

Parameter Estimation for Exponentially Tempered Power Law Distributions

MARK M. MEERSCHAERT¹, PARTHANIL ROY²,
AND QIN SHAO³

¹Department of Statistics and Probability, Michigan State University,
East Lansing, Michigan, USA

²Stat-Math Unit, Indian Statistical Institute, Kolkata, India

³Department of Mathematics and Statistics, University of Toledo,
Toledo, Ohio, USA

Tail estimates are developed for power law probability distributions with exponential tempering, using a conditional maximum likelihood approach based on the upper-order statistics. Tempered power law distributions are intermediate between heavy power-law tails and Laplace or exponential tails, and are sometimes called “semi-heavy” tailed distributions. The estimation method is demonstrated on simulated data from a tempered stable distribution, and for several data sets from geophysics and finance that show a power law probability tail with some tempering.

Keywords Exponential tempering; Heavy tails; Tail estimation; Tempered stable.

Mathematics Subject Classification Primary 62F12; Secondary 60E07.

1. Introduction

Probability distributions with heavy, power law tails are important in many areas of application, including physics (Metzler and Klafter, 2000, 2004; Uchaikin and Zolotarev, 1999), finance (Gorenflo et al., 2001; Mainardi et al., 2000; Raberto et al., 2002; Scalas et al., 2000; Sabatelli et al., 2002), and hydrology (Benson et al., 2000, 2001; Schumer et al., 2001, 2003). Stable Lévy motion with power law tails is useful to model anomalous diffusion, where long particle jumps lead to anomalous superdiffusion (Meerschaert et al., 1999; Meerschaert and Scheffler, 2004). Often the power law behavior does not extend indefinitely, due to some truncation or tapering effects. Truncated Lévy flights were proposed by Mantegna and Stanley (1994, 1995) as a modification of the α -stable Lévy motion, to avoid infinite moments. In that model, the largest jumps are simply discarded. Tempered stable Lévy motion takes

Received July 27, 2010; Accepted December 27, 2010

Address correspondence to Mark M. Meerschaert, Department of Statistics and Probability, Michigan State University, East Lansing, Michigan 48824, USA; E-mail: mcubed@stt.msu.edu

a different approach, exponentially tapering the probability of large jumps, so that all moments exist (Rosiński, 2007). Tempered stable laws were recently applied in geophysics (Meerschaert et al., 2008). The problem of parameter estimation for tempered stable laws remains open. Normal inverse Gaussian distributions have the same asymptotic tail behavior, which Barndorff-Nielsen calls “semi-heavy tails” (Barndorff-Nielsen, 1998). These distributions are important in finance (Barndorff-Nielsen, 1997) and turbulence (Barndorff-Nielsen and Leonenko, 2005). Parameter estimation for the normal inverse Gaussian distribution is a difficult problem (Protassov, 2004). Laplace distributions have also found many applications in engineering, finance, biology, and environmental science. Kotz et al. (2001) contains a comprehensive introduction to the theory and application of Laplace distributions and processes, as well as a number of applications. See Meerschaert et al. (2004) for some additional applications to geophysics.

In practical applications, it is often apparent that data tails are too heavy to admit a Gaussian model. Fitting alternative models with a heavier tail requires a judgment about whether the tails are exponential (Laplace, gamma, Weibull, etc.), power-law (Pareto, stable, geometric stable, etc.), or something in between. This judgment starts with an examination of the empirical tail distribution. If the tail appears to follow a pure exponential, then a Laplace or related model may be appropriate. If it follows a pure power-law, then a stable or related model may suffice. For cases in between, where the tail gradually transitions from power-law to exponential, the methods of this article can be useful. Some further discussion, along with a test for determining the extent of truncation/tempering, appears in Chakrabarty and Samorodnitsky (2009).

This article treats exponentially tempered Pareto distributions $P(X > x) = \gamma x^{-\alpha} e^{-\beta x}$, where γ is a scale parameter, α controls the power law tail, and β governs the exponential tempering. In practical applications, the tempering parameter is relatively small, so that the data seems to follow the power law distribution until the largest values are exponentially cooled. A log-log plot of the data versus rank is linear until the tempering causes the plot to curve downward. Such plots are often observed in real data applications. It is also common that the power law behavior emerges only for large data values, so that the tail of the data is fit to this model. Hence, we will consider parameter estimates based on the largest order statistics. The main technical tool is the Rényi representation for the order statistics, and the mathematical details are similar to Hill’s estimator (Hall, 1982; Hill, 1975) for the traditional Pareto distribution. A related article (Aban et al., 2006) considered parameter estimation for the truncated Pareto distribution, relevant to the original model of Mantegna and Stanley.

2. Estimation

Suppose X_1, X_2, \dots, X_n is a random sample from the tempered Pareto distribution with the survival function

$$\bar{F}_X(x; \theta) = P\{X_1 > x\} = \gamma x^{-\alpha} e^{-\beta x}, \quad x \geq x_0, \quad (2.1)$$

where $\theta := \{\alpha, \beta, \gamma\}$ are unknown parameters and $x_0 > 0$ satisfies $\gamma = x_0^\alpha e^{\beta x_0}$. Clearly, the corresponding density function is given by

$$f_X(x; \theta) = \gamma x^{-\alpha-1} e^{-\beta x} (\alpha + x\beta), \quad x \geq x_0. \quad (2.2)$$

Let $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ be the order statistics of the sample, $z := 1/x$, $z_0 := 1/x_0$, and $Z_i := 1/X_i$ for $i = 1, 2, \dots, n$. Then $Z_{(k)} := 1/X_{(k)} > Z_{(k+1)}$ for all $k = 1, \dots, n - 1$ and we have

$$P\{Z_1 \leq z\} = F_Z(z; \theta) = \gamma z^\alpha e^{-\beta/z}, \quad z \leq z_0,$$

which implies that for $z \leq z_0$,

$$\frac{dF_Z(z; \theta)}{dz} = \gamma z^\alpha e^{-\beta/z} \left(\frac{\alpha}{z} + \frac{\beta}{z^2} \right).$$

By the formula (2.5) in Hill (1975), it follows that that the conditional log-likelihood of $\{Z_{(n-k+1)}, \dots, Z_{(n)}\}$ given $Z_{(n-k+1)} < d_z \leq Z_{(n-k)}$ is proportional to the following:

$$\begin{aligned} & \log[1 - F_Z(d_z, \theta)]^{n-k} + \sum_{i=1}^k \log \frac{dF_Z(z_{(n-i+1)})}{dz_{(n-i+1)}} \\ & \propto (n - k) \log [1 - \gamma d_z^\alpha e^{-\beta/d_z}] + k \log \gamma + \alpha \sum_{i=1}^k \log z_{(n-i+1)} \\ & \quad - \beta \sum_{i=1}^k z_{(n-i+1)}^{-1} + \sum_{i=1}^k \log \left(\frac{\alpha}{z_{(n-i+1)}} + \frac{\beta}{z_{(n-i+1)}^2} \right). \end{aligned}$$

Let $\mathbf{x}_k = \{x_{(n-k+1)}, \dots, x_{(n)}\} := \{z_{(n-k+1)}^{-1}, \dots, z_{(n)}^{-1}\}$ and $d_x = 1/d_z$. Using the change of variable formula, the conditional log-likelihood of $\mathbf{X}_k = \{X_{(n-k+1)}, \dots, X_{(n)}\}$ given $X_{(n-k+1)} > d_x \geq X_{(n-k)}$ is of the form

$$\begin{aligned} \log L_c(\theta; \mathbf{x}_k) & \propto (n - k) \log [1 - \gamma d_x^{-\alpha} e^{-\beta d_x}] + k \log \gamma - (\alpha + 2) \sum_{i=1}^k \log x_{(n-i+1)} \\ & \quad - \beta \sum_{i=1}^k x_{(n-i+1)} + \sum_{i=1}^k \log (\alpha x_{(n-i+1)} + \beta x_{(n-i+1)}^2). \end{aligned} \tag{2.3}$$

The following result gives the normal equations of the conditional likelihood problem with the notation introduced above.

Proposition 2.1.

(a) *The conditional maximum likelihood estimator (MLE) $\hat{\theta} = \{\hat{\alpha}, \hat{\beta}, \hat{\gamma}\}$ of $\theta = \{\alpha, \beta, \gamma\}$ given $X_{(n-k+1)} > d_x \geq X_{(n-k)}$ satisfies the normal equations*

$$\sum_{i=1}^k (\log d_x - \log X_{(n-i+1)}) + \sum_{i=1}^k \frac{1}{\hat{\alpha} + \hat{\beta} X_{(n-i+1)}} = 0, \tag{2.4}$$

$$\sum_{i=1}^k (d_x - X_{(n-i+1)}) + \sum_{i=1}^k \frac{X_{(n-i+1)}}{\hat{\alpha} + \hat{\beta} X_{(n-i+1)}} = 0, \tag{2.5}$$

$$\hat{\gamma} = \frac{k}{n} d_x^{\hat{\alpha}} e^{\hat{\beta} d_x}. \tag{2.6}$$

(b) If the above system of normal equations has a solution $\hat{\theta}$ with $\hat{\alpha} > 0$ and $\hat{\beta} > 0$, then it is the unique conditional MLE.

Proof. (a) Defining $\lambda = \gamma d_x^{-\alpha} e^{-\beta d_x}$, the conditional log-likelihood in (2.3) is simplified to

$$\begin{aligned} \log L_c(\theta; \mathbf{x}_k) &\propto (n-k) \log(1-\lambda) + k \log \gamma - (\alpha+2) \sum_{i=1}^k \log x_{(n-i+1)} \\ &\quad - \beta \sum_{i=1}^k x_{(n-i+1)} + \sum_{i=1}^k \log \left(\alpha x_{(n-i+1)} + \beta x_{(n-i+1)}^2 \right). \end{aligned} \quad (2.7)$$

The estimates $\hat{\theta}$ satisfies the following normal equations obtained by $\frac{\partial \log L_c(\theta; \mathbf{x}_k)}{\partial \theta}$

$$\frac{(n-k)\lambda \log d_x}{1-\lambda} - \sum_{i=1}^k \log x_{(n-i+1)} + \sum_{i=1}^k \frac{1}{\alpha + \beta x_{(n-i+1)}} = 0, \quad (2.8)$$

$$\frac{(n-k)\lambda d_x}{1-\lambda} - \sum_{i=1}^k x_{(n-i+1)} + \sum_{i=1}^k \frac{x_{(n-i+1)}}{\alpha + \beta x_{(n-i+1)}} = 0, \quad (2.9)$$

$$\frac{k-n\lambda}{\gamma(1-\lambda)} = 0. \quad (2.10)$$

From (2.10), we have $\lambda = k/n$ from which (2.6) follows. Thus, (2.8) and (2.9) are simplified to (2.4) and (2.5), respectively. This proves (a).

(b) We plug $\lambda = k/n$ in (2.7) and obtain

$$\begin{aligned} \log L_c^*(\alpha, \beta; \mathbf{x}_k) &\propto (n-k) \log \left(1 - \frac{k}{n} \right) - (\alpha+2) \sum_{i=1}^k \log x_{(n-i+1)} - \beta \sum_{i=1}^k x_{(n-i+1)} \\ &\quad + \sum_{i=1}^k \log \left(\alpha x_{(n-i+1)} + \beta x_{(n-i+1)}^2 \right) + k \log \frac{k}{n} + k\alpha \log d_x + k\beta d_x. \end{aligned}$$

Taking the second partial derivatives of $L_c^*(\alpha, \beta; \mathbf{x}_k)$ yields:

$$\frac{\partial^2 \log L_c^*(\alpha, \beta; \mathbf{x}_k)}{\partial \alpha^2} \propto - \sum_{i=1}^k \frac{1}{(\alpha + \beta x_{(n-i+1)})^2},$$

$$\frac{\partial^2 \log L_c^*(\alpha, \beta; \mathbf{x}_k)}{\partial \beta^2} \propto - \sum_{i=1}^k \frac{x_{(n-i+1)}^2}{(\alpha + \beta x_{(n-i+1)})^2},$$

$$\frac{\partial^2 \log L_c^*(\alpha, \beta; \mathbf{x}_k)}{\partial \alpha \partial \beta} \propto - \sum_{i=1}^k \frac{x_{(n-i+1)}}{(\alpha + \beta x_{(n-i+1)})^2}.$$

Observe that $\frac{\partial^2 \log L_c^*(\alpha, \beta; \mathbf{x}_k)}{\partial \alpha^2} < 0$ and by Cauchy-Schwarz inequality,

$$\left(\frac{\partial^2 \log L_c^*(\alpha, \beta; \mathbf{x}_k)}{\partial \alpha \partial \beta} \right)^2 < \left(\frac{\partial^2 \log L_c^*(\alpha, \beta; \mathbf{x}_k)}{\partial \alpha^2} \right) \left(\frac{\partial^2 \log L_c^*(\alpha, \beta; \mathbf{x}_k)}{\partial \beta^2} \right),$$

for all α, β , and \mathbf{x}_k . Hence, it follows that $L_c(\boldsymbol{\theta}; \mathbf{x}_k)$ has at most one local maximum, which, if exists, has to be the unique global maximum. This completes the proof of (b).

Next, we consider the important question of whether the system of normal Eqs. (2.4) and (2.5) has a positive solution. In order to answer this question, we introduce some notation so that we can eliminate the secondary parameter β and focus on the tail parameter α , which is the main parameter of interest. We start by defining

$$T_1 := \sum_{i=1}^k (\log X_{(n-i+1)} - \log d_x) = \sum_{i=1}^k \log (X_{(n-i+1)}/d_x)$$

and

$$T_2 := \sum_{i=1}^k (X_{(n-i+1)} - d_x).$$

Observe that both T_1 and T_2 are positive. Also, for $n \geq 1$ and $1 \leq k \leq n$, define

$$G_{n,k}(u; \mathbf{x}_k) := \sum_{i=1}^k \frac{x_{(n-i+1)}}{kx_{(n-i+1)} + u(T_2 - T_1x_{(n-i+1)})} - 1, \quad u \in [0, k/T_1] \quad (2.11)$$

and note that $G_{n,k}(0; \mathbf{x}_k) = 0$. With these notation we have the following result which gives the normal equation for $\hat{\alpha}$.

Proposition 2.2.

- (a) For any order statistics \mathbf{x}_k of a given sample with size $n \geq 1$, $G_{n,k}(u; \mathbf{x}_k)$ is a well-defined continuous function of u at every point of the closed interval $[0, k/T_1]$.
- (b) $(\hat{\alpha}, \hat{\beta}) \in (0, \infty) \times (0, \infty)$ satisfies the normal Eqs. (2.4) and (2.5) if and only if $\hat{\alpha} \in (0, k/T_1)$ satisfies

$$G_{n,k}(\hat{\alpha}; \mathbf{X}_k) = 0 \quad (2.12)$$

and $\hat{\beta} = (k - \hat{\alpha}T_1)/T_2$.

- (c) There is at most one $\hat{\alpha} \in (0, k/T_1)$ satisfying (2.12).

Proof. (a) We just need to verify that $G_{n,k}$ has no pole in $[0, k/T_1]$. This is obvious because for each $i = 1, 2, \dots, k$, and $u \in [0, k/T_1]$,

$$kx_{(n-i+1)} + u(T_2 - T_1x_{(n-i+1)}) > (k - uT_1)x_{(n-i+1)} > 0.$$

- (b) If $(\hat{\alpha}, \hat{\beta}) \in (0, \infty) \times (0, \infty)$ satisfies (2.4) and (2.5), then multiplying (2.4) by $\hat{\alpha}$ and (2.5) by $\hat{\beta}$ and adding we have $\hat{\alpha}T_1 + \hat{\beta}T_2 = k$ which gives $\hat{\beta} = (k - \hat{\alpha}T_1)/T_2$. Using this we can eliminate $\hat{\beta}$ from (2.5) to get (2.12). The positiveness of $\hat{\beta}$ implies $\hat{\alpha} \in (0, k/T_1)$.

To prove the converse observe that (2.12) and $\hat{\beta} = (k - \hat{\alpha}T_1)/T_2$ yields $\hat{\beta} > 0$ and (2.5). Multiplying both sides of (2.5) by $\hat{\beta}$, we have

$$\sum_{i=1}^k \frac{\hat{\beta}X_{(n-i+1)}}{\hat{\alpha} + \hat{\beta}X_{(n-i+1)}} = \hat{\beta}T_2 = k - \hat{\alpha}T_1,$$

which yields

$$\sum_{i=1}^k \frac{\hat{\alpha}}{\hat{\alpha} + \hat{\beta}X_{(n-i+1)}} = \hat{\alpha}T_1$$

from which (2.4) follows since $\hat{\alpha} > 0$.

(c) This part follows from part (b) of Proposition 2.1. □

The above result is important for two reasons: firstly, it gives normal equations for the tail parameter (as well as for the parameter β); and secondly it enables us to establish the existence (with high probability for a large sample), consistency, and asymptotic normality of the unconditional MLE, the parameter estimates based on the entire data set.

Remark 2.1. Putting $k = n$ and $d_x = \hat{x}_0 := X_{(1)}$ in (2.4)–(2.6) we obtain the normal equations for the unconditional MLE of θ as follows:

$$\sum_{i=1}^n \frac{1}{\hat{\alpha} + \hat{\beta}X_i} = \sum_{i=1}^n \log \frac{X_i}{\hat{x}_0}, \tag{2.13}$$

$$\sum_{i=1}^n \frac{X_i}{\hat{\alpha} + \hat{\beta}X_i} = \sum_{i=1}^n (X_i - \hat{x}_0), \tag{2.14}$$

$$\hat{\gamma} = \hat{x}_0^{\hat{\alpha}} e^{\hat{\beta}\hat{x}_0}. \tag{2.15}$$

Theorem 2.1.

- (a) *The probability that the normal Eqs. (2.13)–(2.15) of the unconditional MLE for θ has a unique solution $\hat{\theta}_n := \{\hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n\}$ converges to 1 as $n \rightarrow \infty$ and the unconditional MLE $\hat{\theta}_n$ is consistent for θ .*
- (b) *$\hat{\alpha}_n$ and $\hat{\beta}_n$ are asymptotically jointly normal with asymptotic means α and β , respectively, and asymptotic variance–covariance matrix $\frac{1}{n}\mathbf{W}^{-1}$, where*

$$\mathbf{W} := \begin{pmatrix} E((\alpha + \beta X_1)^{-2}) & E(X_1(\alpha + \beta X_1)^{-2}) \\ E(X_1(\alpha + \beta X_1)^{-2}) & E(X_1^2(\alpha + \beta X_1)^{-2}) \end{pmatrix}, \tag{2.16}$$

which is invertible by Cauchy-Schwarz inequality.

In order to prove this theorem, we will need the following lemmas.

Lemma 2.1.

- (a) $E\left(\frac{X_1}{\alpha + \beta X_1}\right) = E(X_1 - x_0)$.

$$(b) E\left(\frac{1}{\alpha + \beta X_1}\right) = E\left(\log \frac{X_1}{x_0}\right).$$

Proof. (a) Using the forms of the density and the survival functions of X_1 , we have

$$\begin{aligned} E\left(\frac{X_1}{\alpha + \beta X_1}\right) &= \int_{x_0}^{\infty} \frac{x}{\alpha + \beta x} \gamma x^{-\alpha-1} e^{-\beta x} (\alpha + \beta x) dx \\ &= \int_{x_0}^{\infty} \gamma x^{-\alpha} e^{-\beta x} dx \\ &= \int_0^{\infty} P(X_1 > x) dx - x_0 = E(X_1 - x_0). \end{aligned}$$

(b) To prove this observe that

$$\begin{aligned} E\left(\log \frac{X_1}{x_0}\right) &= \int_0^{\infty} P\left(\log \frac{X_1}{x_0} > t\right) dt \\ &= \int_0^{\infty} P(X_1 > x_0 e^t) dt \\ &= \int_0^{\infty} \gamma (x_0 e^t)^{-\alpha} \exp(-\beta x_0 e^t) dt, \end{aligned}$$

which by a change of variable $x = x_0 e^t$ becomes

$$\begin{aligned} &= \int_{x_0}^{\infty} \gamma x^{-\alpha-1} e^{-\beta x} dx \\ &= \int_{x_0}^{\infty} \frac{1}{\alpha + \beta x} \gamma x^{-\alpha-1} e^{-\beta x} (\alpha + \beta x) dx = E\left(\frac{1}{\alpha + \beta X_1}\right) \end{aligned}$$

and this completes the proof.

Lemma 2.2. For $\hat{x}_0 = X_{(1)}$ the following hold:

- (a) $\sqrt{n}(\hat{x}_0 - x_0) \xrightarrow{p} 0$;
- (b) $\sqrt{n}(\log \hat{x}_0 - \log x_0) \xrightarrow{p} 0$.

Proof. (a) By the Markov inequality, it is enough to show that $E(\sqrt{n}(\hat{x}_0 - x_0)) \rightarrow 0$ as $n \rightarrow \infty$. Observe that

$$\begin{aligned} 0 \leq E(\sqrt{n}(\hat{x}_0 - x_0)) &= \int_0^{\infty} P(\sqrt{n}(\hat{x}_0 - x_0) > t) dt \\ &= \int_0^{\infty} \left(P\left(X_1 > x_0 + \frac{t}{\sqrt{n}}\right)\right)^n dt \\ &= \int_0^{\infty} \left(1 + \frac{t}{x_0 \sqrt{n}}\right)^{-\alpha n} e^{-\beta t \sqrt{n}} dt \\ &\leq \int_0^{\infty} e^{-\beta t \sqrt{n}} dt = \frac{1}{\beta \sqrt{n}} \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$ and this finishes the proof of part (a).

(b) Part (b) follows from part (a) using the inequality $|\log y - \log x_0| \leq C|y - x_0|$ for all y in a neighborhood of x_0 and for some $C > 0$.

Proof of Theorem 2.1. (a) In this case, $k = n$ (see Remark 2.1). For simplicity of notation, we use G_n to denote $G_{n,n}$, i.e.,

$$G_n(u; \mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^n \frac{X_i}{X_i + u \left(\frac{T_2}{n} - \frac{T_1}{n} X_i \right)} - 1, \quad u \in [0, n/T_1].$$

We will eventually show that for all $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P \left[G_n(u; \mathbf{X}_n) = 0 \text{ has a solution in } (\alpha - \epsilon, \alpha + \epsilon) \right] = 1. \quad (2.17)$$

To prove (2.17), we start by introducing some notation. Define

$$\tilde{G}_n(u; \mathbf{X}_n) := \frac{1}{n} \sum_{i=1}^n \frac{X_i}{X_i + u(B - AX_i)} - 1,$$

where, in view of Lemma 2.1,

$$\begin{aligned} A &:= E \left(\frac{1}{\alpha + \beta X_1} \right) = E \left(\log \frac{X_1}{x_0} \right) > 0, \quad \text{and} \\ B &:= E \left(\frac{X_1}{\alpha + \beta X_1} \right) = E(X_1 - x_0) > 0. \end{aligned} \quad (2.18)$$

Define

$$G(u) := E \left(\frac{X_1}{X_1 + u(B - AX_1)} \right) - 1. \quad (2.19)$$

Since

$$\alpha A + \beta B = 1, \quad (2.20)$$

it follows that $1 - uA > 0$ for all u in a small enough neighborhood $N(\alpha)$ of α , which yields that $\tilde{G}_n(u; \mathbf{X}_n)$ has no pole and hence is well-defined on $N(\alpha)$. Similarly, $G(u)$ is also well defined on $N(\alpha)$ because for all $u \in N(\alpha)$, $-1 < G(u) < E(X_1/uB) - 1 < \infty$.

From (2.20) and (2.18) we obtain

$$G(\alpha) = 0. \quad (2.21)$$

By a dominated convergence argument, we differentiate under the integral sign in (2.19) and obtain

$$G'(\alpha) = E \left[\frac{(AX_1 - B)X_1}{(X_1 + \alpha(B - AX_1))^2} \right] = E(YZ)$$

where

$$Y = \frac{AX_1 - B}{X_1 + \alpha(B - AX_1)} \text{ and } Z = \frac{X_1}{X_1 + \alpha(B - AX_1)}.$$

Since $Z = 1 + \alpha Y$, we have $\text{Cov}(Y, Z) = \alpha \text{Var}(Y) > 0$ and $E(Y) = \alpha^{-1}(E(Z) - 1) = 0$ by (2.21). This shows

$$G'(\alpha) = E(YZ) > E(Y)E(Z) = 0. \tag{2.22}$$

By (2.20), we can find a small enough neighborhood $N^*(\alpha) \subseteq N(\alpha)$ of α and $\delta > 0$ such that $1 - ux > 0$ whenever $u \in N^*(\alpha)$ and $|x - A| < \delta$. We will show that $G_n(u; \mathbf{X}_n) \xrightarrow{p} G(u)$ as $n \rightarrow \infty$ for all $u \in N^*(\alpha)$ in the following. Since, by the weak law of large numbers, $\tilde{G}_n(u; \mathbf{X}_n) \xrightarrow{p} G(u)$ as $n \rightarrow \infty$ for all $u \in N(\alpha)$, it is enough to show that

$$G_n(u; \mathbf{X}_n) - \tilde{G}_n(u; \mathbf{X}_n) = o_p(1) \tag{2.23}$$

for all $u \in N^*(\alpha) \subseteq N(\alpha)$. Since $T_1/n \xrightarrow{p} A$ and $T_2/n \xrightarrow{p} B$ as $n \rightarrow \infty$, we obtain using bivariate mean value theorem that

$$G_n(u; \mathbf{X}_n) - \tilde{G}_n(u; \mathbf{X}_n) = \left(\frac{T_1}{n} - A\right) R_n + \left(B - \frac{T_2}{n}\right) S_n,$$

where

$$R_n = \frac{1}{n} \sum_{i=1}^n \frac{uX_i^2}{(X_i + u(\xi_n - \eta_n X_i))^2}$$

and

$$S_n = \frac{1}{n} \sum_{i=1}^n \frac{uX_i}{(X_i + u(\xi_n - \eta_n X_i))^2}$$

with $\eta_n \xrightarrow{p} A$ and $\xi_n \xrightarrow{p} B$ as $n \rightarrow \infty$. Then in order to show (2.23), it is enough to establish that both R_n and S_n are tight for all $u \in N^*(\alpha)$. Let $\Omega_n := \{|\eta_n - A| < \delta, \xi_n > B/2\}$, where δ is as above. Then on the event Ω_n , we have for all $u \in N^*(\alpha)$,

$$R_n \leq \frac{4}{uB^2} \frac{1}{n} \sum_{i=1}^n X_i^2,$$

which is tight because $n^{-1} \sum_{i=1}^n X_i^2 \xrightarrow{p} E(X_1^2) < \infty$. Similarly on Ω_n , for all $u \in N^*(\alpha)$,

$$|S_n| \leq \frac{4}{uB^2} \frac{1}{n} \sum_{i=1}^n |X_i|,$$

which is also tight. Since $P(\Omega_n) \rightarrow 1$ as $n \rightarrow \infty$, the required tightnesses follow establishing (2.23) and hence $G_n(u; \mathbf{X}_n) \xrightarrow{p} G(u)$ as $n \rightarrow \infty$ for all $u \in N^*(\alpha)$.

We are now all set to complete the proof. By (2.21) and (2.22), we can find $\epsilon_0 > 0$ small enough such that $(\alpha - \epsilon_0, \alpha + \epsilon_0) \subseteq N^*(\alpha)$, $G(u) > 0$ for all $u \in (\alpha, \alpha + \epsilon_0)$ and $G(u) < 0$ for all $u \in (\alpha - \epsilon_0, \alpha)$. Since $G_n(u; X_n) \xrightarrow{p} G(u)$ as $n \rightarrow \infty$ for all $u \in (\alpha - \epsilon_0, \alpha + \epsilon_0)$, (2.17) follows for all $\epsilon \in (0, \epsilon_0)$ and hence for all $\epsilon > 0$. The existence of $\hat{\theta}_n$ (with probability tending to 1 as $n \rightarrow \infty$) follows from (2.17) by continuity of $u \mapsto G_n(u; X_n)$, part (b) of Proposition 2.2 and the observation that $nT_1^{-1} \xrightarrow{p} A^{-1} > \alpha + \epsilon$ for small enough $\epsilon > 0$. The uniqueness is obvious from part (c) of Proposition 2.2. The consistency of $\hat{\alpha}_n$ follows by choosing $\epsilon > 0$ arbitrarily small in (2.17). The consistency of $\hat{\beta}_n$ is straightforward from the equations $\hat{\beta}_n = (n - \hat{\alpha}_n T_1)/T_2$ and (2.20) and finally (2.15) yields the consistency of $\hat{\gamma}_n$.

(b) In order to establish the asymptotic normality of $\hat{\alpha}_n$ and $\hat{\beta}_n$, we start with the following notation. Define, for $s, t, x > 0$,

$$H_n(s, t, x) := \frac{1}{n} \sum_{i=1}^n \frac{1}{s + tX_i} - \frac{1}{n} \sum_{i=1}^n \log \frac{X_i}{x},$$

$$K_n(s, t, x) := \frac{1}{n} \sum_{i=1}^n \frac{X_i}{s + tX_i} - \frac{1}{n} \sum_{i=1}^n (X_i - x).$$

By Remark 2.1, $\hat{\alpha}_n$ and $\hat{\beta}_n$ satisfies $H_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{x}_0) = K_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{x}_0) = 0$. Let $\tilde{\alpha}_n$ and $\tilde{\beta}_n$ be the unconditional MLE of α and β_2 respectively, when x_0 is known. Then, $\tilde{\alpha}_n$ and $\tilde{\beta}_n$ satisfies $H_n(\tilde{\alpha}_n, \tilde{\beta}_n, x_0) = K_n(\tilde{\alpha}_n, \tilde{\beta}_n, x_0) = 0$. When x_0 is known, the support of the density (2.2) does not depend on the parameter values and the corresponding information matrix is given by W as in (2.16). Hence, using the asymptotic properties of the MLE (see, for example, Shorack, 2000, p. 564), it can easily be deduced that $\sqrt{n}(\tilde{\alpha}_n - \alpha, \tilde{\beta}_n - \beta)$ converges in distribution to a bivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix W . To complete the proof it is enough to show that $\sqrt{n}(\tilde{\alpha}_n - \hat{\alpha}_n) \xrightarrow{p} 0$ and $\sqrt{n}(\tilde{\beta}_n - \hat{\beta}_n) \xrightarrow{p} 0$.

To this end, define

$$\tilde{H}_n(s, t) := \frac{1}{n} \sum_{i=1}^n \frac{1}{s + tX_i} - E\left(\log \frac{X_1}{x_0}\right),$$

$$\tilde{K}_n(s, t) := \frac{1}{n} \sum_{i=1}^n \frac{X_i}{s + tX_i} - E(X_1 - x_0)$$

for $s, t > 0$. Then we have

$$\begin{aligned} U_n &:= \sqrt{n} \left(\tilde{H}_n(\hat{\alpha}_n, \hat{\beta}_n) - \tilde{H}_n(\tilde{\alpha}_n, \tilde{\beta}_n) \right) \\ &= \sqrt{n} \left(\tilde{H}_n(\hat{\alpha}_n, \hat{\beta}_n) - H_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{x}_0) \right) + \sqrt{n} \left(H_n(\tilde{\alpha}_n, \tilde{\beta}_n, x_0) - \tilde{H}_n(\tilde{\alpha}_n, \tilde{\beta}_n) \right) \\ &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \log \frac{X_i}{\hat{x}_0} - E\left(\log \frac{X_1}{x_0}\right) \right) + \sqrt{n} \left(E\left(\log \frac{X_1}{x_0}\right) - \frac{1}{n} \sum_{i=1}^n \log \frac{X_i}{x_0} \right) \\ &= \sqrt{n} (\log x_0 - \log \hat{x}_0) \xrightarrow{p} 0 \end{aligned} \tag{2.24}$$

by Lemma 2.2. Similarly, we can show

$$V_n := \sqrt{n} \left(\tilde{K}_n(\hat{\alpha}_n, \hat{\beta}_n) - \tilde{K}_n(\tilde{\alpha}_n, \tilde{\beta}_n) \right) \xrightarrow{p} 0. \tag{2.25}$$

Using bivariate mean value theorem we have

$$\begin{aligned} U_n &= \sqrt{n}(\tilde{\alpha}_n - \hat{\alpha}_n) \frac{\partial \tilde{H}_n}{\partial s}(\rho_n, \zeta_n) + \sqrt{n}(\tilde{\beta}_n - \hat{\beta}_n) \frac{\partial \tilde{H}_n}{\partial t}(\rho_n, \zeta_n), \\ &=: \sqrt{n}(\tilde{\alpha}_n - \hat{\alpha}_n) W_{11}^{(n)} + \sqrt{n}(\tilde{\beta}_n - \hat{\beta}_n) W_{12}^{(n)}, \end{aligned} \tag{2.26}$$

where $\rho_n \xrightarrow{p} \alpha$ and $\zeta_n \xrightarrow{p} \beta$ as $n \rightarrow \infty$, and

$$\begin{aligned} V_n &= \sqrt{n}(\tilde{\alpha}_n - \hat{\alpha}_n) \frac{\partial \tilde{K}_n}{\partial s}(\tilde{\rho}_n, \tilde{\zeta}_n) + \sqrt{n}(\tilde{\beta}_n - \hat{\beta}_n) \frac{\partial \tilde{K}_n}{\partial t}(\tilde{\rho}_n, \tilde{\zeta}_n) \\ &=: \sqrt{n}(\tilde{\alpha}_n - \hat{\alpha}_n) W_{21}^{(n)} + \sqrt{n}(\tilde{\beta}_n - \hat{\beta}_n) W_{22}^{(n)}, \end{aligned} \tag{2.27}$$

where $\tilde{\rho}_n \xrightarrow{p} \alpha$ and $\tilde{\zeta}_n \xrightarrow{p} \beta$ as $n \rightarrow \infty$. Defining $\mathbf{W}_{(n)} := (W_{ij}^{(n)})_{1 \leq i, j \leq 2}$ we get from Eqs. (2.26) and (2.27) that

$$\sqrt{n}(\tilde{\alpha}_n - \hat{\alpha}_n) = \frac{W_{22}^{(n)} U_n - W_{12}^{(n)} V_n}{\det \mathbf{W}_{(n)}} \tag{2.28}$$

and using an argument similar to the proof of (2.23), we get that $\mathbf{W}_{(n)} \xrightarrow{p} -\mathbf{W}$, where \mathbf{W} is as in (2.16). This, in particular, implies $\det \mathbf{W}_{(n)} \xrightarrow{p} \det \mathbf{W} > 0$ by Cauchy–Schwarz inequality. Hence, using (2.24), (2.25), and (2.28), it follows that $\sqrt{n}(\tilde{\alpha}_n - \hat{\alpha}_n) \xrightarrow{p} 0$. By a similar argument we can also show that $\sqrt{n}(\tilde{\beta}_n - \hat{\beta}_n) \xrightarrow{p} 0$ and this completes the proof of Theorem 2.1.

3. Applications

Extensive simulation trials were conducted to validate the estimator developed in the previous section. For simulated data from the exponentially tempered Pareto distribution (2.1) the parameter estimates were generally close to the assumed values, so long as the data range was sufficient to capture both the power law behavior and the subsequent tempering at the highest values. If the tempering parameter β is very large, so that the term $x^{-\alpha}$ in (2.1) hardly varies over the data range, then estimates of α are unreliable. If β is so small that the term $e^{-\beta x}$ in (2.1) hardly varies over the data range, then estimates of β are widely variable. In either case, a simpler model (exponential or power law) is indicated. Naturally the tempered Pareto model (2.1) is only appropriate when both terms are significant. This can be seen in a log-log plot of data versus ranks, where a straight line eventually falls away due to tempering. If the data follows a straight line over the entire tail, then a simpler Pareto model is indicated. If the data follows a straight line on a semi-log plot, then an exponential model is appropriate. Several illustrative examples follow.

Figure 1 illustrates the behavior of the tail estimate $\hat{\alpha}$ from Proposition 2.2 as a function of the number k of upper order statistics used. The simulated data comes from the tempered Pareto distribution (2.1) with lower limit $x_0 = 1$, tail

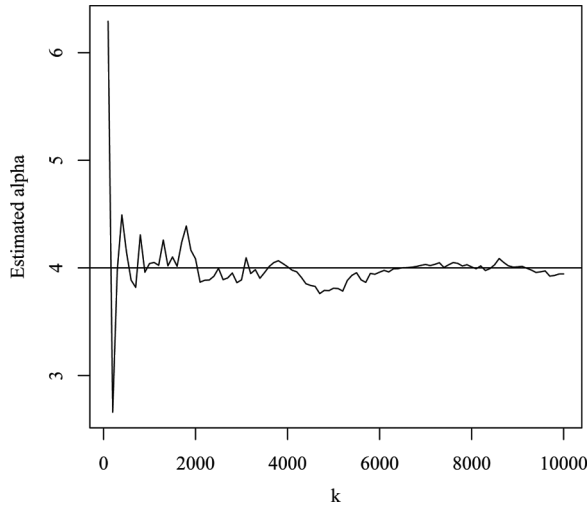


Figure 1. Behavior of α estimate as a function of the number k of upper-order statistics used.

parameter $\alpha = 4$, and tempering parameter $\beta = 0.5$. Simulation was performed using a standard rejection method. It is apparent that, once the number k of upper-order statistics used reaches a few percent of the total sample size of $n = 10,000$, the α estimate settles down to a reasonable value. In practice, any sufficiently large value of k will give a reasonable fit for this data, since the entire distribution follows a tempered Pareto.

Figure 2 shows a histogram and normal quantile–quantile plot for α estimates obtained from 500 replications of the same tempered Pareto simulation. In this case we set $\alpha = 2$, $\beta = 0.5$, and $x_0 = 1$. The sample size is $n = 1,000$ and we use the $k = 500$ largest observations to estimate the distribution parameters. The corresponding plots are similar for various values of the parameters. We conclude that the sampling distribution of the parameters is reasonably well approximated by a

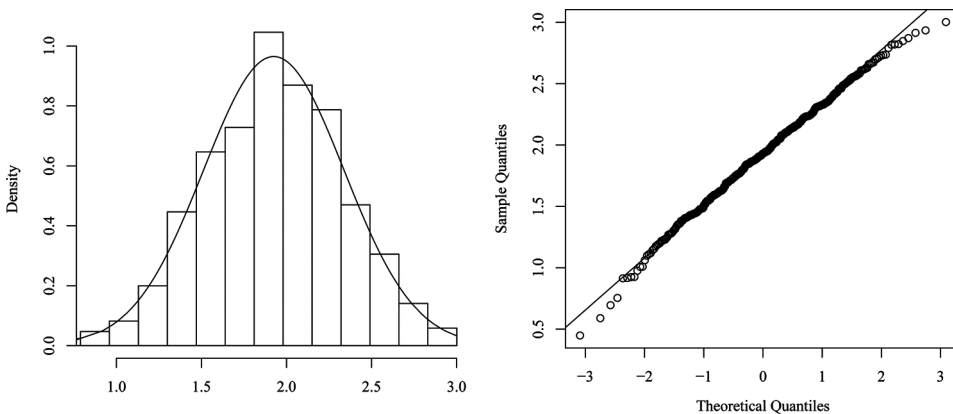


Figure 2. Evidence of normal sampling distribution for α estimates.

normal distribution. Note that the asymptotic normality of the parameter estimates based on the entire data set was established in Theorem 2.1. The asymptotic theory for the general case $k < n$ is much more difficult, since the usual second order regular variation method for Hill-type estimators is not applicable here, due to the exponential tempering.

Tempered stable laws (Rosiński, 2007) have power law tails modified by exponential tempering. Therefore, the exponentially tempered Pareto model (2.1) gives a simple way to approximate the tail behavior, and estimate the parameters. The simple and efficient exponential rejection method of Baeumer and Meerschaert (2010) was used to simulate tempered stable random variates. Figure 3 shows the upper tail of simulated data following a tempered stable distribution. The largest $k = 100$ of the $n = 1,000$ order statistics are plotted. The underlying stable distribution has tail parameter $\alpha = 1.5$, skewness 1, mean 0, and scale $\sigma = 4$ in the usual parameterization (Samorodnitsky and Taqqu, 1994), and the tempering parameter is $\beta = 0.01$. Figure 3 is a log-log plot of the sorted data $X_{(i)}$ vs. rank $(n - i)/n$ exhibiting the power-law tail as a straight line that eventually falls off due to tempering. It is apparent that the tempered Pareto model gives a reasonable fit to the more complicated tempered stable distribution, which has no closed form. Similar results were obtained for other values of the parameters. For smaller values of β the data plot more closely resembles a straight line (power law tail).

Next, we apply the conditional MLE developed in this article to several real data sets. First we consider a data set from hydrology. Hydraulic conductivity K measures the ability of water to pass through a porous medium. This is a function of porosity (percent of the material consisting of pore space) as well as connectivity. We examine a data set with $n = 2,618$ observations of K collected from boreholes at the MAcroDispersion Experiment (MADE) site on an Air Force base near Columbus MS. The data set has been analyzed by several researchers, see for example Benson et al. (2001). Figure 4 shows a log-log plot of the largest 10% of the data, with the best-fitting tempered Pareto model (2.1), where the parameters were fit using Proposition 2.2. Absolute values of K were modeled in order to combine

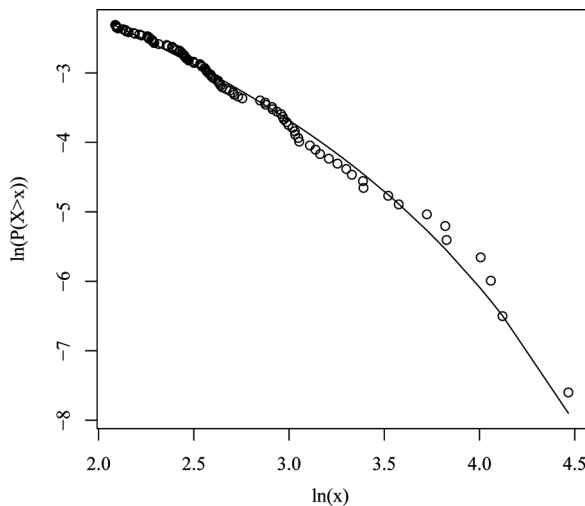


Figure 3. Tempered Pareto fit to the upper tail of simulated tempered stable data.

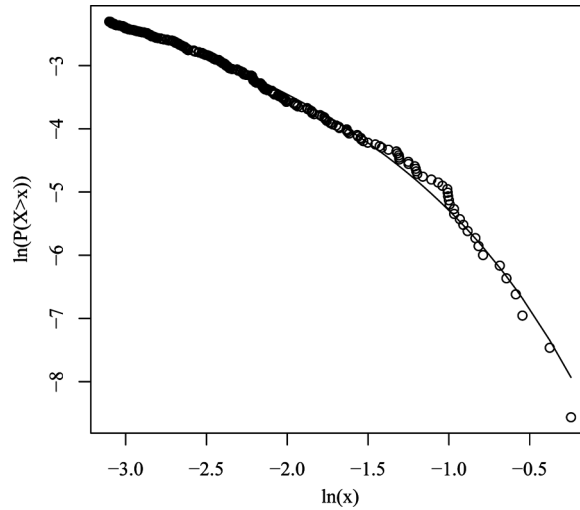


Figure 4. Tempered Pareto model for hydraulic conductivity data.

the heavy tails at both extremes. The largest $k = 262$ values of the data were used (approximately 10%). It is apparent that the tempered Pareto model gives a good fit to the data. Since the data deviates from a straight line on this log-log plot, a simple Pareto model would be inadequate.

Figure 5 shows the constraint function $G_{n,k}(u; \mathbf{x}_k)$ from Proposition 2.2 for the same data set, as a function of u . The vertical line on the graph is the upper bound of $u = k/T_1$. The constraint function has roots at $u = 0$ (by definition) and at $u = \hat{\alpha} = 0.6171$ which is the estimate of the tail parameter. In view of Proposition 2.2, this is the unique solution to the normal equations. The remaining parameter

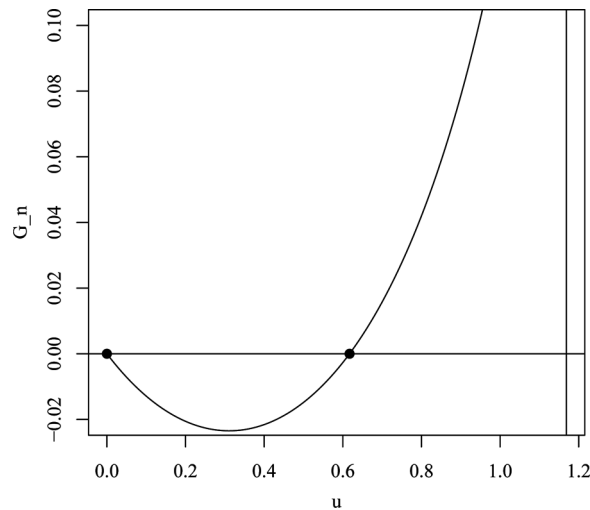


Figure 5. Constraint function $G_{n,k}(u; \mathbf{x}_k)$ for the hydraulic conductivity data, showing unique positive root.

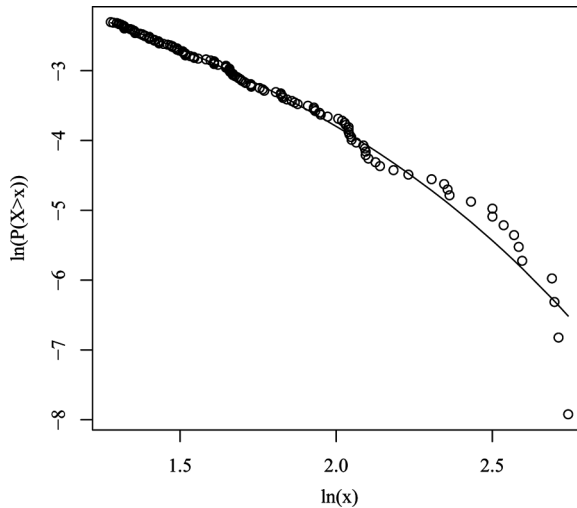


Figure 6. Tempered Pareto model for AMZN stock daily price returns.

estimates are $\hat{\beta} = 5.2397$ and $\hat{\gamma} = 0.0187$. This is a relatively heavy tail with a strong tempering. For any value of k in the range $100 < k < 750$, the tempered Pareto fit to the k largest order statistics of the MADE data appears similar to Fig. 4. However, the individual parameter values $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ vary significantly with k in that range.

Figure 6 fits a tempered Pareto model to absolute log returns in the daily price of stock for Amazon, Inc. The ticker symbol is AMZN. The data ranges from January 1, 1998 to June 30, 2003 ($n = 1,378$). Based on the upper 10% of the data $k = 138$, the best fitting parameter values (conditional MLE from Proposition 2.2) are $\hat{\alpha} = 0.578$, $\hat{\beta} = 0.281$, and $\hat{\gamma} = 0.567$. The data shows a classic power law shape,

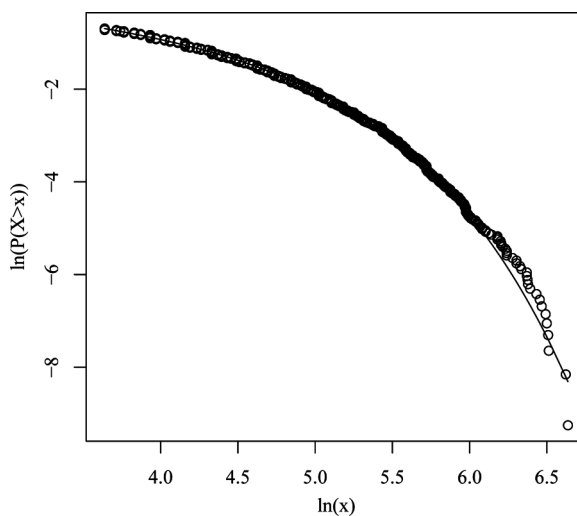


Figure 7. Tempered Pareto model for daily precipitation data.

linear on this log-log plot, but eventually falls off at the largest values. This indicates an opportunity to improve prediction using a tempered model.

Figure 7 shows the tempered Pareto fit to daily precipitation data at Tombstone AZ between July 1, 1893 and December 31, 2001. The fit is based on the largest $k = 2,608$ observations, which constitutes the upper half of the non zero data. The fitted parameters were $\hat{\alpha} = 0.212$, $\hat{\beta} = 0.00964$, and $\hat{\gamma} = 1.56$. The data tail is clearly lighter than a pure power law model (straight line) but is fit well by the tempered model. A semi-log plot (not shown) was examined to rule out a simpler exponential fit.

The same three data sets were also analyzed in Aban et al. (2006) and fit to a truncated Pareto model. In both papers, a log-log plot of the fitted distribution tail shows a straight line that eventually curves downward, matching the data. In our view, the choice between tempered or truncated Pareto in these cases should be based on modeling issues. If the practitioner believes that a fixed upper bound is reasonable, then the truncated Pareto is suitable. Otherwise, a tempered Pareto can offer an attractive alternative, and the fitted parameter β can provide a useful quantification of the tempering effect. For example, the K data in Fig. 5 comes from an instrument that effectively averages hydraulic conductivity, and the parameter β controls the extent to which averaging tempers the largest K values. For additional discussion, see Liu et al. (2009). Note that the fitted α values in such data sets will typically be smaller (indicating a heavier tail) than Hill's estimator. This is because a straight line (power law) fit to the data in Figs. 5–7 for a pure Pareto will require a steeper slope. This provides one explanation for the observation that Hill's estimator can often over-estimate the tail parameter α in practice. For a nice discussion with applications to finance, see McCulloch (1997). We also note that a simple test for tempering/truncation is provided by Remark 5 in Aban et al. (2006). That test uses Hill's estimator and extreme value theory to reject the null hypothesis of a pure Pareto with approximate p -value given by $p = \exp(-n(X_{(1)}/x_0)^{-\alpha})$. For the data sets in Figs. 5–7 the test yields $p < 0.02$, see Aban et al. (2006).

We conclude this section with some practical advice for tail modeling. If a data set exhibits tails that are heavier than Gaussian (e.g., if there are numerous outliers), then it makes sense to consider alternative models. Plotting order statistics of the data $X_{(i)}$ vs. ranks $(n - i)/n$ gives a simple method for initial model selection. For signed data, it is often advisable to begin by examining the absolute values. If the points corresponding to the largest data values follow a straight line on a log-log plot, this indicates a power-law probability tail. If a semi-log plot of the same data appears to follow a straight line, this suggests an exponential tail (e.g., a Laplace model, if the data distribution appears symmetric). Upward or downward curvature on those plots indicates a heavier or lighter tail than the pure power-law or exponential. For example, the downward curvature in Fig. 7 indicates that the data tail is lighter than a power-law. A semi-log plot of the same data (not shown) produced an upward curving shape, indicating that the data tail is heavier than an exponential. Then it makes sense to consider an intermediate model with "semi-heavy" tails, for which the methods of this article may be useful. Some additional discussion of these intermediate models may be found in Chakrabarty and Samorodnitsky (2009).

4. Conclusions

Tempered Pareto distributions are useful to model heavy-tailed data, in cases where a pure power law places too much probability mass on the extreme tail. The simple form of this probability law facilitates the development of a maximum likelihood estimator (MLE) for the distribution parameters, accomplished in this article. Those estimates are proven to be consistent and asymptotically normal. In some practical applications, including the tempered stable model, it is only the upper tail of the data that follows a tempered power law. For that reason, we also develop a conditional MLE in this article, based on the upper tail of the data. The conditional MLE is easily computable. A detailed simulation study was performed to validate the performance of the MLE. In cases where the tempered Pareto model would be appropriate, the conditional MLE is reasonably accurate, and its sampling distribution appears to be well approximated by a normal law. Tempered stable laws are useful models in geophysics, but parameter estimation for this model is an open problem. Simulation demonstrates that the tempered Pareto model is a reasonable approximation, for which efficient parameter estimation can be accomplished, via the methods of this article. Finally, data sets from hydrology, finance, and atmospheric science are examined. In each case, the methods of this article are used to fit a reasonable and predictive tempered Pareto model.

Acknowledgments

Mark M. Meerschaert was partially supported by NSF grants DMS-102548, DMS-0803360, and NIH grant R01-EB012079.

This work was initiated while Professor Shao was visiting the Department of Statistics and Probability at Michigan State University.

References

- Aban, I. B., Meerschaert, M. M., Panorska, A. K. (2006). Parameter estimation methods for the truncated Pareto distribution. *J. Amer. Statist. Assoc. Theor. Meth.* 101:270–277.
- Baeumer, B., Meerschaert, M. M. (2010). Tempered stable Lévy motion and transient super-diffusion. *J. Comput. Appl. Math.* 233:2438–2448.
- Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Statist.* 41(1):1–13.
- Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type. *Fin. Stoch.* 2:41–68.
- Barndorff-Nielsen, O. E., Leonenko, N. N. (2005). Burgers' turbulence problem with linear or quadratic external potential. *J. Appl. Probab.* 42:550–565.
- Benson, D. A., Wheatcraft, S. W., Meerschaert, M. M. (2000). Application of a fractional advection-dispersion equation. *Water Resour. Res.* 36:1403–1412.
- Benson, D. A., Schumer, R., Meerschaert, M. M., Wheatcraft, S. W. (2001). Fractional dispersion, Lévy motions, and the MADE tracer tests. *Transp. Porous Med.* 42:211–240.
- Chakrabarty, A., Samorodnitsky, G. (2009). Understanding heavy tails in a bounded world or, is a truncated heavy tail heavy or not? *Stochastic Models* 28:109–143.
- Gorenflo, R., Mainardi, F., Scalas, E., Raberto, M. (2001). Fractional calculus and continuous-time finance. III. In: Kohlmann, M., Tang, S., eds. *Mathematical Finance* Boston: Birkhäuser Verlag, pp. 171–180.

- Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. B* 44(1):37–42.
- Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* 3:1163–1174.
- Kotz, S., Kozubowski, T. J., Podgórski, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Boston: Birkhäuser.
- Liu, G., Butler, J. J., Jr., Bohling, G. C., Reboulet, E., Knobbe, S., Hyndman, D. W. (2009). A new method for high-resolution characterization of hydraulic conductivity. *Water Resour. Res.* 45:W08202.
- Mainardi, F., Raberto, M., Gorenflo, R., Scalas, E. (2000). Fractional Calculus and continuous-time finance II: the waiting-time distribution. *Physica A* 287:468–481.
- Mantegna, R. N., Stanley, H. E. (1994). Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight. *Phys. Rev. Lett.* 73(22):2946–2949.
- Mantegna, R. N., Stanley, H. E. (1995). Scaling behavior in the dynamics of an economic index. *Nature* 376(6535):46–49.
- McCulloch, J. (1997). Measuring tail thickness to estimate the stable index α : a critique. *J. Bus. Econ. Statist.* 15:74–81.
- Meerschaert, M. M., Benson, D. A., Baeumer, B. (1999). Multidimensional advection and fractional dispersion. *Phys. Rev. E* 59:5026–5028.
- Meerschaert, M. M., Scheffler, H. P. (2004). Limit theorems for continuous time random walks with infinite mean waiting times. *J. Applied Probab.* 41(3):623–638.
- Meerschaert, M. M., Kozubowski, T. J., Molz, F. J., Lu, S. (2004). Fractional Laplace model for hydraulic conductivity. *Geophys. Res. Lett.* 31:L08501.
- Meerschaert, M. M., Zhang, Y., Baeumer, B. (2008). Tempered anomalous diffusion in heterogeneous systems. *Geophys. Res. Lett.* 35:L17403.
- Metzler, R., Klafter, J. (2000). The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* 339:1–77.
- Metzler, R., Klafter, J. (2004). The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics. *J. Phys. A* 37:R161–R208.
- Protassov, R. S. (2004). EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ . *Statist. Comput.* 14(1):67–77.
- Raberto, M., Scalas, E., Mainardi, F. (2002). Waiting-times and returns in high-frequency financial data: an empirical study. *Physica A* 314:749–755.
- Rosiński, J. (2007). Tempering stable processes. *Stoch. Proc. Appl.* 117:677–707.
- Sabatelli, L., Keating, S., Dudley, J., Richmond, P. (2002). Waiting time distributions in financial markets. *Eur. Phys. J. B* 27:273–275.
- Samorodnitsky, G., Taqqu, M. (1994). *Stable Non-Gaussian Random Processes*. New York: Chapman and Hall.
- Scalas, E., Gorenflo, R., Mainardi, F. (2000). Fractional calculus and continuous time finance. *Phys. A* 284:376–384.
- Schumer, R., Benson, D. A., Meerschaert, M. M., Wheatcraft, S. W. (2001). Eulerian derivation of the fractional advection-dispersion equation. *J. Contaminant Hydrol.* 48:69–88.
- Schumer, R., Benson, D. A., Meerschaert, M. M., Baeumer, B. (2003). Multiscaling fractional advection-dispersion equations and their solutions. *Water Resour. Res.* 39:1022–1032.
- Shorack, G. B. (2000). *Probability for Statisticians*. Springer Texts in Statistics. New York: Springer-Verlag.
- Uchaikin, V. V., Zolotarev, V. M. (1999). *Chance and Stability. Stable Distributions and Their Applications*. Utrecht: VSP.