Pareto models for risk management

Arthur Charpentier, Emmanuel Flachaire

HAL Id: hal-02423805

https://hal.archives-ouvertes.fr/hal-02423805

Pareto models for risk management

by

Arthur Charpentier

Université du Québec à Montréal (UQAM)
201, avenue du Président-Kennedy,
Montréal (Québec), Canada H2X 3Y7
arthur.charpentier@uqam.ca

and

Emmanuel Flachaire

Aix-Marseille Université
AMSE, CNRS and EHESS,
5 bd Maurice Bourdet,
13001, Marseille, France,
emmanuel.flachaire@univ-amu.fr

December 2019

Abstract

The Pareto model is very popular in risk management, since simple analytical formulas can be derived for financial downside risk measures (Value-at-Risk, Expected Shortfall) or reinsurance premiums and related quantities (Large Claim Index, Return Period). Nevertheless, in practice, distributions are (strictly) Pareto only in the tails, above (possible very) large threshold. Therefore, it could be interesting to take into account second order behavior to provide a better fit. In this article, we present how to go from a strict Pareto model to Pareto-type distributions. We discuss inference, and derive formulas for various measures and indices, and finally provide applications on insurance losses and financial risks.

JEL: C13; C18; C46; G22; G32

Keywords: EPD; Expected Shortfall; Financial Risks; GPD; Hill; Pareto; Quantile; Rare events; Regular Variation; Reinsurance; Second Order; Value-at-Risk;

1 Introduction

The Gaussian distribution is one of the most popular distribution used initially to model observation errors. It can be used to model individual measurements that are (somehow) centered, such as the height, or the weight of individuals. But not all data exhibit such a property, where the distribution has a peak around a typical value (the median or the mean), such as the distribution of the population among cities, or the distribution of wealth among individuals. Those distribution are known to be right-skewed, where most observations are bulked with small values, a small proportion can reach much higher values than the median (or the mean), leading to long tail on the right of the distribution. For instance, Pareto (1895) introduced the following model for the distribution of income: let $p(x)dx$ denote the proportion of individuals with income between x and $x+dx$; if the histogram is a straight line on the log-log scales, then $\log p(x) = a - b\log(x)$ (since the slope is clearly always negative), or equivalently, $p(x) = cx^{-b}$. Such distributions are said to have a *power law*, or a Pareto distribution since it was introduced by the economist Vilfredo Pareto (see Charpentier and Flachaire (2019) for applications to model income distributions). The constant b might be called the exponent of the power law (especially in the context of networks or in physical applications), but in economic application, the exponent will be the one of the survival distribution $S(x) = \int_x^\infty p(y)dy$, which is also a power function, with index $1 + b$.

The Gaussian distribution is a natural candidate when focusing on the center of the distribution, on average values, because of the central limit theorem: the Gaussian distribution is stable by summing, and averaging. If $\{X_1, \cdots, X_n\}$ are independent Gaussian variables, so is $X_1 + \cdots + X_n$ (and any linear transformation). As mentioned in Jessen and Mikosch (2006) and Gabaix (2009), the power distribution satisfies similar interesting properties: if X_1 is a power distribution with exponent b_1, independent of X_2, another power distribution with exponent b_2, with $b_2 \geq b_1$, then $X_1 + X_2$, $X_1 \cdot X_2$, or $\max\{X_1, X_2\}$ are also power distributed, with exponent b_1. And those properties of invariance and stability are essential in stochastic modeling. As mentioned in Schumpeter (1949), in an obituary article about Vilfredo Pareto, "*few if any economists seem to have realized the possibilities that such invariants hold for the future of our science. In particular, nobody seems to have realized that the hunt for, and the interpretation of, invariants of this type might lay the foundations for an entirely novel type of theory*". If several applications can be found on proportional (random) growth theory (as discussed in Gabaix (2009), with connections to recursive equations), or

on matching and networks, an important application of this Pareto model is risk management.

Insurers and actuaries have used Pareto distribution to model large losses since Hagstroem (1925, 1960). Balkema and de Haan (1974) suggested to use such a distribution in a life-insurance context (to model the remaining lifetimes for very old people), but the Pareto distribution has mainly been used to model (large) insurance losses since reinsurance premiums have simple and analytical expression, as in Vajda (1951), even for more complex treaties than standard stop-loss or excess-of-loss treaties as in Kremer (1984). Beirlant and Teugels (1992), McNeil (1997) (discussed in Resnick (1997)) and Embrechts, Klüppelberg, and Mikosch (1997) discussed more applications in insurance, with connections to extreme value theory.

The Value-at-Risk has been the finance benchmark risk measure for the past 20 years. Estimating a quantile for small probabilities, such as the 1%-quantile of the profit and loss distribution for the next 10 days. Historically, practitioners used a standard Gaussian model, and then multiply by a factor of 3 (as explained in Klüppelberg (2004), this factor 3 is supposed to account for certain observed effects, also due to the model risk; it is based on backtesting procedures and can be increased by the regulatory authorities, if the backtesting proves the factor 3 to be insufficient). A less *ad hoc* strategy is to use a distribution that fits better tails of profit and loss distribution, such as the Pareto one.

Nevertheless, if the Pareto distribution remains extremely popular because several quantities can easily be derived (risk measures or insurance premium), in practice, distributions are only Pareto in very high tails (say above the 99.9% quantile). So it becomes necessary to take into account second order approximation, to provide a better fit. Thus, we will see in this article how to derive risk measures and insurance premiums when distribution are Pareto-type above a (high) threshold u.

In Section 2, we will define the two popular Pareto models, the strict Pareto distribution and the Generalized Pareto distribution (GPD), emphasizing differences between the two. A natural extension will be related to the concept of Pareto-type distribution, using regular variation function. In Section 3, we will get back on the concepts of regular variation, and present the Extended Pareto distribution (EPD), introduced in Beirlant et al. (2009). The inference of Pareto models will be discussed in Section 4, with a discussion about the use of Hill estimator, still very popular in risk management, and Maximum Likelihood estimation. In Section 5 we will define classical measures and indices, and discuss expression and properties when underlying distribution are either strict Pareto, or Pareto-type, with high quantiles

(Q), expected shortfall (ES), and large claim or top share indices (TS). And as Pareto models are usually assumed above a high threshold u we will see, at the end of that section, how analytical expressions of those measures can be derived when Pareto-type distributions are considered only above threshold u (and define Q_u, ES_u, etc). Finally, applications on real data will be considered, with insurance and reinsurance pricing in Section 6 and log-returns and classical financial risk measures in Section 7.

2 Strict Pareto models

Pareto models are obtained when the loss distribution $\mathbb{P}[X > x]$ exhibits a power decay $x^{-\alpha}$. From a mathematical perspective, such a power function is interesting because it is the general solution of Cauchy's functional equation (its multiplicative version), $h(x \cdot y) = h(x) \cdot h(y)$, $\forall x, y$. A probabilistic interpretation is the stability of the conditional distribution X given $X > u$ for any threshold u: the conditional distribution still exhibits a power decay, with the same index. But if we try to formalize more, two distributions will be obtained: a Pareto I distribution will be obtained when modelling the distribution of *relative excesses* X/u given $X > u$ - see Equation (1) - while the GPD (Generalized Pareto Distribution) will be obtained when modelling the distribution of *absolute excesses* $X - u$ given $X > u$ - see Equation (6).

2.1 Pareto I distribution

A Pareto Type I distribution, bounded from below by $u > 0$, with tail parameter α, has probability density function and cumulative density function (CDF) equal to

$$f(x) = \frac{\alpha u^\alpha}{x^{\alpha+1}} \qquad \text{and} \qquad F(x) = 1 - \left(\frac{x}{u}\right)^{-\alpha}, \qquad \text{for } x \geq u \qquad (1)$$

If a random variable X has distribution (1), we will write $X \sim \mathcal{P}_1(u, \alpha)$. This distribution has an attractive property: the average above a threshold is proportional to the threshold, and it does not depend on the scale parameter u,

$$\mathbb{E}(X|X > u') = \frac{\alpha u'}{\alpha - 1}, \qquad \alpha > 1 \qquad (2)$$

where $u' \geq u$. Such a function is related to the *mean excess function* – or *expected remaining lifetime* at age u' when X denotes the (random) life length – defined as

$$e(u') = \mathbb{E}(X - u'|X > u'). \qquad (3)$$

In the case of a Pareto distribution, it is also a linear function

$$e(u') = \frac{u'}{\alpha - 1}, \qquad \alpha > 1 \tag{4}$$

Remark 1 *In several textbooks and articles, Pareto models are defined with tail index $\xi = 1/\alpha$, so that the survival function is proportional to the power function $x^{-1/\xi}$, instead of $x^{-\alpha}$. In that case, we have the confusing expressions*

$$\mathbb{E}(X|X > u') = \frac{u'}{1 - \xi} \quad and \quad e(u') = \frac{\xi u'}{1 - \xi}, \qquad \xi < 1. \tag{5}$$

2.2   Generalized Pareto distribution

A Generalized Pareto Distribution (GPD), bounded from below by $u \geq 0$, with scale parameter σ and tail parameter α, has cumulative density function (CDF)

$$F(x) = 1 - \left[1 + \left(\frac{x - u}{\sigma}\right)\right]^{-\alpha} \qquad \text{for } x \geq u \tag{6}$$

where $\sigma > 0$ and $\alpha \in (0, \infty]$. If a random variable X has distribution (6), we will write $X \sim \mathcal{GPD}(u, \sigma, \alpha)$. The $\mathcal{GPD}(0, \sigma, \alpha)$ distribution is also called "Lomax" in the literature Lomax (1954). The GPD is "general" in the sense that Pareto I distribution is a special case, when $\sigma = u$:

$$\mathcal{GPD}(u, u, \alpha) = \mathcal{P}_1(u, \alpha) \tag{7}$$

Scollnik (2007) suggested an alternative expression, instead of (6),

$$F(x) = 1 - \left[\frac{u + \lambda}{x + \lambda}\right]^{\alpha} \qquad \text{for } x \geq u \tag{8}$$

where $\lambda = \sigma - u$. Note that this function is used in Rigby and Stasinopoulos (2005) to introduce a regression type model, where both λ and α will be the exponential of linear combinations of some covariates, as in the `gamlss` package, in `R`.

Remark 2 *In several textbooks and articles, not only the Generalized Pareto has tail index $\xi = 1/\alpha$, but ξ appears also in the fraction term*

$$F(x) = 1 - \left[1 + \xi\left(\frac{x - u}{\sigma}\right)\right]^{-1/\xi} \qquad for \ x \geq u, \tag{9}$$

which could be seen as a simple alternative expression of the shape parameter σ. This expression is interesting since the limit when ξ tends to 0 can easily be defined, as well as the case where $\xi < 0$, which could be interesting in extreme value theory.

The average above a higher threshold, $u' \geq u$, depends now on all parameters of the distribution,

$$\mathbb{E}(X|X > u') = \frac{\sigma - u}{\alpha - 1} + \frac{\alpha}{\alpha - 1}u', \tag{10}$$

The linearity of this function (as well as the mean excess function) characterizes the GPD class (see Guess and Proschan (1988) and Ghosh and Resnick (2010)). A sketch of the proof is that we want

$$e(u') = \mathbb{E}(X|X > u') = \frac{1}{\overline{F}(u')} \int_{u'}^{\infty} x dF(x) = A + Bu', \tag{11}$$

or, when differentiating with respect to u', we obtain

$$B\overline{F}(u')du' = s - \big[A + (B-1)u'\big]d\overline{F}(u'), \tag{12}$$

thus, up to some change of parameters

$$\frac{d\overline{F}}{\overline{F}} = -\alpha \frac{du}{u + c} \quad \text{with} \quad \alpha = \frac{B}{B - 1} \tag{13}$$

which is a GPD distribution, with tail index α.

One of the most important result in the extreme value theory states that, for most heavy-tailed distributions, the conditional excess distribution function, above a threshold u, converges towards a GPD distribution as u goes to infinity (Pickands (1975), Balkema and de Haan (1974)), for some parameters α and σ,

$$F_u(x) \longrightarrow \text{ GPD (or Pareto II)} \quad \text{as } u \to +\infty \tag{14}$$

where $F_u(x) = P(X - u \leq x|X > u)$. This result is known as the Pickands-Balkema-de Haan theorem, also called the second theorem in extreme value theory (as discussed in footnote 2, page 7). It provides strong theoretical support for modelling the upper tail of heavy-tailed distributions with GPD, also known as the Pareto Type II distribution. In a very general setting, it means that there are α and σ such that F_u can be approximated by the CDF of a $\mathcal{GPD}(u, \sigma, \alpha)$, see Embrechts et al. (1997, Theorem 3.4.13, p.165).

2.3 Threshold selection

Whether to use a Pareto I or GPD model to fit the upper tail is related to the choice of the threshold. From (1) and (6), we have seen that Pareto I is a special case of GPD, when $\sigma = u$. They differ by an affine transformation when $\sigma \neq u$. A key property of Pareto distributions is that, if a distribution is Pareto for a fixed threshold u, it is also Pareto with the same tail parameter α for a higher threshold $u' \geq u$. For GPD, we have

$$\overline{F}_u = \mathcal{GPD}(u, \sigma, \alpha) \quad \Rightarrow \quad \overline{F}_{u'} = \mathcal{GPD}(u', \sigma + u' - u, \alpha). \qquad (15)$$

where $\overline{F}_u$ is the survival excess function above u.[1] Thus, Pareto I and GPD are the same for all $u' \geq u$, if $\sigma = u$. Otherwise, we have $\sigma + u' - u \approx u'$ for very large values of u' only. It follows that a GPD above a threshold will behave approximately as a Pareto I above a *higher* threshold, much higher as σ differs from u. This point will be discussed further in section 5.5.

3 Pareto-type models

Since our interesting in risk management is usually motivated by the description of so-called *downside risks*, or (important) losses, as described in Roy (1952). Thus, we do not need that have a (strict) Pareto distribution for all x's, in $\mathbb{P}[X > x]$, but possibly only when x's are large enough. In this section, we will introduce some regular variation concepts, used to model that asymptotic property, and exhibit a distribution that is not strictly Pareto, by introducing a second order effect: the *Extended Pareto Distribution*.

3.1 First- and second-order regular variation

The tail index α is related to the max-domain of attraction of the underlying distribution, while parameter σ is simply a scaling parameter.[2] The shape of

---

[1]$\overline{F}_{u'}$ is a truncated Pareto distribution, with density equals to $f(x)/(1 - F(u'))$. This property can be observed directly using Equation (8), where both α and λ remain unchanged.

Note that this property is quite intuitive, since the GPD distribution appears as a limit for exceeding distributions, and limit in asymptotic results are always fixed points: the Gaussian family is stable by addition (and appears in the Central Limit Theorem) while Fréchet distribution is max-stable (and appears in the first theorem in extreme value theory).

[2]Historically, extremes were studied through block-maximum - yearly maximum, or maximum of a subgroup of observations. Following Fisher and Tippett (1928), up to some affine transformation, the limiting distribution of the maximum over n i.i.d observations

the conditional excess cumulative distribution function is a power function (the Pareto distribution) if the threshold is large enough. Tails are then said to be Pareto-type, and can be described using so called *regularly varying* functions (see Bingham et al. 1987).

First and second order regular variation were originally used in extreme value theory, to study respectively the tail behavior of a distribution and the speed of convergence of the extreme value condition (see Bingham et al. 1987, de Haan and Stadtmüller 1996, Peng and Qi 2004, or section 2 in de Haan and Ferreira 2006 for a complete survey). A function H is said to be regularly varying (at infinity) with index $\gamma \in \mathbb{R}$ if

$$\lim_{t\to\infty} \frac{H(tx)}{H(t)} = x^\gamma \quad \text{or} \quad \lim_{t\to\infty} x^{-\gamma}\frac{H(tx)}{H(t)} = 1. \tag{16}$$

A function regularly varying with index $\gamma = 0$ is said to be slowly varying. Observe that any regularly varying function of index $-\gamma$ can be written $H(x) = x^{-\gamma}\ell(x)$ where ℓ is some slowly varying function.

Consider a random variable X, its distribution is regularly-varying with index $-\gamma$ if, up-to some affine transformation, its survival function is regularly varying. Hence,

$$\lim_{t\to\infty} x^{-\gamma}\frac{\overline{F}(tx)}{\overline{F}(t)} = 1 \tag{17}$$

or

$$\overline{F}(x) = x^{-\gamma}\ell(x), \tag{18}$$

where $\overline{F}(x) = 1 - F(x)$. A regularly varying survival function is then a function that behaves like a power law function near infinity. Distributions with survival function as defined in (18) are called *Pareto-type distributions*. It means that the survival function tends to zero at polynomial (or power) speed as $x \to \infty$, that is, as $x^{-\gamma}$. For instance, a Pareto I distribution, with survival function $\overline{F}(x) = x^{-\alpha}u^\alpha$, is regularly varying with index $-\alpha$, and the associated slowly varying function is the constant u^α. And a GPD or Pareto II distribution, with survival function $\overline{F}(x) = \left(1 + \sigma^{-1}x\right)^{-\alpha}$, is

---

is either Weibull (observations with a bounded support), Gumbel (infinite support, but light tails, like the exponential distribution) or Fréchet (unbounded, with heavy tails, like Pareto distribution). Pickands (1975) and Balkema and de Haan (1974) obtained further that not only the only possible limiting conditional excess distribution is GPD, but also that the distribution of the maximum on subsamples (of same size) should be Fréchet distributed, with the same tail index γ, if $\gamma > 0$. For instance in the U.S., if the distribution of maximum income per county is Fréchet with parameter γ (and if county had identical sizes), then the conditional excess distribution function of incomes above a high threshold is a GPD distribution with the same tail index γ.

regularly varying with index $-\alpha$, for some slowly varying function. But in a general setting, if the distribution is not *strictly* Pareto, ℓ will not be constant, and it will impact the speed of convergence.

In de Haan and Stadtmüller (1996), a concept of second-order regular variation function is introduced, that can be used to derive a probabilistic property using the quantile function,[3] as in Beirlant et al. (2004). Following Beirlant et al. (2009), we will consider distributions such that an extended version of equation (17) is satisfied,

$$\lim_{t \to \infty} x^{-\gamma} \frac{\overline{F}(tx)}{\overline{F}(t)} = 1 + \frac{x^\rho - 1}{\rho}, \text{ for some } \rho \leq 0, \tag{19}$$

that we can write, up to some affine transformation,

$$\overline{F}(x) = cx^{-\gamma}[1 - x^\rho \ell(x)], \tag{20}$$

for some slowly varying function ℓ and some second-order tail coefficient $\rho \leq 0$. The corresponding class of Pareto-type distributions defined in (20) is often named the Hall class of distributions, referring to Hall (1982). It includes the Singh-Maddala (Burr), Student, Fréchet and Cauchy distributions. A mixture of two strict Pareto-I distributions will also belong to this class.

Since $\rho \leq 0$, $x \mapsto 1 - x^\rho \ell(x)$ is slowly varying, and therefore, a distribution $\overline{F}$ that satisfies (20) also satisfies (18). More specifically, in (20), the parameter ρ captures the rate of convergence of the survival function to a strict Pareto distribution. Smaller is ρ, faster the upper tail behaves like a Pareto, as x increases. Overall, we can see that

- γ is the first-order of the regular variation, it measures the tail parameter of the Pareto distribution,

- ρ is the second-order of the regular variation, it measures how much the upper tail deviates from a strictly Pareto distribution.

In the following, we will write $\mathrm{RV}(-\gamma, \rho)$. There are connections between tail properties of the survival function $\overline{F}$ and the density f (see also Karamata theory for first order regular variation). More specifically, if $\overline{F}$ is $\mathrm{RV}(-\gamma, \rho)$, with $\gamma > 1$, then f is $\mathrm{RV}(-\gamma - 1, \rho)$.

For instance, consider a Singh-Maddala (Burr) distribution, with survival distribution $\overline{F}(x) = [1 + x^a]^{-q}$, then a second order expansion yields

$$\overline{F}(x) = x^{-aq}[1 - qx^{-a} + o(x^{-a})] \text{ as } x \to \infty \tag{21}$$

---

[3]The quantile function U is defined as $U(x) = F^{-1}(1 - 1/x)$.

which is regularly varying of order $-aq$ and with second order regular variation $-a$, that is $\mathrm{RV}(-aq, -a)$.

3.2 Extended Pareto distribution

Beirlant et al. (2009) show that Equation (20) can be approximated by

$$\overline{F}(x) = [x(1 + \delta - \delta x^\tau)]^{-\alpha} \qquad \text{for } x \geq 1 \tag{22}$$

where $\tau \leq 0$ and $\delta > \max(-1, 1/\tau)$.[4] The main feature of this function is that it captures the second-order regular variation of the Hall class of distributions, that is, deviation to a strictly Pareto tail, as defined in Hall (1982),

$$\overline{F}(x) = ax^{-\alpha}\left[1 + bx^{-\beta} + o(x^{-\beta})\right] \tag{23}$$

or, using Taylor's expansion,

$$\overline{F}(x) = ax^{-\alpha}\left[1 + b_1 x^{-\alpha} + \cdots + b_k x^{-k\alpha} + o(x^{-k\alpha})\right] \tag{24}$$

with general notations. For more details, see also Albrecher et al. (2017, section 4.2.1).

From (22), we can define the Extended Pareto Distribution (EPD), proposed by Beirlant et al. (2009), as follows:

$$F(x) = 1 - \left[\frac{x}{u}\left(1 + \delta - \delta\left(\frac{x}{u}\right)^\tau\right)\right]^{-\alpha} \qquad \text{for } x \geq u \tag{25}$$

where $\tau \leq 0$ and $\delta > \max(-1, 1/\tau)$. If a random variable X has (25) as its CDF, we will write $X \sim \mathcal{EPD}(u, \delta, \tau, \alpha)$.

Pareto I is a special case when $\delta = 0$ and GPD is a special case when $\tau = -1$:

$$\mathcal{EPD}(u, 0, \tau, \alpha) = \mathcal{P}_1(u, \alpha) \tag{26}$$
$$\mathcal{EPD}(u, \delta, -1, \alpha) = \mathcal{GPD}(1, u/(1+\delta), \alpha) \tag{27}$$

The mean over a threshold for the EPD distribution has no closed form expression.[5] Numerical methods can be used to calculate it. Since $u' \geq u > 0$, X given $X > u'$ is a positive random variable and

$$\mathbb{E}[X|X > u'] = \int_0^\infty \overline{F}_{u'}(x)dx \tag{28}$$

---

[4] Using the expansion $(1 + y^a)^b \approx 1 + by^a$, for small y^a, in (22) yields (20).

[5] Albrecher et al. (2017, section 4.6) give an approximation, based on $(1 + \delta - \delta y^\tau)^{-\alpha} \approx 1 - \alpha\delta + \alpha\delta y^\tau$, which can be very poor. Thus, we do not recommend to use it.

10

where $\overline{F}_{u'}(x) = \mathbb{P}[X > x | X > u']$ for $x > u$, i.e.

$$\overline{F}_{u'}(x) = \frac{\overline{F}(x)}{\overline{F}(u')} \quad \text{where} \quad \overline{F} \text{ is the s.d.f. of } X \tag{29}$$

Thus

$$\mathbb{E}[X | X > u'] = u' + \frac{1}{\overline{F}(u')} \int_{u'}^{\infty} \overline{F}(x)dx \tag{30}$$

The integral in Equation (30) can be computed numerically. Since numerical integration over a finite segment could be easier, we can make a change of variable $(1/x)$ to obtain an integral over a finite interval:

$$E_{u'} = \int_{u'}^{\infty} \overline{F}(x)dx = \int_{0}^{1/u'} \frac{1}{x^2} \overline{F}\left(\frac{1}{x}\right) dx \tag{31}$$

The Extended Pareto distribution has a stable tail property: if a distribution is EPD for a fixed threshold u, it is also EPD for a higher threshold $u' \geq u$, with the same tail parameter α. Indeed, deriving a truncated EPD distribution, we find

$$\overline{F}_u = \mathcal{EPD}(u, \delta, \tau, \alpha) \quad \Rightarrow \quad \overline{F}_{u'} = \mathcal{EPD}(u', \delta', \tau, \alpha). \tag{32}$$

where $\delta' = \delta(u'/u)^\tau / [1 + \delta - \delta(u'/u)^\tau]$. A plot of estimates of the tail index α for several thresholds would then be useful. If the distribution is Extended Pareto, a stable horizontal straight line should be plotted. This plot is similar to Hill plot for Hill estimates of α from Pareto I distribution. It is expected to be more stable if the distribution is not strictly Pareto. Indeed, Embrechts et al. (1997, p.194) and Resnick (2007, p.87) illustrate that the Hill estimator can perform very poorly if the slowly varying function is not constant in (18). It can lead to very volatile Hill plots, also known as Hill *horror* plots. This will be discussed in section 4.1.

4   Inference based on Pareto type distributions

In this section, we briefly recall two general techniques used to estimate the tail index, and additional parameters: Hill estimator and maximum likelihood techniques. For further details, see Embrechts et al. (1997), de Haan and Ferreira (2006) or Beirlant et al. (2004), and references therein.

4.1 Hill's Estimator

In the case where the distribution above a threshold u is supposed to be $\mathcal{P}(u, \alpha)$, Hill's estimator of α is

$$\widehat{\alpha}_u = \left(\frac{1}{n_u} \sum_{i=1}^{n_u} \log(x_{n+1-i:n}) - \log(u) \right)^{-1} \tag{33}$$

This estimator is the maximum likelihood estimator of α.

From Theorem 6.4.6 in Embrechts et al. (1997) and section 4.2 in Beirlant et al. (2004), if we assume that the distribution of X is Pareto-type, with index α, $\overline{F}(x) = x^{-\alpha}\ell(x)$, if $u \to \infty$, $n \to \infty$ and $n_u/n \to 0$, then $\widehat{\alpha}_u$ is a (strongly) consistent estimator of α when observations are i.i.d., and further, under some (standard) technical assumptions,

$$\sqrt{n_u}(\widehat{\alpha}_u - \alpha) \overset{\mathcal{L}}{\to} \mathcal{N}(0, \alpha^2). \tag{34}$$

This expression can be used to derive confidence interval for α, but also any index mentioned in the previous section (quantile, expected shortfall, top shares, etc) using the Δ-method. For instance

$$\frac{\sqrt{n_u}Q(1-p)}{\sqrt{1 + (\log[n_u/n] - \log[p])^2}} \left(\widehat{Q}_u(1-p) - Q(1-p) \right) \overset{\mathcal{L}}{\to} \mathcal{N}(0, \alpha^{-2}), \tag{35}$$

as shown in Section 4.6. of Beirlant et al. (2004).

Nevertheless, as discussed in Embrechts et al. (1997) and Beirlant et al. (2004), this asymptotic normality is obtained under some appropriate choice of u (as a function of n) and ℓ : u should tend to infinity sufficiently slowly, otherwise, $\widehat{\alpha}_u$ could be a biased estimator, even asymptotically. More specifically, in section 3.1, we introduced second order regular variation: in Equation (20), assume that $\ell(x) = -1$, so that

$$\overline{F}(x) = cx^{-\alpha}[1 + x^{\rho}] \tag{36}$$

From Theorem 6.4.9 in Embrechts et al. (1997), if $n_u = o(n^{(2\rho)/(2\rho+\alpha)})$, then the asymptotic convergence of Equation (35) is valid. But if $n^{-(2\rho)/(2\rho+\alpha)}n_u$ tends to λ as $n \to \infty$, then

$$\sqrt{n_u}(\widehat{\alpha}_u - \alpha) \overset{\mathcal{L}}{\to} \mathcal{N}\left(\frac{\alpha^3\lambda}{\rho - \alpha}, \alpha^2 \right), \tag{37}$$

has an asymptotic bias. Such a property makes Hill estimator dangerous to use for Pareto-type distributions.

4.2 Maximum Likelihood Estimator

For the GPD, a popular is to use the maximum likelihood estimator. When threshold u is given, the idea is to fit a $\mathcal{GPD}(0, \sigma, \alpha)$ distribution on the sample $\{y_1, \cdots, y_{n_u}\} = \{x_{n+1-n_u:n} - u, \cdots, x_{n:n} - u\}$. The density being

$$f(y; \sigma, \alpha) = \frac{\alpha}{\sigma} \left(1 + \frac{y}{\sigma}\right)^{-\alpha-1} \tag{38}$$

the log-likelihood is here

$$\log \mathcal{L}(\sigma, \alpha; \boldsymbol{y}) = -n \log \frac{\alpha}{\sigma} - (\alpha + 1) \sum_{i=1}^{n_u} \left(1 + \frac{y_i}{\sigma}\right). \tag{39}$$

As expected, the maximum likelihood estimator of (σ, α) is asymptotically Gaussian, as shown in Section 6.5 of Embrechts et al. (1997) (with a different parameterization of the GPD). In the case of the Extended Pareto distribution, there might be numerical complications, but (Beirlant, Joossens, and Segers 2009) provided theoretical properties, and Reynkens (2018) provided R codes, used in Albrecher et al. (2017).

Note that since α is usually seen as the most important parameter (and σ is more a nuisance parameter) it can be interesting to use profile likelihood techniques to derive some confidence interval, as discussed in Section 4.5.2 in Davison (2003). Consider some Pareto type model, with parameter $(\alpha, \boldsymbol{\theta})$. Let

$$(\widehat{\alpha}, \widehat{\boldsymbol{\theta}}) = \underset{\alpha, \boldsymbol{\theta}}{\operatorname{argmax}} \left\{\log \mathcal{L}(\alpha, \boldsymbol{\theta})\right\} \tag{40}$$

denote the maximum likelihood estimator. From the likelihood ratio test, under technical assumptions,

$$2\big(\log \mathcal{L}(\widehat{\alpha}, \widehat{\boldsymbol{\theta}}) - \log \mathcal{L}(\alpha, \boldsymbol{\theta})\big) \to \chi^2\big(\dim(\alpha, \boldsymbol{\theta})\big)$$

where $I_n(\alpha, \boldsymbol{\theta})$ is Fisher information. The idea of the profile likelihood estimator is to define, given α

$$\widehat{\boldsymbol{\theta}}_\alpha = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left\{\log \mathcal{L}(\alpha, \boldsymbol{\theta})\right\} \tag{41}$$

and then

$$\widehat{\alpha}_p = \underset{\alpha}{\operatorname{argmax}} \left\{\log \mathcal{L}(\alpha, \boldsymbol{\theta}_\alpha)\right\} \tag{42}$$

Then

$$2\big(\log \mathcal{L}(\widehat{\alpha}_p, \widehat{\boldsymbol{\theta}}_{\widehat{\alpha}_p}) - \log \mathcal{L}(\alpha, \boldsymbol{\theta})\big) \to \chi^2(1).$$

Thus, if $\log \mathcal{L}_p(\alpha)$ denotes the profile likelihood, defined as $\log \mathcal{L}_p(\alpha) = \log \mathcal{L}(\alpha, \widehat{\boldsymbol{\theta}}_\alpha)$, then a 95% confidence interval can be obtained,

$$\left\{ \alpha : \log \mathcal{L}_p(\alpha) \geq \log \mathcal{L}_p(\widehat{\alpha}_p) - \frac{1}{2} q_{\chi^2(1)}(95\%) \right\} \tag{43}$$

where $q_{\chi^2(1)}(95\%)$ is the 95% quantile of the $\chi^2(1)$ distribution.

4.3 Application on Simulated Data

In this section, three distributions where simulated: a strict Pareto on Figure 1 with tail index $\alpha = 1.5$, a Generalized Pareto on Figure 2 with tail index $\alpha = 1.5$, and an Extented Pareto on Figure 3 with tail index $\alpha = 1.5$. Each time, on the left, the Hill plot is plotted, i.e. the plot $u \mapsto \widehat{\alpha}_u$. The dotted lines are bounds of the confidence interval (at level 95%). The vertical segment (—) is the confidence interval for α when u is the 80%-quantile of the sample. On the right, the profile likelihood of a GPD distribution is plotted, including the horizontal line (—) that defines the 95%-confidence interval.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

In the case of the EPD, the smaller the value of τ, the more bias Hill estimator has (see Charpentier and Flachaire (2019) for a detailed discussion). When generating a Pareto type distribution, using the top 80% observations does not guarantee anymore that Hill estimator could be a relevant estimator. On Figure 4, we used simulated samples of size $n = 1,000$ with the same Monte Carlo seed (on the left), and another one (on the right), as on Figure 3, and we simply change the value of τ. The red curve (—) is Hill estimator $\widehat{\alpha}_u$ obtained when u is the 60% quantile, while the blue curve (—) is obtained when u is the 90% quantile. The green curve (—) is the profile likelihood estimator of a GPD distribution. Horizontal segments are confidence intervals. For Hill estimator, observe that if the threshold u is not large enough, estimator can be substantially biased. For the GPD, the fit is better, but the confidence interval is very large. Observe further that when $\tau \in (-2, 0)$, Hill estimator in the higher tail can actually be overestimated.

[Figure 4 about here.]

14

5 Modeling large events

Given n i.i.d. random losses $X_1, \cdots, X_n$, let S_n and M_n respectively denote the sum and the maximum

$$S_n = X_1 + \cdots + X_n \quad \text{and} \quad M_n = \max\{X_1, \cdots, X_n\}. \tag{44}$$

A classical concept is the *Probable Maximum Loss* (section 5.1) which is the distribution of M_n, or an affine transformation (since the distribution depends on n, the number of claims in insurance, or the time horizon in finance). It is possible, as in Section 3 of Beirlant, Goegebeur, Segers, and Teugels (2004), to look "*close to the maximum*", by studying the limiting distribution of the k-th largest value[6] $X_{k:n}$, when n goes to infinity, as well as k. If $k/n = O(1)$, we study the asymptotic distribution of some high quantile (Section 5.2). It is also possible to focus on large losses with respect to the entire portfolio, i.e. S_n. This will yield the concepts of *Sub-Exponential* distribution (section 5.3) when comparing M_n and S_n, and comparing a high quantile, or the sum above a high quantile will be related to *Top Share* (section 5.4), also called *Large Claim Index* in insurance applications.

5.1 Probable Maximum Loss

Paul Levy extended the Central Limit Theorem to variables with a non-finite variance by considering non-degenerate distributions G such that

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{\overline{X}_n - b_n}{a_n} \leq x \right) = G(x) \tag{45}$$

where $\overline{X}_n = n^{-1} S_n$. It is an extension of the Central Limit Theorem in the sense that, if X_i's have a finite variance and if $b_n = \mathbb{E}[X]$ while $a_n = \sqrt{n^{-1} \text{Var}[X]}$, then G is the $\mathcal{N}(0,1)$ distribution, as well as any sequences such that $b_n \to \mathbb{E}[X]$ and $\sqrt{n}\, a_n \to \sqrt{\text{Var}[X]}$, as $n \to \infty$. In the case of variables with infinite variance, then G is called a stable distribution.

At the same period, Fisher-Tippett investigated a rather similar problem, searching for possible limiting distribution for a standardized version of the maximum

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x \right) = G(x). \tag{46}$$

---

[6]Given a sample $\{x_1, \cdots, x_n\}$, let $\{x_{1:n}, \cdots, x_{n:n}\}$ denote the ordered version, with $x_{1:n} = \min\{x_1, \cdots, x_n\}$, $x_{n:n} = \max\{x_1, \cdots, x_n\}$ and $x_{1:n} \leq \cdots x_{n-1:n} \leq x_{n:n}$.

It was shown - in Gnedenko (1943) - that the only possible limit was the so-called extreme value distribution

$$G_\alpha(x) = \exp(-x^{-\alpha}) \tag{47}$$

including the limiting case $G_0(x) = \exp(-e^{-x})$, on some appropriate interval. For instance, assume that X is $\mathcal{P}(1, \alpha)$ distributed, let $b_n = 0$ and $a_n = U(n) = n^{1/\alpha}$, then

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} [F(a_n x + b_n)]^n = \lim_{n\to\infty} \left[F(n^{1/\alpha}x)\right]^n \tag{48}$$

$$= \lim_{n\to\infty} \left(1 - \frac{x^{-\alpha}}{n}\right)^n = \exp(-x^{-\alpha}) \tag{49}$$

More generally, assume that X is Pareto type with tail index α, and consider $b_n = 0$ and $a_n \sim U(n)$ as $n \to \infty$. From Proposition 3.3.2 in Embrechts et al. (1997),

$$\lim_{n\to\infty} [F(a_n x + b_n)]^n = \exp\left(\lim_{n\to\infty} -n\overline{F}(a_n x + b_n)\right) \tag{50}$$

and since

$$n\overline{F}(a_n x) \sim \frac{\overline{F}(a_n x)}{\overline{F}(a_n)} \to x^{-\alpha} \text{ as } n \to \infty \tag{51}$$

therefore, we can obtain

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = \exp(-x^{-\alpha}) \tag{52}$$

as previously. Such a distribution is know as the Fréchet distribution with index α.

5.2   High Quantiles and Expected Shorfall

Given a distribution F, the quantile of level p of the distribution is

$$Q(p) = \inf\left\{y \in \mathbb{R} : F(y) \geq p\right\}. \tag{53}$$

Since we focus on high quantiles, i.e. $p \to 1$, a natural related function is

$$U(x) = Q\left(1 - \frac{1}{x}\right), \tag{54}$$

and to study properties of that function as $x \to \infty$.

16

In the previous section, we studied limiting behavior of $a_n^{-1}(M_n - b_n)$ as $n \to \infty$. Let $\widehat{F}_n$ denote the empirical cumulative distribution of $\{X_1, \cdots, X_n\}$,

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_i \le x). \tag{55}$$

Observe that $M_n = \widehat{F}_n^{-1}(1 - 1/n) = \widehat{U}_n(n)$, so M_n should be 'close' to $U(n)$. So it could make sense to study

$$\lim_{n \to \infty} \frac{U(nx) - U(n)}{a_n} \tag{56}$$

Beirlant et al. (2009, section 3.3) show that the only possible limit is

$$h(x) = \frac{x^\gamma - 1}{\gamma} \tag{57}$$

If U is the quantile function of $\mathcal{P}(1, \alpha)$, $U(x) = x^{1/\alpha}$, with auxiliary function[7] $a(t) = \alpha^{-1} t^{1/\alpha}$, then

$$\frac{U(tx) - U(t)}{a(t)} = \alpha(x^{1/\alpha} - 1) \tag{58}$$

for all x, i.e. $\gamma = \alpha^{-1}$.

Assume that X is $\mathcal{P}(u, \alpha)$, then

$$Q(p) = u \cdot (1 - p)^{-1/\alpha} \tag{59}$$

and if X is $\mathcal{GPD}(u, \sigma, \alpha)$, then

$$Q(p) = u + \sigma \big[(1 - p)^{-1/\alpha} - 1 \big] \tag{60}$$

For extended Pareto distributions quantiles are computed numerically.[8]

Another important quantity is the average of the top $p\%$, called expected shortfall,

$$\mathrm{ES}(p) = \mathbb{E}(X | X > Q(1 - p)) = \frac{1}{p} \int_{1-p}^{1} Q(u) du. \tag{61}$$

---

[7]The study of the limiting distribution of the maximum of a sample of size n made us introduce a normalizing sequence a_n. Here, a continuous version is considered - with $U(t)$ instead of $U(n)$ - and the sequence a_n becomes the auxiliary function $a(t)$.

[8]See the R packages `ReIns` or `TopIncomes`.

This quantity is closely related to the mean excess function, since $\mathrm{ES}(p) = Q(1-p) + e(Q(1-p))$, where e is the mean excess function. Assume that X is $\mathcal{P}(u, \alpha)$, then

$$\mathrm{ES}(p) = \frac{\alpha}{\alpha - 1} u(1-p)^{-1/\alpha} = \frac{\alpha}{\alpha - 1} Q(p), \tag{62}$$

and if X is $\mathcal{GPD}(u, \sigma, \alpha)$, then $X | X > Q(1-p)$ is $\mathcal{GPD}\left(Q(1-p), \sigma + Q(1-p) + u, \alpha\right)$ from Equation (15), and therefore

$$\mathrm{ES}(p) = \frac{\alpha Q(1-p) + \sigma - u}{\alpha - 1}. \tag{63}$$

For the extended Pareto model, only numerical approximations can be obtained.

On Figure 5 we have the evolution of $Q(1-p)$ for various values of p, from $1/10$ down to $1/1000$. A strict Pareto (——) with tail index $\alpha = 1.5$ is considered. On the left, we plot the quantile function of a GPD distribution (- - -), for small and large σ's, and on the right, we consider an EPD distribution (- - -) with different values of τ.

[Figure 5 about here.]

On Figure 6 we have the evolution of $ES(p)$ for various values of p, from $1/10$ down to $1/1000$. The strict Pareto distribution (——) still has tail index $\alpha = 1.5$. On the left, we plot the expected shortfall of a GPD distribution (- - -), for small and large σ's, and on the right, we consider an EPD distribution (- - -) with different values of τ.

[Figure 6 about here.]

Finally, on Figure 7 we compare $Q(1-p)$ and $ES(p)$ for various values of p, and plot the line $(Q(1-p), ES(p))$. As expected, it is a straight line in the strict Pareto case (——), above the first diagonal (- - -), meaning that there can be substantial losses above any quantile. Two GPD distributions with the same tail index α are considered on the left (- - -), and an EPD case is on the right (- - -). The GPD case is not linear, but tends to be linear when p is large enough (in the upper tail, we have a strict Pareto behavior). The EPD case is linear, with an Expected Shorfall smaller than the strict Pareto case.

[Figure 7 about here.]

18

5.3 Maximum, Sum and Subexponential distributions

In insurance, one possible way to define *large* losses is to say that a single claim is large when the total amount (over a given period of time) is predominantly determined by that single claim. The dominance of the largest value can be written

$$\mathbb{P}[S_n > x] \sim \mathbb{P}[M_n > x] \text{ as } x \to \infty. \tag{64}$$

Heuristically, the interpretation is that the sum will exceed a large value x only when (at least) one single loss X_i also exceeds x. If losses X_i are i.i.d. with CDF F we will say that F belongs to the class of subexponential distributions, $F \in \mathcal{S}$. As proved in Albrecher et al. (2017, Section 3.2) strict Pareto and Pareto-type distributions belong to class $\mathcal{S}$. Hence, this class can be seen as an extension of the Pareto models. See Goldie and Klüppelberg (1998) for applications of such distributions in a risk-management context.

In sections 6.2.6 and 8.2.4 of Embrechts et al. (1997), another quantity of interest is introduced, the ratio of the maximum and the sum, M_s/S_n. As proved in O'Brien (1980), $\mathbb{E}(X)$ is finite if and only if the ratio converges (almost surely) towards 0. A non-degenerated limit is obtained if and only if X is Pareto type, with index $\alpha \in (0, 1)$. An alternative would be to consider the ratio of the sum of the largest claims (and not only the largest one), and the total sum.

5.4 Top Share and Large Claim Index

In section 5.2, we discussed how to move from M_n to the average of the top $p\%$ of the distribution. It could be possible to normalize by the total sum. The top p-% loss share can be defined as

$$\text{TS}(p) = \frac{p \, \mathbb{E}(X|X > Q(1-p))}{\mathbb{E}(X)} = \frac{p\text{ES}(p)}{\mathbb{E}(X)} = 1 - L(1-p) \tag{65}$$

where L denote the Lorenz curve, i.e.

$$L(u) = \frac{1}{\mathbb{E}(X)} \int_0^{Q(u)} \overline{F}(y)dy = \frac{1}{\mathbb{E}(X)} \int_0^u Q(y)dy. \tag{66}$$

The empirical version of TS_p is

$$\widehat{\text{TS}}(p) = \frac{x_{n-[np]:n} + x_{n:n}}{S_n} \tag{67}$$

where $(x_{i:n})$ denote the order version of sample (x_i), with $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$. See section 8.2.2 of Embrechts et al. (1997) for convergence properties of that estimator. In the case where X is a strict Pareto $\mathcal{P}(u, \alpha)$, then

$$\text{TS}(p) = p^{(\alpha-1)/\alpha}, \tag{68}$$

while if X is $\mathcal{GPD}(u, \sigma, \alpha)$,

$$\text{TS}(p) = 1 - \frac{u(1-p) + \alpha \cdot \left(I_1(\alpha-2, 2) - I_{p^{1/\alpha}(\alpha-1,2)}\right)}{u + \sigma\alpha I_1(\alpha-1, 2)} \tag{69}$$

as in Section 6 of Arnold (2008), where $I_z(a, b)$ is the incomplete Beta function,

$$I_z(a, b) = \int_0^z t^{a-1} t^{(}1-t)^{b-1} dt. \tag{70}$$

No simple expression can be derive for the extended Pareto case, but numerical computations are possible, as on Figure 8, where strict Pareto distribution with tail index $\alpha = 1.5$ is plotted against Pareto type distributions, GPD and EPD, with the same tail index.

[Figure 8 about here.]

5.5 From Pareto to Pareto Type Models

The concepts introduced earlier were based on the assumption that the distribution F of the observations was known. But actually, for small enough p, we might only have a Pareto distribution in the upper tail, above some threshold u, with $\mathbb{P}[X > u] \geq p$, as suggested in Smith (1987). And let q_u denote the probability to exceed that threshold q_u.

Note that if $x > u$, $\mathbb{P}[X > x] = \mathbb{P}[X > x | X > u] \cdot \mathbb{P}[X > u] = q_u \mathbb{P}[X > x | X > u]$. Thus, if a $\mathcal{P}(u, \alpha)$ distribution is considered in the tails,

$$\mathbb{P}[X > x] = q_u \left(\frac{x}{u}\right)^{-\alpha} \tag{71}$$

and therefore,

$$Q_u(1-p) = u \cdot \left(\frac{p}{q_u}\right)^{-1/\alpha}, \quad \text{where } q_u = \mathbb{P}[X > u], \tag{72}$$

and

$$\text{ES}_u(p) = \frac{\alpha}{\alpha - 1} Q_u(1-p). \tag{73}$$

20

If a $\mathcal{GPD}(u, \sigma, \alpha)$ distribution is considered in the tails,

$$Q_u(1 - p) = u \cdot \sigma \left[\left(\frac{p}{q_u} \right)^{-1/\alpha} - 1 \right]^{-1} , \text{ where } q_u = \mathbb{P}[X > u], \qquad (74)$$

and

$$\mathrm{ES}_u(p) = \frac{\alpha Q_u(1 - p) + \sigma - u}{\alpha - 1}. \qquad (75)$$

Nevertheless, if a simple update of formulas was necessary for downside risk measures (such as a quantile and an expected shortfall), expression can be more complicated when they rely on the entire distribution, such as the Large Claim Index TS. From Equation (65), if p is small enough, write

$$\mathrm{TS}_u(p) = \frac{p\mathrm{ES}_u(p)}{\mathbb{E}(X)} \qquad (76)$$

where $\mathbb{E}(X)$ is either approximated using $\overline{x}$, or a more robust version would be $(1 - q_u)\overline{x}_u + q_u\mathrm{ES}_u(q_u)$ where $\overline{x}_u$ is the empirical average of values below threshold u, while $\mathrm{ES}_u(q_u)$ is the parametric mean from a Pareto-type model, about threshold u.

6   Insurance and Reinsurance

Classical quantities in a reinsurance context are the return period (that can be related to quantiles, and the financial Value-at-Risk) and a stop-loss premium (related to the expected-shortfall).

6.1   Return Period and Return Level

Consider $Y_1, \cdots, Y_n$, a collection of i.i.d. random variables, with distribution F. Consider the sequence of i.i.d. Bernoulli variables $X_i = \mathbf{1}_{Y_i > z(t)}$ for some threshold $z(t)$, such that $\overline{F}(z(t)) = 1/t$. The first time threshold $z(t)$ is reached is the random variable N_t defined as

$$N_t = \min \big\{ i \in \mathbb{N}_* : Y_i > z(t) \big\}. \qquad (77)$$

Then N_t has a geometric distribution,

$$\mathbb{P}\big[N_t = k\big] = \frac{1}{(1 - t)^{k-1}} \frac{1}{t}. \qquad (78)$$

The return period of the events $\{Y_i > z(t)\}$ is then $\mathbb{E}[N_t] = t$, or conversely, the threshold that is reached - on average - every t events is $z(t) = Q(1 - 1/t) = U(t)$, called also return level.

Hence, with strict Pareto losses $\mathcal{P}(u, \alpha)$

$$z(t) = ut^{1/\alpha}, \tag{79}$$

while with $\mathcal{GPD}(u, \sigma, \alpha)$ losses,

$$z(t) = u + \sigma \left[t^{1/\alpha} - 1\right]. \tag{80}$$

In the case of non-strict Pareto distribution, where only the top $q\%$ is $\mathcal{P}(u, \alpha)$ distributed,

$$z_u(t) = u(q_u t)^{1/\alpha}, \quad \text{where} \quad q_u = \mathbb{P}[X > u]. \tag{81}$$

Thus, if we assume that the top 10% is Pareto distributed, the return level of a centennial event ($t = 100$) corresponds to the return level of a decennial event ($q_u t = 10$) for the top 10% observations. If the top $q\%$ is $\mathcal{GPD}(u, \sigma, \alpha)$ distributed,

$$z_u(t) = u + \sigma\left((q_u t)^{1/\alpha} - 1\right), \quad \text{where} \quad q_u = \mathbb{P}[X > u]. \tag{82}$$

From an empirical perspective, consider a dataset with n observations, i.i.d., $y_1, \cdots, y_n$. Assume that above threshold u, the (conditional) distribution is $\mathcal{GPD}(u, \sigma, \alpha)$. Denote n_u the number of observations above u, so that $\widehat{F}_n(u) = 1 - n_u/n$. Then

$$\widehat{z}_u(t) = u + \widehat{\sigma}\left[\left(\frac{n_u}{n}t\right)^{1/\widehat{\alpha}} - 1\right]. \tag{83}$$

6.2   Reinsurance Pricing

Let Y denote the loss amount of an accident. Consider a contract with deductible d, so that the indemnity paid by the insurance company is $(Y - d)_+$, where $\cdot_+$ denotes the positive part, i.e. $(y - d)_+ = \max\{0, y - d\}$. The pure premium of such a contract, with deductible d, is

$$\pi(d) = \mathbb{E}\left[(Y - d)_+\right] \tag{84}$$

and therefore, if $e(d)$ denotes the mean excess function,

$$\pi(d) = \mathbb{E}\left[Y - d | Y > d\right] \cdot \mathbb{P}(Y > d) = e(d) \cdot \mathbb{P}(Y > d). \tag{85}$$

22

If we assume that losses above some threshold u have a $\mathcal{GPD}(u, \sigma, \alpha)$ distribution, for any $d > u$,

$$e_u(d) = \left(\frac{\sigma - u}{\alpha - 1} + \frac{\alpha}{\alpha - 1} d \right) \tag{86}$$

and therefore

$$\pi_u(d) = \frac{n_u}{n} \left[1 + \left(\frac{d - u}{\sigma} \right) \right]^{-\alpha} \cdot \left(\frac{\sigma - u}{\alpha - 1} + \frac{\alpha}{\alpha - 1} d \right). \tag{87}$$

If we plugin estimators of σ and α, we derive the estimator of the pure premium $\pi_u(d)$. Note that in Section 4.6. in Albrecher et al. (2017), approximations are given for Extended Pareto distributions.

As previously, consider a dataset with n observations, i.i.d., $y_1, \cdots, y_n$, assume that above threshold u, the (conditional) distribution is $\mathcal{GPD}(u, \sigma, \alpha)$ and let n_u denote the number of observations above u. Then if we fit a GPD distribution, the estimated premium would be

$$\widehat{\pi}_u(d) = \frac{n_u}{n} \left[1 + \left(\frac{d - u}{\widehat{\sigma}} \right) \right]^{-\widehat{\alpha}} \cdot \left(\frac{\widehat{\sigma} - u}{\widehat{\alpha} - 1} + \frac{\alpha}{\widehat{\alpha} - 1} d \right), \tag{88}$$

for some deductible d, higher than u, our predefined threshold.

6.3 Application on Real Data

In order to illustrate, two datasets are considered, with a large fire loss dataset (the 'Danish dataset' studied in Beirlant and Teugels 1992 and McNeil 1997)[9] and large medical losses (from the SOA dataset studied in Cebrián et al. 2003). On Figure 9, Pareto plots are considered for losses above some threshold u, with $u = 10$ for fire losses, and $u = 1$ for medical losses (here in \$'$00,000$), i.e.

$$\left(x_{i:n}, 1 - \frac{i - (n - n_u)}{n_u} \right)_{i = n - n_u + 1, \cdots, n} \tag{89}$$

on a log-log scatterplot. In Pareto models, points should be on a straight line, with slope $-\alpha$. The plain line (——) is the Pareto distribution (fitted using maximum likelihood techniques) and the dotted line (- - -) corresponds to a Generalized Pareto distribution.

---

[9]It is the `danishuni` dataset in the `CASdatasets` package, available from http://cas.uqam.ca/

[Figure 9 about here.]

On Figure 10 we can visualize estimates of α, as a function of the number of tail events considered. The strong line (—) is the Pareto distribution, the thin line (—) is the GPD distribution while the dotted line (- - -) corresponds to a Extended Pareto distribution. Here the three distributions were fitted using maximum likelihood techniques, and only $\widehat{\alpha}$ is ploted.

[Figure 10 about here.]

On Figure 11, we can visualize return levels for various return periods (on a log-scale), when Generalized Pareto distributions are fitted above level u ($u = 10$ for fires and $u = 2$ for medical claims). Those values are quantiles, then return periods are interpreted as probabilities (e.g. a return period of 200 is a quantile of level 99.5%).

[Figure 11 about here.]

On Figure 12, we can visualize the mean excess function of individual claims, for fire losses on the left, and medical claims on the right (in \$ '00,000). The dark line is the empirical version (average above some threshold) and the line (—) is the GPD fit

$$\widehat{e}_u(d) = \left(\frac{\widehat{\sigma} - u}{\widehat{\alpha} - 1} + \frac{\widehat{\alpha}}{\widehat{\alpha} - 1} d \right) \tag{90}$$

respectively with $u = 10$ on the left and $u = 2$ on the right, for various values of the threshold d ($\geq u$).

[Figure 12 about here.]

On Figure 13, we can visualize the stability of estimators with respect to the choice of the threshold u, with the estimation of $e(d)$, with $d = 20$ for Danish fires and $d = 12$ for Medical losses. The empirical average is the horizontal dashed line (- - -), the estimator derived from a Pareto distribution above threshold u is the plain strong line (—), the one from a Generalized Pareto model above threshold u is the plain line (—) and the Extended Pareto model above threshold u is the dashed line (- - -), as a function of u. The estimator obtained from the Extended Pareto model is quite stable, and close to the empirical estimate.

[Figure 13 about here.]

Finally, on Figure 14, we have the evolution of the (annual) pure premium, which is simply the mean excess function, multiplied by the (annual) probability to have a loss that exceed the deductible. For the Danish fire losses, since we have 10 years of data, the premium takes into account this 10 factor (which is consistent with a Poisson process assumption for claims occurrence): on average there are 11 claims per year above 10, and 2.5 above 25. For the SOA medical losses, there are 213 claims above $ 500,000 ($u = 5$) for a mean excess function slightly below 3, so the pure premium is close to 600, for a deductible (per claim) of 5.

[Figure 14 about here.]

7 Finance and risk-measures

7.1 Downside Risk Measures

Classical problems in risk management in financial applications involve (extreme) quantile estimation, such as the Value-at-Risk and the Expected Shortfall. Let Y denote the negative returns of some financial instrument, then the Value-at-Risk $\text{VaR}(p)$ is defined as the quantile of level $1 - p$ of the distribution,

$$\text{VaR}(p) = Q(1 - p) = \inf \big\{ y \in \mathbb{R} : F(y) \geq 1 - p \big\}, \tag{91}$$

while the expected shortfall (or tail conditional expectation) is the potential size of the loss exceeding the Value-at-Risk,

$$\text{ES}(p) = \mathbb{E}\big[X \big| X > \text{VaR}(p) \big] \tag{92}$$

If X_t is the (random) loss variable of a financial position from time t to time $t + h$ (with a given time horizon h), classical distributions have been intensively consider in financial literature, starting from the Gaussian case, $X_t \sim \mathcal{N}(\mu_t, \Sigma_t)$. Then

$$\text{VaR}(p) = \mu_t + \Phi^{-1}(1 - p)\Sigma_t \quad \text{and} \quad \text{ES}_p = \mu_t + \frac{\phi(\Phi^{-1}(1 - p))}{p}\Sigma_t, \tag{93}$$

where ϕ and Φ denote respectively the density and the c.d.f. of the centered standard Gaussian distribution.

Since Pareto models have a simple expression to model losses above a threshold, with a GPD above a threshold u, for $y > u$,

$$\widehat{F}_u(y) = \mathbb{P}[Y \leq y]$$
$$= \mathbb{P}[Y \leq y | Y > u] \cdot \mathbb{P}[Y > u] + \mathbb{P}[Y \leq u]$$
$$= G_{\widehat{\sigma}, \widehat{\alpha}}(y - u) \cdot \frac{n_u}{n} + \frac{n - n_u}{n}$$

where n_u denotes the number of observations above threshold u, and

$$G_{\sigma, \alpha}(x) = 1 - \left[1 + \left(\frac{x}{\sigma}\right)\right]^{-\alpha} \qquad \text{for } x \geq u, \tag{94}$$

is the $\mathcal{GPD}(0, \sigma, \alpha)$ CDF. Then for an event rare enough - i.e. a probability small enough $(p < n_u/n)$ - the Value-at-Risk VaR_p can be approximated by

$$\widehat{\text{VaR}}_u(p) = u \cdot \widehat{\sigma} \left[\left(\frac{p}{q_u}\right)^{-1/\widehat{\alpha}} - 1\right]^{-1}, \text{ where } q_u = \mathbb{P}[Y > u], \tag{95}$$

as well as the expected-shortfall,

$$\widehat{\text{ES}}_u(p) = \frac{\widehat{\alpha}\widehat{\text{VaR}}_u(p) + \widehat{\sigma} - u}{\widehat{\alpha} - 1}. \tag{96}$$

7.2 Application on Real Data

In finance, log-returns are rarely independent, and it is necessary to take into account the dynamics of the volatility. Consider some GARCH-type process for log-returns, $y_t = \mu_t + \Sigma_t x_t$, where (x_t) are i.i.d. variables, and where (σ_t) is the stochastic volatility. Then, as mentioned in McNeil and Frey (2000)

$$\widehat{\text{VaR}}_{Y_t | \Sigma_t}(p) = \mu_t + \Sigma_t \cdot \widehat{\text{VaR}}_{X_t}(p) \tag{97}$$
$$\widehat{\text{ES}}_{Y_t | \Sigma_t}(p) = \mu_t + \Sigma_t \cdot \widehat{\text{ES}}_{X_t}(p) \tag{98}$$

To estimate (Σ_t), several approaches can be considered, from the exponential weighted moving average (EWMA), with

$$\widehat{\Sigma}_{t+1}^2 = \beta\widehat{\Sigma}_t + (1 - \beta)y_t^2, \text{ for some } \beta \in (0, 1), \tag{99}$$

to the GARCH(1,1) process,

$$\widehat{\Sigma}_{t+1}^2 = \alpha_0 + \alpha_1 y_t^2 + \beta_1\widehat{\Sigma}_t \text{ for some } \alpha_1, \beta_1 \in (0, 1), \text{ with } \alpha_1 + \beta_1 < 1. \tag{100}$$

To illustrate, we use returns of the Brent crude price of oil[10], extracted from the North Sea prices and comprises Brent Blend, Forties Blend, Oseberg and Ekofisk crudes.

On Figure 15, for the plain line (GARCH+GPD, —) we use

$$\widehat{\text{VaR}}_{Y_t|\Sigma_t}(p) = \widehat{\Sigma}_t \cdot \widehat{\text{VaR}}_{X_t}(p) \tag{101}$$

where

$$\widehat{\text{VaR}}_{X_t}(p) = u + \frac{\widehat{\sigma}}{\widehat{\alpha}}\left[\left(\frac{n}{n_u}(1-q)\right)^{-\widehat{\alpha}} - 1\right] \tag{102}$$

with $u = 2\%$ and $\widehat{\Sigma}_t$ is the fitted volatility with a GARCH(1,1) process. The red line is obtained using a Generalized Pareto fit on a sliding window $[t \pm 150]$.

[Figure 15 about here.]

On Figure 16, an extended Pareto model is considered for $\widehat{\text{VaR}}_{X_t}(p)$.

[Figure 16 about here.]

Finally, on Figure 17, we compare the GARCH + Pareto models, with either a Generalized Pareto model (as in McNeil and Frey (2000)) or an Extended Pareto model. Observe that the 0.05% Value-at-Risk is always over estimated with the Generalized Pareto model. Nevertheless, with Pareto models estimated on sliding windows (on $[t \pm 150]$) the difference can be negative, sometimes. But overall, it is more likely that the Generalized Pareto overestimates the Value-at-Risk.

[Figure 17 about here.]

References

Albrecher, H., J. Beirlant, and J. L. Teugels (2017). *Reinsurance: Actuarial and Statistical Aspects*. Wiley Series in Probability and Statistics.

Arnold, B. C. (2008). Pareto and generalized Pareto distributions. In D. Chotikapanich (Ed.), *Modeling Income Distributions and Lorenz Curves*, Chapter 7, pp. 119–146. Springer.

---

[10] Available from https://www.nasdaq.com/market-activity/commodities/bz%3Anmx

Balkema, A. and L. de Haan (1974). Residual life time at great age. *Annals of Probability 2*, 792–804.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics.

Beirlant, J., E. Joossens, and J. Segers (2009). Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference 139*, 2800–2815.

Beirlant, J. and J. L. Teugels (1992). Modeling large claims in non-life insurance. *Insurance: Mathematics and Economics 11*(1), 17 – 29.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1987). *Regular Variation*. Encyclopedia of Mathematics and its Applications. Cambridge University Press.

Cebrián, A. C., M. Denuit, and P. Lambert (2003). Generalized pareto fit to the society of actuaries' large claims database. *North American Actuarial Journal 7*(3), 18–36.

Charpentier, A. and E. Flachaire (2019). Pareto models for top incomes. hal id: hal-02145024.

Davison, A. (2003). *Statistical Models*. Cambridge University Press.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An introduction*. Springer Series in Operations Research and Financial Engineering.

de Haan, L. and U. Stadtmüller (1996). Generalized regular variation of second order. *Journal of the Australian Mathematical Society 61*, 381–395.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin Heidelberg: Springer-Verlag.

Fisher, R. A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Proceedings of the Cambridge Philosophical Society*, Volume 24, pp. 180–290.

Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics 1*(1), 255–294.

Ghosh, S. and S. Resnick (2010). A discussion on mean excess plots. *Stochastic Processes and their Applications 120*(8), 1492 – 1517.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics 44* (3), 423–453.

Goldie, C. M. and C. Klüppelberg (1998). Subexponential distributions. In R. J. Adler, R. E. Feldman, and M. S. Taqqu (Eds.), *A Practical Guide to Heavy Tails*, pp. 436–459. Birkhäuser.

Guess, F. and F. Proschan (1988). 12 mean residual life: Theory and applications. In *Quality Control and Reliability*, Volume 7 of *Handbook of Statistics*, pp. 215 – 224. Elsevier.

Hagstroem, K. G. (1925). La loi de pareto et la reassurance. *Skandinavisk Aktuarietidskrift 25*.

Hagstroem, K. G. (1960). Remarks on pareto distributions. *Scandinavian Actuarial Journal 60* (1-2), 59–71.

Hall, P. (1982). On some simple estimate of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B 44*, 37–42.

Jessen, A. H. and T. Mikosch (2006). Regularly varying functions. *Publications de l'Institut Mathématique 19*, 171 – 192.

Klüppelberg, C. (2004). Risk management with extreme value theory. In B. Finkenstädt and H. Rootzén (Eds.), *Extreme Values in Finance, Telecommunications, and the Environment*, Chapter 3, pp. 101–168. Oxford: Chapman & Hall/CRC.

Kremer, E. (1984). *Rating of Non Proportional Reinsurance Treaties Based on Ordered Claims*, pp. 285–314. Dordrecht: Springer Netherlands.

Lomax, K. S. (1954). Business failures: Another example of the analysis of failure data. *Journal of the American Statistical Association 49* (268), 847–852.

McNeil, A. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin 27* (27), 117–137.

McNeil, A. J. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance 7* (3), 271 – 300. Special issue on Risk Management.

O'Brien, G. L. (1980). A limit theorem for sample maxima and heavy branches in galton-watson trees. *Journal of Applied Probability 17* (2), 539–545.

Pareto, V. (1895). La legge della domanda. In Pareto (Ed.), *Ecrits d'économie politique pure*, Chapter 11, pp. 295–304. Genève: Librairie Droz.

Peng, L. and Y. Qi (2004). Estimating the first- and second-order parameters of a heavy-tailed distribution. *Australian & New Zealand Journal of Statistics 46*, 305–312.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics 23*, 119–131.

Resnick, S. I. (1997). Discussion of the danish data on large fire insurance losses. *ASTIN Bulletin 27*(1), 139–151.

Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.* Springer Series in Operations Research and Financial Engineering.

Reynkens, T. (2018). *ReIns: Functions from "Reinsurance: Actuarial and Statistical Aspects".* R package version 1.0.8.

Rigby, R. A. and D. M. Stasinopoulos (2005). Generalized additive models for location, scale and shape,(with discussion). *Applied Statistics 54*, 507–554.

Roy, A. D. (1952). Safety first and the holding of assets. *Econometrica 20*(3), 431–449.

Schumpeter, J. A. (1949). Vilfredo pareto (1848–1923). *The Quarterly Journal of Economics 63*(2), 147–173.

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

Smith, R. L. (1987, 09). Estimating tails of probability distributions. *Ann. Statist. 15*(3), 1174–1207.

Vajda, S. (1951). Analytical studies in stop-loss reinsurance. *Scandinavian Actuarial Journal 1951*(1-2), 158–175.
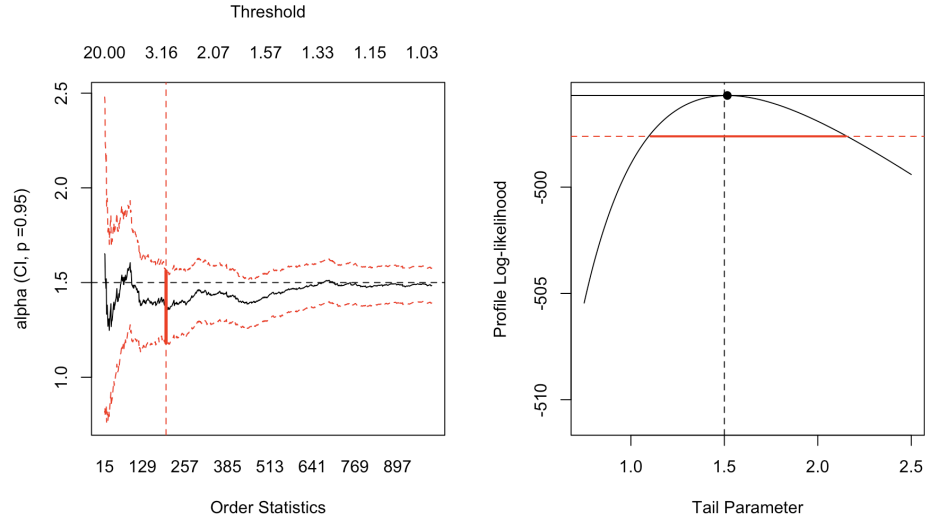
Figure 1: Hill estimator (on the left) for α and profile likelihood for α from a GPD model (on the right) on the top 20% of the observations, when $n = 1,000$ observations where generated from a (strict) Pareto distribution, with $\alpha = 1.5$.
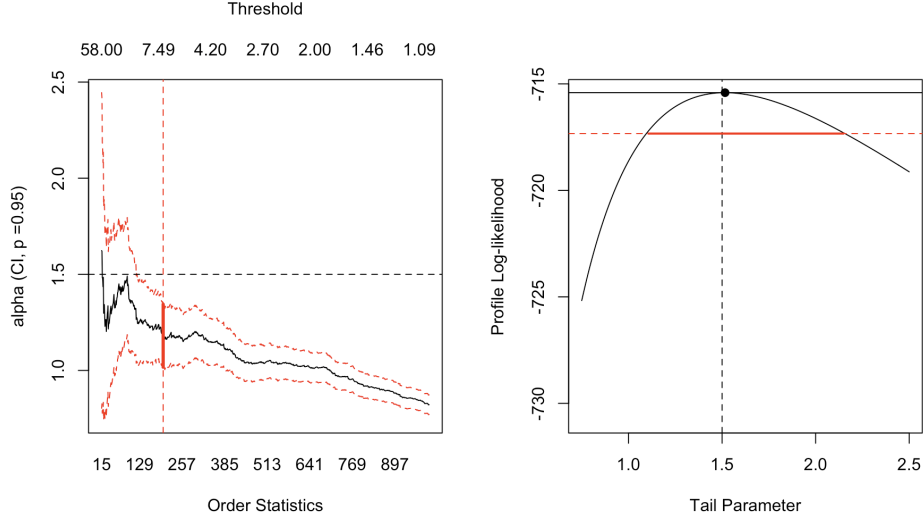
Figure 2: Hill estimator (on the left) for α and profile likelihood for α from a GPD model (on the right) on the top 20% of the observations, when $n = 1,000$ observations where generated from a Generalized Pareto distribution, with $\alpha = 1.5$, and $\sigma = 1$.
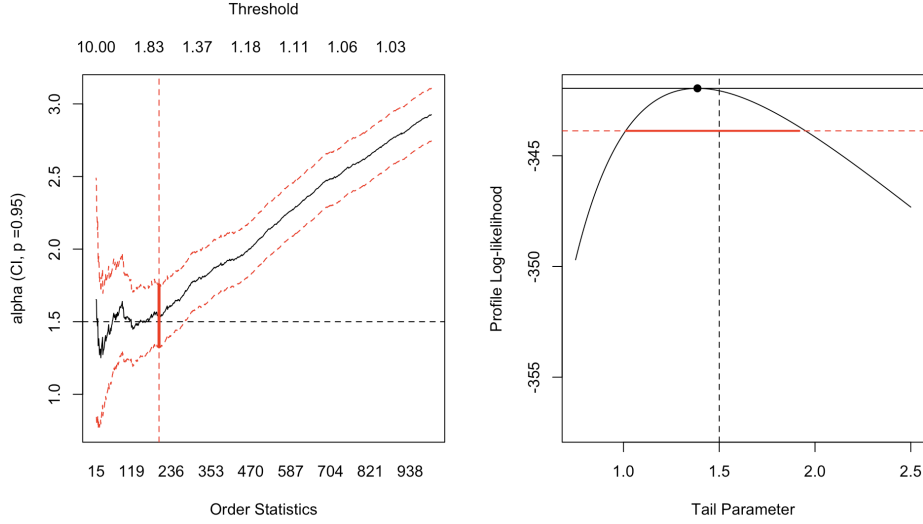
Figure 3: Hill estimator (on the left) for α and profile likelihood for α from a GPD model (on the right) on the top 20% of the observations, when $n = 1,000$ observations where generated from an extended Pareto distribution, with $\alpha = 1.5$, and $\tau = -2$.

32
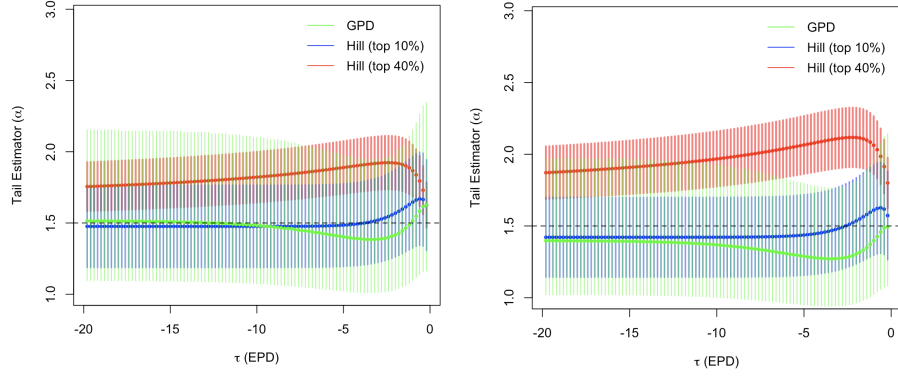
Figure 4: Simulation of EPD samples (with the same monte-carlo seed on the left, and another seed on the right), with $\alpha = 1.5$, and τ varying from -20 to 0. The red curve (—) is Hill estimator $\widehat{\alpha}_u$ obtained when u is the 60% quantile, while the blue curve (—) is obtained when u is the 90% quantile. The green curve (—) is the profile likelihood estimator of a GPD distribution.
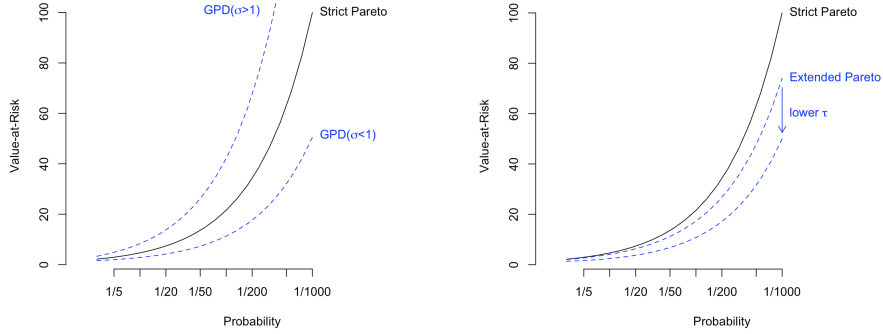
Figure 5: Quantile $Q(1-p)$ as a function of the probability p (on a log scale), with a strict Pareto (—) with tail index $\alpha = 1.5$, and a GPD distribution (- - -) on the left, and an EPD distribution (- - -) on the right.

33
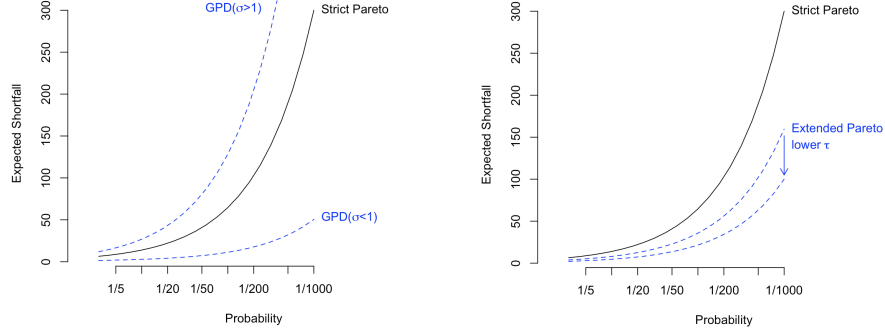
Figure 6: Expected Shortfall ES(p) as a function of the probability (on a log scale), with a strict Pareto (——) with tail index $\alpha = 1.5$, and a GPD distribution (- - -) on the left, and an EPD distribution (- - -) on the right.
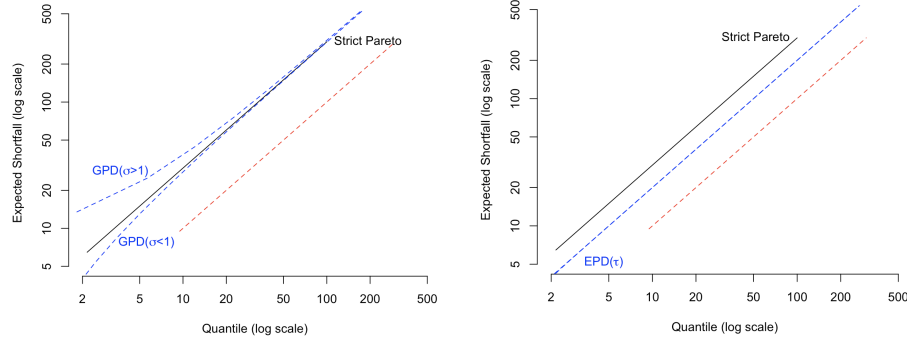
Figure 7: Expected Shortfall against quantile for different values of the probability (i.e. $(Q(1-p)), ES(p))$, with a strict Pareto (——) with tail index $\alpha = 1.5$, and a GPD distribution (- - -) on the left, and an EPD distribution (- - -) on the right. The straight line below (- - -) is the identity curve, where the Expected Shortfall would be equal to the associated quantile.

34
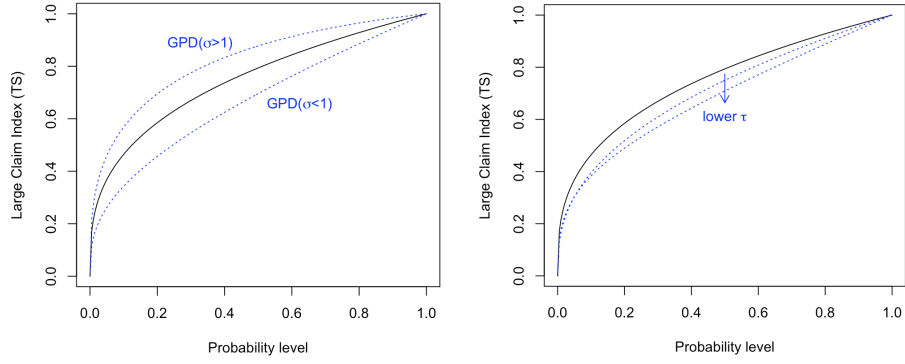
Figure 8: Large Claim Index TS(p) as a function of the probability p, with a strict Pareto (——) with tail index $\alpha = 1.5$, and a GPD distribution (---) on the left, and an EPD distribution (---) on the right.
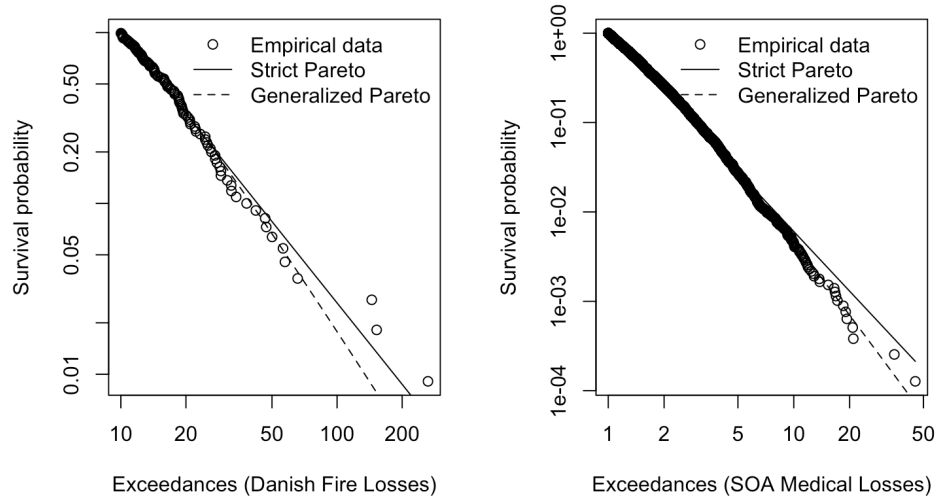
Figure 9: Pareto plot, with the Danish Fire data on the left, and the SOA Medical claims data on the right.
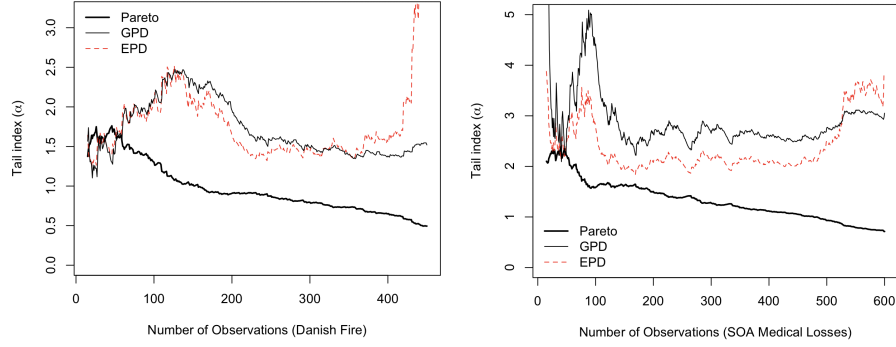
Figure 10: Estimation of α, with a Pareto model (——), a Generalized Pareto model (——) and an Extended Pareto Model (- - -), as a function of n_u, the number of tail events considered.
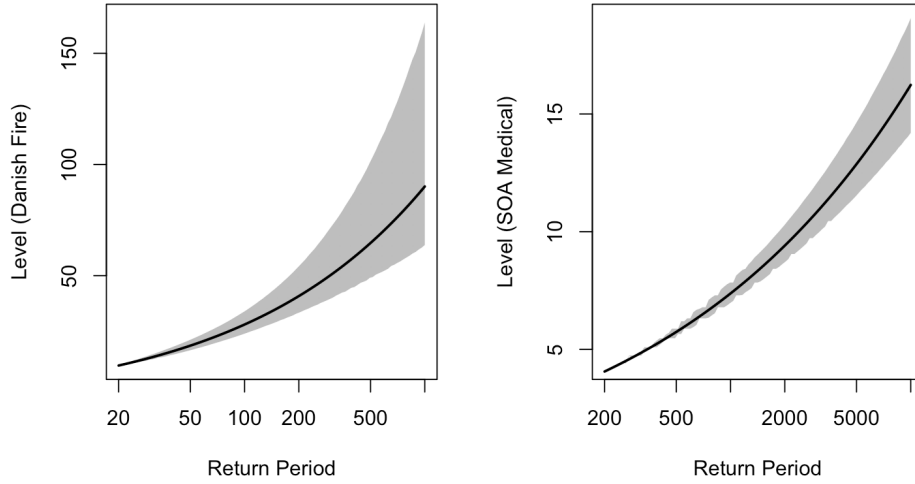
Figure 11: Estimation return levels for various return periods, for with the Danish Fire data on the left, and the SOA Medical claims data on the right. The grey area is the confidence region.
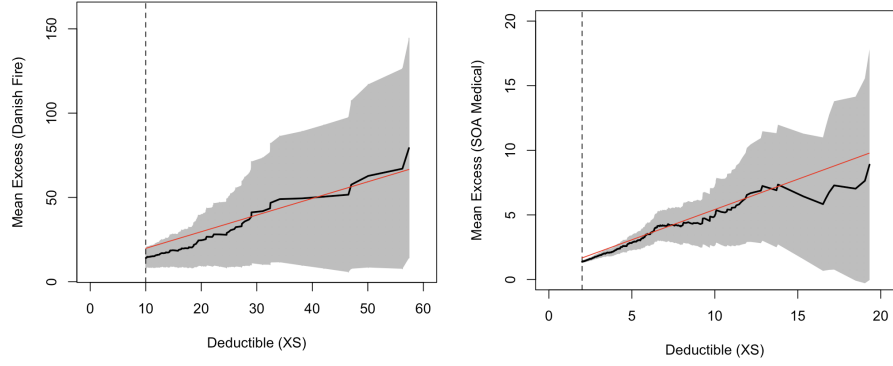
Figure 12: Mean excess functions Danish Fire data on the left, and the SOA Medical claims data on the right, with confidence intervals on the empirical version. The line (——) is the GPD fit, for a given threshold (respectively 10 and 2).
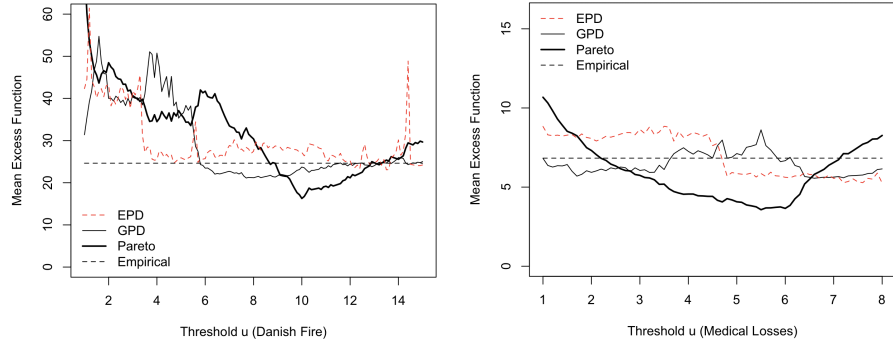
Figure 13: Estimation of $e(d)$, with $d = 20$ for Danish fires and $d = 12$ for Medical losses, with the empirical average (- - -), the estimator derived from a Pareto distribution above threshold u (——), from a Generalized Pareto model above threshold u (——) and from an Extended Pareto Model above threshold u (- - -), as a function of u.
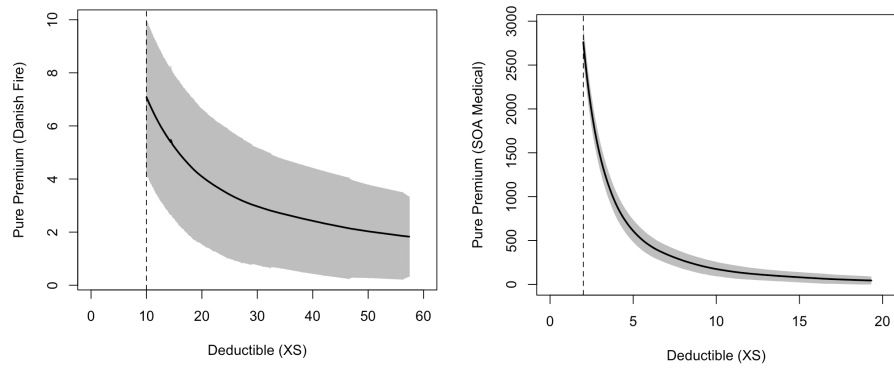
Figure 14: Annual Pure Premium for Danish Fire data on the left, and the SOA Medical claims data on the right.
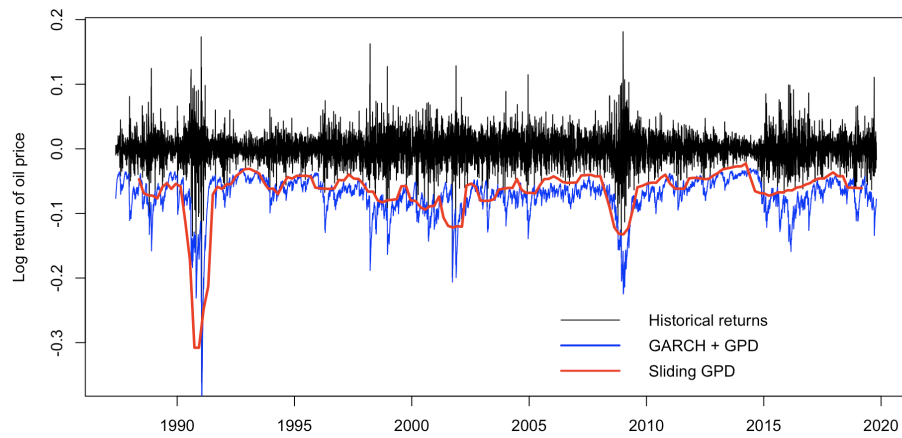

Figure 15: Value-at-Risk at level 0.5% for the daily log-return of oil prices (——), with a GARCH+GPD model (——) and a sliding window estimator GPD model (——)).
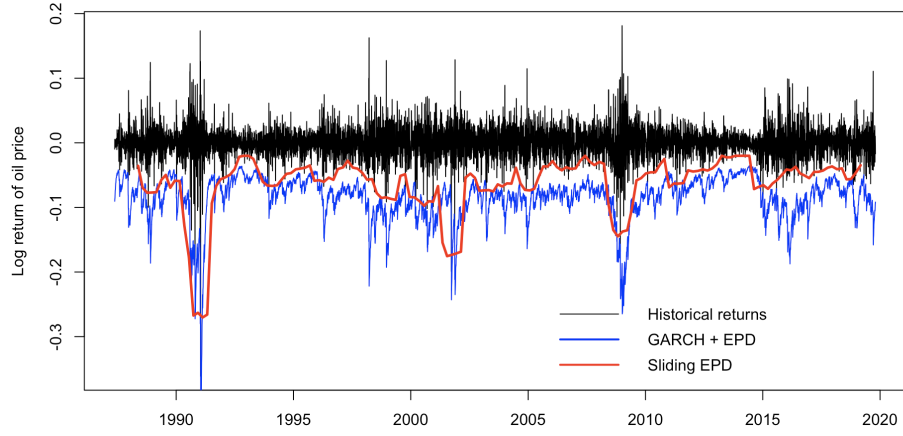
Figure 16: Value-at-Risk at level 0.5% for the daily log-return of oil prices
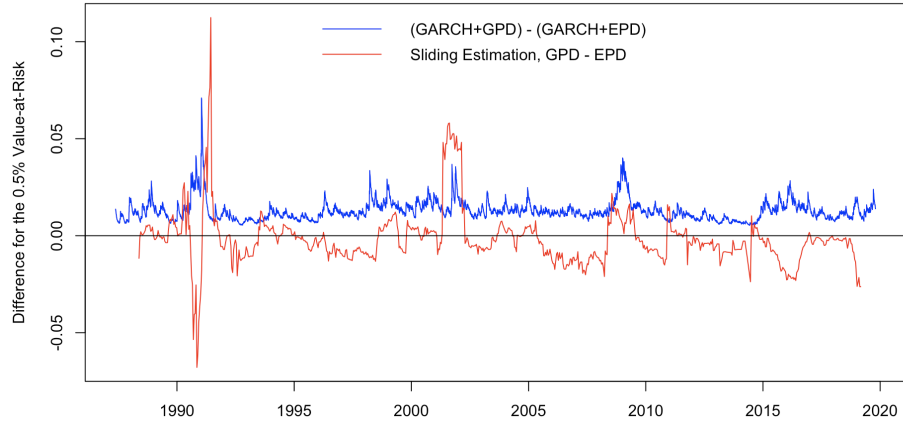(—), with a GARCH+EPD model (—) and a sliding window estimator
EPD model (—).

Figure 17: Difference of the Value-at-Risk at level 0.5% for the daily log-
return of oil prices between GARCH+Pareto models (—) and a sliding
window estimator of Pareto models (—).