Distinguishing between a power law and a Pareto distribution

Joan del Castillo [*] and Pedro Puig [†]
Departament de Matemàtiques, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

(Received 27 January 2023; accepted 27 April 2023; published 13 June 2023)

This paper introduces the location Pareto distribution as a natural extension of the power law distribution and gives a likelihood ratio test for choosing between the two models. Some properties of the distribution and test are thoroughly investigated, and applications to real data are provided. For large values of the observations the two models perform similarly; this explains why some classical approaches are very insensitive to the differentiation between them. The likelihood ratio test between the two models is simple to use and has a high level of discrimination power. It is recommended when the complementary cumulative distribution function exhibits linearity on a log-log scale.

DOI: 10.1103/PhysRevE.107.064113

I. INTRODUCTION

Pareto models have the great appeal of the underlying simplicity of the complementary cumulative distribution function (tail function) which takes the form of a straight line when plotted on a log-log graph. The study of these models for analyzing examples of a wide range of disciplines began with the work by Pareto [1,2] to describing the distribution of wealth in a society; see [3,4].

The approaches used in works published in physics and statistics journals diverge in several ways because of differences between the notions of power law, or scale-free, and Pareto distributions. This article should serve as a starting point to clarify this issue and to discuss a common methodology for studying Pareto models. According to Sec. III, our analysis extends beyond the generalized Pareto distribution (GPD) to include the power law model (see Fig. 2).

The extreme value theory (EVT) has a long history in statistics that begins in the work by Fisher and Tippett [5] and the characterization of the limit distributions of maxima and minima by Gnedenko [6]. Pickands [7], Balkema, and de Haan [8] are the first to direct the theory to the study of all observations obtained above a particular threshold that lead to GPD. The EVT is explained in [9–15] among many others.

On articles published in journals of physics, we consider the widely cited works by Newman [16] and Clauset *et al.* [17] as our starting point. The significant impact of these articles on complex systems physics has extended their methodologies in multiple directions; see [18–22]. Here we will employ power law as is commonly used in the physics articles and Pareto in a different sense, which we will explain later. However, in physics, Pareto is also used as a synonym for power law, like in [23,24] when discussing "Pareto versus the lognormal," which has been the subject of an interesting controversy. The first cited authors approach the problem through the sample coefficient of variation following [25], and the second uses a maximum entropy test.

Section II, in addition to providing the precise definitions utilized in this work, introduces from a physical point of view the location Pareto distribution (LPD) and its expression from the δ parameter, which is interpreted as the displacement of the origin of the scale of measure in Eq. (3). When $\delta = 0$ is used, the LPD model matches the power law distribution, and consequently it is not included in GPD. This allows to use the same language for researchers working on this distribution in complex systems physics as for researchers working on the Pareto distribution in statistical theory of extreme values.

When a sufficiently high threshold is reached, LPD is confused with a power law distribution (see Fig. 1 and Proposition II.1) and is observed in practice, for instance, in Sec. V C on customers affected by electrical blackouts in the United States. The Pickands-Balkema–de Haan theorem in Sec. III provides a mechanism for generating the LPD. This theorem and Proposition II.1 also form a mechanism for generating power law distributions, making LPD a prominent alternative to the power law distribution (which is a particular case); thus it is crucial to know how to distinguish between the two models.

In IV A the likelihood ratio test between LPD and a power law distribution has been studied in detail, showing that the techniques proposed [17] to validate the power law distribution are especially weak in front of a LPD alternative; see Table III, Table IV, and Table V in the Appendix. Section V A, on daily rainfall accumulations in southwest England, and Sec. V C, on the numbers of customers affected in electrical blackouts, clearly show the different accuracy between the two tests; however, in Sec. V B, on Danish fire insurance, the two methodologies complement each other in the selection of the threshold. The examples frequently show the interest in examining them using different approaches.

The software on extreme value analysis that can be used to duplicate the outcomes found by various authors is reviewed in the final portion of Sec. III. Both the previous software and the software we built are used in the real-world examples that we investigate in Sec. V. We also include an evolution of the code by [26] to estimate the location Pareto model as well as for the likelihood ratio test introduced here.

[*]joan.delcastillo@uab.cat
[†]Alternate affiliation: Centre de Recerca Matemàtica, Cerdanyola del Vallès, Spain; pere.puig@uab.cat

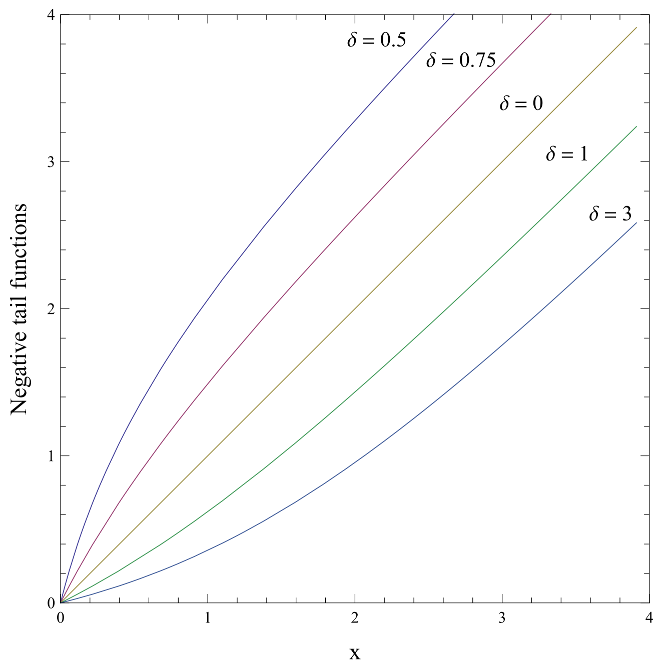


FIG. 1. Tail functions of a LPD against the tail function of a PL ($\delta = 0$) in log-log scale for $\mu = 1$, $\alpha = 1$, and several values of δ.

II. LOCATION PARETO MODEL

The classical Pareto (type I) distribution, following Arnold [3], is the two-parameter family of continuous probability density functions,

$$f_{\mathrm{PL}}(x;\alpha,\mu) = \frac{\alpha}{\mu}\left(\frac{\mu}{x}\right)^{\alpha+1} \quad (x \geqslant \mu), \tag{1}$$

where the shape parameter $\alpha > 0$ is known as the *tail index*, and the parameter $\mu > 0$ is the *minimum* value of the distribution support (μ, ∞). This distribution, originally introduced by [1], occurs in many situations of scientific interest and has significant consequences for understanding natural and man-made phenomena. In this article, following [16] and [17], we will call the distribution (1) a *power law*, and we will represent it by $\mathrm{PL}(\alpha, \mu)$. A random variable X power law distributed is expressed as $X \sim \mathrm{PL}(\alpha, \mu)$, and so on.

The main feature of this model is its characterization given by Pareto in 1896 [1] in terms of the linear relationship to the *log-log* scale of its *tail function* (complementary of the cumulative distribution function). Specifically, for $F(x) \equiv F_{\mathrm{PL}}(x;\alpha,\mu)$ it is fulfilled:

$$\log[1 - F(x)] = -\alpha\,\log(x) + \alpha\,\log(\mu). \tag{2}$$

From a physical point of view it is convenient to extend this model in order to include possible changes in the origin of the unit of measurement of x. This can occur, for instance, by measuring temperatures, where in degrees Celsius the freezing point is 0, in Fahrenheit 32, and in Kelvin -273.15. At this point, we begin to explore distribution functions $F_\delta(x)$ that satisfy the linear relation

$$\log[1 - F_\delta(x)] = -\alpha\,\log(x+\delta) + \alpha\,\log(\mu+\delta). \tag{3}$$

From this equation follows a characterization and parametrization of the Pareto type II distribution [3]. That is, the tail function that satisfies Eq. (3) must necessarily have the expression $F_\delta(x) = F_{\mathrm{LPD}}(x;\alpha,\mu,\delta)$ where

$$1 - F_{\mathrm{LPD}}(x;\alpha,\mu,\delta) = \left(\frac{\mu+\delta}{x+\delta}\right)^{\alpha} \quad (x \geqslant \mu), \tag{4}$$

with the probability density function

$$f_{\mathrm{LPD}}(x;\alpha,\mu,\delta) = \frac{\alpha}{\mu+\delta}\left(\frac{\mu+\delta}{x+\delta}\right)^{\alpha+1} \quad (x \geqslant \mu), \tag{5}$$

where the shape parameter $\alpha > 0$ will be called the *tail index*, the parameter $\mu \geqslant 0$ is the *minimum value* of the distribution support (μ, ∞). Since $f_1(\mu;\alpha,\mu,\delta) = \alpha/(\mu+\delta)$, it follows that $(\mu+\delta) > 0$ must be satisfied and $\delta > -\mu$ is the change in the unit of measurement's origin. Although negative μ values are allowed in Pareto type II, we do not address this option here. We will call the model (5) the *location Pareto distribution*, and we will represent it by $\mathrm{LPD}(\alpha, \mu, \delta)$. We remark that the δ parameter we have introduced can also be interpreted as the "distance" to the PL distribution, as we will see in (8).

We use here the *Pareto distribution*, which is also known as a Pareto type II distribution with $\mu = 0$ in [3], or the Lomax distribution [27]. It is denoted by $\mathrm{PD}(\alpha, \delta)$, and its probability

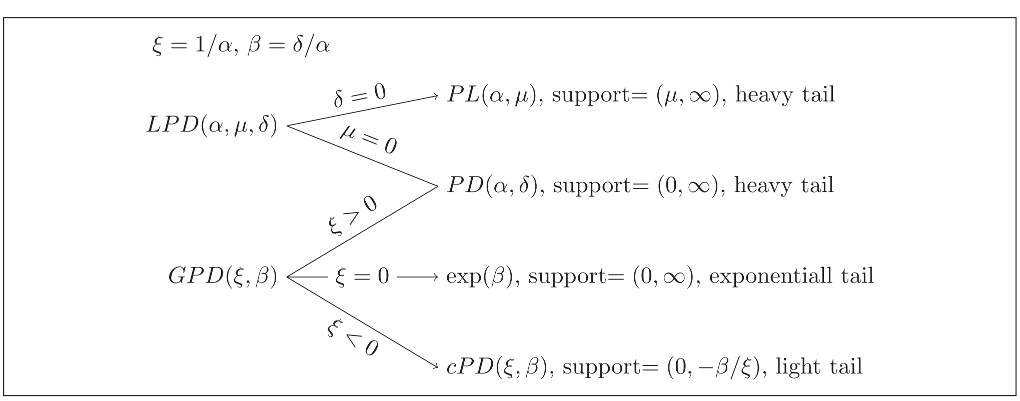


FIG. 2. Considered generalizations of the Pareto distribution (PD) in Sec. II and Sec. III.

density function is

$$f_{\rm PD}(x;\alpha,\delta)=\frac{\alpha}{\delta}\left(\frac{\delta}{x+\delta}\right)^{\alpha+1}\ (x\geqslant 0), \qquad (6)$$

where the shape parameter $\alpha > 0$ is the *tail index*, and $\delta > 0$ is scale.

Note that the LPD includes the Pareto distribution ($\mu = 0$) and the power law ($\delta = 0$). That is, we can write

$$\begin{aligned}\mathrm{PD}(\alpha,\delta)&=\mathrm{LPD}(\alpha,0,\delta)\\ \mathrm{PL}(\alpha,\mu)&=\mathrm{LPD}(\alpha,\mu,0).\end{aligned} \qquad (7)$$

The Pareto distribution is also the submodel of distributions with power decreasing tail functions of the generalized Pareto distribution, which will be discussed in the next section; see Fig. 2.

Remark II.1. The following result may appear to be a paradox: The transformation $Y = X - \mu$ for $X \sim \mathrm{PL}(\alpha, \mu)$ gives $Y \sim \mathrm{PD}(\alpha, \mu)$. That is, there is a one-to-one correspondence between a power law distribution (1) and Pareto distribution (6). However, the two statistical models are really different, since there is no unique transformation for all probability density functions of the model (the transformation depends on the parameter μ, which depends on the variable X). The estimation of the parameters is then not comparable.

For large x, in a log-log scale, the tail functions of LPD and PL look like parallel straight lines; see the negative tail functions in Fig. 1 for $\mu = 1$, $\alpha = 1$, and several values of δ. Only for $\delta = 0$ is the graph a true straight line, but for large x the lines look parallel with the origin's ordinate displaced by $\alpha\ \log(1 + \delta/\mu)$. The next proposition explains how the LPD model tends to the PL approach for large enough values, as suggested in Fig. 1.

Proposition II.1. For large values of x, when it tends to infinity, an LPD distribution resembles a PL distribution.

Proof. Taking the difference between (3) and (2)

$$\begin{aligned}\log[1-F_\delta(x)-\log[1-F(x)]&=-\alpha\ \log(1+\delta/x)\\ &\quad+\alpha\ \log(1+\delta/\mu),\end{aligned}$$

and from the Taylor formula and large x the last expression is approximately

$$=\alpha\ \log(1+\delta/\mu)-\alpha\,\delta/x+O\big(1/x^2\big)\approx\alpha\ \log(1+\delta/\mu).$$

Hence, for large enough x, (2) and (3) look like parallel straight lines.

A. Location, scaling, truncation, and residual distribution

The LPD is a family of probability distributions closed by location-scale transformations. That is, for any random variable X whose probability distribution function belongs to LPD, for any real ν and any positive λ, the random variable $Y = \lambda X + \nu$ belongs to LPD. The family is also closed taking *exceedances (truncation)* and*excesses (residual)* over a threshold, as will be reviewed in this section.

A power law distribution can be considered synonymous with a *scale-free* distribution; the equivalence is demonstrated in [16]. The requirement of scale free merely tells us that PL is closed by scale transformation. As the following proposition demonstrates, the result generalizes to LPD.

Proposition II.2. If X is $\mathrm{LPD}(\alpha, \mu, \delta)$ distributed, then, for any positive λ and real ν, $\lambda X + \nu$ is $\mathrm{LPD}(\alpha, \lambda\mu + \nu, \lambda\delta - \nu)$ distributed.

Proof. From (4), it is self-evident that λX is $\mathrm{LPD}(\alpha, \lambda\mu, \lambda\delta)$ distributed, for any positive λ. Hence, the proof can then be reduced to the case $\lambda = 1$.

Transforming the tail function of X, $\overline{F}_X = [(\mu+\delta)/(x+\delta)]^\alpha$, by $Y = X + \nu$, we obtain

$$\overline{F}_Y(y)=\left(\frac{\mu+\nu+\bar\delta}{y+\bar\delta}\right)^\alpha,$$

where $\delta = \nu + \bar\delta$ [note that $\bar\delta > -(\mu + \nu)$]. From this it follows that Y is $\mathrm{LPD}(\alpha, \mu + \nu, \bar\delta)$ distributed. ■

In particular, the above proposition says

$$\begin{aligned}\mathrm{LPD}(\alpha,\mu,\delta)+\delta&=\mathrm{PL}(\alpha,\mu+\delta)\\ \mathrm{LPD}(\alpha,\mu,\delta)-\mu&=\mathrm{PD}(\alpha,\mu+\delta)\\ \mathrm{PD}(\alpha,\delta)+\mu&=\mathrm{LPD}(\alpha,\mu,\delta-\mu).\end{aligned} \qquad (8)$$

The first formula in (8) provides an interpretation of the parameter δ as the distance from LPD to the PL. Equation (8) explains how by relocating a sample from a LPD we can obtain a sample of a power law (adding δ) or a sample of a Pareto distribution (subtracting μ). On the other hand, small displacements of a sample from a power law or a Pareto distribution always give a sample of a LPD. This could be the answer to the article's title question.

If X is $\mathrm{LPD}(\alpha, \mu, \delta)$ distributed, then $X/\mu \sim \mathrm{LPD}(\alpha, 1, \delta/\mu)$, as Proposition II.2 shows. Moreover, from (8), it follows

$$X\sim\mathrm{LPD}(\alpha,\mu,\delta)\Leftrightarrow(X+\delta)/(\mu+\delta)\sim\mathrm{PL}(\alpha,1). \qquad (9)$$

Definition II.1. Let X be a continuous non-negative random variable with distribution function $F(x)$. For any threshold $\nu > 0$ the distribution of *exceedances* of X conditional on $X > \nu$, denoted as $X_\nu = \{X \mid X > \nu\}$, is called the *truncated distribution* (of X over ν). The tail function of X_ν, $\overline{F}_\nu(x)$, is given by

$$\overline{F}_\nu(x)=\overline{F}(x)/\overline{F}(\nu). \qquad (10)$$

The distribution function of *excesses*$X - \nu$ conditional on $X > \nu$, denoted as $X_{+\nu} = \{X - \nu \mid X > \nu\}$, is called the *residual distribution* (of X over ν); see [10]. The cumulative distribution function of $X_{+\nu}$, $F_{+\nu}(x)$, is given by $F_{+\nu}(x) = F_\nu(x + \nu)$.

Proposition II.3. The truncated distribution of a LPD over a threshold $\nu > \mu$ is simply the change of μ by ν:

$$\mathrm{LPD}_\nu(\alpha,\mu,\delta)=\mathrm{LPD}(\alpha,\nu,\delta).$$

The residual distribution of a LPD over a threshold $\nu > \mu$ always leads to a PD distribution, according to the formula

$$\mathrm{LPD}_{+\nu}(\alpha,\mu,\delta)=\mathrm{LPD}(\alpha,0,\delta+\nu)=\mathrm{PD}(\alpha,\delta+\nu).$$

Proof. Equation (10) is simpler in terms of tail functions; then from (4)

$$\overline{F}_{\nu}(x) = \overline{F}(x)/\overline{F}(\nu) = \left(\frac{\mu+\delta}{x+\delta}\right)^{\alpha}\left(\frac{\mu+\delta}{\nu+\delta}\right)^{-\alpha}$$
$$= \left(\frac{\nu+\delta}{x+\delta}\right)^{\alpha} \sim \mathrm{LPD}(\alpha,\nu,\delta),$$

and in the same way

$$\overline{F}_{+\nu}(x) = \overline{F}(x+\nu)/\overline{F}(\nu) = \left(\frac{\nu+\delta}{x+\nu+\delta}\right)^{\alpha}$$
$$\sim \mathrm{LPD}(\alpha,0,\delta+\nu).$$

■

Since the parameters α and δ of LPD do not change by truncation, if $\delta = 0$ the first part of Proposition II.3 extends to $\mathrm{PL}(\alpha,\mu)$.

The Pareto distribution is closed by taking a residual distribution, from Proposition II.3, according to

$$\mathrm{PD}_{+\nu}(\alpha,\delta) = \mathrm{PD}(\alpha,\delta+\nu). \quad (11)$$

B. Estimation for LPD and PL

Given a sample of a random variable X $\mathrm{LPD}(\alpha,\mu,\delta)$ distributed of size n the log-likelihood function is

$$l(\alpha,\mu,\delta) = n\log(\alpha) - n\log(\mu+\delta) - (\alpha+1)\sum_{i=1}^{n}\log\left(\frac{x_i+\delta}{\mu+\delta}\right)$$
$$= n\log(\alpha) + n\alpha\log(\mu+\delta) - (\alpha+1)\sum_{i=1}^{n}\log(x_i+\delta), \quad (12)$$

provided $\mu < x_i$ for all observations in the sample.

From the second expression it is clear that with fixed α and δ the maximum of $l(\alpha,\mu,\delta)$ is attained at the minimum of the sample $\hat{\mu} = x_m = \min\{x_i\}$; in other words, μ should be as big as possible, but less than or equal to the minimum, x_m. Deriving the log-likelihood in the α direction we obtain

$$\partial_{\alpha} l = \frac{n}{\alpha} - \sum_{i=1}^{n}\log\left(\frac{x_i+\delta}{x_m+\delta}\right)$$
$$= 0 \;\Leftrightarrow\; \frac{1}{\alpha} = \frac{1}{n}\sum\log\left(\frac{x_i+\delta}{x_m+\delta}\right).$$

That is, $\hat{\alpha} = 1/\xi(\delta)$,where

$$\xi(\delta) = \frac{1}{n}\sum\log\left(\frac{x_i+\delta}{x_m+\delta}\right). \quad (13)$$

Substituting the estimated values into the log-likelihood function we obtain the *profile likelihood*

$$l_p(\delta) = l(\hat{\alpha},\hat{\mu},\delta)$$
$$= -n\{\log[\xi(\delta)] + \log(x_m+\delta) + \xi(\delta) + 1\}. \quad (14)$$

The $\hat{\delta}$ value that maximizes $l_p(\delta)$ provides the maximum of the log-likelihood. The maximum likelihood estimator (MLE) is $(\hat{\alpha},\hat{\mu},\hat{\delta})$, and the log-likelihood value at this point is $l_p(\hat{\delta})$, where

$$\hat{\mu} = \min\{x_i\},\;\; \hat{\alpha} = 1/\xi(\hat{\delta}),\;\; \hat{\delta} = \arg\{\max\{l_p(\delta)\}\}.$$

In particular, making $\delta = 0$ on Eq. (13), the MLE of the parameters of a $\mathrm{PL}(\alpha,\mu)$ distribution is $\hat{\mu} = x_m$ and $\hat{\alpha}_0 = 1/\xi(0)$.

On the other hand, the second formula of (8) says that subtracting μ from the data of a $\mathrm{LPD}(\alpha,\mu,\delta)$ we get a $\mathrm{PD}(\alpha,\mu+\delta)$. Then, once the data are relocated to 0, subtracting μ, we can use a code that estimates the parameters of the Pareto distribution $\mathrm{PD}(\alpha,\sigma)$, with $\sigma = \mu+\delta$. For instance, developing the code by [26], the following R code provides the algorithm we use to obtain a MLE of the parameters of a LPD from a numerical standpoint:

```
# MLE estimation for LPD
eLPD<-function(xdat, thres = NA){xm<-ifelse(!is.na(thres), thres, min(xdat));
xdt<-xdat[xdat >= xm]; max<-max(xdt);
xi<-function(delta){mean(log((xdt + delta)/(xm + delta)))};
int<-c(-xm, 100 * max);
fp<-function(delta){-length(xdt) * (log(xi(delta) * (xm + delta)) + xi(delta) + 1)};
out<-optimize(fp, interval = int, maximum = T);
delta<-out$maximum; loglik<-out$objective;
list(alpha = 1/xi(delta), min = xm, delta = delta, lLPD = loglik)}
```

Note that the functions *xi* and *fp* in the code correspond to Eqs. (13) and (14), respectively.

Sometimes the threshold is known and does not need to be estimated. It is also not necessary to know μ to compare PL and LPD, because in both cases the estimate is the same, $\hat{\mu} = \min\{x_i\}$. However, in most examples determining μ is not an estimation problem but a threshold selection problem, as we will see in the Sec. IV C.

On the other hand, to calculate confidence intervals for the α and δ parameters of LPD we can reduce to the PD distribution, using the second formula of (8) as we have already done, and we continue to work conditioning by $\hat{\mu}$. The expected Fisher information matrix for PD(α, σ) is calculated directly from the log-likelihood function and can be found in [3, Ch. 5]:

$$I(\alpha, \sigma) = \begin{pmatrix} \alpha^{-2} & -(\alpha+1)^{-1}\sigma^{-1} \\ -(\alpha+1)^{-1}\sigma^{-1} & \alpha(\alpha+2)^{-1}\sigma^{-2} \end{pmatrix}.$$

The expected Fisher information matrix for the α and δ parameters of LPD with the change ($\alpha = \alpha$, $\delta = \sigma - \mu$) is exactly $I(\alpha, \delta)$, and the standard errors turn out to be

$$\begin{aligned} \text{se}(\widehat{\alpha}) &= \alpha(\alpha+1)/\sqrt{n}, \\ \text{se}(\widehat{\delta}) &= (\alpha+1)(\alpha+2)^{1/2}\alpha^{-1/2}\delta/\sqrt{n}. \end{aligned} \tag{15}$$

Finally, to quantify the uncertainty for $\hat{\mu}$ a parametric *bootstrap* method can be used. Note that, as mentioned in Sec. IV C, applications frequently face a threshold selection problem for μ rather than an estimating problem.

III. GENERALIZED PARETO DISTRIBUTION AND FITTING TECHNIQUES

A relevant application of Pareto's ideas emerged in the 1970s with the fundamental results in extreme value theory by Pickands [7] and Balkema and de Haan [8]. The Pickands-Balkema–de Haan theorem (PBdH) characterizes the asymptotic behavior of the residual distribution over a high threshold under widely applicable regularity conditions; see McNeil *et al.* [11, Ch. 7] and Coles [10, Ch. 4]. The generalized Pareto distribution (GPD) examined in this section is exactly the limit distributions determined by the PBdH theorem. Because all of the GPD distributions are supported on positive numbers with the minim value $\mu = 0$, neither the PL nor the LPD is included in the GPD model; see Fig. 2.

The cumulative distribution function of a GPD is given by

$$F_{\text{GPD}}(x;\xi,\beta) = \begin{cases} 1-(1+\xi\, x/\beta)^{-1/\xi}, & \xi \neq 0 \\ 1-\exp(-x/\beta), & \xi = 0 \end{cases}, \tag{16}$$

where $\xi \in \mathbb{R}$ is called the *extreme value index* and $\beta > 0$ is a scale parameter. If $\xi < 0$ the distribution has bounded support $0 \leqslant x \leqslant -\beta/\xi$, and if $\xi \geqslant 0$, it is $x \geqslant 0$.

If $\xi < 0$, we say that the distribution is *light tailed*. If $\xi = 0$, the GPD is the exponential distribution with mean β. When $\xi > 0$ a GPD(ξ, β) is a PD(α, δ) where $\alpha = 1/\xi$ and $\delta = \beta/\xi$, in this case the distribution has a power decreasing tail function and finite moments of order k if $k < \alpha$, and we say that it is *heavy tailed*. The mean of a GPD is $\beta/(1-\xi)$, and the variance is $\beta^2/[(1-\xi)^2(1-2\xi)]$ provided $\xi < 1$ and $\xi < 1/2$, respectively.

The (ξ, β) parametrization, due to Von Mises [28], has the outstanding property of unifying in a single model the three different behaviors that can present the tails of the distributions: heavy, exponential, and light tailed. Then, when estimating the parameters, the type of distribution does not need to be specified *a priori*. By extension we will call heavy, exponential, and light-tailed distributions those that have by limit a GPD distribution with parameter ξ positive, zero, or negative, respectively.

The next result shows a duality of heavy and light tails (changing from ξ to $-\xi$) that is very interesting and will be used for the examples in the Sec. V. Castillo *et al.* [29] introduced it; however, the demonstration is placed here because it hasn't been published yet.

Proposition III.1. If $\xi > 0$, $\beta > 0$ and $\delta = \beta/\xi$, then a random variable X is GPD(ξ, β) distributed if and only if its transformed variable $t(X) = \delta X/(X+\delta)$ is GPD($-\xi, \beta$) distributed.

Proof. For $\delta > 0, t(x) = \delta\, x/(x+\delta)$ is a strictly increasing function that applies $(0, \infty)$ to $(0, \delta)$. Its inverse is $t^{-1}(z) \equiv x(z) = \delta\, z/(\delta - z)$, from $(0, \delta)$ to $(0, \infty)$. Let $Z = t(X)$, its tail function is

$$\begin{aligned} \overline{F}_Z(z) = \overline{F}_{gpd}(x(z)) &= \left(1 + \frac{\xi\, z}{\beta - \xi\, z}\right)^{-1/\xi} \\ &= \left(\frac{\beta}{\beta - \xi\, z}\right)^{-1/\xi} = (1 - \xi\, z/\beta)^{1/\xi}, \end{aligned}$$

that is, the tail function of a GPD($-\xi, \beta$). The converse is a consequence of the same method applied to $z(x) \equiv t(x)$. ■

Under our parametrization, the Pareto distribution (PD) of Eq. (6) appears as a GPD(ξ, β) with $\xi > 0$,

$$\begin{aligned} &\text{PD}(\alpha, \delta) = \text{LPD}(\alpha, 0, \delta) = \text{GPD}(\xi = 1/\alpha, \beta = \delta/\alpha), \\ &\quad \alpha > 0, \delta > 0. \end{aligned} \tag{17}$$

Similarly, after Proposition III.1, the GPD with $\xi < 0$ can be thought of as the complement of the PD and denoted cPD; see Fig. 2. It is also observed that PD generalizes to LPD by including the location parameter μ and generalizes to GPD by allowing parameter ξ to be zero or negative.

All distributions in LPD have power decreasing tails (or heavy tails) and support (μ, ∞), see (4). Proposition II.3 shows that the residual distribution of a LPD(α, μ, δ) over *any* threshold is exactly the PD(α, δ) (its asymptotic limit). All of the examples we will look at in this paper have been conjectured to have power decreasing tails. Then *asymptotically* the residual distribution for a large enough threshold μ must be a PD, that is, a GPD with $\xi > 0$, by the PBdH theorem. Equivalently the truncated distribution above this threshold (relocated at μ) must be as a LPD. This last point gives mathematical support to LPD, providing a natural asymptotic mechanism for the observation of LPD in real-world examples.

Remark III.1. Gnedenko [6] characterized the domains of attraction of the maxima in terms of the regularly varying functions, properly extending to the tails. The approximations obtained from Gnedenko and the PBdH theorem are not the same [11, Ch. 7]. A regularly varying function, which approximates a distribution in the Fréchet attraction domain, is the product of a PL(α, 1) with a slowly varying function, unknown in practice. The PBdH theorem is more accurate because the asymptotic approximation of the residual distribution in this case is exactly a PD distribution ($\xi > 0$) and more generally a GPD. The approximation that is derived usually in practice for a truncated distribution for a large enough threshold, say, μ, is in the first case a PL(α, μ) distribution for $x > \mu$, since the slowly varying function is forgotten, and, in the second case, a LPD(α, μ, δ) distribution for $x > \mu$. In the first case the data are fitted for a family of a single parameter and in the second case for a biparametric family that contains the previous one and will have a better adjustment. Proposition II.1 explains how the LPD approach leads to the PL approach.

A review of the available software on extreme value analysis appears in [30]. R software [31] contains some useful packages for dealing with EVT. The R package *evir* [32] provides maximum likelihood estimation at the same time for the block maxima and threshold model approaches. The R package *poweRlaw* [33] enables power law and other heavy-tailed distributions to be fitted using the techniques proposed by [17]. The R package *ercv* [34] based on the residual coefficient of variation is a complement, and often an alternative, to the available software on EVT; see [29]. The R package *ismev* [35] allows fitting parameters of a generalized Pareto distribution depending on covariates and offers diagnostics such as qqplots and return level plots with confidence bands.

IV. TESTING THE POWER LAW HYPOTHESIS

Clauset *et al.* [17] in the abstract of their paper write, "Commonly used methods for analyzing power law data, such as least-squares fitting, can produce substantially inaccurate estimates of parameters for power law distributions, and even in cases where such methods return accurate answers they are still unsatisfactory because they give no indication of whether the data obey a power law at all." They introduce a methodology based on an estimate of the parameters by MLE to determine the distance to the model by Kolmogorov-Smirnov distance and validate the PL model with a bootstrap procedure. These techniques, also called here the Clauset methodology, are available in the R package *poweRlaw* [33].

The PBdH theorem gives an asymptotic mechanism for generating the LPD that is a natural extension of PL and that for a high threshold performs similarly to PL; see Proposition II.1. Hence, LPD is a prominent alternative to PL, and it is crucial to distinguish between the two alternatives. Clauset's methodology has been very helpful to identify when the PL model is acceptable in view of the numerous occasions in which roughly straight-line behavior on a doubly logarithmic plot for the cumulative distribution function has been observed, usually only for values above some lower bound. These techniques can be applied to testing PL against an alternative LPD, but the results are unsatisfactory, as we will see later.

A. Likelihood ratio test for PL against alternative LPD

The likelihood-ratio test (LRT), which can be justified by the Neyman-Pearson lemma, is the best way to test the goodness of fit of a nested competing model under regularity conditions. It can be used to test the null hypothesis of PL against the alternative LPD with the *deviance statistic*

$$D = 2[l(\hat{\alpha}, \hat{\mu}, \hat{\delta}) - l(\hat{\alpha}_0, \hat{\mu}, 0)], \tag{18}$$

where the log-likelihood function (12) is evaluated in $(\hat{\alpha}, \hat{\mu}, \hat{\delta})$, the maximum likelihood estimator for LPD, and in $(\hat{\alpha}_0, \hat{\mu}, 0)$, the maximum likelihood estimator for PL; see [10, Sec. 2.6.5] or [36, Sec. 8.2]. The asymptotic distribution of the deviance statistic is in this case a chi-square distribution with one degree of freedom, $D \sim \chi_1^2$. Following a detailed simulation examination, it is evident that for sample sizes larger than 100 and α greater than 2, the asymptotic distribution is a good approximation. The critical values are obtained through simulation for smaller sample sizes and various α values.

The approximate critical points are obtained by simulation of the PL distribution, computing the LRT from the LPD and PL estimates. Since the test is invariant for monotone transformations the calculations made for $\mu = 1$ are valid for any other value. The simulations are all run with 10^5 samples. Table II in the Appendix shows the critical points for the LRT (18) for values of the α parameter 1, 2, 4, and 8, and for sample sizes 10, 25, 50, 100, 250, and 500, corresponding to the 90th and 95th percentiles, as well as the values obtained with the asymptotic distribution. In most practical examples α is between 1 and 8, and above this value the exponential model is already a comparable alternative.

The power estimates are obtained by simulation of the LPD distribution with 10^4 samples, testing PL against LPD by LRT at significance level $p = 0.10$. Tables III and IV show the power of the LRT for values of the α parameter 1, 2, 4, and 8, values of the δ parameter -0.5, 1, 2, 4, and 8, and for sample sizes 25, 50, 100, 250, and 500. The power increases with δ (distance to the null hypothesis) and the sample size, n, and decreases slightly with α. The power is greater than 80% in all cases for samples of size 100 or larger and for delta values equal to or greater than 4.

While being widely used, Clauset's technique is not free from criticism. For instance, Ref. [37] mentions that the algorithm inappropriately rejects PL in a particular case. We obtained guarantees for the procedure while creating Table V by simulating the empirical significance levels and achieving values that did not surpass the nominal level of significance ($p = 0.10$). However, the techniques in [17] are very insensitive to differentiating between PL and LDP (see Table V) in front of the high levels of power obtained with the LRT, as shown in Tables III and IV. The power of Clauset's methodology is similar to the level of significance for α parameters and δ parameters between 2 and 8, and sample size $n = 100$, and is therefore useless. The power of Clauset's methodology is less than 20% and for LRT more than 77% in the same range of parameters for sample size $n = 250$, Values become less than 26% and higher than 96% for LPD for sample size $n = 500$.

B. Validating the LPD hypothesis

Clauset's methods can be used to verify that a dataset is suitable for the PL model. The stronger power of the likelihood ratio test, on the other hand, may lead to the rejection of PL in favor of LPD for the same dataset, but this does not guarantee that the LPD model is true. It would also be necessary to test the LPD hypothesis.

For a nonnegative random variable X, the coefficient of variation is the ratio of the standard deviation to the mean, $\mathrm{CV}(X) = \mathrm{sd}(X)/\mathrm{mean}(X)$. The coefficient of variation offers the best comparison test between the exponential distribution and the truncated normal alternative [25] and the best compar-

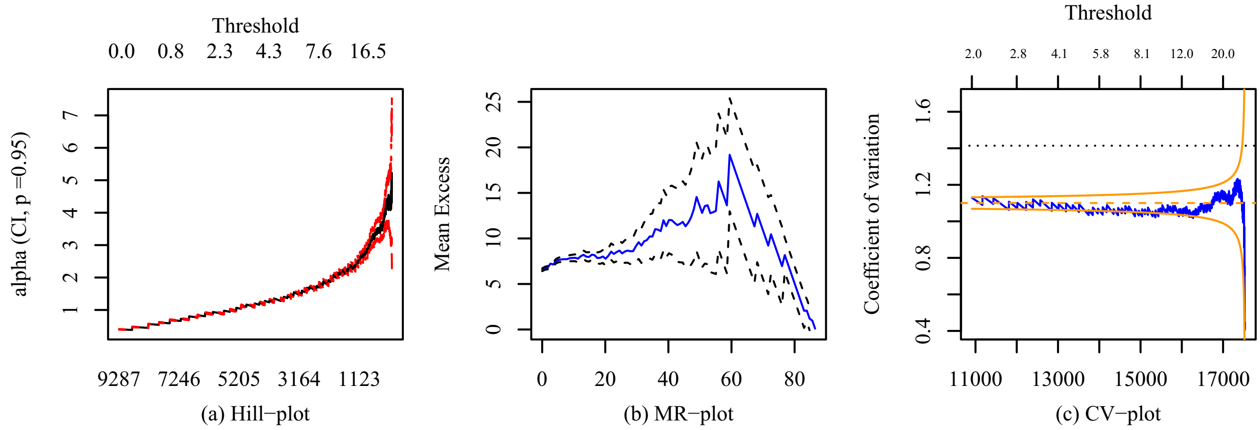


FIG. 3. Exploratory data analysis for threshold selection on the daily rainfall accumulation data with the graphs of Hill, residual mean, and residual coefficient of variation.

ison test between a power law and the truncated log-normal alternative [23].

The *residual* coefficient of variation, $\mathrm{cv}(t)$, was introduced in [38,39] as a tool for describing data sets with tail behavior similar to GPD. This is the coefficient of variation of the residual distribution of X over the threshold $t \geqslant 0$; see Sec. II A. If the raw data consist of a sequence of independent and identically distributed measurements $x_1, \ldots, x_n$, extreme events can be detected using the *sample* $\mathrm{CV}(t)$,

$$t \to \mathrm{cv}(t) = \mathrm{sd}\{x_j - t \mid x_j > t\}/\mathrm{mean}\{x_j - t \mid x_j > t\}. \tag{19}$$

The graph of $\mathrm{cv}(t)$ with pointwise error limits is called a CV plot; see Fig. 3(c) and Fig. 5(c). The R package *ercv* provides exploratory data analysis, a threshold selection method, and model validation approaches established in [29,38,39], from the asymptotic distribution of the $\mathrm{cv}(t)$ as a random process indexed by the threshold. This practice is referred to here as*residual CV methodology*.

The LPD hypothesis can be tested using the residual CV methodology (*Tm* function in R package *ercv*), with the following reasoning: Given a sample that comes from an LPD distribution, the residual distribution from any value, for example, the minimum of the sample, must be PD, by Proposition II.3. Then PD is contrasted with the sample shifted to the origin. If the contrast accepts PD, any displacement of this distribution is LPD, by Proposition II.2.

C. Threshold selection algorithms

In general there are few doubts about the estimation methods and algorithms to implement them when the model with which the data are analyzed is well defined. The difficulty arises in specifying the model in the context of the various Pareto models and, especially, because it is usually the case that empirical data, if they follow some kind of Pareto distribution, do so only for values above some unknown lower bound. The standard practice in threshold selection is to adopt as low a threshold as possible, subject to the fact that the Pareto approximation is reasonable for the selection of higher data.

The threshold selection is an important aspect in the asymptotic approximation by Pareto distributions, but there is no definitively established algorithm. The Hill method is the best-known procedure and one of the best-studied estimators in the EVT literature. The Hill estimator is exactly the MLE estimator for the PL model (Sec. II B), and the Hill-plot *hill* function in R package *evir* gives the estimates for various values of the truncated ordered sample. The general strategy for selecting the threshold is to find a stable region where the estimates of different order statistics are quite similar. Novak [15] recommends that, in these cases, one select the threshold at which the estimator achieves the average value of the estimators in the stable region because, despite fluctuations in the region's boundary, the resulting threshold will be nearly the same.

Coles [10] recommends using GPD distributions with MLE methods to extrapolate from observed levels to unobserved levels of a dataset of independent observations above a threshold, and he uses a *mean residual* MR-plot (*mrl.plot* function in R package *ismev*) for threshold selection. He proposes revising threshold selection by locating a stable region where estimates of the shape and modified scale parameters over a range of thresholds are produced (*gpd.fitrange* function in the *ismev* R package). Instead, in some cases, he stays with the original decision despite the evidence to the contrary; see Sec. V A. In [40] an automated threshold selection technique is developed by locating a stable region of the shape and modified scale parameters using the Pearson's chi-square test of goodness of fit for the normality of the observed differences.

The techniques of Clauset *et al.* [17] present an objective method of discerning and quantifying PL behavior in empirical data. The first step of their methodology is to choose the threshold μ that minimizes the Kolmogorov-Smirnov distance between the truncated ordered sample above this value and the PL distribution (*estimate_xmin* function in R package *poweRlaw*). This estimated value is considered as the threshold that selects the subset of data that could be fitted to a PL model. This method can be considered free of subjectivity.

Usually, the selection of an appropriate threshold is performed on a visual basis of several graphs, each with pointwise error limits for a wide range of thresholds. Hence, some kind of global test is needed to avoid the risk of misuse for multiple testing, such as the *Tm* function in R package *ercv* already mentioned. In [38] an automatic threshold selection algorithm is provided that truly reduces the multiple testing problem (*thrselect* function in R package *ercv*). An automatic method is an "objective" test that helps researchers. Automatic methods need not be used in an uncritical way; they can of course be used as a starting point for fine tuning. It may well be that a subjective approach is in reality the most useful one.

D. Tail plots

Before fitting Pareto models to a dataset, it is common to run an exploratory data analysis to see if the linear relationship (2) holds empirically. For this it is sufficient to draw on a log-log scale the empirical tail funcition. Keep in mind that the LPD model is more general than the PL model and that it is likewise expressed on a log-log scale with a linear relationship (3). It is convenient to draw the locus of points

$$\left\{\left[\log\left(\frac{x_{r,n}+\hat{\delta}}{\hat{\mu}+\hat{\delta}}\right), \log\left(1-\frac{r}{n+1}\right)\right], \quad r=1,2,\dots,n\right\}, \tag{20}$$

where $x_{r,n}$ is the r-order statistic of a size n dataset with a lower limit μ, to which a displacement δ of the origin has been applied. This graph is used to visualize if the data can correspond to an LPD model, that in the case $\delta = 0$ would be the PL; see Eq. (9). In the general situation, the LPD model's parameter δ must be estimated, which can be done using the *evir* or *ismev* packages we discussed before, or by choosing the δ value that produces the greatest correlation between the two variables of Eq. (20).

Sometimes a roughly straight-line behavior on a doubly logarithmic plot with $\delta = 0$ is observed only for values above some lower limit. This could lead to ignoring data below this limit and adjusting a PL for the sample above the lower limit. It can also happen that estimating δ the graph shows straight-line behavior for all data, suggesting the LPD adjustment for the whole sample, and avoiding removing an important part of the sample. This is shown in Sec. V C as well as Fig. 5.

However, in practice, identifying Pareto distributions by the approximately straight-line behavior of a sample is not enough. It should always be interpreted and validated using models, using either Clauset's approach for PL, the LRT, the residual CV methodology for LPD, or better yet, all options at once.

V. DATA ANALYSIS OF REAL-WORLD EXAMPLES

Three well-known examples will be discussed since they show disparities in the application of the approaches described, emphasizing the importance of weighing the different indicators before deciding on the final model. The analyzed examples present distributions that have been conjectured to have heavy tails. For this reason, no previous studies have been presented to rule out exponentiality or light tails.

As explained in part by Proposition II.1 and Fig. 1, it is habitual for LPD to be accepted with a low threshold and PL to be accepted only for a high threshold. Adopting a model with little data, on the other hand, is easier than accepting one with a large number. As a result, the more data with which we can accept a model, the more evidence we have that it is valid. We usually specify the standard error in parenthesis for the the estimators of parameters. However, when referring to a threshold next to a value, we usually specify the number of observations in parentheses that are larger (*exceedances*), for instance, "$\mu = 30$ (152 exceed)." Standard errors are calculated using this sample size from (15).

A. Daily rainfall

This example is based on the rainfall dataset *rain* in the R package *ismev*. The vector contains 17 531 daily rainfall accumulations in mm (48 years) at a location in southwest England over the period 1914 to 1962. These data form part of a study made in [41] and was discussed in [10, Sec. 4.4]. The author, in the latest publication, states that a *residual mean MR plot* for these data suggests a threshold of $\mu = 30$ (152 exceed); see Fig. 3(b). Then the parameters of LPD are estimated by $\hat{\alpha} = 5.420$ (2.822), $\hat{\delta} = 10.327$ (6.292).

Clauset 's methodology indicates that the dataset is consistent (at the tail) with the PL hypothesis, with p-value 0.63, for the threshold $\mu = 27.7$ (210 exceed) and tail index estimated by $\hat{\alpha}_0 = 4.170$ (1.487). The chosen threshold is very close to the one suggested in [10, Sec. 4.4], and the LRT does not reject PL against LPD. On the other hand, PL is rejected when the LRT is applied to $\mu = 25$ (284 exceed), with a p-value of 0.047, and to $\mu = 20$ (554 exceed), with a p-value of 0.001. These details cannot be disclosed with the Clauset technique. The Hill plot, Fig. 3(a), explains why PL has been rejected with lower thresholds, since the evolution of the MLE of α parameter for the PL model as a function of the threshold grows steadily and stabilizes only in a "small" interval in the part top right of the graph.

The methodologies of Coles (package *ismev*) and Clauset (package *poweRlaw*) are coincident with respect to threshold selection, although in the first case the LPD model is estimated and in the second the PL model. In the first case the choice of the threshold is a subjective interpretation of the interval where a linear growth of the MR plot is observed; see Fig. 3(b). In the second, is derived from an automatic algorithm that may be used to any dataset.

The residual CV methodology (package *ercv*) takes a different approach. The plot of the residual CV above the threshold $\mu = 2$ [Fig. 3(c)] shows a stable behavior that is equivalent to the fact that the α estimate in the LPD model is also stable. This fact can be verified in [10, Fig. 4.1], or using function *gpd.fitrange* in R package *ismev*, where the MLE estimate is approximately constant. Nevertheless, surprisingly, the author accepts the threshold $\mu = 30$, contrary to his own recommendations; see Sec. IV C.

The LPD hypothesis can be tested as explained in Sec. IV B. The *Tm* function (package *ercv*) is applied from $\mu = 2$ (6619 exceed) averaging the CV for 100 approximately equally separated thresholds. The test allows us to accept LPD, with the p-value 0.245. The MLE estimator is $\hat{\alpha} = 7.527$ (0.789) and $\hat{\delta} = 42.671$ (5.031). The LRT does not deny that the data are supported by a PL above the threshold $\mu = 30$. This is clearly compatible with accepting LPD above the

TABLE I. The parameter estimates (in bold) for daily rainfall accumulations made with three methodologies, as well as estimates of the expected value over thresholds for a period of 48 years. To provide confidence intervals, the models have been reestimated from $\hat{\alpha}$ estimates plus (up) and minus (down) twice its standard error.

			LPD parameters			Expectation over threshold			
Method	Model	size	$\hat{\alpha}$	μ	$\hat{\delta}$	60	70	80	100
Coles	up2	152	11.06	30.0	58.51	6.0	2.5	1.1	0.2
Coles	**at**	**152**	**5.42**	**30.0**	**10.33**	**7.5**	**3.6**	**1.9**	**0.6**
Coles	down2	152	−0.22	30.0	−33.95				
Clauset	up2	210	7.14	27.7	0.00	0.8	0.3	0.1	0.0
Clauset	**at**	**210**	**4.17**	**27.7**	**0.00**	**8.4**	**4.4**	**2.5**	**1.0**
Clauset	down2	210	1.19	27.7	0.00	83.4	69.4	59.1	45.3
Resid. CV	up2	6619	9.11	2.0	53.31	9.7	4.5	2.2	0.6
Resid. CV	**at**	**6619**	**7.53**	**2.0**	**42.67**	**12.6**	**6.3**	**3.0**	**1.1**
Resid. CV	down2	6619	5.95	2.0	32.10	17.9	9.7	5.6	2.1
Empirical		17 531		0.0		6	5	3	0

$\mu = 2$, as explained in Proposition II.1 and Fig. 1. However, accepting the null hypothesis of a test done with 6619 data points gives more credibility than accepting the hypothesis of another test with 152 or 210. The standard error of the estimated parameters for LPD clearly shows an improvement in accuracy with sample size.

Table I shows (in bold) the parameter estimates for daily rainfall accumulations made with the three methodologies we call "Coles," based on the classic EVT methodology, "Clauset," and "residual CV," as well as estimates of the expected value of observations over the 60, 70, 80, and 100 mm thresholds for a period of 48 years, equal to the observed dataset; see Sec. IV C. Moreover, to provide confidence intervals, we have reestimated the models and expected values from the estimates of $\hat{\alpha}$ plus (up) and minus (down) twice its standard error. The last row of the table corresponds to the observations made empirically, the maximum of which is 86.6 mm.

When we compare with the empirical results in Table I, we can see that the estimations of the expected value for 70 and 80 mm are similar in all three models, allowing us to extrapolate above 100 mm. The Coles model, on the other hand, understates the expected values slightly and shows a negative value $\alpha = -0.22$ when its estimation is reduced by two standard errors. This result is not compatible with the LPD model and is interpreted within the GPD model leading to the estimation of a light tail model with a support bounded by 33.95 mm ($-\delta$), which makes it impossible to set values above this threshold The Clauset model, which assumes the PL distribution ($\delta = 0$), provides very wide confidence intervals for the expected values over the thresholds. The limited sample size explains these findings, of course.

The first two methodologies use 152 and 210 observations that represent the values above the empirical quantiles 99.13% and 98.80%. With these models we estimate the observations above the quantiles 99.96 (60 mm), 99.97 (70 mm), and 99.98 (80 mm). These models cannot be used to estimate values above the usual quantiles 90, 95, 97.5, and 99. On the other hand, the 6619 observations (larger than 2 mm) represent 62.24% of the sample and allow us to estimate with the residual CV methodology the quantiles for these probabilities as well. It represents well in general the daily accumulation of rainfall and can be used to see if significant differences are observed over time.

B. Danish fire insurance

The Danish fire insurance dataset, "*danish*" in the R package *evir*, is a highly heavy-tailed, infinite-variance model used to illustrate the basic ideas of extreme value theory; see [9], [11, Example 7.23], and [15, Example 9.8]. Using Clauset 's methodology, the LRT, and the results of [15] the study on dataset *danish* in [29] is expanded below.

In the last work [29] is reported that the automatic threshold selection algorithm from *ercv* (*thrselect* function) chooses for *danish* the threshold 9.2 (116 exceed) with the estimate $\hat{\alpha} = 2.242$ (0.675). In this case, the technique of Proposition III.1 was applied on data that had been transformed using the strictly increasing function *tdata* (*ercv*) because this duality allows us to work with finite moments. The result is compared to that obtained in [11] by MLE above threshold 10 (109 exceed) with the estimate $\hat{\alpha} = 2.0$ (0.56). Although they are not the same, the results are not significantly different. Figure 4 compares the estimation of the complementary cumulative distribution function in log-log scale to the previously discussed estimations for Danish fire insurance data.

The tail plot of Eq. (20) presents for the whole dataset *danish* a roughly straight-line behavior. The Clauset's methodology indicates that the dataset is consistent with the PL hypothesis (p-value 0.52), for the threshold $\mu = 1.375$ (1564 exceed), and the LRT clearly accepts PL against LPD (p-value 0.90) with the estimate $\hat{\alpha}_0 = 1.403$ (0.085). Surprisingly this value of $\hat{\alpha}$ practically coincides with the estimate obtained in [15, Example 9.8] of $\hat{\alpha} = 1.41$ with the *ratio estimator* introduced there; Fig. 4 also compares all approximations. These latest results, acquired with 1564 exceedances in a thoroughly studied example, introduce a unique approach. Extrapolating with the PL model appears to be preferable to the LPD model in this case.

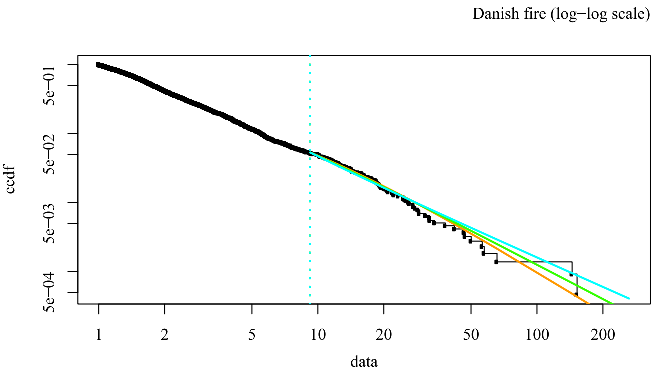


FIG. 4. Empirical tail function of Danish fire insurance data in log-log scale, adjusted first from McNeil (lower curve), second from residual CV (middle curve), and third, simultaneously, from Novak and Clauset's methodology (upper curve).

C. Electrical blackouts

The dataset consists of 210 observations of the numbers of customers affected in electrical blackouts in the United States between 1984 and 2002. The values range from 646 to 7 499 000 with a median of 91 000. This example was studied in [17], which does not reject PL for observations above 229 000 (59 exceed) with p-value 0.62.

The tail plot (20) in Fig. 5(a) for the whole dataset does not behave in a straight line, therefore PL can be considered only on a higher threshold. In contrast, the tail plot in Fig. 5(b) with parameter δ estimated for model LPD shows for the whole dataset a roughly straight-line behavior, corresponding to Eq. (3). It is well known that empirical observation alone is insufficient to validate LPD, and an appropriate technique should be used.

When the computations are repeated with the R package *powerRlaw*, PL is accepted with the estimated threshold of 93 285 (104 excess) and p-value of 0.61. The LRT applied to the 104 exceedances rejects PL with a p-value of 0.002, but PL is accepted with LRT applied to the 59 exceedances (p-value 0.68), with the estimate of $\hat{\alpha}_0 = 1.27$ (0.375).

This example has very heavy tails, and the methodology of the residual coefficient of variation (*ercv*) must be applied again by transforming the data with *tdata* function (*ercv*); see Proposition III.1. Then the multiple thresholds test *Tm* accepts LPD for the whole 210 observations (p-value 0.799). The MLE for LPD parameters are $\hat{\alpha} = 1.764$ (0.337), $\delta = 201\,504$ (56 146), for the whole dataset with minimum value $\mu = 646$ (210 exceed). Figure 5(c) shows a stable behavior of the residual CV for the whole sample, according to what the *Tm* test suggests.

The methodology of the residual CV guarantees adjust ingLPD to all dataset, and Clauset's methodology guarantees adjusting PL to the 59 largest values, above the threshold 229 000. Figure 5(a) shows the tail plot for the whole dataset, a dotted vertical line at the minimum of the 59 largest values, and the fit of these exceedances by PL. Morever, the dashed vertical line at $\delta = 201\,504$ relates the displacement required to adjust LPD to the threshold from which PL adjustment is possible (the x-axis point 229 000/646 and 201 504/646, normalizing to the minimum). Note that to adjust PL we must discard 151 observations below the minimum of exceedances, and to adjust LPD we can use all the data and simply move them by δ (almost the same amount) to obtain a linear relationship to log-log scale as shown in Fig. 5(b).

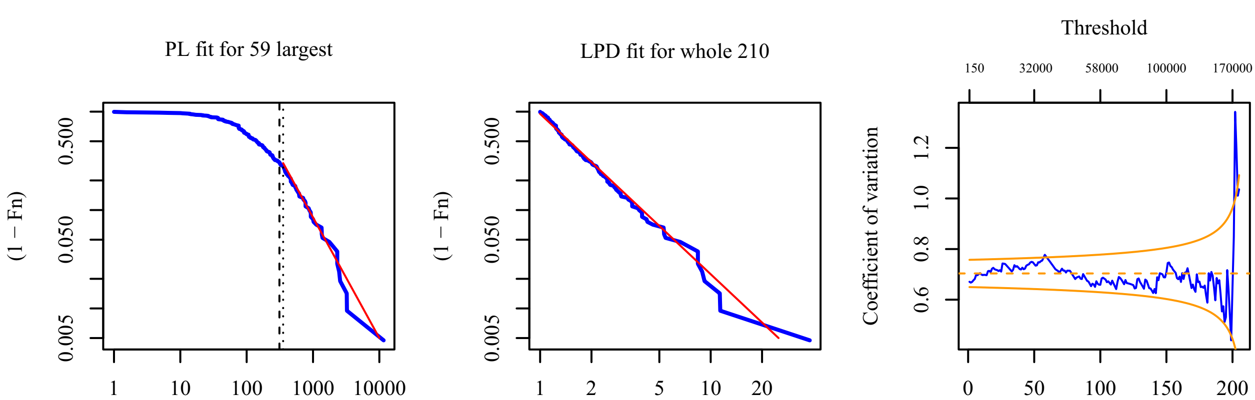
The contribution of the LPD against PL model is shown in this example, especially in the first two plots in Fig. 5. The first model is broader, based on an asymptotic mathematical solution that can be validated with a larger sample size (making it more representative), and fits the PL model for higher thresholds in real-world examples with power decreasing tails.

VI. CONCLUSIONS

(1) By adding an additional physical parameter to the power law distribution (PL) the LPD model (location Pareto distribution) is introduced. Moreover, LPD is a generalization of the Pareto distribution (PD) that is not part of the GPD.

(2) LPD allows for comparison between GPD (extreme value theory) and PL at tail models. When a sufficiently high threshold is considered, PL and LPD are confused.

(3) The likelihood ratio test clearly differentiates between LPD and PL and shows an improvement over Clauset's methodology.



(a) Tail plot, delta = 0 (b) Tail plot, delta estimated (c) CV−plot

FIG. 5. Tail plots for the electrical blackouts data with parameter $\delta = 0$ (a) and with parameter δ estimated (b). The residual CV for the whole data shows a stable behavior (c).

(4) The residual coefficient of variation approach is the best tool to validate the LPD model.

ACKNOWLEDGMENTS

This work was funded by the Grant No. RTI2018-096072-B-I00 from the Spanish Ministry of Science, Innovation and Universities and by the Spanish State Research Agency, through the Severo Ochoa and María de Maeztu Program for Centers and Units of Excellence in R&D (CEX2020-001084-M).

APPENDIX: TABLES SHOWING TEST POWER

The tables for the Sec. IV A with connection to the LRT test between LPD and PL distributions are included in this appendix. All tables are obtained by simulation. Table II shows the critical points of the test at the significance levels $p = 0.10$ and $p = 0.05$, using 10^5 samples of a PL distribution. Tables III, IV, and V show the tests power at signicance levels $p = 0.10$, using 10^4 samples of a LPD distribution. The first two show the power for the LRT. The last one for the Clauset's methodology. The specifics that show how LRT is superior to Clauset's methods are covered in Sec. IV A.

TABLE II. The critical points for the LRT for sample size 10 to 500, and asymptotic values, tail index $\alpha = 0.1$ to 8, corresponding to the 90th and 95th percentiles.

α	Probability	10	25	50	100	250	500	Asymptotic
1,0	90	3.434	3.099	2.869	2.76	2.717	2.719	2.706
1,0	95	4.745	4.396	4.076	3.924	3.877	3.852	3.841
2,0	90	2.600	2.974	2.905	2.795	2.757	2.748	2.706
2,0	95	3.776	4.007	4.105	3.996	3.914	3.900	3.841
4,0	90	1.996	2.295	2.701	2.845	2.765	2.735	2.706
4,0	95	3.271	3.037	3.567	3.959	3.911	3.894	3.841
8,0	90	1.722	1.621	2.000	2.454	2.760	2.722	2.706
8,0	95	3.030	2.528	2.670	3.217	3.852	3.842	3.841

TABLE III. Power of the LRT for sample size $n = 25$ to 500, $\delta = -0.5$ to 8, tail index $\alpha = 1$ and 2, at significance level of $p = 10\%$.

	$\alpha = 1$					$\alpha = 2$				
δ	$n = 25$	50	100	250	500	25	50	100	250	500
−0.5	0.258	0.400	0.620	0.917	0.995	0.179	0.271	0.43	0.741	0.941
1.0	0.167	0.322	0.583	0.930	0.998	0.171	0.276	0.495	0.859	0.989
2.0	0.298	0.616	0.914	1.000	1.000	0.297	0.551	0.859	0.998	1.000
4.0	0.520	0.890	0.997	1.000	1.000	0.521	0.855	0.994	1.000	1.000
8.0	0.742	0.986	0.999	1.000	1.000	0.764	0.982	1.000	1.000	1.000

TABLE IV. Power of the LRT for sample size $n = 25$ to 500, $\delta = -0.5$ to 8, tail index $\alpha = 4$ and 8, at significance level of $p = 10\%$.

	$\alpha = 4$					$\alpha = 8$				
δ	$n = 25$	50	100	250	500	25	50	100	250	500
−0.5	0.143	0.164	0.237	0.435	0.671	0.139	0.138	0.146	0.215	0.335
1.0	0.190	0.223	0.333	0.645	0.901	0.182	0.219	0.239	0.370	0.600
2.0	0.303	0.423	0.674	0.969	1.000	0.292	0.354	0.461	0.769	0.965
4.0	0.500	0.744	0.958	1.000	1.000	0.454	0.608	0.819	0.992	1.000
8.0	0.744	0.955	0.999	1.000	1.000	0.677	0.882	0.988	1.000	1.000

TABLE V. Power of the Clauset methodology for sample size $n = 100$ to 500, $\delta = 2$ to 8 and tail index $\alpha = 2$ to 8, at significance level of $p = 10\%$.

	$\alpha = 2$			$\alpha = 4$			$\alpha = 8$		
δ	$n = 100$	250	500	100	250	500	100	250	500
2	0.081	0.141	0.16	0.080	0.157	0.227	0.056	0.107	0.211
4	0.096	0.138	0.148	0.113	0.196	0.203	0,087	0.202	0.261
8	0.105	0.113	0.161	0.105	0.154	0.181	0.115	0.188	0.227

[1] V. Pareto, *Cours d'économie politique. Professe a l'Universite de Lausanne* (F. Rouge, Lausanne, Switzerland, 1896) (in French).

[2] V. Pareto, *The Mind and Society. Trattato di sociologia generale* (Harcourt Brace, New York, 1935) (in Italian).

[3] B. Arnold, *Pareto Distribution* (International Cooperative Publishing House, Fairland, MD, 1983).

[4] B. Boghosian, Kinetics of wealth and the Pareto law, Phys. Rev. E **89**, 042804 (2014).

[5] R. Fisher and L. Tippett, Limiting forms of the frequency distribution of the largest or smallest member of a sample, Math. Proc. Cambridge Philos. Soc. **24**, 180 (1928).

[6] B. Gnedenko, Sur la distribution limite du terme maximum d'une série aléatoire, Ann. Math. **44**, 423 (1943).

[7] J. Pickands, Statistical inference using extreme order statistics, Ann. Stat. **3**, 119 (1975).

[8] A. T. Balkema and L. de Haan, Residual life time at great age, Ann. Probab. **2**, 792 (1974).

[9] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance* (Springer, Berlin, 1997).

[10] S. Coles, *An Introduction to Statistical of Extremes Values* (Springer, London, 2001).

[11] A. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton Series in Finance (Princeton University Press, Princeton, 2005).

[12] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels, *Statistics of Extremes: Theory and Applications* (John Wiley & Sons, Chichester, UK, 2006).

[13] L. de Haan and A. Ferreira, *Extreme Value Theory: An Introduction* (Springer, New York, 2007).

[14] N. Markovich, *Nonparametric Analysis of Univariate Heavy-Tailed Data*, Wiley Series in Probability and Statistics (Wiley, Chichester, UK, 2007).

[15] S. Novak, *Extreme Value Methods with Applications to Finance* (CRC Press, Boca Raton, FL, 2012).

[16] M. Newman, Power laws, Pareto distributions and Zipf's law, Contemp. Phys. **46**, 323 (2005).

[17] A. Clauset, C. Shalizi, and M. Newman, Power-law distributions in empirical data, SIAM Rev. **51**, 661 (2009).

[18] M. Newman, Resource letter CS–1: Complex systems, Am. J. Phys. **79**, 800 (2011).

[19] A. Deluca, P. Puig, and A. Corral, Testing universality in critical exponents: The case of rainfall, Phys. Rev. E **93**, 042301 (2016).

[20] A. Broido and A. Clauset, Scale-free networks are rare, Nat. Commun. **10**, 1017 (2019).

[21] A. Corral, I. Serra, and R. Ferrer i Cancho, Distinct flavors of Zipf's law and its maximum likelihood fitting: Rank-size and size-distribution representations, Phys. Rev. E **102**, 052113 (2020).

[22] B. Ahundjanov, S. Akhundjanov, and B. Okhunjanov, Power law in COVID-19 cases in China, J. R. Stat. Soc. A **185**, 699 (2022).

[23] Y. Malevergne, V. Pisarenko, and D. Sornette, Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities, Phys. Rev. E **83**, 036111 (2011).

[24] M. Bee, M. Riccaboni, and S. Schiavo, Pareto versus lognormal: A maximum entropy test, Phys. Rev. E **84**, 026104 (2011).

[25] J. del Castillo and P. Puig, The best test of exponentiality against singly truncated normal alternatives, J. Am. Stat. Assoc. **94**, 529 (1999).

[26] J. del Castillo and I. Serra, Likelihood inference for generalized Pareto distribution, Comput. Stat. Data Anal. **83**, 116 (2015).

[27] K. Lomax, Business failures: Another example of the analysis of failure data, J. Am. Stat. Assoc. **49**, 847 (1954).

[28] R. Von Mises, La distribution de la plus grande de n valuers, Rev. Math. Union Interbalc. **1**, 141 (1936).

[29] J. del Castillo, D. Moriña, M. Padilla, and I. Serra, Fitting tails by the empirical residual coefficient of variation: The ercv package, R J. **11**, 56 (2019).

[30] E. Gilleland, M. Ribatet, and A. Stephenson, A software review for extreme value analysis, Extremes **16**, 103 (2013).

[31] R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2017).

[32] B. Pfaff and A. McNeil, evir: Extreme Values, R package version 1.7-4, 2018, https://CRAN.R-project.org/package=evir.

[33] C. Gillespie, Fitting heavy tailed distributions: The poweRlaw package, J. Stat. Softw. **64**, 1 (2015).

[34] J. del Castillo, D. M. Soler, and I. Serra, ercv: fitting tails by the empirical residual coefficient of variation, R package version 1.0.1, 2019, https://CRAN.R-project.org/package=ercv.

[35] J. Heffernan and A. Stephenson, ismev: an introduction to statistical modeling of extreme values, R package version 1.42, 2018, https://CRAN.R-project.org/package=ismev.

[36] G. Casella and R. Berger, *Statistical Inference*, 2nd ed. (Duxbury Press, Pacific Grove, 2002).

[37] A. Deluca and A. Corral, Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions, Acta Geophys. **61**, 1351 (2013).

[38] J. del Castillo, J. Daoudi, and R. Lockhart, Methods to distinguish between polynomial and exponential tails, Scand. J. Stat. **41**, 382 (2014).

[39] J. del Castillo and M. Padilla, Modelling extreme values by the residual coefficient of variation, Stat. Oper. Res. T. **40**, 303 (2016).

[40] P. Thompson, Y. Cai, D. Reeve, and J. Stander, Automated threshold selection methods for extreme wave analysis, Coast. Eng. **56**, 1013 (2009).

[41] S. Coles and J. Tawn, Modelling extremes of the areal rainfall process, J. R. Stat. Soc. B **58**, 329 (1996).