Parameter Estimation for the Truncated Pareto Distribution

Inmaculada B. ABAN, Mark M. MEERSCHAERT, and Anna K. PANORSKA

The Pareto distribution is a simple model for nonnegative data with a power law probability tail. In many practical applications, there is a natural upper bound that truncates the probability tail. This article derives estimators for the truncated Pareto distribution, investigates their properties, and illustrates a way to check for fit. These methods are illustrated with applications from finance, hydrology, and atmospheric science.

KEY WORDS: Maximum likelihood estimator; Order statistics; Pareto distribution; Tail behavior; Truncation.

1. INTRODUCTION

A random variable, X, has a Pareto distribution if $P(X > x) = Cx^{-\alpha}$ for some $\alpha > 0$ (see Johnson, Kotz, and Balakrishnan 1994; Arnold 1983). This distributional model is important in applications, because many datasets are observed to follow a power law probability tail, at least approximately, for large values of x. Stable distributions with index α and max-stable type II extreme value distributions with index α are also asymptotically Pareto in their probability tails, and this fact has been frequently used to develop estimators for those distributions. In some applications there is a natural upper bound on the probability tail that truncates the Pareto law. In other cases there is empirical evidence that a truncated Pareto gives a better fit to the data. Because both sums (if $\alpha < 2$) and extremes (for any $\alpha > 0$) fall in different domains of attraction, depending on whether a dataset follows a Pareto or truncated Pareto distribution, it is also useful to test for possible truncation in a dataset with evidence of power law tails. In this article, we develop parameter estimates for the truncated Pareto distribution that are easy to compute, prove their consistency (and, in some cases, asymptotic normality), and propose a way to compare the fit of truncated and unbounded Pareto distributional models on the basis of data. These methods should be useful to practitioners in areas of science and engineering where power law probability tails are prevalent.

Heavy-tailed random variables are important in applications in finance (Embrechts, Klüppelberg, and Mikosch 1997; Fama 1965; Jansen and de Vries 1991; Loretan and Phillips 1994; Mandelbrot 1963; McCulloch 1996; Meerschaert and Scheffler 2003; Rachev and Mittnik 2000), physics (Barkai, Metzler, and Klafter 2000; Klafter, Blumen, and Shlesinger 1987; Kotulski 1995; Meerschaert, Benson, Scheffler, and Becker-Kern 2002; Metzler and Klafter 2000), hydrology (Anderson and Meerschaert 1998; Benson, Schumer, Meerschaert, and Wheatcraft 2001; Benson, Wheatcraft, and Meerschaert 2000; Hosking and Wallis 1987; Lu and Molz 2001; Schumer, Benson, Meerschaert, and Wheatcraft 2001), engineering (Nikias and Shao 1995; Resnick 1997; Resnick and Stărică 1995; Uchaikin and Zolotarev 1999), and many other fields (Adler, Feldman, and Taqqu 1998; Feller 1971; Samorodnitsky and Taqqu 1994). Recently, Burroughs and Tebbens (2001b, 2002) surveyed evidence of truncated power law distributions in datasets on earthquake magnitudes, forest fire areas, fault lengths (on Earth and on Venus), and oil and gas field sizes. Earthquake fault sizes are limited by physical (or terrain) considerations (see Scholtz and Contreras 1998; Pacheco, Scholtz, and Sykes (1992). The size of forest fires may be naturally limited by the availability of fuel and climate (see Malamud, Morein, and Turcotte 1998). Additional applications of truncated power law distributions in finance, groundwater hydrology, and atmospheric science are given in Section 4.

Burroughs and Tebbens estimated parameters of the truncated Pareto distribution by least squares fitting on a probability plot (2001a) and by minimizing mean squared error fit on a plot of the tail distribution function (2001b). Minimum variance unbiased estimators for the parameters of a truncated Pareto law were developed by Beg (1981, 1983). Beg (1981) also provided maximum likelihood estimates of the lower truncation parameter, scale, and probability of exceedance for a truncated Pareto distribution. The maximum likelihood estimator (MLE) of α when the lower truncation limit is known was presented by Cohen and Whitten (1988), with some recommendations for the case when the lower truncation limit is not known. In this article we develop MLEs for all parameters of a truncated Pareto distribution. We prove the existence and uniqueness of the MLE under certain easy-to-check conditions that are shown to hold with probability approaching 1 as the sample size increases. A simple formula for the MLE is obtained that can be easily computed. Asymptotic normality is established for the estimator of the tail parameter α. We also consider the case where only the upper tail of the data fits a truncated Pareto distribution, and we compute the conditional MLE based on the upper-order statistics. These results can be used for robust tail estimation for truncated models with power law tails (e.g., stable data). Examples from hydrology, climatology, and finance illustrate the utility and practical application of these results. For ease of reading, proofs are deferred to the Appendix.

Inmaculada B. Aban is Assistant Professor, Department of Biostatistics, University of Alabama, Birmingham, AL 35294 (E-mail: *caban@uab.edu*). Mark M. Meerschaert is Professor, Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand (E-mail: *mcubed@maths.otago.ac.nz*). Anna K. Panorska is Assistant Professor, Department of Mathematics and Statistics, University of Nevada, Reno, NV 89557-0045 (E-mail: *ania@unr.edu*). This work was supported in part by National Science Foundation grants DMS-01-39927, DMS-04-17869, and ATM-0231781 and by the Marsden Foundation in New Zealand. The authors thank Alexander Gershunov, Scripps Institution of Oceanography, for his help in obtaining and understanding the precipitation data in Figure 4, and Fred Molz of Clemson University for his help in understanding the depositional process and measurement procedures at the MADE site.

Journal of the American Statistical Association
March 2006, Vol. 101, No. 473, Theory and Methods
DOI 10.1198/016214505000000411

2. MAXIMUM LIKELIHOOD ESTIMATION FOR THE ENTIRE SAMPLE

A random variable W has a Pareto distribution function if

$$P(W > w) = \gamma^{\alpha} w^{-\alpha} \quad \text{for } w \geq \gamma > 0 \text{ and } \alpha > 0. \qquad (1)$$

An upper-truncated Pareto random variable, X, has distribution

$$1 - F_X(x) = P(X > x) = \frac{\gamma^{\alpha}(x^{-\alpha} - \nu^{-\alpha})}{1 - (\gamma/\nu)^{\alpha}} \qquad (2)$$

and density

$$f_X(x) = \frac{\alpha \gamma^{\alpha} x^{-\alpha-1}}{1 - (\gamma/\nu)^{\alpha}} \qquad (3)$$

for $0 < \gamma \leq x \leq \nu < \infty$, where $\gamma < \nu$. Consider a random sample $X_1, X_2, \ldots, X_n$ from this upper-truncated distribution and let $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$ denote its order statistics. We now derive the MLE for the unknown parameters. We start by considering the simplest case, where γ and ν are both known.

Theorem 1. The MLE, $\tilde{\alpha}$, of α for an upper-truncated Pareto defined in (2), where γ and ν are known, solves the equation

$$\frac{n}{\tilde{\alpha}} + \frac{n(\gamma/\nu)^{\tilde{\alpha}} \ln(\gamma/\nu)}{1 - (\gamma/\nu)^{\tilde{\alpha}}} - \sum_{i=1}^{n} \left[\ln X_{(i)} - \ln \gamma\right] = 0. \qquad (4)$$

Furthermore, $\sqrt{n}(\tilde{\alpha} - \alpha)$ converges to a normal distribution with mean 0 and variance

$$\left\{ \frac{1}{\alpha^2} - \frac{(\gamma/\nu)^{\alpha}[\ln(\gamma/\nu)]^2}{[1 - (\gamma/\nu)^{\alpha}]^2} \right\}^{-1}.$$

Remark 1. The MLE for α presented in (4) was first given by Cohen and Whitten (1988) for the case where the lower truncation limit γ is known. These authors also made recommendations for computing the estimate when γ is unknown.

Next we obtain the MLEs for the parameters of an upper-truncated Pareto distribution where α, γ, and ν are all unknown.

Theorem 2. Consider a random sample $X_1, X_2, \ldots, X_n$ from an upper-truncated Pareto distribution defined in (2) where α, γ, and ν are unknown. The MLEs for the parameters in this model are given by $\hat{\gamma} = X_{(n)} = \min(X_1, X_2, \ldots, X_n)$, $\hat{\nu} = X_{(1)} = \max(X_1, X_2, \ldots, X_n)$, and $\hat{\alpha}$ solves the equation

$$\frac{n}{\hat{\alpha}} + \frac{n[X_{(n)}/X_{(1)}]^{\hat{\alpha}} \ln[X_{(n)}/X_{(1)}]}{1 - [X_{(n)}/X_{(1)}]^{\hat{\alpha}}} - \sum_{i=1}^{n} \left[\ln X_{(i)} - \ln X_{(n)}\right] = 0. \qquad (5)$$

When γ and ν are unknown, the upper-truncated Pareto distribution has support that depends on these unknown parameters, and hence it is not a member of a multiparameter exponential family of distributions. Furthermore, it does not satisfy the regularity conditions necessary to directly apply the asymptotic properties of MLEs to derive the asymptotic distribution of $\hat{\alpha}$. We instead use a Taylor series expansion to obtain the asymptotic properties of $\hat{\alpha}$. Note that it is easy to show, using basic results in probability, that $\hat{\gamma}$ and $\hat{\nu}$ are consistent estimators of γ and ν.

Theorem 3. Under the conditions of Theorem 2, the MLE, $\hat{\alpha}$, of α has an asymptotic normal distribution with asymptotic mean α.

Remark 2. The asymptotic variance of $\hat{\alpha}$ is not the same as the asymptotic variance of $\tilde{\alpha}$, because adjustments need to be made on the asymptotic variance of $\hat{\alpha}$ because γ and ν were estimated. Furthermore, the standard theory of obtaining asymptotic variance using the Fisher information matrix is no longer applicable.

The next result shows that the MLE optimization problem for α typically has a unique solution.

Theorem 4. The probability that a solution to the MLE equation (4) or (5) exists tends to 1 as $n \to \infty$, and if it exists, then the solution is unique.

We compare the performance of the proposed estimators with that of the estimators of Hill and Beg. We generate $m = 1{,}000$ random samples of size $n = 100$ from a truncated Pareto distribution with $\alpha = .8$, $\gamma = 1$, and $\nu = 10$. Figure 1 summarizes the observed sampling distributions of the estimators based on this simulation. The proposed estimators and Beg's estimators perform well, with Beg's estimators performing a little better in terms of the bias. It is a well-known fact that MLE is often biased, but this bias disappears as the sample size increases. Hill's estimator performed well only in estimating ν and poorly estimated α, which we attribute to model misspecification. When we simulated data (results not presented here) from a regular Pareto distribution, Hill's estimators performed best; that is, the center of the simulated distribution was near the true value, and the observed spread was the least. The latter property results from the fact that Hill's estimator estimates only two parameters, and hence there is no additional uncertainty due to estimating the upper truncation parameter. But for samples from a truncated Pareto distribution, we recommend using Beg's estimators or our proposed estimators. One advantage of our estimators is that they may also be extended to the case where the true distribution is not a truncated Pareto but the tail behaves like a truncated Pareto. We discuss this in the next section.

3. MAXIMUM LIKELIHOOD ESTIMATION FOR THE TAIL

In many practical applications, a truncated Pareto distribution may be fit to the upper tail of the data if, for sufficiently large $x > 0$,

$$P(X > x) \approx \frac{\gamma^{\alpha}(x^{-\alpha} - \nu^{-\alpha})}{1 - (\gamma/\nu)^{\alpha}}, \qquad 0 < \gamma \leq x \leq \nu.$$

In this case we estimate the parameters by obtaining the conditional MLE based on the $(r+1)$ $(0 \leq r < n)$ largest-order statistics representing only the portion of the tail where the truncated Pareto approximation holds. This extends the well-known Hill estimator (Hall 1982; Hill 1975) to the case of a truncated Pareto distribution.

Theorem 5. When $X_{(r)} > X_{(r+1)}$, the conditional MLE for the parameters of the upper-truncated Pareto distribution in (2) based on the $(r+1)$ largest-order statistics is given by $\hat{\nu} = X_{(1)}$,

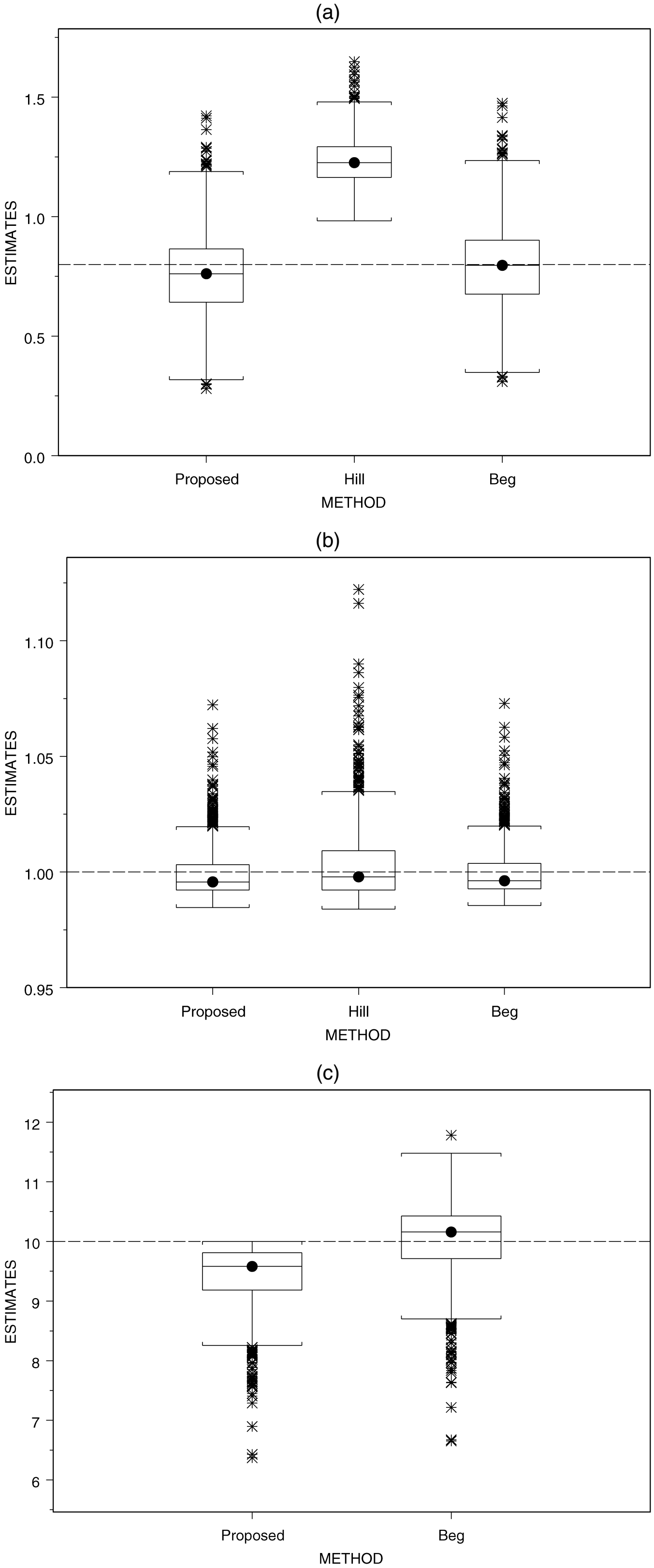

Figure 1. Boxplots of the Observed Sampling Distributions of the Estimates of (a) α, (b) γ, and (c) ν Based on $m = 1{,}000$ Simulated Datasets of Size $n = 100$ From a Truncated Pareto With $\alpha = .8$, $\gamma = 1$, and $\nu = 10$. Broken horizontal lines represent the true values of the parameters.

$\hat{\gamma} = r^{1/\hat{\alpha}}(X_{(r+1)})[n - (n-r)(X_{(r+1)}/X_{(1)})^{\hat{\alpha}}]^{-1/\hat{\alpha}}$, and $\hat{\alpha}$ solves the equation

$$0 = \frac{r}{\hat{\alpha}} + \frac{r(X_{(r+1)}/X_{(1)})^{\hat{\alpha}} \ln(X_{(r+1)}/X_{(1)})}{1 - (X_{(r+1)}/X_{(1)})^{\hat{\alpha}}} - \sum_{i=1}^{r}\bigl[\ln X_{(i)} - \ln X_{(r+1)}\bigr]. \quad (6)$$

Remark 3. The conditional MLE, $\hat{\alpha}_{TP}$, of α given in Theorem 5 is smaller than the Hill estimator, $\hat{\alpha}_H$ of α. This is easy to see after noting that the second term on the right side of (6) is always negative.

Remark 4. In some applications it may be natural to use a truncated and possibly shifted exponential distribution. All of the results in this section and the previous section also apply to that case, because $Y = \ln X$ has a truncated shifted exponential distribution with

$$P(Y > y) = \frac{e^{-\alpha(y - \ln\gamma)} - (\gamma/\nu)^{\alpha}}{1 - (\gamma/\nu)^{\alpha}} \quad \text{for } \ln\gamma \le y \le \ln\nu$$

if and only if X has a truncated Pareto distribution. If $\gamma = 1$, then Y has a truncated exponential distribution with no shift.

Remark 5. For a dataset that graphically exhibits a power law tail, we propose a simple test to check whether a Pareto model is appropriate, based on simple results from extreme value theory. Our proposed asymptotic level-q test ($0 < q < 1$) rejects the null hypothesis $H_0 : \nu = \infty$ (Pareto) if and only if $X_{(1)} < [(nC)/(-\ln q)]^{1/\alpha}$, where $C = \gamma^{\alpha}$. The corresponding approximate p value of this test is given by $p = \exp\{-nCX_{(1)}^{-\alpha}\}$. In practice, we use Hill's estimator (Hill 1975),

$$\hat{\alpha}_H = \left[r^{-1} \sum_{i=1}^{r} \bigl\{ \ln X_{(i)} - \ln X_{(r+1)} \bigr\} \right]^{-1} \quad \text{and}$$

$$\hat{C} = \frac{r}{n}\bigl(X_{(r+1)}\bigr)^{\hat{\alpha}_H},$$

to estimate the parameters C and α. Note that a small p value in this case indicates that the Pareto model is not a good fit. But in itself this is not sufficient to indicate goodness of fit of the truncated Pareto distribution. To check whether a truncated Pareto distribution is a reasonably good fit, the test should always be supplemented by a graphical check of the data tail, as illustrated in the next section.

4. APPLICATIONS

Figure 2 shows the upper tail for a dataset of absolute daily price changes in U.S. dollars for Amazon, Inc. stock from January 1, 1998 to June 30, 2003 ($n = 1{,}378$), plotted on a log–log scale. The apparent downward curve is characteristic of a truncated Pareto distribution. Hill's estimator gives $\hat{\alpha}_H = 2.343$ and $\hat{C} = 2.203$, whereas the truncated Pareto estimator from Theorem 5 gives $\hat{\alpha}_{TP} = 1.681$, $\hat{\gamma} = .963$, and $\hat{\nu} = 15.53$ (estimated $C = \gamma^{\alpha} \doteq .936$). Both estimates are based on the upper $r = 100$ order statistics. Using the test in Remark 5, the computed p value of approximately .007 is strong evidence that a regular Pareto is not a good fit to the data. In contrast, Figure 2 shows that the truncated Pareto model fits the data very well

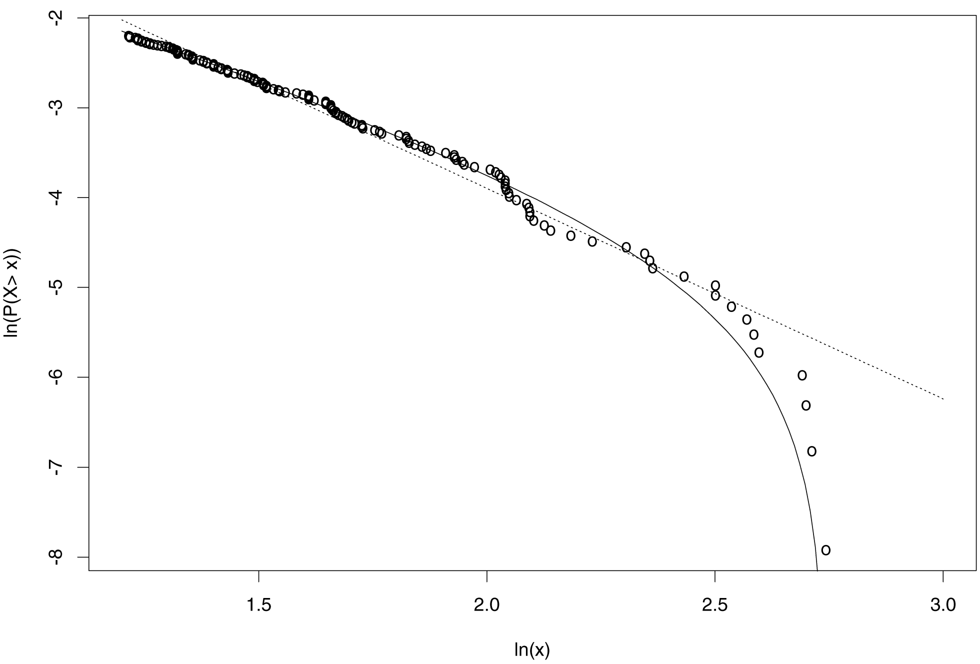

Figure 2. Log–Log Plot of the Largest Absolute Values of Daily Price Changes in Amazon, Inc. Stock, With Best-Fitting Pareto (----) and Truncated Pareto (——) Tail Distributions.

and much better than a regular Pareto. One possible explanation for the fit of the truncated Pareto in this application is that the stock exchange uses automatic mechanisms to slow trading in the event of extreme price changes due to automatic trading by large mutual funds.

Figure 3 shows the upper tail of the survival function for absolute differences from a series of 2,618 measurements of hydraulic conductivity in K cm/sec taken at 15-cm intervals in vertical boreholes at the MAcroDispersion Experiment (MADE) site on the Columbus Air Force Base in northeastern Mississippi (see Rehfeldt, Boggs, and Gelhar 1992), along with the best-fitting Pareto and truncated Pareto models. The truncated Pareto model seems to be a good fit to the data, whereas the Pareto model is not a good fit. The estimated parameters of the truncated Pareto model based on the 100 upper-order statistics are $\hat{\gamma} = .008$, $\hat{\nu} = .783$, and $\hat{\alpha}_{TP} = 1.196$. The estimated parameters of the Pareto model are $\hat{C} = .0011$ and $\hat{\alpha}_H = 1.598$, so that the estimated $\gamma = C^{1/\alpha} \doteq .014$. The p value from the test in

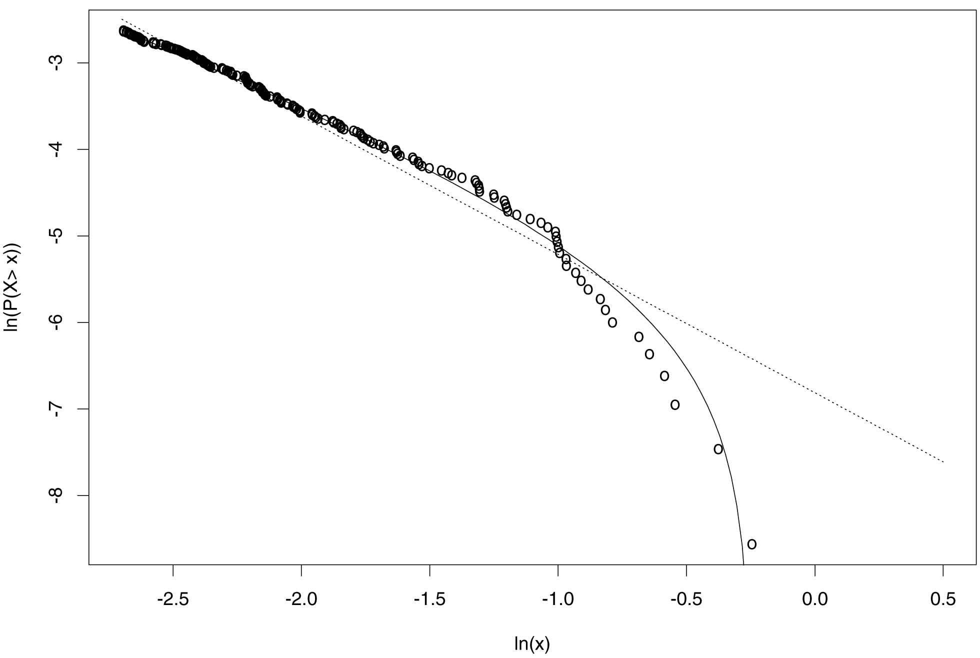

Figure 3. Log–Log Plot of the Empirical Survival Function for Absolute Differences of Hydraulic Conductivity at the MADE Site, With Best-Fitting Pareto (----) and Truncated Pareto (——) Tail Distributions.

Table 1. Estimates of Truncated Pareto Parameters for Hydraulic Conductivity at the MADE Site (n = 2,618), and p Values, for Varying r Values

(r + 1) largest-order statistics	Estimates					Pareto test p value
	$\hat{\gamma}$	$\hat{\nu}$	$\hat{\alpha}$	Hill's $\hat{\alpha}$	Hill's $\hat{C}$	
50	.0073	.7828	1.1663	1.9000	.0008	.04068
100	.0079	.7828	1.1958	1.5985	.0011	.0120
200	.0057	.7828	1.0562	1.3160	.0020	.0009
300	.0042	.7828	.9434	1.1496	.0028	.0001

Remark 5 of approximately .012 confirms that Pareto model is not a good fit. The α estimate from the truncated Pareto model is also in good agreement with the analysis of Benson et al. (2001). A possible explanation for truncation is that the depositional process that formed the MADE aquifer created high-velocity channels that account for the largest K values. Over time, these high-flow channels are occluded with sedimentation, truncating the high K values. Another truncation effect comes from the K measurement process, which averages values over a cylindrical region. The highest K values occur in a relatively small sector of this cylinder and thus are attenuated by combination with the more prevalent lower K values. Table 1 shows how the parameter estimates and the p value for the Pareto test of Remark 5 vary with the number r of upper-order statistics used. As a general guide, the value of r should be chosen on the basis of a log–log plot similar to Figure 3 so that r is as large as possible as long as the model fit is adequate.

Figure 4 shows the 100 largest observations of total daily precipitation in units of .1 mm at Tombstone, Arizona between July 1, 1893 and December 31, 2001 (from Groisman et al. 2004), along with the best-fitting Pareto and truncated Pareto models. Based on the $n = 5{,}216$ nonzero observations, the estimated parameters of the truncated Pareto model are $\hat{\gamma} = 88.025$, $\hat{\nu} = 762$, and $\hat{\alpha}_{TP} = 2.933$. The estimated parameters of the Pareto model are $\hat{C} = 7.59478 \times 10^7$ and $\hat{\alpha}_H = 3.813$, so that the estimated $\gamma = C^{1/\alpha} \doteq 117$. The p value from the test in Remark 5 of approximately .017 is evidence against the regular Pareto model. The fact that the tail of the data curves downward in Figure 4 is evidence in support of a truncated Pareto model. Note that because our parameter estimates are based on the nonzero observations, we are actually modeling the conditional distribution of precipitation given that some precipitation occurs. It is generally assumed that precipitation data have an upper bound, called the "probable maximum precipitation," computed by various methods (see, e.g., Douglas and Barros 2003; Thompson and Tomlinson 1995; World Meteorological Organization 1986), so that a truncated Pareto is more consistent with standard practice in hydrometerology than the unbounded model.

Another alternative model, in which the parameter estimates $\alpha = 2.933$ and $C = 88.025^{2.933}$ from the truncated Pareto are used in the Pareto, is appropriate if the investigator believes that precipitation follows a power law that has been truncated by some natural or observational mechanism. For illustration purposes, we present estimates of the .999th quantile for three models for daily precipitation. For the Pareto model, the quan-

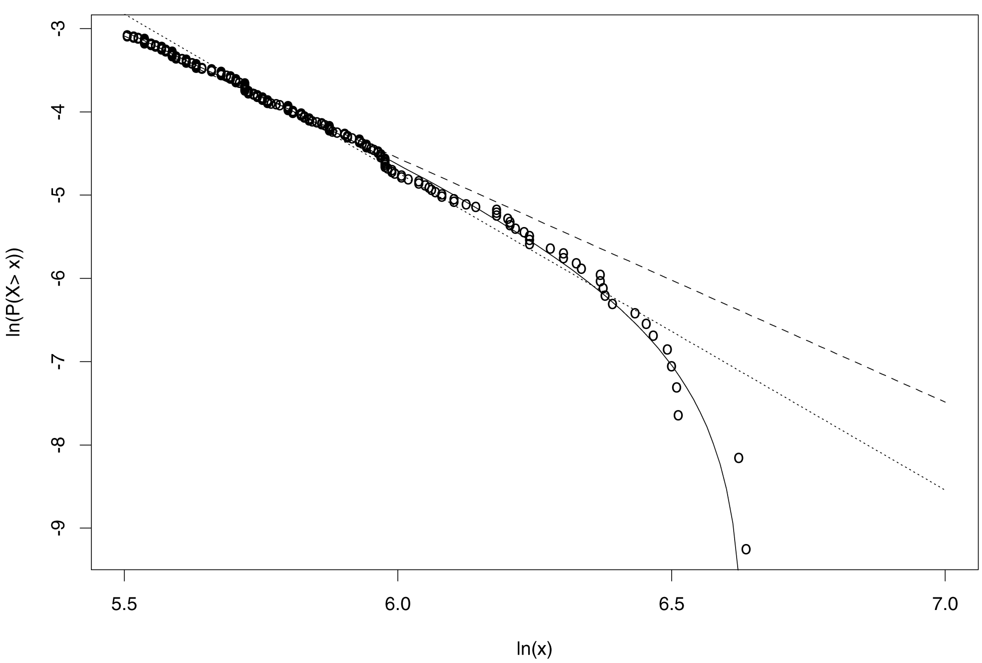

Figure 4. Log–Log Plot of the Empirical Survival Function for the 100 Largest Observations of Positive Daily Total Precipitation in Tombstone, AZ Between July 1, 1893 to December 31, 2001 With Best-Fitting Pareto (·····), Pareto With Truncated Pareto Parameters (----), and Truncated Pareto (——) Tail Distributions.

tile is estimated as 714 (71.4 mm or 2.81 inches of precipitation); for the truncated Pareto model, it is 655; and for the Pareto model with truncated Pareto parameters, it is 928. Because precipitation was recorded on only 13% of the days in this period, the .999th percentile actually represents the level of daily precipitation that one could expect on any given day with probability (.13 × .001), or about once every 20 years. This example is typical in that for extreme upper values, the truncated Pareto model gives the lowest estimate, and the Pareto model with truncated Pareto parameters gives the highest estimate.

APPENDIX: PROOFS

Proof of Theorem 1

The likelihood function under this setting is given by

$$L(\gamma,\nu,\alpha)=\frac{\alpha^n\gamma^{n\alpha}}{[1-(\gamma/\nu)^\alpha]^n}\times\prod_{i=1}^n X_i^{-\alpha-1}\prod_{i=1}^n I\{0<\gamma\le X_i\le\nu<\infty\},\quad\text{(A.1)}$$

where $\gamma<\nu$. Maximizing this likelihood, we get the MLE for α. Next, note that the density in (3) when γ and ν are fixed and known is a one-parameter exponential family of distributions. Hence the asymptotic distribution of the MLE $\tilde{\alpha}$ is obtained by applying the results on the asymptotic properties of the MLE for a one-parameter exponential family (see, e.g., thm. 5.3.5 in Bickel and Doksum 2001).

Proof of Theorem 2

Referring to likelihood function in (A.1) with γ and ν unknown, note that $\prod_{i=1}^n I\{0<\gamma\le X_i\le\nu<\infty\}=1$ if and only if $X_{(n)}\ge\gamma$ and $X_{(1)}\le\nu$. Because the likelihood function L is an increasing function of γ for $\gamma\le X_{(n)}$ and a decreasing function of ν for $\nu\ge X_{(1)}$, for a fixed α, the likelihood is maximized when $\hat{\gamma}=X_{(n)}$ and $\hat{\nu}=X_{(1)}$. One can then easily derive (5).

Proof of Theorem 3

For a given γ and ν, suppose that $h\equiv h(\gamma,\nu)$ solves the equation

$$G=\frac{1}{h}+\frac{(\gamma/\nu)^h\ln(\gamma/\nu)}{1-(\gamma/\nu)^h}-\frac{1}{n}\sum_{i=1}^n\bigl(\ln X_{(i)}-\ln\gamma\bigr)=0\quad\text{(A.2)}$$

for a fixed γ and ν. Then, by Theorem 1, $\tilde{\alpha}=h(\gamma_0,\nu_0)$, where γ_0 and ν_0 are the true values of γ and ν, and by Theorem 2, $\hat{\alpha}=h(\hat{\gamma},\hat{\nu})$, where $\hat{\gamma}=X_{(n)}$ and $\hat{\nu}=X_{(1)}$. By a Taylor series expansion about γ_0 and ν_0,

$$\hat{\alpha}=h(\hat{\gamma},\hat{\nu})\approx h(\gamma_0,\nu_0)+h_\gamma(\gamma_0,\nu_0)(\hat{\gamma}-\gamma_0)+h_\nu(\gamma_0,\nu_0)(\hat{\nu}-\nu_0),\quad\text{(A.3)}$$

where $h_\gamma(\gamma_0,\nu_0)$ and $h_\nu(\gamma_0,\nu_0)$ are the partial derivatives of h with respect to γ and ν, evaluated at γ_0 and ν_0. Because $\hat{\nu}$ and $\hat{\gamma}$ are consistent estimators of ν and γ, the last two terms on the right side go to 0 in probability as $n\to\infty$, provided that $h_\gamma(\gamma_0,\nu_0)$ and $h_\nu(\gamma_0,\nu_0)$ are bounded. Applying Slutsky's theorem (see, e.g., Bickel and Doskum 2001), it then follows that $\hat{\alpha}$ has asymptotic mean α and converges in distribution to an asymptotic normal distribution.

We complete the proof by showing that the partial derivatives $h_\gamma(\gamma_0,\nu_0)$ and $h_\nu(\gamma_0,\nu_0)$ are bounded. If (A.2) holds, then $dG/d\gamma=\partial G/\partial\gamma+(\partial G/\partial h)(\partial h/\partial\gamma)=0$, implying that $h_\gamma(\gamma_0,\nu_0)=\partial h/\partial\gamma=(-\partial G/\partial\gamma)/(\partial G/\partial h)$ and, similarly, $h_\nu=(-\partial G/\partial\nu)/(\partial G/\partial h)$. Because neither $\partial G/\partial\gamma$ nor $\partial G/\partial\nu$ has any singularities for $0<\gamma<\nu$ and $h>0$, both $h_\gamma(\gamma_0,\nu_0)$ and $h_\nu(\gamma_0,\nu_0)$ are bounded if we can show that $\partial G/\partial h\neq 0$. Let $b=\gamma/\nu$. In fact, we have

$$\frac{\partial G}{\partial h}=\frac{-(1-b^h)^2+h^2b^h(\ln b)^2}{h^2(1-b^h)^2}<0\quad\text{for every } 0<b<1 \text{ and } h>0.\quad\text{(A.4)}$$

To see this, first note that

$$e^x-e^{-x}>2x\quad\text{for all } x>0.\quad\text{(A.5)}$$

Now (A.4) is equivalent to $h^2b^h(\ln b)^2<(1-b^h)^2$, so that $-hb^{h/2}(\ln b)<1-b^h$. Using (A.5), where $x=(-h/2)\ln b$, we get $-h\ln b<b^{-h/2}-b^{h/2}$ for all $h>0$.

Proof of Theorem 4

First, we consider the case where γ and ν are known. In this case the MLE $\tilde{\alpha}=h$, where $h>0$ solves the equation

$$G(h)=\frac{1}{h}+\frac{b^h\ln b}{1-b^h}-\frac{1}{n}\sum_{i=1}^n\bigl(\ln X_{(i)}-\ln\gamma\bigr)=0,$$

where $b=\gamma/\nu\in(0,1)$. Write $M=(1/n)\sum_{i=1}^n(\ln X_{(i)}-\ln\gamma)$ and compute $\lim_{h\to 0}G(h)=(-\ln b)/2-M$ and $\lim_{h\to\infty}G_1(h)=0-M$. In view of (A.4), we have that $\partial G/\partial h<0$ for all $h>0$, so that G is monotone. Hence the solution to $G(h)=0$ is unique if it exists, and it exists if and only if $M<(-\ln b)/2$. As $n\to\infty$, we have $M\to E\{\ln(X/\gamma)\}$ in probability by the law of large numbers, where $W=\ln(X/\gamma)$ has a truncated exponential distribution with density $g(w)=\alpha\exp\{-\alpha w\}/[1-\exp\{-\alpha(-\ln b)\}]$ supported on $0<w<-\ln b$. Because the density of W is monotone decreasing, a simple geometrical argument shows that the mean of W must lie in the left half of the interval $[0,-\ln b]$, and hence $EW<-(1/2)\ln b$. Then $P\{M<(-\ln b)/2\}\to 1$ as $n\to\infty$, so that the MLE $\tilde{\alpha}$ exists with probability approaching 1 as $n\to\infty$.

In the case where all three parameters are estimated, the MLE $\hat{\alpha}=h$ solves $G(h)=1/h+[(b^h\ln b)/(1-b^h)]-(1/n)\sum_{i=1}^n(\ln X_{(i)}-\ln X_{(n)})=0$, where $b=X_{(n)}/X_{(1)}\in(0,1)$ with probability 1. Write $M'=(1/n)\sum_{i=1}^n(\ln X_{(i)}-\ln X_{(n)})$ and compute $\lim_{h\to 0}G(h)=(-\ln b)/2-M'$ and $\lim_{h\to\infty}G_1(h)=0-M'$ as before. Now (A.4) shows that G is monotone, and hence the solution $\hat{\alpha}=h$ to $G(h)=0$ is unique if it exists, and it exists if and only if $M'=M+\ln(\gamma/X_{(n)})<(-\ln b)/2$. Because $X_{(n)}\to\gamma$ in probability, the continuous mapping theorem implies that $\ln(\gamma/X_{(n)})\to 0$ in probability. Because $M\to E\{\ln(X/\gamma)\}$ in probability, another application of the continuous mapping theorem shows that $M'\to E\{\ln(X/\gamma)\}$ in probability as well, and the same argument as before shows that the MLE $\hat{\alpha}$ exists with probability approaching 1 as $n\to\infty$.

Proof of Theorem 5

The proof is similar to that of the Hill estimator (Hill 1975). It is convenient to transform the data, taking $Z_{(i)}=(X_{(n-i+1)})^{-1}$ so that $G(z)=P(Z\le z)=\gamma^\alpha(z^\alpha-\nu^{-\alpha})/[1-(\gamma/\nu)^\alpha]$ and $Z_{(1)}\ge\cdots\ge Z_{(n)}$. Because $U_{(i)}=G(Z_{(i)})$ are (decreasing) order statistics from a uniform distribution, $E_{(i)}=-\ln U_{(i)}$ are (increasing) order statistics from a unit exponential. Following David (1981, pp. 20–21), we let $Y_i=(n-i+1)(E_{(i)}-E_{(i-1)})$, and hence we can easily check that $\{Y_i, i=1,\ldots,n\}$ are independent and identically distributed unit exponential. Define $Y^*=nE_{(n-r+1)}=n(Y_1/n+\cdots+Y_{n-r+1}/r)$ so that $Y_n,\ldots,Y_{n-r+2},Y^*$ are mutually independent with joint density $\exp(-y_{(n)}-\cdots-y_{(n-r+2)})p(y^*)$, where $p(y^*)$ is the density of Y^*. Because $U_{(n-r+1)}$ has density $K_1u^{r-1}(1-u)^{n-r}$, it follows that $Y^*=-n\ln U_{(n-r+1)}$ has density $p(y)=K_2\exp(-y/n)^r(1-\exp(y/n))^{n-r}$, where $\{K_j, j=1,2\}$ are positive constants. Use the fact that $Y_{(i)}=(n-i+1)[-\ln G(Z_{(i)})+\ln G(Z_{(i-1)})]=(n-$

$i+1)[\ln(Z_{(i-1)}^{\alpha} - \nu^{-\alpha}) - \ln(Z_{(i)}^{\alpha} - \nu^{-\alpha})]$, for $i = n-r+2, \ldots, n$, and $\exp(-Y^*/n) = G(Z_{(n-r+1)}) = \gamma^{\alpha}(Z_{(n-r+1)}^{\alpha} - \nu^{-\alpha})/[1-(\gamma/\nu)^{\alpha}]$, to obtain the likelihood conditional on the values of the $(r+1)$ smallest-order statistics, $Z_{(n)} = z_{(n)}, \ldots, Z_{(n-r+1)} = z_{(n-r+1)}$, as

$$K\left[\prod_{i=1}^{r} \frac{\alpha z_{(n-i+1)}^{\alpha-1}}{z_{(n-i+1)}^{\alpha} - \nu^{-\alpha}}\right] \times \exp\left[-\sum_{i=1}^{r-1} i\left[\ln\left(z_{(n-i)}^{\alpha} - \nu^{-\alpha}\right) - \ln\left(z_{(n-i+1)}^{\alpha} - \nu^{-\alpha}\right)\right]\right] \times \left[\frac{\gamma^{\alpha}(z_{(n-r+1)}^{\alpha} - \nu^{-\alpha})}{1-(\gamma/\nu)^{\alpha}}\right]^{r}\left[1 - \frac{\gamma^{\alpha}(z_{(n-r+1)}^{\alpha} - \nu^{-\alpha})}{1-(\gamma/\nu)^{\alpha}}\right]^{n-r} \times \prod_{i=1}^{r} I\left\{\frac{1}{\nu} \le z_{(n-i+1)} \le \frac{1}{\gamma}\right\},$$

where $K > 0$ does not depend on the data or on the parameters and the first product term is the Jacobian for the transformations defined by $Y_n, \ldots, Y_{n-r+2}, Y^*$. Note that

$$\sum_{i=1}^{r-1} i\left[\ln\left(z_{(n-i)}^{\alpha} - \nu^{-\alpha}\right) - \ln\left(z_{(n-i+1)}^{\alpha} - \nu^{-\alpha}\right)\right] = r\ln\left(z_{(n-r+1)}^{\alpha} - \nu^{-\alpha}\right) - \sum_{i=1}^{r} \ln\left(z_{(n-i+1)}^{\alpha} - \nu^{-\alpha}\right).$$

Next, condition on $Z_{(n-r+1)} \le d < Z_{(n-r)}$, which multiplies the conditional likelihood by a factor of

$$\left[1 - \frac{\gamma^{\alpha}(d^{\alpha} - \nu^{-\alpha})}{1-(\gamma/\nu)^{\alpha}}\right]^{n-r}\left[1 - \frac{\gamma^{\alpha}(z_{(n-r+1)}^{\alpha} - \nu^{-\alpha})}{1-(\gamma/\nu)^{\alpha}}\right]^{-(n-r)}.$$

Then the conditional likelihood function simplifies to

$$K\alpha^{r}\frac{\gamma^{r\alpha}[1-(\gamma d)^{\alpha}]^{n-r}}{[1-(\gamma/\nu)^{\alpha}]^{n}}\left[\prod_{i=1}^{r} z_{(n-i+1)}^{\alpha-1}\right]\prod_{i=1}^{r} I\left\{\frac{1}{\nu} \le z_{(n-i+1)} \le \frac{1}{\gamma}\right\}.$$

Substitute $\beta = (\gamma d)^{\alpha}$ to obtain

$$K\alpha^{r}\frac{\beta^{r} d^{-r\alpha}(1-\beta)^{n-r}}{[1-\beta(d\nu)^{-\alpha}]^{n}}\left[\prod_{i=1}^{r} z_{(n-i+1)}^{\alpha-1}\right]\prod_{i=1}^{r} I\left\{\frac{1}{\nu} \le z_{(n-i+1)} \le \frac{1}{\gamma}\right\}.$$

In terms of the original data, this conditional likelihood becomes

$$K\alpha^{r}\frac{\beta^{r} D^{r\alpha}(1-\beta)^{n-r}}{[1-\beta(D/\nu)^{\alpha}]^{n}}\left[\prod_{i=1}^{r} x_{(i)}^{-\alpha+1}\right]\left[\prod_{i=1}^{r} x_{(i)}^{-2}\right]\prod_{i=1}^{r} I\{\gamma \le x_{(i)} \le \nu\},$$

where $d = D^{-1}$ and the product term $[\prod_{i=1}^{r} x_{(i)}^{-2}]$ is the Jacobian associated with the transformations $X_{(i)} = (Z_{(n-i+1)})^{-1}$ for $i = 1, \ldots, r$. Note that $\prod_{i=1}^{r} I\{\gamma \le x_{(i)} \le \nu\}$ if and only if $x_{(1)} \le \nu$ and $x_{(r)} \ge \gamma$. Consequently, whenever $x_{(1)} \le \nu$ and $x_{(r)} \ge \gamma$, the conditional log-likelihood is

$$\ln L = K_0 + r\ln\alpha + r\ln\beta + r\alpha\ln D + (n-r)\ln(1-\beta) - n\ln\left(1-\beta(D/\nu)^{\alpha}\right) - (\alpha+1)\sum_{i=1}^{r}\ln\left(x_{(i)}\right),$$

and we seek the global maximum over the parameter space consisting of all (α, β) for which $\alpha > 0$ and $0 < \beta < 1$. Because this conditional log-likelihood is a decreasing function of ν for $\nu \ge X_{(1)}$, the conditional MLE for ν is $\hat{\nu} = X_{(1)}$. Substituting the $\hat{\nu}$ for ν and taking partial derivatives of the conditional log-likelihood with respect to α and β, we get

$$\frac{\partial \ln L}{\partial \alpha} = \frac{r}{\alpha} + \frac{n\beta(D/X_{(1)})^{\alpha}\ln(D/X_{(1)})}{1-\beta(D/X_{(1)})^{\alpha}} - \sum_{i=1}^{r}\left[\ln X_{(i)} - \ln D\right] \tag{A.6}$$

and

$$\frac{\partial \ln L}{\partial \beta} = \frac{r}{\beta} - \frac{n-r}{1-\beta} + \frac{n(D/X_{(1)})^{\alpha}}{1-\beta(D/X_{(1)})^{\alpha}}. \tag{A.7}$$

We obtain the MLE for $\hat{\alpha}$ and $\hat{\beta}$ by setting these partial derivatives to 0 and solving. The solution to (A.6) and (A.7) depends on the value of D, which will be unknown in practice. Because we assume $X_{(r)} \ge D > X_{(r+1)}$ we can estimate D by $X_{(r)}$ or $X_{(r+1)}$ or some point in between. We use the estimate $D = X_{(r+1)}$ to be consistent with standard usage for Hill's estimator (Anderson and Meerschaert 1997; Fofack and Nolan 1999; Hall 1982; Jansen and de Vries 1991; Loretan and Phillips 1994; Resnick 1997; Resnick and Stărică 1995), MLE for α in a Pareto. Setting the equations in (A.6) to 0 and solving, we find that $\hat{\beta} = r[n-(n-r)(X_{(r+1)}/X_{(1)})^{\hat{\alpha}}]^{-1}$ and $\hat{\alpha}$ solves the equation

$$0 = \frac{r}{\hat{\alpha}} + \frac{r(X_{(r+1)}/X_{(1)})^{\hat{\alpha}}\ln(X_{(r+1)}/X_{(1)})}{1-(X_{(r+1)}/X_{(1)})^{\hat{\alpha}}} - \sum_{i=1}^{r}\left[\ln X_{(i)} - \ln X_{(r+1)}\right].$$

We complete the proof of the theorem by using $\gamma = D\beta^{1/\alpha}$.

[Received May 2004. Revised March 2005.]

REFERENCES

Adler, R., Feldman, R., and Taqqu, M. (1998), *A Practical Guide to Heavy Tails*, Boston: Birkhäuser.

Anderson, P., and Meerschaert, M. M. (1997), "Periodic Moving Averages of Random Variables With Regularly Varying Tails," *The Annals of Statistics*, 25, 771–785.

——— (1998), "Modeling River Flows With Heavy Tails," *Water Resources Research*, 34, 2271–2280.

Arnold, B. C. (1983), *Pareto Distributions*, Burtonsville, MD: International Cooperative Publishing House.

Barkai, E., Metzler, R., and Klafter, J. (2000), "From Continuous Time Random Walks to the Fractional Fokker–Planck Equation," *Physics Review*, Ser. E, 61, 132–138.

Beg, M. A. (1981), "Estimation of the Tail Probability of the Truncated Pareto Distribution," *Journal of Information & Optimization Sciences*, 2, 192–198.

——— (1983), "Unbiased Estimators and Tests for Truncation and Scale Parameters," *American Journal of Mathematical and Management Sciences*, 3, 251–274.

Benson, D., Schumer, R., Meerschaert, M. M., and Wheatcraft, S. (2001), "Fractional Dispersion, Lévy Motions, and the MADE Tracer Tests," *Transport in Porous Media*, 42, 211–240.

Benson, D., Wheatcraft, S., and Meerschaert, M. M. (2000), "Application of a Fractional Advection-Dispersion Equation," *Water Resources Research*, 36, 1403–1412.

Bickel, P., and Doksum, K. (2001), *Mathematical Statistics Basic Ideas and Selected Topics*, Vol. 1 (2nd ed.), Upper Saddle River, NJ: Prentice-Hall.

Burroughs, S. M., and Tebbens, S. F. (2001a), "Upper-Truncated Power Law Distributions," *Fractals*, 9, 209–222.

——— (2001b), "Upper-Truncated Power Laws in Natural Systems," *Journal of Pure and Applied Geophysics*, 158, 741–757.

——— (2002), "The Upper-Truncated Power Law Applied to Earthquake Cumulative Frequency-Magnitude Distributions," *Bulletin of the Seismological Society of America*, 92, 2983–2993.

Cohen, A. C., and Whitten, B. J. (1988), *Parameter Estimation in Reliability and Life Span Models*, New York: Marcel Dekker.

David, H. (1981), *Order Statistics* (2nd ed.), New York: Wiley.

Douglas, E. M., and Barros, A. P. (2003), "Probable Maximum Precipitation Estimation Using Multifractals: Application in the Eastern United States," *Journal of Hydrometeorology*, 4, 1012–1024.

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

Fama, E. (1965), "The Behavior of Stock Market Prices," *Journal of Business*, 38, 34–105.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol. II (2nd ed.), New York: Wiley.

Fofack, H., and Nolan, J. P. (1999), "Tail Behavior, Modes and Other Characteristics of Stable Distribution," *Extremes*, 2, 39–58.

Groisman, P. Y., Knight, R. W., Karl, T. R., Easterling, D. R., Sun, B., and Lawrimore, J. H. (2004), "Contemporary Changes of the Hydrological Cycle Over the Contiguous United States: Trends Derived From in Situ Observations," *Journal of Hydrometeorology*, 5, 64–85.

Hall, P. (1982), "On Some Simple Estimates of an Exponent of Regular Variation," *Journal of the Royal Statistical Society*, Ser. B, 44, 37–42.

Hill, B. (1975), "A Simple General Approach to Inference About the Tail of a Distribution," *The Annals of Statistics*, 3, 1163–1173.

Hosking, J., and Wallis, J. (1987), "Parameter and Quantile Estimation for the Generalized Pareto Distribution," *Technometrics*, 29, 339–349.

Jansen, D., and de Vries, C. (1991), "On the Frequency of Large Stock Market Returns: Putting Booms and Busts Into Perspective," *Review of Economics and Statistics*, 73, 18–24.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994), *Continuous Univariate Distributions* (2nd ed.), New York: Wiley.

Klafter, J., Blumen, A., and Shlesinger, M. F. (1987), "Stochastic Pathways to Anomalous Diffusion," *Physics Review*, Ser. A, 35, 3081–3085.

Kotulski, M. (1995), "Asymptotic Distributions of the Continuous Time Random Walks: A Probabilistic Approach," *Journal of Statistic Physics*, 81, 777–792.

Loretan, M., and Phillips, P. (1994), "Testing the Covariance Stationarity of Heavy-Tailed Time Series," *Journal of Empirical Finance*, 1, 211–248.

Lu, S. L., and Molz, F. J. (2001), "How Well Are Hydraulic Conductivity Variations Approximated by Additive Stable Processes?" *Advances in Environmental Research*, 5, 39–45.

Malamud, B. D., Morein, G., and Turcotte, D. L. (1998), "Forest Fires: An Example of Self-Organized Critical Behavior," *Science*, 281, 1840–1842.

Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394–419.

McCulloch, J. (1996), "Financial Applications of Stable Distributions," in *Statistical Methods in Finance*, eds. G. Madfala and C. R. Rao, Amsterdam: Elsevier, pp. 393–425.

Meerschaert, M. M., Benson, D. A., Scheffler, H. P., and Becker-Kern, P. (2002), "Governing Equations and Solutions of Anomalous Random Walk Limits," *Physics Review*, Ser. E, 66, 102–105.

Meerschaert, M. M., and Scheffler, H. P. (2003), "Portfolio Modeling With Heavy-Tailed Random Vectors," in *Handbook of Heavy-Tailed Distributions in Finance*, ed. S. T. Rachev, New York: Elsevier, pp. 595–640.

Metzler, R., and Klafter, J. (2000), "The Random Walk's Guide to Anomalous Diffusion: A Fractional Dynamics Approach," *Physics Report*, 339, 1–77.

Nikias, C., and Shao, M. (1995), *Signal Processing With Alpha-Stable Distributions and Applications*, New York: Wiley.

Pacheco, J., Scholtz, C. H., and Sykes, L. R. (1992), "Changes in Frequency–Size Relationship From Small to Large Earthquakes," *Nature*, 355, 71–73.

Rachev, S., and Mittnik, S. (2000), *Stable Paretian Models in Finance*, Chichester, U.K.: Wiley.

Rehfeldt, K. R., Boggs, J. M., and Gelhar, L. W. (1992), "Field Study of Dispersion in a Heterogeneous Aquifer. 3: Geostatistical Analysis of Hydraulic Conductivity," *Water Resources Research*, 28, 3309–3324.

Resnick, S. (1997), "Heavy Tail Modeling and Teletraffic Data," *The Annals of Statistics*, 25, 1805–1869.

Resnick, S., and Stărică, C. (1995), "Consistency of Hill's Estimator for Dependent Data," *Journal of Applied Probability*, 32, 139–167.

Samorodnitsky, G., and Taqqu, M. (1994), *Stable Non-Gaussian Random Processes*, New York: Chapman & Hall.

Scholz, C. H., and Contreras, J. C. (1998), "Mechanics of Continental Rift Architecture," *Geology*, 26, 967–970.

Schumer, R., Benson, D., Meerschaert, M. M., and Wheatcraft, S. (2001), "Eulerian Derivation of the Fractional Advection-Dispersion Equation," *Journal of Contaminant Hydrology*, 38, 69–88.

Thompson, C. S., and Tomlinson, A. I. (1995), "A Guide to Probable Maximum Precipitation in New Zealand," Science and Technolody Series Technical Note ST19 (*http://www.niwa.co.nz/pubs/st*), National Institute of Water & Atmospheric Research of New Zealand.

Uchaikin, V., and Zolotarev, Z. (1999), *Chance and Stability: Stable Distributions and Their Applications*, Utrecht: VSP.

World Meteorological Organization (1986), *Manual for Estimation of Probable Maximum Precipitation*, Geneva: Author.