Three-layer problems and the GPD (working paper)

Michael Fackler
Independent actuary
Munich, Germany

2019-04-02

Abstract

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

We use the Generalized Pareto distribution to build consistent underlying models from 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 constellations 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 data points.

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, Solvency

1 Introduction

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 the output you need from the analytic 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. Important example for this situation: large losses that have to be modeled for solvency issues or the pricing of reinsurance 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: One could 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) quotation of a 2 xs 1 layer, but are insufficient to assess higher layers. However, for the layer 5 xs 5 there is e.g. a market benchmark BC rate (common for Motor Liability in several countries) or a reliable expected loss estimate from a geophysical simulation model for natural disasters (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") Payback approach, i.e. you assign a return period of say 200 years to the 20 (million USD) loss.

This is the challenge that we are going to explore in this work: how to assign a consistent and plausible model to (about) 3 bits of information, namely risk premiums of layers or 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, see e.g. [Embrechts et al., 2013]) and by widespread practical experience.

1.1 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 very 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 GPD 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 (2 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.

1.2 Outline

In Section 2 we introduce compact layer notation, give our main result (the three-layer problem), and explore a number of variants and applications, including proposals about what to do if one has a bit more or less than 3 data inputs. In Section 3 we show some non-parametric inequalities helping explore 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 Three-layer problems

2.1 Layers

We borrow notation and basic results for layers from [Riegel, 2018], generalizing a bit.

Definition 2.1. A (re)insurance layer $c \,\mathrm{xs}\, a$ (*c in excess of a*) pays the part of each loss X that exceeds the *attachment point* $a \geq 0$, up to a maximum (*cover*) of $c > 0$, which mathematically means paying $\min\left((X-a)^+, c\right)$.

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

If e is the risk premium of the layer, we call $r \coloneqq e/c$ 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 equality is only possible in the (rather unrealistic) case that all losses hitting the layer ($X > a$) are total layer losses reaching even beyond the layer ($X > b$). The RRoL can be seen as a kind of average loss frequency across the interval defining the layer.

We include two extremes. In case $\mathrm{E}(X) < \infty$, 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 infinitesimally 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 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 losses hit this layer with probability > 0, e or r is positive (if not both). The respective information provided by them is either the risk premium (for 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 ($a_2 < b_1$).

A *tower of layers* is a sequence of ordered non-overlapping layers $[a_i, b_i]$ attaching exactly at each other, i.e. $a_{i+1} = b_i$.

In the situation of two ordered layers we have $f_1 \geq f_2$, $g_1 \geq g_2$, and notably $r_1 \geq r_2$.

In the tower case we have more strongly a decreasing sequence of altering frequencies and RRoLs:

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

2.2 Generalized Pareto

Definition 2.2. The GPD is a model for severity distribution tails, *starting at* a known threshold $s \geq 0$. Notice that $s = 0$ is possible and yields a ground-up model as we need it when the lowest layer starts at 0. The *survival function* (1 minus the cdf) in its common form, see e.g. [Embrechts et al., 2013], reads

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

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 any tail distribution $\overline{F}_X\left(x \mid X > t\right)$ starting from a higher tail $t > s$ is again GPD with the same ξ, but the second parameter changes. There are alternative GPD parameterizations which are tail-shifting-invariant in both parameters and thus more practical and intuitive for various purposes, see Section 6.5 in [Fackler, 2017] and Sections 2-3 in [Fackler, 2013]. However, for our scope the standard parameterization works well. For completeness we mention that for $s > 0$ the GPD embraces the *single-parameter Pareto* tail model as special case: for $\xi = \frac{\sigma}{s} > 0$ we get with $\alpha = 1/\xi$

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

If for a layer $[a, b]$ we have $a \geq s$, such that the whole layer is located in the area of the GPD tail, and if the expected loss frequency at s equals λ, then we can express the risk premium of the layer with the tail severity (see [Riegel, 2018]) as follows:

$$\lambda \int_a^b \overline{F}_X\left(x \mid X > s\right) dx = \lambda \int_a^b \left(\left(1 + \xi \frac{x-s}{\sigma}\right)^+\right)^{-\frac{1}{\xi}} dx =: \lambda\, E\left(s, a, b, \xi, \sigma\right)$$

Here we have introduced a compact notation for the integral term. If we relate the risk premiums of two layers to each other, λ drops out, the ratio depends only on integrals over the survival function.

2.3 Main result

Definition 2.3. The *general three-layer problem* is as follows: If we have three layers $[a_i, b_i]$ with respective risk premiums e_i and RRoLs r_i, is there a severity distribution consistent with this information?

In practice one will mostly be interested in 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 will be about layers that are disjoint, having in common only endpoints, if at all. This case can be solved with the GPD.

Theorem 2.4. *Suppose you have given three (different) limited layers $[a_i, b_i]$ with their respective risk premiums e_i and RRoLs r_i, which are ordered and don't overlap, i.e.*

$$0 \leq a_1 \leq b_1 \leq a_2 \leq b_2 \leq a_3 \leq b_3 < \infty$$

If $r_1 > r_2 > r_3 > 0$, there is a unique GPD tail model starting at a_1 that together with a (unique) loss frequency at a_1 yields the given risk premiums and RRoLs. This result extends to the case of an unlimited top layer ($b_3 = \infty$), where we have $r_3 = 0$ and require in addition $e_3 > 0$ (to ensure that in any case all layers (thresholds) are hit by losses with probability > 0).

Proof. Very short sketch of proof: First restrict to proper layers ($a_i < b_i$), the cases where some b_i equal a_i follow (with some effort) as limiting cases. For proper layers the conditions on the RRoLs are equivalent to $0 < e_2/e_1 < c_2/c_1$, $0 < e_3/e_2 < c_3/c_2$. 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 $\sigma > -\xi\left(a_3 - s\right)$. We now work with $s = a_1$.

Some (lengthy but elementary) analysis shows that for fixed $a_{i,}, b_i$ the mapping

$$\left\{(\xi, \sigma) \in \mathbb{R}^2 \mid \sigma > \left(-\xi\left(a_3 - a_1\right)\right)^+\right\} \ni (\xi, \sigma) \mapsto \left(\frac{E\left(a_1, a_2, b_2, \xi, \sigma\right)}{E\left(a_1, a_1, b_1, \xi, \sigma\right)}, \frac{E\left(a_1, a_3, b_3, \xi, \sigma\right)}{E\left(a_1, a_2, b_2, \xi, \sigma\right)}\right) \in \left(0, \frac{c_2}{c_1}\right) \times \left(0, \frac{c_3}{c_2}\right)$$

is well-defined, bijective, and continuous in both directions; for $b_3 = \infty$ this holds with the additional constraint $\xi < 1$. Thus, we have unique parameters ξ, σ such that the corresponding GPD tail model starting at a_1 matches the given ratios e_2/e_1, e_3/e_2. Setting finally $\lambda = e_1/E(a_1, a_1, b_1, \xi, \sigma)$ we are done. □

As for uniqueness, note that GPD tail models starting at some lower threshold $t < a_1$ cannot yield further solutions – their upper tail starting at a_1 would again be the GPD tail model found in the theorem.

Let us illustrate some special cases, where some layers are thresholds, i.e. shrunk to a point.

- If the top layer is a threshold, we get 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 $0 \le a_1 < a_2 < a_3$ are matched by a unique GPD.
- If bottom and top layer are thresholds that equal attachment and detachment point, respectively, of the middle layer, we get a very particular tower of layers: $[a, a] < [a, b] < [b, b]$. 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 GPD model can be found in [Aviv, 2018], together with an example (frequencies and RRoLs 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 with the given input $f_i > r_i > g_i$. This yields a piecewise GPD (n pieces).

Notice that the theorem cannot by extended to all straightforward limiting cases. In particular, it is not difficult to see that the case of a loss-free top layer/threshold ($e_3 = 0$, $r_3 = 0$) has GPD solutions, but they are not unique. Solutions must have a supremum loss that is finite and not greater than a_3, i.e. $\xi < 0$ and $a_1 + \frac{\sigma}{-\xi} \le a_3$. To get uniqueness this supremum loss has to be specified.

2.4 Applications

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: layers with risk premiums, thresholds with frequencies, or a mixture of both.

We 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 the single-parameter Pareto distribution would be the preferred choice. In many lines of business benchmarks for the parameter α are known.

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 yielding, see e.g. [Riegel, 2018],

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

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

Now we sketch procedures how to construct a severity according to the number of given data inputs. Finding subsequently the correct frequency is as simple as in the 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: Use the GPD, solve the three-layer problem numerically. Instead of proceeding as in the proof of the theorem, one can try to find λ, ξ, σ in one step, by solving the system of three equations $e_i = \lambda E(a_1, a_i, b_i, \xi, \sigma)$. The uniqueness of the solution and practical experience suggest that this is not a hard numerical problem, choice of algorithm and starting values are apparently not critical. As for the latter, for $a_1 > 0$ one could start with say $\lambda = r_1$ and GPD parameters $\xi = 1/\alpha$, $\sigma = \xi a_1$ yielding a Pareto, where α is calculated from thresholds a_1, a_2 and frequencies r_1, r_2.

≥4, exact: Use GPDs. Solve the three-layer problem for the three highest layers. (GPD is most plausible for high tails.) This 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; use instead the layer-endpoints approach for each layer.

≥4, approximate: Use the GPD and solve the system of equations $e_i = \lambda E(a_1, a_i, b_i, \xi, \sigma)$ approximately, according to some deviation metric. In real-world constellations one often gets surprisingly close.

Overall we have built severity models for a variety of situations where the data are too scarce (or otherwise not suitable) for classical parameter inference.

3 A note on model risk

Of course, 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 (approximately) know the risk premium (first moment) and a bit more, there is not much further uncertainty about the second moment, such that different parametric severity models must lead to similar results.

To this end, we explore the collective model of risk theory, using 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 further, just assume finite first and second moment. Then Wald's equations for the aggregate layer loss $S = \sum_{k=1}^{N} X_k$ yield

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

Let us now use an alternative parameter that is introduced and explored in Chapter 5 of [Fackler, 2017].

Definition 3.1. The *contagion* of a counting random variable N having finite first and second moment is

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

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 $\kappa \coloneqq \mathrm{Ct}(N)$ 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, (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 rewriting the formula again, now using the RRoL $r = e/c$ and the quantity

$$\tau_X \coloneqq \frac{\mathrm{E}\left(X^2\right)}{c\,\mathrm{E}(X)}$$

that somehow reflects the geometry of the cdf of X. We get

$$\mathrm{CV}^2(S) = \kappa + \frac{\tau_X}{r}$$

Can we "estimate" τ_X without knowing much about the distribution of X? Only $\tau_X \leq 1$ is obvious (equality means that all layer losses are total losses), however, we will see that in some constellations it is possible to narrow down τ_X surprisingly well, in particular for heavy tails.

Proposition 3.2. *Suppose we have a (proper) limited layer $[a, b]$ being affected by layer losses represented by X, which leads to figures $f \geq r \geq g \geq 0$, where $r > 0$ is the RRoL, f is the frequency of layer losses and g is the frequency of total layer losses. Then we have*

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

If the cdf of X is concave (as is often assumed for distribution tail areas), 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 second interval is contained in the first one being smaller by a third at least. Both inequalities are sharp, i.e. cannot be improved. The first one also holds if f is replaced by an upper bound and/or g by a lower bound.

Proof. The first formula is essentially Formula (2.1) in [Aviv, 2018], slightly amended and generalized – and profoundly rearranged in order to be more easily interpretable.

Sketch of proof: Both formulae can be proved by a kind of variational approach. Recall that X represents the *excess* loss to the layer (not the ground-up loss), thus takes on values between 0 and $c = b - a$. Its survival function $\overline{F}(x)$ decreases on $[0, c]$ from 1 to $\frac{g}{f}$ (then jumps down to 0) and the integral over $[0, c]$ yields

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

i.e., the average of $\overline{F}$ over $[0, c]$ equals $\frac{r}{f}$. Now, with values at interval endpoints and average fixed, how can the decreasing function $\overline{F}$ look like? More to the point, how must it look like in order to get a value as high (low) as possible for the integral

$$\int_0^c 2x\overline{F}(x)\,dx = \int_0^\infty 2x\overline{F}(x)\,dx = \mathrm{E}\left(X^2\right)$$

giving the second moment? This variational problem yields the first formula, while the second one comes from the analogous problem having the convexity of $\overline{F}$ (equivalent to the concavity of the cdf) on $[0, c]$ as additional constraint.

The rest of the proof is simple analysis, starting with the heuristical idea that the integral of $x\overline{F}(x)$ over $[0, c]$ is highest (lowest) if $\overline{F}$ takes on values as low (high) as possible at the left end of the interval and as high (low) as possible at the right end, while meeting the boundary conditions. □

Whether the above inequalities 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, 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 ($r - g \ll g$) 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, the partial losses have a (possibly much) higher probability to be small than to be large.

If one wants to (approximately) calculate τ_X, one needs r and the frequencies f, g. For layers in the middle of a tower this works even when the frequencies are not available, provided you have the RRoLs of the adjacent layers: According to the proposition, in the first inequality f can be replaced by an upper bound, so one can use the RRoL of the next layer below. Analogously, instead of g one can use the RRoL of the next layer above.

We have only looked at the second moment here, but it is plausible that higher moments can be narrowed down similarly.

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 can choose to work with the GPD simply for *practical* reasons: It solves the three-layer problem and is easy and intuitive to work with in many ways, see Section 3 of [Fackler, 2013] for details. Why look for alternative (possibly less handy) parametric models when they will anyway yield similar statistical output?

4 Wrap up

We can conclude that the building of models from very scarce data input, via the GPD (or in extreme cases the Pareto) as described in this work, is not just powerful in the sense of getting things done (which is what is expected from practitioners in the first place), but makes also sense in terms of statistical modeling.

References

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

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

[Embrechts et al., 2013] Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events for insurance and finance*. Springer.

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

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

[Neth and Gigerenzer, 2015] 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, 2008] Riegel, U. (2008). Generalizations of common ILF models. *Blätter der DGVFM*, 29(1):45–71.

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