inner domain

$$\frac{d}{d\xi}\left(\ln\left(\tilde{E}_i\right)\right) = {}^{\tilde{F}}\overline{\tilde{\psi}_\xi}^{[a_i, b_i]}$$

The corresponding $\tilde{R}_i$ has the same inner domain and logarithmic derivative. Further, $\tilde{R}_i$ embraces degenerate layers, where the same formula holds. Overall the logarithmic derivatives, wherever they exist, are $\tilde{F}$-weighted averages of $\tilde{\psi}_\xi$ over the respective layer intervals, such that (using an abbreviated notation for these weighted averages)

$$\frac{d}{d\xi}\left(\ln\left(\frac{\tilde{E}_{i+1}}{\tilde{E}_i}\right)\right) = \overline{\tilde{\psi}_\xi}^{i+1} - \overline{\tilde{\psi}_\xi}^{i} > 0$$

and analogously for the other terms appearing as components of $\Phi^{(**)}(\xi, 0)$. Hence, the latter are strictly increasing functions of ξ on their respective inner domains, on the lower end of which they must have the infimum 0, in order to attach continuously to the area below the inner domain where the value 0 is taken on.

As for the upper limit of $\Phi_2^{(**)}(\xi,0)$ for $\xi \nearrow \xi_{\text{sup}}$, it is sufficient to look at $\Phi_2^{*(*)}(\xi,0)$, from which the result for $\Phi_2(\xi,0)$ immediately follows. For any interval $[a,b]$ such that $a_1 < a \leq b < \infty$, we have

$$\tilde{R}(a_1,a,b,\xi) = \overline{(x-a_1)^{-\frac{1}{\xi}}}^{[a,b]} \underset{\xi\nearrow\infty}{\longrightarrow} 1$$

while in the case $a_1 = a < b < \infty$ we have (for $\xi = 1$ via l'Hôspital's rule)

$$\tilde{R}(a_1,a_1,b,\xi) = \frac{(b-a_1)^{1-\frac{1}{\xi}}}{(b-a_1)\left(1-\frac{1}{\xi}\right)} = \frac{(b-a_1)^{-\frac{1}{\xi}}}{1-\frac{1}{\xi}} \underset{\xi\nearrow\infty}{\longrightarrow} 1$$

For an unlimited third layer we have with $a_3 > a_1$ that

$$\tilde{E}_3 = \tilde{E}(a_1,a_3,\infty,\xi) = \frac{(a_3-a_1)^{1-\frac{1}{\xi}}}{\frac{1}{\xi}-1} \underset{\xi\nearrow 1}{\longrightarrow} \infty$$

By combining these formulae we can treat all situations having nonempty inner domains and get

$$\lim_{\xi\nearrow\infty}\frac{\tilde{R}_3}{\tilde{R}_2} = 1, \qquad \lim_{\xi\nearrow 1}\frac{\tilde{E}_3}{\tilde{R}_2} = \infty$$

which are indeed the limits of $\Phi_2^{*(*)}(\xi,\sigma)$ for $\xi \nearrow \xi_{\text{sup}}$ as calculated in Proposition B.9. □

Lemma B.13. *For any given, strongly ordered and weakly overlapping, layer triple, each contour line of the second component of any layer ratio function, i.e., each nonempty set*

$$\left\{(\xi,\sigma)\in\Delta \,\middle|\, \Phi_2^{(**)}(\xi,\sigma) = t\right\}$$

can be written as $\{(g(\sigma),\sigma) \mid \sigma \in (0,\infty)\}$ *with a strictly decreasing C1 function g, which has the limits*

$$\lim_{\sigma\nearrow\infty} g(\sigma) = -\infty, \qquad (0,\infty) \ni \lim_{\sigma\searrow 0} g(\sigma) = \begin{cases} 1, & b_3 = \infty, a_2 = a_1 \\ \left(\Phi_2^{(**)}(\_,0)\right)^{-1}(t), & b_3 > \infty \text{ or } a_2 > a_1 \end{cases}$$

Proof. To define a contour line, t must lie in the respective image of $\Phi_2^{(**)}$, which according to Proposition B.10 is $(0, t_{\text{sup}})$, where t_{sup} equals $\frac{c_3}{c_2}$ for Φ_2, 1 for Φ_2^*, ∞ for Φ_2^{**}.

Let us fix such a t and set $q = \ln(t)$. Now we can proceed exactly as in the proof of Proposition B.11 and define, for any $\sigma > 0$, $g(\sigma) := g_q(\sigma)$ as the (existing and unique) ξ such that $(\xi,\sigma) \in \Delta$ and $\Phi_2^{(**)}(\xi,\sigma) = e^q = t$. The resulting function g is C1 and strictly decreasing, having in particular limits for $\sigma \searrow 0$ and $\sigma \nearrow \infty$.

We show first that $\lim_{\sigma\nearrow\infty} g(\sigma) = -\infty$. For an indirect proof, assume there is an $m < 0$ such that $\lim_{\sigma\nearrow\infty} g(\sigma) \geq m > -\infty$. Then we would for any large $\sigma > -m(a_3 - a_1)$ have $(m,\sigma) \in \Delta$ and

$$t = \Phi_2^{(**)}(g(\sigma),\sigma) \geq \Phi_2^{(**)}(m,\sigma)$$

where the last inequality follows from the monotonicity of $\Phi_2^{(**)}$ in its variables, see Proposition B.10. According to Proposition B.9, the limit of the RHS for $\sigma \nearrow \infty$ equals t_{sup}, which implies $t \geq t_{\text{sup}}$, which is a contradiction.

Let us now look at the limit to the left, which we call

$$\xi_q := \lim_{\sigma\searrow 0} g_q(\sigma)$$

As all $g(\sigma) \in (0,\xi_{\text{sup}})$, we have either $\xi_q = \xi_{\text{sup}}$ or $\xi_q \in [0,\xi_{\text{sup}})$, the interval on the ξ-axis which the layer ratio functions were continuously extended to in Lemma B.8. First consider the case $b_3 = \infty, a_2 = a_1$, where according to the lemma $\Phi_2^{(**)}(\xi,0) \equiv 0$. If we had $\xi_q \in [0,\xi_{\text{sup}}) = [0,1)$, the continuity of the extended layer ratio functions would imply $\Phi_2^{(**)}(\xi_q, 0) = \lim_{\sigma\searrow 0} \Phi_2^{(**)}(g(\sigma), \sigma) = t > 0$, which is a contradiction. So, we must have $\xi_q = 1$.

In the remaining cases, where $a_2 > a_1$ and/or $b_3 < \infty$, $\Phi_2^{(**)}(\_, 0)$ has an inner domain, on which it is invertible. Due to the preceding lemma, $\lim_{\xi\nearrow\xi_{\sup}} \Phi_2^{(**)}(\xi, 0) = t_{\sup} > t$. Here $\xi_q = \xi_{\sup}$ is impossible, otherwise the monotonicity properties of the layer ratio functions would imply

$$t = \lim_{\sigma\searrow 0} \Phi_2^{(**)}(g(\sigma), \sigma) \geq \lim_{\sigma\searrow 0} \Phi_2^{(**)}(g(\sigma), 0) = \lim_{\xi\nearrow\xi_{\sup}} \Phi_2^{(**)}(\xi, 0) = t_{\sup}$$

which is a contradiction. Hence, $\xi_q \in [0, \xi_{\sup})$ and due to the continuity of $\Phi_2^{(**)}$ at $(\xi_q, 0)$, ξ_q must be the (existing and unique) $\xi > 0$ such that $\Phi_2^{(**)}(\xi, 0) = t > 0$. □

Proposition B.14. *Suppose you have got a strongly ordered layer triple $[a_i, b_i]$ where top and middle layer do not overlap. If the bottom layer is a threshold and/or the top layer is unlimited, the applicable layer ratio functions (in particular Φ^{**}) are surjective in the sense that they have the images*

$$\Phi(\Delta) = \left(0, \frac{c_2}{c_1}\right) \times \left(0, \frac{c_3}{c_2}\right), \qquad \Phi^*(\Delta) = (0,1)^2, \qquad \Phi^{**}(\Delta) = (0,1) \times (0, \infty)$$

which are the maximum possible intervals that the risk premiums/RRoL's of three ordered layers can attain, apart from cases where a layer has loss probability zero or the same RRoL as another layer. Here the applicable layer ratio functions are C1 diffeomorphisms on the domain

$$\Delta = \{(\xi, \sigma) \in (-\infty, \xi_{\sup}) \times (0, \infty) \mid \sigma > -\xi(a_3 - a_1)\}, \qquad \xi_{\sup} = \begin{cases} \infty, & b_3 < \infty \\ 1 & b_3 = \infty \end{cases}$$

which is likewise the maximum possible parameter space where the GP model starting at a_1 assigns positive loss probabilities and finite expectations to all three given layers.

In the remaining case having a proper first *layer and a* limited third *layer, Φ (if applicable) and Φ^* are C1 diffeomorphisms on Δ too, but have somewhat smaller images, being restricted by the following* technical condition*:*

$$\Phi^*(\Delta) = \left\{(z_1, z_2) \in (0,1)^2 \mid z_1 > \varrho_1\left(\varrho_2^{-1}(z_2)\right)\right\}, \qquad \varrho_i = \frac{\tilde{R}_{i+1}}{\tilde{R}_i}$$
$$\Phi(\Delta) = \left\{(y_1, y_2) \in \left(0, \frac{c_2}{c_1}\right) \times \left(0, \frac{c_3}{c_2}\right) \middle| \, y_1 > \frac{c_2}{c_1}\varrho_1\left(\varrho_2^{-1}\left(\frac{c_2}{c_3}y_2\right)\right)\right\}$$

Briefly, the GPD solves the three-layer problem, yielding a unique solution, for strongly ordered layer triples with a non-overlapping top layer, under the following constraints on the given premiums/RRoL's: $r_1 > r_2 > r_3 \geq 0$ (always), further $r_3 > 0$ and/or $e_3 > 0$, plus the above technical condition where applicable.

Proof. The comments on maximum possible ranges are obvious. For three ordered layers the inequality $r_1 \geq r_2 \geq r_3 \geq 0$ is a mathematical necessity and any equivalence implies that a layer has RRoL zero, which for limited layers means loss probability zero, or that two layers have the same RRoL, which for limited layers means that all losses affecting the lower one are total losses for both layers (a case being possible in theory, but very implausible in practice).

Now, for a given layering, choose any input of the three-layer problem with strict inequalities, i.e., $r_1 > r_2 > r_3 > 0$. For infinite b_3 choose $e_3 > 0$ instead of r_3, which here must equal 0. Let us write $u_3 > 0$ for the relevant data input about the third layer, i.e., r_3 or e_3. We want to find $(\xi, \sigma) \in \Delta$ such that $\Phi^{*(*)}(\xi, \sigma) = \left(\frac{r_2}{r_1}, \frac{u_3}{r_2}\right)$, from which for proper layers $\Phi(\xi, \sigma) = \left(\frac{e_2}{e_1}, \frac{e_3}{e_2}\right)$ would immediately follow. Note that the layer triple is weakly overlapping, such that we can apply all preceding results.

Let Λ be the component-wise logarithm of the function $\Phi^{*(*)}$ and (p, q) that of $\left(\frac{r_2}{r_1}, \frac{u_3}{r_2}\right)$. The function $g(\sigma) = g_q(\sigma)$ giving the contour line for the "level" $t = \frac{u_3}{r_2}$ is, due to proof of the preceding lemma, for all $\sigma > 0$ well defined and fulfills $(g(\sigma), \sigma) \in \Delta$ and $\Lambda_2(g(\sigma), \sigma) = q$ or equivalently $\Phi_2^{*(*)}(g(\sigma), \sigma) = \frac{u_3}{r_2}$. This means that, for any $\sigma > 0$, $(g(\sigma), \sigma)$ solves half of the three-layer problem, by matching $\frac{u_3}{r_2}$.

Now we consider, as in Proposition B.11, $h(\sigma) := h_q(\sigma) = \Lambda_1(g_q(\sigma), \sigma)$, which is also well defined for all $\sigma > 0$ and has a continuous and positive derivative, which carries over to its exponential $\Phi_1^{*(*)}(g(\sigma), \sigma)$.

If we find a σ such that $h(\sigma) = p$ or equivalently $\Phi_1^{*(*)}(g(\sigma), \sigma) = \frac{r_2}{r_1}$, we have $\Lambda(g(\sigma), \sigma) = (p, q)$ or equivalently $\Phi^{*(*)}(g(\sigma), \sigma) = \left(\frac{r_2}{r_1}, \frac{u_3}{r_2}\right)$, thus have solved the three-layer problem. As $h(\sigma)$ and its exponential $\Phi_1^{*(*)}(g(\sigma), \sigma)$ are (strictly) increasing functions, their images (ranges of values taken) are determined by their limits for $\sigma \searrow 0$ and $\sigma \nearrow \infty$. Let us calculate these limits for the latter function.

Recall from the preceding lemma that $\lim_{\sigma \nearrow \infty} g(\sigma) = -\infty$. We show that for $\sigma \nearrow \infty$

$$\Phi_1^{*(*)}(g(\sigma), \sigma) = \frac{R_2}{R_1}(g(\sigma), \sigma) = \frac{\overline{\left(\left(1 + g(\sigma)\frac{x-a_1}{\sigma}\right)^+\right)^{-\frac{1}{g(\sigma)}}}^{[a_2,b_2]}}{\overline{\left(\left(1 + g(\sigma)\frac{x-a_1}{\sigma}\right)^+\right)^{-\frac{1}{g(\sigma)}}}^{[a_1,b_1]}}$$

tends to 1, or more strongly that this holds both for numerator and denominator of the last expression. We have $(g(\sigma), \sigma) \in \Delta$, which means for large σ, where ultimately $g(\sigma) < 0$, that $\frac{\sigma}{-g(\sigma)} > a_3 - a_1$. This implies for $x \geq a_1$

$$0 \leq \frac{-g(\sigma)}{\sigma}(x - a_1) \leq \frac{x - a_1}{a_3 - a_1}, \qquad 1 \geq \left(1 + g(\sigma)\frac{x - a_1}{\sigma}\right)^+ \geq \left(1 - \frac{x - a_1}{a_3 - a_1}\right)^+ = \left(\frac{a_3 - x}{a_3 - a_1}\right)^+$$

Now consider a layer $[a, b]$ located somewhere between a_1 and a_3, e.g. the first or the second layer. (Here we use, for the first and only time, that the third layer does not overlap with the lower ones.) If $a_1 \leq a \leq b < a_3$, we have for large σ

$$1 \geq \overline{\left(\left(1 + g(\sigma)\frac{x - a_1}{\sigma}\right)^+\right)^{-\frac{1}{g(\sigma)}}}^{[a,b]} \geq \left(\left(1 + g(\sigma)\frac{b - a_1}{\sigma}\right)^+\right)^{-\frac{1}{g(\sigma)}} \geq \left(\frac{a_3 - b}{a_3 - a_1}\right)^{-\frac{1}{g(\sigma)}} \longrightarrow 1$$

if $g(\sigma) \to \infty$, because $\frac{a_3 - b}{a_3 - a_1} > 0$. If $a_1 \leq a < b \leq a_3$, we get analogously

$$\begin{aligned} 1 \geq \overline{\left(\left(1 + g(\sigma)\frac{x - a_1}{\sigma}\right)^+\right)^{-\frac{1}{g(\sigma)}}}^{[a,b]} &\geq \overline{\left(\frac{a_3 - x}{a_3 - a_1}\right)^{-\frac{1}{g(\sigma)}}}^{[a,b]} \\ &= \frac{1}{b - a}\frac{(a_3 - a_1)^{\frac{1}{g(\sigma)}}}{1 - \frac{1}{g(\sigma)}}\left[(a_3 - a)^{1 - \frac{1}{g(\sigma)}} - (a_3 - b)^{1 - \frac{1}{g(\sigma)}}\right] \longrightarrow 1 \end{aligned}$$

These two cases cover all possible locations of the first layer, such that we have $\lim_{\sigma \nearrow \infty} R_1(g(\sigma), \sigma) = 1$. The same limit holds for R_2, but here it remains to prove the situation $a_2 = b_2 = a_3 < b_3$, which requires some extra reasoning about $g(\sigma)$. This function was constructed such that $\Phi_2^{*(*)}(g(\sigma), \sigma) = \frac{u_3}{r_2}$. For a proper third layer this can be equivalently written as $\frac{E_3}{R_2}(g(\sigma), \sigma) = \frac{e_3}{r_2}$, notably for both Φ_2^* and Φ_2^{**}. For large σ, where ultimately $g(\sigma) < 0$, we have

$$E_3(g(\sigma), \sigma) \leq E(a_1, a_3, \infty, g(\sigma), \sigma) = \frac{\sigma}{1 - g(\sigma)}\left(1 + g(\sigma)\frac{a_3 - a_1}{\sigma}\right)^{1 - \frac{1}{g(\sigma)}}$$

and get with

$$R_2(g(\sigma), \sigma) = \left(1 + \frac{g(\sigma)}{\sigma}(a_3 - a_1)\right)^{-\frac{1}{g(\sigma)}}$$

$$\frac{e_3}{r_2} = \frac{E_3}{R_2}(g(\sigma), \sigma) \leq \frac{\sigma}{1 - g(\sigma)}\left(1 + g(\sigma)\frac{a_3 - a_1}{\sigma}\right) < \frac{\sigma}{-g(\sigma)}\left(1 + g(\sigma)\frac{a_3 - a_1}{\sigma}\right) = \frac{\sigma}{-g(\sigma)} - (a_3 - a_1)$$

or equivalently $\frac{\sigma}{-g(\sigma)} > \frac{e_3}{r_2} + a_3 - a_1$. So, we have

$$\begin{aligned} 1 \geq R_2(g(\sigma), \sigma) &= \left(1 - \frac{-g(\sigma)}{\sigma}(a_3 - a_1)\right)^{-\frac{1}{g(\sigma)}} \\ &> \left(1 - \frac{a_3 - a_1}{e_3/r_2 + a_3 - a_1}\right)^{-\frac{1}{g(\sigma)}} = \left(\frac{e_3/r_2}{e_3/r_2 + a_3 - a_1}\right)^{-\frac{1}{g(\sigma)}} \longrightarrow 1 \end{aligned}$$

if $g(\sigma) \to \infty$, because $\frac{e_3/r_2}{e_3/r_2+a_3-a_1} > 0$. Overall we have shown that the upper limit of $\frac{R_2}{R_1}$ on the contour line equals 1, which is the supremum of the possible given RRoL ratios $\frac{r_2}{r_1} \in (0,1)$. The first half of the surjectivity of $\Phi^{(**)}$ is proved.

To finalize, let us now look at the opposite endpoint. We have calculated the limit $\xi_q = \lim_{\sigma \searrow 0} g(\sigma)$ in the preceding lemma and found it to be always finite. Thus, we can use the continuous extension of $\Phi_1^{*(*)}$ to the ξ-axis bordering Δ^* as specified in Lemma B.8, and get

$$\lim_{\sigma \searrow 0} \Phi_1^{*(*)}(g(\sigma), \sigma) = \Phi_1^{*(*)}(\xi_q, 0) = \frac{\tilde{R}_2}{\tilde{R}_1}(\xi_q)$$

This notably embraces the case $b_3 = \infty, a_2 = a_1$, where $\xi_q = 1$. Here Φ^{**} is restricted to $\xi < \xi_{\sup} = 1$, but Φ_1^{**} is well defined for $\xi \geq 1$ and continuously extendable to the whole non-negative ξ-axis.

Let us evaluate $\Phi_1^{*(*)}(\xi_q, 0)$, recalling from Lemma B.8 in particular the values ξ where $\Phi_1^{*(*)}(\xi, 0)$ equals 0. It turns out that $\Phi_1^{*(*)}(\xi_q, 0) = 0$, or equivalently the surjectivity of $\Phi^{(**)}$, holds in a number of cases, but not always, namely:

If the first layer is a threshold ($a_1 = b_1$), $\Phi_1^{*(*)}(\xi_q, 0) = 0$ irrespective of ξ_q, such that $\Phi^{*(*)}$ is surjective.

If the third layer is unlimited, we have $\xi_q \leq 1$ and thus $\Phi_1^{**}(\xi_q, 0) = 0$. So, Φ^{**} is surjective.

In the remaining case of a proper bottom layer and a limited top layer, where $a_1 < b_1, b_3 < \infty$ and thus $a_1 < a_2$ and $u_3 = r_3$, we have with $\varrho_i = \tilde{R}_{i+1}/\tilde{R}_i$, due to the preceding lemma:

$$\Phi_1^*(\xi_q, 0) = \varrho_1(\xi_q) = \frac{\tilde{R}_2}{\tilde{R}_1}\left(\left(\frac{\tilde{R}_3}{\tilde{R}_2}\right)^{-1}\left(\frac{r_3}{r_2}\right)\right) = \varrho_1\left(\varrho_2^{-1}\left(\frac{r_3}{r_2}\right)\right)$$

To evaluate this, recall from Lemmas B.8 and B.12 that in the present case we have the following situation: The function $\varrho_2(\xi)$ is strictly increasing on $[0, \infty)$, having the image $[0, 1)$. $\xi_0 = \varrho_2^{-1}\left(\frac{r_3}{r_2}\right)$ is well defined and positive. The function $\varrho_1(\xi)$ equals 0 for $\xi \in [0, 1]$ and increases strictly on $[1, \infty)$. So, $\Phi_1^*(\xi_q, 0)$ may equal 0, namely iff $\varrho_2^{-1}\left(\frac{r_3}{r_2}\right) \leq 1$ or equivalently

$$\frac{r_3}{r_2} \leq \varrho_2(1) = \frac{\tilde{R}_3(1)}{\tilde{R}_2(1)}, \qquad \tilde{R}_i(1) = \overline{(x-a_1)^{-1}}^{[a_i, b_i]} = \begin{cases} \frac{1}{b_i - a_i} \ln \frac{b_i - a_1}{a_i - a_1} & a_i < b_i \\ \frac{1}{a_i - a_1} & a_i = b_i \end{cases}$$

This is a sufficient and easily verifiable condition for the existence of a solution of the three-layer problem. So, if $\frac{r_3}{r_2} \leq \varrho_2(1) < 1$, which qualitatively means that $\frac{r_3}{r_2}$ is not too large, the three-layer problem can be solved irrespective of $\frac{r_2}{r_1}$. Things are more complex for rather large $\frac{r_3}{r_2}$, namely if $\frac{r_3}{r_2} > \varrho_2(1)$. Then $\Phi_1^*(\xi_q, 0) > 0$, such that the three-layer problem has a solution only when $\frac{r_2}{r_1} > \Phi_1^*(\xi_q, 0)$, which qualitatively means that $\frac{r_2}{r_1}$ must be rather large too. Overall we can infer (very qualitatively) that the three-layer problem is hard to solve if $\frac{r_3}{r_2}$ is large and $\frac{r_2}{r_1}$ is small, or equivalently if r_2, which is located in the interval (r_3, r_1), is much closer to r_3 than to r_1.

Summing up, in the case $a_1 < b_1, b_3 < \infty$ the applicable layer ratio functions, i.e., Φ^* and for proper layers also Φ, are not surjective; here for a solution of the three-layer problem the given r_i must meet a *technical condition*, which reads

$$\frac{r_2}{r_1} > \Phi_1^*(\xi_q, 0) = \varrho_1\left(\varrho_2^{-1}\left(\frac{r_3}{r_2}\right)\right)$$

This yields the image of Φ^*, namely $\Phi^*(\Delta) = \left\{(z_1, z_2) \in (0,1)^2 \mid z_1 > \varrho_1\left(\varrho_2^{-1}(z_2)\right)\right\}$, a proper subset of $(0,1)^2$. If we convert, for proper layers, RRoL's into risk premiums, we get analogously $\Phi(\Delta)$ as asserted above. □

Remark B.15. There is an intuitive interpretation for the function ϱ_2^{-1} in the technical condition. Recall Remark 3.5 stating that $\tilde{R}_i$ reflects the RRoL of the shifted layer $[a_i - a_1, b_i - a_1]$ for (single-parameter) Pareto distributed losses with parameter $\alpha = \frac{1}{\xi} > 0$, plus a formal extension thereof for layers having attachment point 0. In the case where the technical condition applies, we have $a_2 > a_1$, such that for the shifted second and third layer the Pareto model is properly applicable. So, if $\frac{r_3}{r_2}$ is the given RRoL ratio of third and second layer, then $\varrho_2^{-1}\left(\frac{r_3}{r_2}\right)$ is the (unique) parameter ξ for which the Pareto model with parameter $\alpha = \frac{1}{\xi}$ reproduces this ratio for the shifted third and second layer. In reinsurance terminology: $\varrho_2^{-1}\left(\frac{r_3}{r_2}\right)$ is the Pareto ξ (the inverse of the mostly used Pareto α) *between the layers* (Riegel, 2008).

References

Aviv, R. (2018). An extreme-value theory approximation scheme in reinsurance and insurance-linked securities. *ASTIN Bulletin*, 48(3):1157–1173.

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

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events for insurance and finance*. Berlin Heidelberg: Springer.

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

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

FINMA (2006). *Technical document on the Swiss Solvency Test*. FINMA (Federal Office of Private Insurance).

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

Neth, H. and Gigerenzer, G. (2015). Heuristics: Tools for an uncertain world. In *R. Scott and S. Kosskyn (Eds.), Emerging trends in the social and behavioral sciences: An Interdisciplinary, Searchable, and Linkable Resource*, pages 1–18. Wiley Online Library.

Riegel, U. (2008). Generalizations of common ILF models. *Blätter der DGVFM*, 29(1):45–71.

Riegel, U. (2010). On fire exposure rating and the impact of the risk profile type. *ASTIN Bulletin*, 40(2):727–777.

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

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

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60.

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

Schwepcke, A., editor (2004). *Reinsurance: Principles and State of the Art – A Guidebook for Home Learners*. Verlag Versicherungswirtschaft.

Scollnik, D. P. M. (2007). On composite lognormal-Pareto models. *Scandinavian Actuarial Journal*, 2007(1):20–33.