Reinventing Pareto: Fits for both small and large losses

Michael Fackler
Independent Actuary
Munich, Germany
E-mail: michael_fackler@web.de

Abstract

Fitting loss distributions in insurance is sometimes a dilemma: either you get a good fit for the small / medium losses or for the very large losses. To be able to get both at the same time, this paper studies generalizations and extensions of the Pareto distribution. This leads not only to a classification of potentially suitable functions, but also to new insights into tail behavior and exposure rating.

Keywords

Loss severity distribution; spliced model; Lognormal; Pareto; GPD; power curve; exposure rating; Riebesell

Reinventing Pareto: Fits for both small and large losses

Michael Fackler

1 Introduction

Loss severity distributions and aggregate loss distributions in insurance often have a shape being not so easy to model with the common distributions implemented in software packages. In the range of smaller losses and around the mean the observed densities often look somewhat like asymmetric bell curves, being positively skewed with one positive mode. This is not a problem in itself as well-known models like the Gamma or the Lognormal distribution have exactly this kind of geometry. Alternatively distributions like Exponential are available for cases where a strictly decreasing density seems more adequate. However, it often occurs that the traditional models, albeit incorporating the desired kind of skewness towards the right, have a less heavy tail than what the data indicates – if we restrict the fit on the very large losses the Pareto distribution or variants thereof frequently seem the best choice. But those typical heavy-tailed distributions rarely have the right shape to fit nicely below the tail area.

In practice bad fits in certain areas can sometimes be ignored. If we are mainly focused on the large majority of small and medium losses we can often accept a bad tail fit and work with e.g. the Lognormal distribution. It might have a too light tail, thus we will underestimate the expected value, however, often the large losses are such rare that their numerical impact is extremely low. Conversely, in case we are focused on extreme quantiles like the 200 year event, or want to rate a policy with a high deductible or a reinsurance layer, we only need an exact model for the large losses. In such situations we could work with a distribution that models smaller losses wrongly (or completely ignores them). There is a wide range of situations where the choice of the model can be made focusing just on the specific task to be accomplished, while inaccuracy in unimportant areas is wittingly accepted.
However, exactness over the whole range of loss sizes, from the many smaller to the very few large ones, is increasingly required. E.g. according to a modern holistic risk management / solvency perspective we do not only regard the average loss (often depending mainly on the smaller losses) but also want to derive the probability of a very bad scenario (depending heavily on the tail) – namely out of the same model. Further, it becomes more and more popular to study various levels of retentions for a policy, or for a portfolio to be reinsured. (A traditional variant of this methodology is what reinsurers call exposure rating.) For such analyses one needs a distribution model being very accurate both in the low loss area, which is where the retention applies, and in the tail area, whose impact on the expected loss becomes higher the higher the retention is chosen.

Such situations require actuaries to leave the distribution models they know best and proceed to somewhat more complex ones. There is no lack of such models, neither in the literature nor in software packages. E.g. Klugman et al. provide generalizations of the Gamma and the Beta distribution having up to 4 parameters. However, despite the availability of such models, actuaries tend to stick to their traditional distributions. This is not (just) due to nostalgia – it has to do with a common experience of actuarial work: lack of data. In an ever changing environment it is not easy to gather a sufficient number of reliable and representative data points to fit distributions with several parameters. A way to detect and possibly avoid big estimation errors is to check the estimated parameters with what is called market experience: analogous results calculated from other data sets covering similar business.
To be able to do this it would be ideal to work with distributions looking initially like one of the traditional distributions but having a tail shape similar to well-known tail models like Pareto. Such models exist indeed and are not too difficult to deal with, however, they are a bit scattered over various literature. The main scope of this paper is to assemble and generalize a number of them, which results a general framework of variants and extensions of the Pareto distribution family.

Section 2 explains why reinsurers love the Single-parameter Pareto distribution so much, collecting some results helping gather intuition about distribution tails in general. Section 3 presents a less common but extremely useful parametrisation of the Generalized Pareto Distribution, which will make reinsurers love this model as well. Section 4 explains how more Pareto variants can be created, catering in particular for a more flexible modeling of smaller losses. Section 5 provides an inventory of spliced Lognormal-Pareto models, embracing two such models introduced by different authors and explaining how they are related to each other. Section 6 provides an overview of analogous models employing various distributions, referring to real world applications. The last section revisits two old exposure rating methods in the light of the methodology developed so far.

A few technical remarks:

In most of the following we will not distinguish between loss severity and aggregate loss distributions. Technically the fitting task is the same, and the shapes being observed for the two types of distributions overlap. For aggregate losses, at least in case of large portfolios and not too many dependencies among the single risks, it is felt that distributions should mostly have a unique positive mode (maximum density) like the Normal distribution, however, considerable skewness and heavy tails cannot be ruled out at all. Single losses are felt to be more heavy-tailed; here a priori both a strictly decreasing density and a positive mode are plausible scenarios, let alone multimodal distributions requiring very complex modeling.

For any single loss or aggregate loss distribution let $\overline{F}(x) = 1 - F(x) = P(X > x)$ be the **survival function**, f(x) the probability density function (if existent), i.e. the derivative of the cumulative distribution function F(x). As this is often more intuitive (and possibly a bit more general) we will formulate as many results as possible in terms of **cdf** (cumulative distribution function) instead of **pdf** (probability density function), mainly working with the survival function as this often yields simpler formulae than the cdf.

Unless otherwise specified, the parameters in this paper are positive real numbers.

2 Pareto – Reinsurer's old love

One could call it the *standard model of reinsurance actuaries*:

Pareto = Pareto Type I = European Pareto = Single-parameter Pareto: $\overline{F}(x) = \left(\frac{\theta}{x}\right)^{\alpha}$, θ<x

Remark: In this paper (sticking to old Continental European style) we reserve the name "Pareto" for this specific model, bearing in mind that is occurs to be used for other variants of the large Pareto family as well.

Does the Pareto model have one or two parameters? It depends – namely on what the condition θ<x means. It may mean that no losses between 0 and θ exist, or alternatively that nothing shall be specified about potential losses between 0 and θ. Unfortunately this is not always clearly mentioned when the model is used. In precise wording we have:

Situation 1: There are no losses below the threshold θ:

$\overline{F}(x) = 1$ for 0<x<θ, for θ≤x $\overline{F}(x) = \left(\frac{\theta}{x}\right)^{\alpha}$.

This model has two parameters α und θ. Here θ is not just a parameter, it is indeed a *scale parameter* (as defined e.g. in Klugman et al.) of the model. We call it **Pareto-only**, referring to the fact that there is no area of small losses having a distribution shape other than Pareto. This model is quite popular despite its hardly realistic shape in the area of low losses, whatever θ be. (If θ is large there is an unrealistically large gap in the distribution. If θ is small, say θ = 1 Euro, the gap is negligible but a Pareto-like shape for losses in the range from 1 to some 10000 Euro is felt rather implausible.)

Situation 2: Only the tail is modeled, thus to be precise we are dealing with a conditional distribution: $\overline{F}(x|X > \theta) = \left(\frac{\theta}{x}\right)^{\alpha}$

This model has the only parameter α. θ is the *known* lower threshold of the model.

Situation 1 implies Situation 2, but not vice versa. We will later see distributions combining a Pareto tail with a totally different distribution of small losses.

A memoryless property

Why is the **Pareto** model so popular among reinsurers, then? The most useful property of the Pareto tail model is without doubt the **parameter invariance** when modeling upper tails: If we have $\overline{F}(x|X > \theta) = \left(\frac{\theta}{x}\right)^{\alpha}$ and derive the model for a higher threshold d≥θ we get:

$$\overline{F}(x|X > d) = \left(\frac{\theta}{x}\right)^{\alpha} \Big/ \left(\frac{\theta}{d}\right)^{\alpha} = \left(\frac{d}{x}\right)^{\alpha}$$

We could say, when going "upwards" to model larger losses the model "forgets" the original threshold θ, which is not needed any further – instead the new threshold comes in. That implies:

- If a function has a Pareto tail and we only need to work with quite large losses, we do not need to know exactly where that tail starts. As long as we are in the tail (let's call it **Pareto area**) we always have the same parameter α, whatever the threshold be.
- It is possible to compare data sets with different (reporting) thresholds. Say for a MTPL portfolio we know all losses greater than 2 million Euro, for another one we only have the losses exceeding 3 million available. Despite these tail models having different thresholds we can judge whether the underlying portfolios have similar tail behavior or not, according to whether they have similar Pareto alphas. Such comparisons of tails starting at different thresholds are extremely useful in the reinsurance practice, where typically to get an overview per line of business one assembles data from several reinsured portfolios, all possibly having different reporting thresholds.
- This comparability can lead to *market values* for Pareto alphas, being applicable as benchmarks. Say we observe that a certain type of Fire portfolio in a certain country frequently has Pareto tails starting somewhere between 2 and 3 million Euro, having an alpha typically in the range of 1.8.

We recall some useful basic facts about losses in the Pareto tail. These are well known, however, we will get to some less-known generalizations later.

Pareto extrapolation: To relate *frequencies* at different thresholds d_1, $d_2>\theta$ the Pareto model yields a famous very simple formula: $\dfrac{frequency-at-d_2}{frequency-at-d_1}=\left(\dfrac{d_1}{d_2}\right)^{\alpha}$

Structure of layer premiums:

Regard a layer C xs D in the Pareto area, i.e. D>θ. Infinite C is admissible for α>1. The average layer loss equals

$E(\min(X-D,C)|X>D)=\dfrac{D}{\alpha-1}\left(1-\left(1+\dfrac{C}{D}\right)^{1-\alpha}\right)$, for α=1 we have $D\ln\left(1+\dfrac{C}{D}\right)$.

If η is the loss frequency at θ, the frequency at D equals $\eta\left(\dfrac{\theta}{D}\right)^{\alpha}$. Thus the *risk premiums* of layers have a particular structure, equaling up to a constant a function $D^{1-\alpha}\psi(D/C)$. From this we quickly obtain the

Pareto extrapolation formula for layers:

$\dfrac{risk-premium-of-C_2xsD_2}{risk-premium-of-C_1xsD_1}=\dfrac{(C_2+D_2)^{1-\alpha}-D_2^{1-\alpha}}{(C_1+D_1)^{1-\alpha}-D_1^{1-\alpha}}$, in case α=1 we have $\dfrac{\ln\left(1+\dfrac{C_2}{D_2}\right)}{\ln\left(1+\dfrac{C_1}{D_1}\right)}$.

Distributions with nice properties only help if they provide good fits in practice. From (re)insurance data it is clear that not all tails in the world of finance are Pareto distributed, in particular the model often seems to be somewhat too heavy-tailed at the very large end. However, nevertheless Pareto can be a good model for a wide range of loss sizes. E.g., if it fits well between 1 und 20 Mio., one can use it for layers in that area independently of whether beyond 20 million one needs a different model or not.

To quickly check whether an empirical distribution is well fit by the Pareto model, at least for a certain range of loss sizes, there is a well-known graphical method available:

$\overline{F}(x)$ *is Pareto* is equivalent to

$\overline{F}(x)$ *is a straight line on double-logarithmic paper* (having slope –α)

Hence, if the log-log-graph of an empirical distribution is about a straight line for a certain range of loss sizes, in that area a Pareto fit is reasonable.

Thinking of quite small intervals of loss sizes being apt for Pareto fitting, we come to a generalization being applicable to any smooth distribution (see Riegel, 2008): the *local Pareto alpha*. Mathematically it is the linear approximation of $\overline{F}(x)$ in the log-log world, to be exact the negative derivative.

Local Pareto alpha at d: $\alpha_d = -\frac{d}{dt}\Big|_{t=\ln(d)} \ln\left(\overline{F}(e^t)\right) = d \cdot \frac{f(d)}{\overline{F}(d)}$

If α_d is constant on an interval, this interval is a Pareto distributed piece of the distribution. In practice one often, but not always, observes that for very large d (say in the million Euro range) α_d is an increasing function of d – in a way the tail gradually becomes less heavy.

The above Pareto extrapolation formula for frequencies finally yields an *intuitive interpretation* of the (possibly local) Pareto alpha: it is the *speed* of the decrease of the loss frequency as a function of the threshold. One sees quickly that if we increase a threshold d by p% with a very small p, the loss frequency decreases by approximately α_dp%.

3 Generalized Pareto – Reinsurer's new love?

Now we study a well-known generalization of the Pareto model (see Embrechts et al.). Arguably less known is that it shares some of the properties making the Pareto model so popular:

GPD (*Generalized Pareto Distribution*): $\overline{F}(x|X>\theta) = \left(1+\xi\frac{x-\theta}{\tau}\right)^{-1/\xi}$

The parameter ξ can take any real value. However, for negative ξ x is bounded from above, a case less interesting for application in insurance. ξ=0 is the well-known Exponential distribution. We only treat the case ξ>0 here.

The above parametrisation comes from Extreme Value Theory, which suggests that GPD is an adequate model for many situations of data exceeding large thresholds. This is a good reason to work with this model. A further good reason turns out from a parameter change proposed by Scollnik.

Set $\alpha := 1/\xi > 0$, $\lambda := \alpha\tau - \theta > -\theta$. Now we have $\overline{F}(x|X>\theta) = \left(\frac{\theta+\lambda}{x+\lambda}\right)^{\alpha}$

This is a **tail model** in two parameters α und λ; θ is the *known* model threshold. However, θ is the third parameter in the corresponding **GPD-only** model, having no losses between 0 and θ analogously to the Pareto case.
The parameter space is a bit intricate as λ may take on (not too large) negative values. Maybe for parameter estimation one better makes a detour to a different parametrisation. However, apart from this complication the chosen representation will turn out to be extremely handy, revealing in particular a lot of analogies to the Pareto model.

At a glance we notice two well-known special cases:

λ=0: This is the Pareto tail model from above.

λ>0, θ=0: This is not a tail model but a **ground-up model** for losses of any range. In the literature it is often called Pareto as well. However, some more specific names have been introduced:
Pareto Type II = American Pareto = Two-parameter Pareto = **Lomax**

Let us look briefly at a third kind of model. Any tail model $X|X>\theta$ has a corresponding **excess model** $X-\theta|X>\theta$. If the former is GPD as above, the latter has the survival function $\left(\frac{\theta+\lambda}{x+\theta+\lambda}\right)^{\alpha}$, which is Lomax with parameters α and θ+λ>0. In the Pareto case we have $\left(\frac{\theta}{x+\theta}\right)^{\alpha}$, which appears to be Two-parameter Pareto as well but strictly speaking is not: Here θ is the *known* threshold – this model has the only parameter α.

The names *Single* vs. *Two-parameter Pareto* (apart from anyway not being always consistently used) are somewhat misleading – as we have seen, both models have variants with 1 or 2 parameters. Whatever the preferred name, when using a Pareto variant it is essential to make always clear whether one treats it as a ground-up, a tail, or an excess model.

Coming back to the GPD tail model, if we as above derive the model for a higher tail starting at d>θ, we get

$$\overline{F}(x|X>d)=\left(\frac{\theta+\lambda}{x+\lambda}\right)^{\alpha}\Big/\left(\frac{\theta+\lambda}{d+\lambda}\right)^{\alpha}=\left(\frac{d+\lambda}{x+\lambda}\right)^{\alpha}$$

As in the Pareto case the model "forgets" the original threshold θ, replacing it by the new one. Again the parameter α remains unchanged, but also the second parameter λ. Both are thus invariants when modeling higher tails. Other common parametrisations of the GPD have only the invariant parameter α (or the inverse ξ, respectively), the second parameter changes in a complicate way when shifting from a tail threshold to another one.

The invariance of the GPD parameters α and λ yields the same advantages for tail analyses as the Pareto model: There is no need to know exactly where the tail begins, one can compare tails starting at different thresholds, and finally it might be possible to derive market values for the two parameters in certain business areas. Thus the potential range of application of the GPD should be the same as that of the Pareto model. The additional parameter adds flexibility – while on the other hand requiring more data for the parameter estimation.

Here too it is possible to interpret the two parameters in an intuitive way:

λ is a "shift" from the Pareto model with the same alpha, having the same dimension as the losses and layer parameters. We could think of starting with a Pareto distribution having threshold θ+λ, then all losses are shifted by λ to the left (by subtracting λ) and we obtain the GPD. Thus in graphs (with linear axes) GPD tails have exactly the same shape as Pareto tails, only their location on the x-axis is different.

The parameter α, apart from belonging to the corresponding Pareto model, is the local Pareto alpha at infinite: $\alpha_\infty = \alpha$. More generally, one sees quickly that $\alpha_d = \frac{d}{d+\lambda}\alpha$.

The behavior of α_d as a function of d is as follows:

λ>0: α_d rises (often observed)
λ=0: Pareto
λ<0: α_d decreases

For any d≥θ we easily get $\overline{F}(x|X>d) = \left(1+\frac{\alpha_d}{\alpha}\left(\frac{x}{d}-1\right)\right)^{-\alpha}$, which is an alternative GPD parametrisation focusing on the local alphas, see Riegel (2008).

Bearing in mind that GPD is essentially Pareto with the x-axis shifted by λ, we get without any further calculation very easy formulae being very similar to the Pareto case:

$$\frac{frequency-at-d_2}{frequency-at-d_1} = \left(\frac{d_1+\lambda}{d_2+\lambda}\right)^{\alpha}$$

$$E(\min(X-D,C)|X>D) = \frac{D+\lambda}{\alpha-1}\left(1-\left(1+\frac{C}{D+\lambda}\right)^{1-\alpha}\right),\quad \text{case } \alpha=1\text{:}\quad (D+\lambda)\ln\left(1+\frac{C}{D+\lambda}\right)$$

$$\frac{risk-premium-of-C_2xsD_2}{risk-premium-of-C_1xsD_1} = \frac{(C_2+D_2+\lambda)^{1-\alpha}-(D_2+\lambda)^{1-\alpha}}{(C_1+D_1+\lambda)^{1-\alpha}-(D_1+\lambda)^{1-\alpha}},\quad \text{case } \alpha=1\text{:}\quad \frac{\ln\left(1+\frac{C_2}{D_2+\lambda}\right)}{\ln\left(1+\frac{C_1}{D_1+\lambda}\right)}$$

Summing up, GPD is nearly as easy to handle as Pareto, having two advantages: greater flexibility and the backing of Extreme Value Theory making it the preferred candidate for the modeling of high tails.

As for the estimation of the GPD parameters, see Brazauskas & Kleefeld (2009) studying various fitting methods, from the traditional to newly developed ones. (Note that their parametrisation uses the exponent $\gamma := -\xi = -1/\alpha$.)

4 With or without a knife – construction of distribution variants

We strive after further flexibility in our distribution portfolio. Before focusing on the most critical area, namely the small losses, we have a brief look at the opposite side, the very large losses:
Sometimes losses greater than a certain maximum are impossible (or not interesting at all): X ≤ Max. There are two easy ways to adapt distributions with infinite support to this: *Censoring* and *Truncation*. We follow the terminology of Klugman et al., noting that in the literature we occasionally found the two names interchanged.

Right censoring: $\overline{F}_{cs}(x) = \overline{F}(x)$ for x<Max, for Max≤x $\overline{F}_{cs}(x) = 0$

Properties of the survival function:

- mass point (jump) at Max with probability $\overline{F}(Max)$
- below Max same shape as original model

A mass point at the maximum loss is indeed plausible in some practical cases. Say there is a positive probability for a total loss (100% of the sum insured) in a fire homeowners policy, which occurs if the insured building completely burns down.

Right truncation: $\overline{F}_{tr}(x) = \dfrac{\overline{F}(x) - \overline{F}(Max)}{1 - \overline{F}(Max)}$ for x<Max, for Max≤x $\overline{F}_{tr}(x) = 0$

Properties of the survival function:

- equals the conditional distribution of $X | X \le Max$
- continuous at Max, no mass point
- shape below Max is a bit different from original model, tail is thinner, however, the numerical impact of this deviation is low for small/medium losses

Of course both variants yield finite expectation even if the expected value of the original model is infinite, so working with such models, e.g. GPD tails with α<1, will not cause any problems.
Left censoring and left truncation are analogous. We have already applied the latter above – an upper tail is formally a left truncation of the model it is derived from. Both ways to disregard the left or right end of the distribution can be combined and applied to any distribution, including the Pareto family. Right truncation is in particular a way to get tails being somewhat thinner than Pareto in the area up to Max.

It shall be noted that the right-censored/truncated versions of models with GPD/Pareto tail preserve the memoryless property stated above. For censoring this is trivial – the only change is that the tail ends in a jump at Max. As for truncating, let $\overline{F}$ be a survival function with such a tail, i.e. $\overline{F}(x) = \overline{F}(\theta)\left(\dfrac{\theta+\lambda}{x+\lambda}\right)^{\alpha}$ for x≥θ. As each higher tail is again GPD with the same parameters, for any x≥d≥θ we have $\overline{F}(x) = \overline{F}(d)\left(\dfrac{d+\lambda}{x+\lambda}\right)^{\alpha}$, leading to

$$\overline{F}_{tr}(x|X>d) = \frac{\overline{F}_{tr}(x)}{\overline{F}_{tr}(d)} = \frac{\overline{F}(x) - \overline{F}(Max)}{\overline{F}(d) - \overline{F}(Max)} = \frac{\left(\dfrac{d+\lambda}{x+\lambda}\right)^{\alpha} - \left(\dfrac{d+\lambda}{Max+\lambda}\right)^{\alpha}}{1 - \left(\dfrac{d+\lambda}{Max+\lambda}\right)^{\alpha}}.$$

The original threshold θ disappears again; each truncated GPD/Pareto tail model has the same parameters α, λ, and Max.

Now we start investigating ground-up models having a more promising shape for smaller losses than Pareto-only with its gap between 0 and θ. In the following we always display the survival function $\overline{F}(x)$.

One example we have already seen – as a special case of the GPD:

Lomax: $\left(\frac{\lambda}{x+\lambda}\right)^{\alpha} = \left(\frac{1}{1+x/\lambda}\right)^{\alpha}$

This is a ground-up distribution with two parameters, the exponent and a scale parameter. It can be generalized via transforming, see Klugman et al. The resulting three-parameter distribution is called

Burr: $\left(\frac{1}{1+\left(x/\lambda\right)^{\tau}}\right)^{\alpha}$

For large x this model asymptotically tends to Pareto-only with exponent ατ, however, in the area of the small losses there is much more flexibility. While Burr with τ<1 and Lomax (τ=1) always have strictly decreasing density, thus mode (maximum density) at zero, for the Burr variants with τ>1 the (only) mode is positive. This is our first example of a unimodal distribution having a density looking roughly like an asymmetric bell curve, as well as a tail being very similar to Pareto.

More examples can be created via combining distributions. There are two handy options for this, see Klugman et al.

Mixed distributions

A mixed distribution (we only regard finite mixtures) is a weighted average of two (or more) distributions, being in particular an elegant way to create a bimodal distribution out of two unimodal ones, etc. Generally speaking the intuitive idea behind mixing is as follows: We have two kinds of losses, e.g. material damage and bodily injury in MTPL, having different distributions. Then it is most natural to model them separately and combine the results, setting the weights according to the frequency of the two loss types. The calculation of cdf, pdf, (limited) expected value, and many other figures is extremely easy – just take the weighted average of the figures describing the original models.

A classical example is a mixture of two Lomax distributions, the so-called

5-parameter-Pareto: $r\left(\frac{\lambda_1}{x+\lambda_1}\right)^{\alpha_1} + (1-r)\left(\frac{\lambda_2}{x+\lambda_2}\right)^{\alpha_2}$

Even more popular seems to be **4-parameter-Pareto**, essentially the same model with the number of parameters reduced via the formula $\alpha_1 = \alpha_2 + 2$.

Sometimes mixing is used even in case there is no natural separation into various loss types. The idea is: There is a model describing the smaller losses very well but underestimating the tail. If this model is combined with a quite heavy-tailed model, giving the latter only a tiny weight, the resulting mixture will for small losses be very close to the first model, whose

impact will though quickly fade out for larger losses, giving the second model the chance to take over and yield a good tail fit.
Pursuing this idea more strictly one naturally gets to spliced, i.e. piecewise defined distributions. Basic idea: Why not just stick pieces of two (or more) different models together?

Spliced distributions

Splicing is frequently defined in terms of densities, however, in order to make it a bit more general and intuitive we formulate it via the survival function.

The straightforward approach is to replace the tail of a model with another one: Define

$$\overline{F}(x) = \overline{F}_1(x) \text{ for } 0<x<\theta, \text{ for } \theta\leq x \quad \overline{F}(x) = \overline{F}_1(\theta)\overline{F}_2(x)$$

with survival functions $\overline{F}_1(x)$ and $\overline{F}_2(x)$. Note that to enable continuous functions the latter must be a tail model starting at θ while the former is a model for the whole range of loss sizes, whose tail is ignored.
We could let the survival function have a jump (mass point) at the threshold θ, however, jumps in the middle of an elsewhere continuous function are felt implausible in practice – typically one glues continuous pieces to obtain a continuous function (apart from maybe an mass point at a maximum, see the censoring procedure above). More than continuous the pieces often are (more or less) smooth, so it seems even natural to demand some smoothness at θ too. We will see a range of examples soon.

Splicing can be a bit more general than the above *tail replacement*. One starts with two distributions which do not intersect: The **body distribution** $\overline{F}_b(x)$ has all loss probability ≤θ, i.e. $\overline{F}_b(x) = 0$ for θ≤x. The **tail distribution** has all probability above θ, i.e. $\overline{F}_t(x) = 1$ for x≤θ. The spliced distribution is simply the weighted average of the two, which means that splicing is a special case of mixing, yielding the same easy calculations. See the overview:

Weight	*Range*	*Name*	*Description*
r	0≤x<θ	Body	small/medium loss distribution
1–r	θ≤x	Tail	large loss distribution

Note that here the weights can be chosen arbitrarily in order to let a lower or higher percentage of the losses be large: r is the probability of a loss being not greater than the threshold θ.
Our first approach (tail replacement) is the special case $r = F_1(\theta) = 1 - \overline{F}_1(\theta)$: Formally it is a weighted average of the right truncation of $\overline{F}_1(x)$, equaling $\dfrac{\overline{F}_1(x)}{1-\overline{F}_1(\theta)}$ for x<θ and 0 for θ≤x, and the F_2-only analogue of Pareto-only, equaling 1 for x<θ and $\overline{F}_2(x)$ for θ≤x.

Although looking a bit technical, splicing has a number of advantages. Firstly the interpretation (smaller vs. large losses) is very intuitive. Further, by combining suitable types of distributions we can get the desired geometries in the body and the tail area, respectively, without having an area where the two models interfere. In particular we can combine traditional tail models with distributions known to be apt for smaller losses. Finally, if we have a clear idea about θ, i.e. where the tail starts, we have the option to split the empirical data and do the parameter estimation of body and tail completely separately.

5 The Lognormal-Pareto world

Let us now apply the splicing procedure to the Lognormal and the GPD distribution. Starting from the most general case and successively adding constraints we get a hierarchy (more precisely a partially ordered set) of distributions. As before we always display the survival function $\overline{F}(x)$. Here's the starting point:

LN-GPD-0: $1-\frac{r}{\Phi\left(\frac{\ln(\theta)-\mu}{\sigma}\right)}\Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)$ for 0≤x<θ, for θ≤x $(1-r)\left(\frac{\theta+\lambda}{x+\lambda}\right)^{\alpha}$

This is a function in six parameters, inheriting μ and σ from Lognormal, α and λ from GPD, plus the splicing point θ and the mixing weight r. As for the parameter space, μ can take any real value; σ, θ, α>0; λ>–θ; 0<r<1. Limiting cases are Lognormal (r=1, θ=∞) and for r=0 GPD-only.

To simplify the notation about the Normal distribution we will sometimes write shortly $\Phi_x=\Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)$, $\phi_x=\phi\left(\frac{\ln(x)-\mu}{\sigma}\right)$. The body part of LN-GPD-0 then reads $1-\frac{r}{\Phi_\theta}\Phi_x$.

From this basic function can be can derive special cases, having less parameters, in three straightforward ways:

Tail: We can choose a Pareto tail, i.e. set λ=0. This is always possible whatever values the other parameters take. We call the resulting model **LN-Par-0**.

Distortion: As mentioned the distribution in the body area is generally not exactly Lognormal, instead it is "distorted" via the weight r. If we want it to be exactly Lognormal we have to choose $r=\Phi_\theta=\Phi\left(\frac{\ln(\theta)-\mu}{\sigma}\right)$.
This choice is always possible whatever values the other parameters take. We call this model **pLN-GPD-0** with "pLN" meaning **proper Lognormal**.

Smoothness: If we want the distribution to be smooth we can demand that the pdf be continuous, or more strongly the derivative of the pdf too, and so on. Analogously to the classes C0, C1, C2, … of more or less smooth functions we would call the resulting distributions **LN-GPD-1**, **LN-GPD-2**, …, according to how many derivatives of the cdf are continuous.
How many smoothness conditions can be fulfilled must be analyzed step by step. For C1 we must have that the pdf at θ– and θ+ be equal. Some algebra yields
$\frac{r\phi_\theta}{\Phi_\theta\sigma\theta}=\frac{(1-r)\alpha}{\theta+\lambda}$, or equivalently $\frac{\alpha\theta}{\theta+\lambda}=\frac{r}{1-r}\frac{\phi_\theta}{\sigma\Phi_\theta}$, or equivalently $\alpha=\frac{\theta+\lambda}{\theta}\frac{\phi_\theta}{\sigma\Phi_\theta}\frac{r}{1-r}$.
The second equation describes the local Pareto alpha at θ– and θ+; the third one makes clear that one can always find an α fulfilling the C1-condition whatever values the other parameters take.
Note that all LN-GPD variants with continuous pdf must be unimodal: The GPD density is strictly decreasing, thus the pdf of any smooth spliced model must have negative slope at θ. Hence, the mode of the Lognormal body must be smaller than θ and is thus also the mode of

the spliced model. This gives the pdf the (often desired) shape of an asymmetric bell curve with a heavy tail.
If the pdf is instead discontinuous at θ the model could be bimodal: Say the Lognormal mode is smaller than θ and the pdf of the GPD takes a very high value at θ. Then both points are local maximums of the density.

We have not just found three new distributions – the underlying conditions for λ, r, and α can be combined with each other, which yields intersections of the three defined function subspaces. All in all we get eight distributions, forming a three-dimensional grid. We label them according to the logic used so far. Here's an overview, showing also the parameter reduction in the tree dimensions:

LN-GPD-0 μ, σ, θ, r, α, λ	➔ λ	**LN-Par-0** μ, σ, θ, r, α		
↓ α		↓		
<u>LN-GPD-1</u> μ, σ, θ, r, λ	➔	**<u>LN-Par-1</u>** μ, σ, θ, r		
		↘ r		
		pLN-GPD-0 μ, σ, θ, α, λ	➔	**pLN-Par-0** μ, σ, θ, α
		↓		↓
		pLN-GPD-1 μ, σ, θ, λ	➔	**<u>pLN-Par-1</u>** μ, σ, θ

The three highlighted distributions have been published, in two papers apparently not referring to each other:

pLN-Par-1 was introduced 2003 by Knecht & Küttel naming it **Czeledin distribution**. (*Czeledin* is the Czech translation of the German word *Knecht*, meaning *servant*. Precisely, having all accents available, it would be spelt Čeledín.)
Recall this is a Lognormal distribution with a Pareto tail attached, having continuous pdf. The original motivation for this model was apparently the fitting of aggregate loss distributions (or equivalently loss ratios). However, that does not mean that it could not be suitable as a model for single losses.

As the Czeledin model is quite popular in reinsurance we rewrite its survival function:

$1-\Phi_x$ for $0 \le x < \theta$, for $\theta \le x$ $\quad (1-\Phi_\theta)\left(\frac{\theta}{x}\right)^\alpha \quad$ with $\quad \alpha = \frac{\phi_\theta}{\sigma(1-\Phi_\theta)}$

LN-GPD-1 and **LN-Par-1** were introduced 2007 by Scollnik as *third* and *second composite Lognormal-Pareto model*. (An initial *first* model turned out to be too inflexible in practice.) He also derived the condition for **LN-GPD-2**: f'(θ–)=f'(θ+) leads to $\ln(\theta)-\mu=\sigma^2\frac{\alpha\theta-\lambda}{\theta+\lambda}$, with the right hand side simplifying to $\sigma^2\alpha$ in the **LN-Par-2** case. Note that this formula is not independent of the parameter r, which comes in via the C1-condition on α above. This system of equations yields solutions, thus LN-GPD-2 and LN-Par-2 exist.

Testing whether a C3-function is possible Scollink derived the equation $\ln(\theta)-\mu-1=\sigma^2\left[(\alpha+1)\left(\frac{\theta}{\theta+\lambda}\right)^2-1\right]$. In the Pareto case the right hand side again simplifies to $\sigma^2\alpha$, which is inconsistent with the preceding equation. Thus the model LN-Par-3 does not exist; for **LN-GPD-3** only non-Pareto constellations are possible.

To conclude with the Lognormal-GPD world let us compare two of the models already present in the literature: Czeledin (pLN-Par-1) and the C2 variant of Scollnik's second model (LN-Par-2). Both have three parameters (thus as for complexity are similar to the Burr distribution), being special cases of the model LN-Par-1 having parameters μ, σ, θ, r. See the relevant part of the grid:

LN-Par-1
μ, σ, θ, r

↓ ↘

LN-Par-2	**pLN-Par-1**
σ, θ, r	μ, σ, θ

Are these functions the same, at least for some parameter constellation? Or are they fundamentally different? If yes, how?
In other words: Can we attach a Pareto tail to a (proper) Lognormal distribution in a C2 (twice continuously differentiable) manner?

Lemma: For any real number z we have $z<\frac{\phi(z)}{1-\Phi(z)}$.

Proof: Only the case z>0 is not trivial. Recall that the Normal density fulfils the differential equation $\phi'(x)=-x\phi(x)$. From this we get

$$z(1-\Phi(z))=z\int_z^\infty\phi(x)dx=\int_z^\infty z\phi(x)dx<\int_z^\infty x\phi(x)dx=-\int_z^\infty\phi'(x)dx=\phi(z)\qquad \text{q e d.}$$

Now recall that the C1 condition for α in the Pareto case reads $\alpha=\frac{\phi_\theta}{\sigma\Phi_\theta}\frac{r}{1-r}$. This is fulfilled in both models we are comparing. LN-Par-2 in addition meets the C2 condition $\ln(\theta)-\mu=\sigma^2\alpha$. Plugging in α and rearranging we get $\frac{\ln(\theta)-\mu}{\sigma}\frac{\Phi_\theta}{\phi_\theta}=\frac{r}{1-r}$.

If we apply the lemma on $z=\frac{\ln(\theta)-\mu}{\sigma}$ we see that the left hand side of this equation is smaller than $\frac{\Phi_\theta}{1-\Phi_\theta}$. Thus we have $\frac{r}{1-r}<\frac{\Phi_\theta}{1-\Phi_\theta}$, being equivalent to $r<\Phi_\theta$. That means: The weight r of the body of LN-Par-2 is always smaller than Φ_θ, which is the body weight in all proper Lognormal cases, including the Czeledin function.

We can conclude: The two functions LN-Par-2 and pLN-Par-1 are fundamentally different. If we want to attach a Pareto tail to a Lognormal, the best we can have is a C1 function (continuous pdf). If we want to attach it in a smoother way, the Lognormal must be distorted, more exactly: distorted in such a way that the part of small/medium losses gets less probability weight while the large losses get more weight than in the proper Lognormal case.

6 Variants and applications

This section collects variants of the models discussed so far, referring in particular to fits on real data. Apart from the Czeledin function above, applications apparently focus on loss severity distributions. This is not surprising as here, if lucky, one can work with thousands of losses per year, enabling more complex modeling, while fits on aggregate losses typically have only one data point per year available, which is a bit scarce for any kind of fit.

The potentially richest data sources for parametric modeling purposes are arguably institutions routinely pooling loss data on behalf of whole markets. E.g. in the USA for most non-life lines of business this is ISO (Insurance Services Office), in Germany it is the insurers' association GdV. Such institutions would typically, due to confidentiality reasons, neither disclose their data nor the details of their analyses (like parameters of fits) to the general public, however, in many cases their general methodology is disclosed and may even be part of the actuarial education. So for certain classes/lines of business it is widely known to the industry which type of model the market data collectors found useful. Although from a scientific viewpoint one would prefer to have the data and the details of the analyses available, such reduced information can nevertheless help get an idea of which distributions might be apt for what kind of business.
As for ISO, it is known that they have successfully applied the above 4-parameter Pareto model to classes of general liability business, see Klugman et al. In other businesses a much less heavy-tailed but otherwise very flexible model comes into play: a mixture of several Exponential distributions.
GdV supports insurers with a model providing the risk premium discount for deductibles, being applicable to certain segments of property fire policies. The parametric fit is only partial here: the body is Lognormal up to about 5% of the sum insured. The empirical tail is heavier than what Lognormal would yield, resulting in an average loss exceeding that of the Lognormal model by a certain percentage. This model, albeit not a complete fit, is sufficient to yield the desired output for the deductibles offered in practice. However, if one wanted to rate deductibles higher than 5% SI or layer policies one would need to extend the Lognormal fit. The LN-GPD family described in the past section would be a suitable candidate for this.

Coming back to spliced models in general, it is straightforward to test alternative distributions for the body. In the literature (at least) three models analogous to the LN-GPD family can be found, replacing Lognormal by another common distribution with cdf $F_1(x)$. The resulting survival function is $1-\frac{r}{F_1(\theta)}F_1(x)$ for 0≤x<θ, for θ≤x $(1-r)\left(\frac{\theta+\lambda}{x+\lambda}\right)^{\alpha}$.

Weibull-GPD

This model is analyzed by Scollnik & Sun, proceeding analogously to Scollnik. They derive formulae for **Wei-GPD-1/2/3** and **Wei-Par-1/2** (no. 3 here doesn't exist either) and apply these and their Lognormal counterparts to a data set being very popular in the actuarial literature, the *Danish fire data*, adding comments on how parameter inference can be realized.

Note that, like Burr, the Weibull model, according to the value of one of the parameters, either has a strictly falling density or a unique positive mode (skewed bell shape), see Klugman et al. Thus for Weibull-GPD both geometries are possible.

Exponential-GPD

Teodorescu & Vernic also follow the path of Scollnik and derive formulae for the functions **Exp-GPD-2** and **Exp-Par-2**, showing that further smoothness is not possible in either case. These models are less complex, having one parameter less, than their Lognormal and Weibull counterparts. The Exponential pdf is strictly decreasing, thus Exp-GPD cannot provide bell-shaped densities.
Riegel (2010) uses **pExp-Par-1** as an option to fit various Property market data. This model has the nice property that in the body area the local Pareto alpha increases linearly from 0 to the alpha of the tail.
Practitioners remember that **pExp-Par-0**, the continuous-only variant, some time ago was an option for some ISO data.

Power function-GPD

The power function $\left(\frac{x}{\theta}\right)^{\beta}$, β>0 can be seen as a cdf concentrated between 0 and θ, which makes it a perfect body candidate for the body in a spliced model. Thus we can define the survival function

Pow-GPD-0: $1-r\left(\frac{x}{\theta}\right)^{\beta}$ for 0≤x<θ, for θ≤x $(1-r)\left(\frac{\theta+\lambda}{x+\lambda}\right)^{\alpha}$

It has as many parameters as Exp-GPD, however, the shape of the density below θ is very flexible: For β>1 rising, for β<1 decreasing, for β=1 we have the uniform distribution.
The special case **Pow-Par-1** is well known (far beyond the actuarial world), appearing in the literature as **(asymmetric) Log-Laplace** or **double Pareto** distribution, see Kozubowski & Podgórski for a comprehensive overview.

Generally wee have $f(\theta-)=\frac{r\beta}{\theta}$ and $f(\theta+)=\frac{(1-r)\alpha}{\theta+\lambda}$, thus the C1 condition for the Pareto case reads $\frac{r}{1-r}=\frac{\alpha}{\beta}$ or equivalently $r=\frac{\alpha}{\alpha+\beta}$. More generally for Pow-GPD this condition reads $\frac{r}{1-r}=\frac{\alpha_\theta}{\beta}$ with the local Pareto alpha at θ taking the place of α.

The C2 condition turns out to be $\beta = 1 - (\alpha + 1)\frac{\theta}{\theta + \lambda}$, which can be fulfilled by a positive β iff $\lambda > \alpha\theta$. Thus the parameter space of **Pow-GPD-2** is restricted and in particular Pow-Par-2 does not exist.

The distinction of proper and distorted bodies, as set out for LN-GPD, is meaningless in the power case – here each function is proper: For x<θ we have $r\left(\frac{x}{\theta}\right)^{\beta} = \left(\frac{x}{\zeta}\right)^{\beta}$ with $\zeta = \theta r^{-\frac{1}{\beta}} > \theta$, thus we can interpret Pow-GPD as being derived from a power function having parameters β and ζ, being cut at θ<ζ, and equipped with a GPD tail.

???-GPD

Analogous functions with other body distributions can be easily constructed, and most probably some of them have already been used. For each combination of a body with a GPD tail there is a grid of models, analogously to the LN-GPD case discussed in detail, linking the most general continuous function **…-GPD-0** with its subclasses Pareto / proper / C1, 2, … and intersections thereof. It is yet to be discovered how many of these functions exist – and how many are useful in practice.
Notice that however rich (and maybe confusing) the class of spliced functions with GPD tail will get – all of them are comparable i.r.o. tail behavior via the parameters α and λ.

A word about the question how many parameters are adequate, which in spliced models is tied to the degree of smoothness. There is a trade-off. Smoothness reduces parameters, which is good for the frequent cases of scarce data. On the other hand it links the geometries of body and tail, reducing the flexibility the spliced models are constructed for. Bearing in mind that all those readily available 2-parameter-distributions are only occasionally flexible enough for good fits over the whole range of loss sizes, it is plausible that a minimum of 3-4 parameters is necessary. However, the risk of overparametrisation goes always on the actuary's side…

The great advantage of spliced models with GPD or Pareto tail over other models with similar complexity (same number of parameters) is what was highlighted in sections 3 and 4: From other analyses we might have an idea of what range of values α (and λ, if applicable) in practice take on. Although being possibly vague, this kind of knowledge can be (formally or at least informally via subsequent cross checking) incorporated into the parameter estimation, which should enable actuaries to work with more parameters than they would feel comfortable with had they to rely on the data to be fitted only.

A quite moderate way to increase flexibility, being applicable in case of limited loss size, is right truncation as explained above. It can be applied to all models discussed so far, making them somewhat less heavy tailed. Truncation does add the additional parameter Max, however, being a linear transformation of the original curve it does not affect the overall geometry of the distribution too much. Thus truncating should generally be less sensitive than other extensions of the parameter space. Moreover in practice there are many situations where the parameter inference is greatly eased by the fact that the value of Max is (approximately) known, resulting from insurance policy conditions or other knowledge.

7 A new look at old exposure rating methods

Focusing on the mathematical core, is exposure rating essentially the calculation of the limited expected value $LEV(C) = E(\min(X,C)) = \int_0^C \overline{F}(x)dx$ of a loss severity distribution, for varying C, see Mack & Fackler.
As the various distributions presented in this paper are promising candidates to model the severity of insurance risks, they could all make their way into the world of exposure rating models. Some have been there long-since.
We present two of them, being possibly the oldest parametric models in their respective area.

An industrial fire exposure curve

Mack (1980) presents an exposure rating model derived from loss data of large industrial fire risks. As is common in the property lines (and various hull business too), the losses are not modeled in Dollar amounts but as loss degrees, i.e. in percent of a figure describing the insured value, here being the sum insured. The applied model in our terminology is

right truncated Pareto-only: $\overline{F}(x) = \frac{(\theta/x)^\alpha - (\theta/Max)^\alpha}{1-(\theta/Max)^\alpha}$ for $\theta \le x < Max$,

with a very small θ =0.01% and Max = 100% (of the SI).

For α values of 0.65 und even lower are proposed. Interestingly for this Pareto variant the parameter space of α can be extended to arbitrary real values. $\alpha = -1$ is the uniform distribution between θ and Max, for lower α higher losses are more probable than lower ones (which makes this area implausible for practical use).

Left truncated distributions like this can emerge if the underlying business has high deductibles cutting off the whole body of smaller losses. If in such a case one wanted to model the same risks with possibly lower or no deductibles, one could extend it to a spliced model having the original model as its tail.

The Riebesell model for liability policies

This liability exposure model, also called **power curve** model, dates back as far as 1936, however, initially it was an intuitive pricing scheme. Much later it turned out to have an underlying stochastic distribution model, namely a spliced model with Pareto tail, see Mack & Fackler (2003). We rewrite a part of their findings in the terminology introduced here, enhancing them and commenting on various practical issues.

The **Riebesell rule** states that if the limit of a liability policy is doubled, the risk premium increases by a fixed percentage z, whatever the policy limit. Let us call z the **doubled limits surcharge** (DLS), following Riegel (2008). A typical DLS in practice is 20%, however, values vary greatly according to the kind of liability coverage, almost exhausting the interval of reasonable values ranging from 0 to 100%.

As the risk premium is the product of frequency and (limited) expected value, this rule can be formulated equivalently in terms of LEV: $LEV(2C) = LEV(C)(1+z)$. The consistent extension of this rule to arbitrary multiples of C is $LEV(aC)/LEV(C) = (1+z)^{ld(a)} = a^{ld(1+z)}$ with *ld* being the logarithm to the base 2.
Hence, the function *LEV*(*x*) up to a constant equals $x^{ld(1+z)}$. Taking the derivative we get the survival function, which up to a factor equals $x^{-\alpha}$ with $\alpha = 1 - ld(1+z) > 0$.

Thus the loss severity is Pareto distributed with α<1. (This part of the story was widely known to the reinsurance industry decades before the Mack & Fackler paper emerged.) Now two problems seem to arise.
Firstly, Pareto starts at a threshold θ>0, which means that the Riebesell rule cannot hold below. However, if the threshold were very low, say 10 Euro or US$, this would be no material restriction as such low deductibles in practice hardly exist.
Secondly, the distribution has infinite expectation. However, as almost all liability policies have a limit (rare exceptions being Personal Lines TPL and MTPL in some European countries) one could in principle insure a risk with infinite expected value. (Perhaps this is unwittingly done in some real world cases.) Moreover it could be that a severity distribution is perfectly fit by a Pareto with α<1 up to a rather high value Max being larger than the limits needed in practice, while beyond it has a much lighter tail or even a maximum loss. In this case the well-fitting Pareto distribution is an adequate model for practical purposes, always bearing in mind its limited range of application.

Summing up, it is possible to slightly generalize the deduction of the Pareto formula:

Assume the Riebesell rule with a DLS 0<z<1 holds for all policy limits contained in an open interval (θ, Max), θ<Max≤∞. Then we have the following *necessary* condition:

(NC1) $$\overline{F}(x|X>\theta) = \left(\frac{\theta}{x}\right)^{\alpha}, \quad \theta<x<\text{Max}, \quad \text{with } \alpha = 1 - ld(1+z).$$

The proof is the same as above, noting that only local properties were used. Notice that as *LEV*(*x*) is continuous the Riebesell rule will hold on the closed interval from θ to Max.

In order to find a sufficient condition recall that the unconditioned survival function can be written as a spliced model with a survival function $\overline{F}_1(x)$ being cut at θ:

$\overline{F}_1(x)$ for 0<x<θ, for θ≤x $(1-r)\left(\frac{\theta}{x}\right)^{\alpha}$ with $r = 1 - \overline{F}_1(\theta)$

Whether or not $\overline{F}(x)$ can be (made) continuous is yet to be found out. For θ≤x<Max we have

$$LEV(x) = LEV(\theta) + \overline{F}(\theta)E(\min(X-\theta, x-\theta)|X>\theta) = LEV(\theta) + (1-r)\frac{\theta}{\alpha-1}\left(1-\left(\frac{x}{\theta}\right)^{1-\alpha}\right) =$$

$$= LEV(\theta) - \theta\frac{1-r}{1-\alpha} + \theta\frac{1-r}{1-\alpha}\left(\frac{x}{\theta}\right)^{1-\alpha}$$

Thus for the Riebesell rule to hold we have the sufficient condition

(SC) $$LEV(\theta) = \theta\frac{1-r}{1-\alpha}.$$

To see whether, and how this can be fulfilled note that $0 \le LEV(\theta) \le \theta$. The right hand side of (SC) is non-negative, to let it be not greater than θ we must fulfill a further – and somewhat hidden – necessary condition:

(NC2) $\alpha \le r$

Altogether we have the parameter constellation $0 < \alpha \le r < 1$. Note that in practice α is often rather close to 1, being the closer the smaller the DLS is. E.g. for z=20% we have α=0.737. In other words, $1 - r = \overline{F}(x)$, the probability of a loss being in the tail starting at θ, must be rather small, namely not greater than $1 - \alpha = ld(1+z)$. Thus in terms of probabilities the Pareto tail is only a small part of the overall distribution: This model is very different from the Pareto-only model, here we must have plenty of losses ≤θ, namely 100α% or more.
This makes clear that in realistic situations θ, the lowest figure allowing the application of the Riebesell model, cannot be as low as 10 Euro, it must be a good deal larger. While this does not need to narrow the application of the rule to the typically large sums insured (e.g. comparison of the risk premiums for the limits 2 and 3 million Euro), we cannot hope that the Riebesell rule be adequate for the calculation of the rebate to be given for a deductible of say 200 Euro, although this mathematically being the same calculation as that involving million Euro figures.

In a way the Riebesell rule is a trap. The formula seems so simple, having one parameter only with no obvious limitation of the range of policy limits the rule can be applied to. The attempt to construct in a general way severity distributions fulfilling the rule has revealed (more or less hidden) constraints. That does not mean that it is impossible to find a Riebesell model having realistic parameters. It simply means that this model is much more complex than a one-parameter model and that the properties of the model automatically confine the applicability of the Riebesell rule.

How many parameters does the general Riebesell model have, then? Let us first look at the case r=α. Now (SC) yields $LEV(\theta) = \theta$, thus all losses ≤θ must equal θ, there are no losses below θ. This is a model in two parameters θ and α. The survival function is discontinous at θ having there a mass point with probability α, whereupon the Pareto tail starts:

1 for 0≤x<θ, for θ≤x $(1-\alpha)\left(\frac{\theta}{x}\right)^{\alpha}$

This model can be found in Riegel (2008) who, apart from providing a comprehensive theory of LEV functions, generalizes the Riebesell model in various ways, focusing on the higher tail of the severity distribution.
Here instead we take a closer look at the body area, namely at the conditional body $X|X \le \theta$.
For any r≥α we want to rewrite (SC) in terms of an intuitive quantity: the **average smaller loss** $\gamma = E(X|X \le \theta) = LEV(\theta|X \le \theta)$. As X is the mixture of $X|X \le \theta$ and $X|X > \theta$ we have $LEV(\theta) = rLEV(\theta|X \le \theta) + (1-r)LEV(\theta|X > \theta) = r\gamma + (1-r)\theta$. Plugging this into (SC) we get

(SC*) $$\gamma = \theta \frac{\alpha}{1-\alpha} \frac{1-r}{r}.$$

(NC2) ensures that γ is always in the admissible range of values between 0 and θ, including the maximum in case r=α. Thus no further necessary conditions arise and we can give an intuitive classification of the severity distributions leading to the Riebesell model for risk premiums:

Theorem (Riebesell distribution): Assume that for a risk *X* the Riebesell rule with a DLS $0<z<1$ holds for all policy limits contained in the open interval (θ, Max), θ<Max≤∞.

(NC1) Then with $\alpha = 1 - ld(1+z)$ the survival function is a spliced model equaling

$$\overline{F}(x) = (1-r)\left(\frac{\theta}{x}\right)^{\alpha} \quad \text{for} \quad \theta<x<\text{Max}.$$

(NC2) For r, being the percentage of *smaller losses* not exceeding θ, we have

$$\alpha \le r < 1.$$

(SC*) The distribution of these smaller losses $X|X \le \theta$ is such that for their average we have

$$\gamma = E(X|X \le \theta) = \theta \frac{\alpha}{1-\alpha} \frac{1-r}{r}.$$

The Riebesell distribution has the parameters **α** (or equivalently z), **r**, **θ**, **Max** (unless infinite), plus additional degrees of freedom for the distribution beyond Max, plus additional degrees of freedom for the distribution up to θ (unless r=α, which lets it be concentrated at θ).

If *I* is the closed interval between θ and Max, i.e. [θ, Max] or [θ, ∞), for all policy limits C contained in *I* we have:

$$LEV(x) = \theta \frac{1-r}{1-\alpha}\left(\frac{x}{\theta}\right)^{1-\alpha} = \gamma \frac{r}{\alpha}\left(\frac{x}{\theta}\right)^{1-\alpha}$$

If we interpret policies having limit C and no deductible as special layers Cxs0 and further for any layer CxsD call D the **attachment point** and C+D the **detachment point**, the Pareto extrapolation formula for layers

$$\frac{risk-premium-of-C_2xsD_2}{risk-premium-of-C_1xsD_1} = \frac{(C_2+D_2)^{1-\alpha} - D_2^{1-\alpha}}{(C_1+D_1)^{1-\alpha} - D_1^{1-\alpha}}$$

holds for all layers *detaching* in *I* and *attaching* either also in *I* or at 0. □

It is remarkable that despite the Riebesell formula itself does not apply to the area (0, θ) of smaller losses, the **risk premium extrapolation formula** holds for policies insuring these losses, provided they cover them completely. Like the Riebesell rule this extrapolation formula depends on α (or z) only, with θ and Max coming in only indirectly via the admissible range of policy parameters.

Notice that in the theorem it is not assumed that θ be optimal, i.e. be the minimum threshold for application of the Riebesell rule with DLS z to the risk *X*. In fact it is obvious that if the rule holds for a θ, it holds all thresholds between θ and Max, each leading to a different spliced representation of the Riebesell function with different parameters θ, r, γ – only α is invariant.
This point is of practical interest. When working with real data it might be impossible to find the exact threshold where the Pareto area starts: Frequently one can only say that an empirical distribution looks very much like Pareto from a certain threshold θ onwards, while somewhat below it could be the same, however, this cannot be verified as here the single loss records are unavailable or arguably incomplete.
Nevertheless in such a data situation it could be possible to verify whether for a liability segment the Riebesell rule holds (albeit it might be impossible to find the lowest possible threshold for this risk type). A practical procedure could be as follows: In a first step tail data is analyzed, which is the kind of data reinsurers routinely collect. If such data yields values

$\alpha<1$ and θ fulfilling (NC1), r can be estimated by relating large loss (losses $>\theta$) counts with overall loss counts – such data could come from primary insurers or market statistics. If the derived r fulfils (NC2), the condition (SC*) for γ can be formulated, which can be checked with smaller loss data, available in the databases of primary insurers or alternatively in market statistics – even grouped data should be sufficient for this.

Summing up, the conditions of the theorem are such that it can be verified involving quite diverse data sources: (NC1) affects the tail distribution only. (NC2) relates the overall loss frequency with that at the threshold, involving nothing more about the shape of the distribution. (SC*) connects the tail parameters with the body distribution, however, from the latter only the average loss is needed, not the exact shape.

In order to gather more intuition about the smaller losses, let now be $r>\alpha$, the only case leaving room for speculation about how these are distributed and which parametric models could be applied. The average smaller loss $\gamma=\theta\frac{\alpha}{1-\alpha}\frac{1-r}{r}$ can take on any value between 0 and θ. Looking at the extremes one gets:

Border case 1: If r is close to α, γ is close to θ, yielding a function having most small losses concentrated just to the left of the threshold and leaving much less probability over for smaller losses. Although it is in principle possible to find very smooth functions fulfilling this condition, the distribution will be anyway similar to the case $r=\alpha$ with its mass point at the threshold. Such a distribution may be difficult to model with well-established distributions, though.
Border case 2: If r is closer to 1, γ is smaller, leaving more options for nice and well-known functions to be applied. However, very large r means that almost all losses are below the threshold θ, which in practice will not work with low thresholds. On the other hand, low values θ are more appealing as this extends the range of application of the Riebesell rule.

There is a trade-off in this. Riebesell functions having a wide interval of validity of the Riebesell rule will have to pay with a somewhat uneven shape of the body distribution.

To conclude we give two examples of how the body distribution of a Riebesell function could look like, writing down as usual the survival function.

Example 1: If we interpret the constant random variable as a distribution having a scale parameter, being the mass point of probability one, we get the following spliced model, being a straightforward generalization of the case $r=\alpha$:

Const-Par 1 for $0\le x<\gamma$, $1-r$ for $\gamma\le x<\theta$, and for $\theta\le x$ $(1-r)\left(\frac{\theta}{x}\right)^{\alpha}$

(This model is indeed continuous at the splicing point θ, however, not so at $\gamma<\theta$.)

The LEV is a continuous function, being piecewise linear in the body area. We have

$LEV(x)=x$ for $0\le x\le\gamma$ and $LEV(x)=r\gamma+(1-r)x$ for $\gamma\le x\le\theta$.

Example 2: To get an at least continuous survival function we try a power curve body as introduced above.

Pow-Par-0: $1-r\left(\frac{x}{\theta}\right)^{\beta}$ for 0≤x<θ, for θ≤x $(1-r)\left(\frac{\theta}{x}\right)^{\alpha}$

The conditioned body function has the survival function $1-\left(\frac{x}{\theta}\right)^{\beta}$ with expected value $\theta\frac{\beta}{\beta+1}$.

Then (SC*) reads $\frac{\beta}{\beta+1}=\frac{\alpha}{1-\alpha}\frac{1-r}{r}$, yielding $\beta=\alpha\frac{1-r}{r-\alpha}$.

This value is always positive, thus for any parameter constellation 0<α<r<1 there is a (unique) power curve creating a continuous Riebesell distribution. This is a model in three parameters α, r, θ. Notice that the exponent β has an intuitive interpretation: The ratio of β and α equals the ratio of the distances of r from 1 and α, respectively. If r tends to α, β becomes very large, having as limiting case the discontinuous function of the r=α case.

The Riebesell Pow-Par function is C0 but not C1: One quickly gets $f(\theta-)=r\frac{\beta}{\theta}=r\frac{\alpha}{\theta}\frac{1-r}{r-\alpha}$.

Recall $f(\theta+)=(1-r)\frac{\alpha}{\theta}$. The former divided by the latter yields $\frac{r}{r-\alpha}$, which is greater (in practice often much greater) than 1. With the pdf being larger just before the discontinuity at θ than just thereafter it is clear that the area where the losses are overall most concentrated is the left neighborhood of θ.

The LEV is a C1 function, in the area 0≤x≤θ equaling $x-\theta\frac{r}{\beta+1}\left(\frac{x}{\theta}\right)^{\beta+1}=x-\theta\frac{r-\alpha}{1-\alpha}\left(\frac{x}{\theta}\right)^{r\frac{1-\alpha}{r-\alpha}}$.

8 Conclusion – reinventing Pareto

In this paper we looked at the Pareto and the GPD distribution in a particular way, interpreting them essentially as tails of comprehensive distributions modeling losses of any size. This makes the Pareto family much larger, yielding *tail models*, *excess models*, and notably *ground-up models*. Among the latter we drew the attention to the rich group of continuous spliced models with GPD tail, offering a great deal of flexibility while at the same time all being comparable among each other in terms of tail-behavior via the threshold-invariant GPD parameters α and λ.

We developed a framework ordering all spliced distributions being constructed out of the same model for the body of small/medium losses, according to three criteria: *tail shape* (Pareto or not), *body shape* (original or distorted), and *smoothness*. This yields a hierarchy (more precisely a three-dimensional grid) of distributions having a decreasing number of parameters, with a C0 model on the top.

Among a number of practical applications we in particular revisited the traditional Riebesell (or power curve) model for the exposure rating of liability business. We specified and illustrated the necessary and sufficient conditions leading to this model, constructing finally as a special case a spliced C0 PowerFunction-Pareto model.

We hope to have inspired the reader to share the view on the Pareto world outlined here, and to apply some of the presented models (data permitting).

References

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

Embrechts P, Klüppelberg C, Mikosch T (1997) *Modelling extremal events for insurance and finance.* Springer, Berlin Heidelberg.

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

Knecht, M & Küttel S (2003) The Czeledin distribution function. *ASTIN Colloquium Berlin.*

Kozubowski T J & Podgórski K (2003) Log-Laplace distributions. *Internat. Math. J. 3*, 467–495.

Mack T (1980) Über die Aufteilung des Risikos bei Vereinbarung einer Franchise. *Blätter der DGVFM 14:4*, 631–650.

Mack T & Fackler M (2003) Exposure-Rating in liability reinsurance. *Blätter der DGVFM 26:2*, 229–238.

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

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

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

Scollnik D P A & Sun C (2012) Modeling with Weibull-Pareto models. *NAAJ 16:1,* 260–272.

Teodorescu, S & Vernic R (2009) Some composite Exponential-Pareto models for actuarial prediction. *Romanian J. Econom. Forecasting 12:4*, 82–100.