Robust and Efficient Estimation for the Generalized Pareto Distribution

Sergio F. Juárez
Faculty of Statistics and Informatics
Veracruzana University, Xalapa, Ver, México
email: sejuarez@uv.mx
and
William R. Schucany
Department of Statistical Science
Southern Methodist University, Dallas, TX, USA
email: schucany@smu.edu

July 9, 2004

Abstract

In this article we implement the *minimum density power divergence estimator* (MDPDE) for the shape and scale parameters of the generalized Pareto distribution (GPD). The MDPDE is indexed by a constant $\alpha \geq 0$ that controls the trade-off between robustness and efficiency. As α increases, robustness increases and efficiency decreases. For $\alpha = 0$ the MDPDE is equivalent to the maximum likelihood estimator (MLE). We show that for $\alpha > 0$ the MDPDE for the GPD has a bounded influence function. For $\alpha < 0.2$ the MDPDE maintains good asymptotic relative efficiencies, usually above 90%. The results from a Monte Carlo study agree with these asymptotic calculations. The MDPDE is asymptotically normally distributed if the shape parameter is less than $(1 + \alpha)/(2 + \alpha)$, and estimators for standard errors are readily computed under this restriction. We compare the MDPDE, MLE, Dupuis' *optimally-biased robust* estimator (OBRE), and Peng and Welsh's *Medians* estimator for the parameters. The simulations indicate that the MLE has the highest efficiency under uncontaminated GPDs. However, for the GPD contaminated with gross errors OBRE and MDPDE are more efficient than the MLE. For all the simulated models that we studied the Medians estimator had poor performance.

Keywords: Excesses over high thresholds, extreme values, minimum distance, optimally-biased robust estimator, medians estimator.

1 Introduction

The generalized Pareto distribution (GPD) is an extreme value model widely used for excesses over high thresholds. The most common estimators for the shape and scale parameters of the GPD are maximum likelihood, moments, and probability weighted moments. Other estimators have been proposed motivated by disadvantages of the common estimators, other than their lack of robustness. Even though the non-robustness of maximum likelihood and probability weighted moments was pointed out in Davison and Smith (1990), only recently have robustness issues come into consideration for fitting the GPD. To the best of our knowledge, robust estimation for the GPD was not addressed until the articles by Dupuis (1998) and by Peng and Welsh (2002). Dupuis implements Hampel's optimally-biased robust estimator (OBRE). Peng and Welsh derive estimators (denoted by Medians here) from equating medians of sample and population score functions. In this article we implement a general class of *minimum-divergence* type estimators. The class contains robust and efficient estimators for the parameters of the GPD.

In extreme value modeling the extremes are the important part of the data, as opposed to other statistical procedures where extreme observations are removed or their influence downweighted. In a sample of excesses, the larger extremes may be the most informative observations for estimating the parameters of the GPD, or other quantities derived from the fit of this distribution. Apparently, the implication is that the "extreme extremes" should not be downweighted with the use of robust procedures. However, it is precisely in the "extreme extremes" where contamination by gross errors most likely will occur. Since procedures such as maximum likelihood or probability weighted moments are non-robust, a single extreme that is not consistent with the bulk of extremes may jeopardize the inferences drawn. On the other hand, a robust fit to a sample of excesses would be relatively insensitive to departures from the model. This is valuable in those real problems where it is desirable to have a model that fits well for the bulk of the data without being affected by outlying observations. Clearly, there are situations where a more complex modeling approach has to be followed, for example using mixture models or changepoint models. Another good reason for a robust fit for the GPD, is that it automatically takes into account departures from the postulated model. This is especially good with large data sets where contaminating observations are almost certain to occur and a detailed diagnostic analysis is not feasible. Even when a detailed diagnostic analysis is possible, robust procedures are valuable too. They serve as diagnostic tools for standard techniques, such as maximum likelihood, because they can indirectly help to validate model assumptions. For instance, if the MLE fit does not differ from a robust fit, then one may be confident that there are not grossly outlying observations affecting the MLE.

The article is organized as follows. In Section 2 we introduce the MDPDE, some asymptotic and finite sample properties with technical details relegated to the appendix. In Section 3 we report the results of a simulation comparing MDPDE, OBRE, Medians, and MLE. In Section 4 we apply the previous estimators to 100 years of monthly rainfall data. Finally, in Section 5 we conclude by itemizing the competitiveness of MDPDE with MLE.

2 Minimum Density Power Divergence Estimators

Let $X_1, \ldots, X_n$ be a random sample from the distribution G with density g, and let $\mathcal{F} = \{f(x;\theta), x \in \mathcal{X} : \theta \in \Theta \subseteq \mathbb{R}^p\}$ be a parametric family of densities. Assume this family is identifiable in the sense that the set $\{x \in \mathcal{X} : f(x;\theta_1) \neq f(x;\theta_2)\}$ has positive Lebesgue measure if $\theta_1 \neq \theta_2$.

Basu, Harris, Hjort, and Jones (1998), BHHJ henceforth, introduce the *density power divergence* (DPD) between the densities f and g as

$$d_\alpha(g, f) = \int_{\mathcal{X}} \left\{ f^{1+\alpha}(x) - \left(1 + \frac{1}{\alpha}\right) g(x) f^\alpha(x) + \frac{1}{\alpha} g^{1+\alpha}(x) \right\} \mathrm{d}x$$

for positive α, and for $\alpha = 0$ as

$$d_0(g, f) = \lim_{\alpha \to 0} d_\alpha(g, f) = \int_{\mathcal{X}} g(x) \log[g(x)/f(x)] \mathrm{d}x.$$

For fixed $\alpha > 0$, the minimum DPD functional T_α at G is defined as the point in the parameter space Θ corresponding to the element in $\mathcal{F}$ *closest* to g so that

$$d_\alpha(g, f(\cdot; T_\alpha(G))) = \inf_{\theta \in \Theta} d_\alpha(g, f(\cdot;\theta)).$$

This minimum DPD functional $\theta_0 = T_\alpha(G)$ is the target parameter. BHHJ define the *minimum density power divergence* estimator (MDPDE) as the point $\hat{\theta}_{\alpha,n} = T_\alpha(G_n) \in \Theta$ associated with the element of $\mathcal{F}$ that is closest to the empirical density function g_n. Here G_n denotes the empirical distribution function. That is, $\hat{\theta}_{\alpha,n}$ minimizes $d_\alpha(g_n, f)$ with respect to θ. Thus, the MDPDE is obtained by minimizing with respect to θ

$$H_\alpha(\theta) = \int_{\mathcal{X}} f^{1+\alpha}(x;\theta) \mathrm{d}x - \left(1 + \frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^{n} f^\alpha(X_i;\theta), \tag{1}$$

over the parameter space Θ. When $\alpha = 0$, $T_0(G_n)$ maximizes $(1/n)\sum_{i=1}^n \log f(X_i;\theta)$, and $T_0(G_n)$ is the MLE. When $\alpha = 1$, $T_1(G_n)$ minimizes $\int_{\mathcal{X}} f^2(x;\theta)\mu(\mathrm{d}x) - (2/n)\sum_{i=1}^n f(X_i;\theta)$, and $T_1(G_n)$ is the L^2 estimator. BHHJ show that under certain regularity conditions the MDPDE is a consistent estimator of θ_0 and its distribution, properly normalized, is asymptotically normal.

In our case, $X_1, X_2, \ldots, X_n$, is a random sample of excesses from an unknown excess distribution with density g. The parametric model is given by the GPD family

$$f(x;\xi,\beta) = \frac{1}{\beta}\left(1-\xi\frac{x}{\beta}\right)^{\xi^{-1}-1}, x \in D(\xi,\beta), (\xi,\beta) \in \Theta,$$

where $D(\xi,\beta) = [0,\infty)$ if $\xi \leq 0$, and $D(\xi,\beta) = [0,\beta/\xi]$ if $\xi > 0$, and $\Theta = \{(\xi,\beta) \in \mathbb{R}^2 : \xi \in \mathbb{R}, \beta > 0\}$. Let (ξ_0,β_0) be the target parameter. That is, $\inf_{(\xi,\beta)\in\Theta} d_\alpha(g,f) = d_\alpha(g, f(\cdot;\xi_0,\beta_0))$. Thus, for fixed $\alpha > 0$ the MDPDE for the GPD is the value $(\hat{\xi}_\alpha, \hat{\beta}_\alpha)$ that minimizes

$$H_\alpha(\xi,\beta) = \frac{1}{\beta^\alpha(1+\alpha-\alpha\xi)} - \left(1+\frac{1}{\alpha}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\beta^\alpha}\left(1-\xi\frac{X_i}{\beta}\right)^{(\xi^{-1}-1)\alpha}$$

over

$$\left\{(\xi,\beta) \in \Theta : \beta > 0, \max_{1\leq i\leq n}\{X_i\}\xi < \beta, -\infty < \xi < 0, \text{ and } 0 < \xi < (1+\alpha)/\alpha\right\}.$$

The restriction $\max_{1\leq i\leq n}\{X_i\}\xi < \beta$ is due to the dependence of the support on the parameters. The restriction $\xi < (1+\alpha)/\alpha$ is needed to assure the existence of the integral in (1). The limiting case $\xi = 0$ is omitted because it corresponds to the exponential distribution, which is treated in BHHJ.

2.1 Asymptotic Distribution

Consider the score function and the information matrix of the GPD, $S(x;\xi,\beta)$, and $i(x;\xi,\beta)$, respectively. Their expressions are in Appendix I. Define the symmetric 2×2 matrices K_α and J_α as follows

$$K_\alpha(\xi,\beta) = \int_{D(\xi,\beta)} S(x;\xi,\beta)S^t(x;\xi,\beta)f^{1+2\alpha}(x;\xi,\beta)\mathrm{d}x - U_\alpha(\xi,\beta)U_\alpha^t(\xi,\beta), \tag{2}$$

where

$$U_\alpha(\xi,\beta) = \begin{bmatrix} \int_{D(\xi,\beta)} S_\xi(x;\xi,\beta)f^{\alpha+1}(x;\xi,\beta)\mathrm{d}x \\ \int_{D(\xi,\beta)} S_\beta(x;\xi,\beta)f^{\alpha+1}(x;\xi,\beta)\mathrm{d}x \end{bmatrix},$$

and

$$\begin{aligned} J_\alpha(\xi,\beta) &= \int_{D(\xi,\beta)} S(x;\xi,\beta)S^t(x;\xi,\beta)f^{1+\alpha}(x;\xi,\beta)\mathrm{d}x \\ &+ \int_{D(\xi,\beta)} \left\{i(x;\xi,\beta) - \alpha S(x;\xi,\beta)S^t(x;\xi,\beta)\right\}[g(x)-f(x;\xi,\beta)]f^\alpha(x;\xi,\beta)\mathrm{d}x. \end{aligned} \tag{3}$$

The asymptotic bivariate normality of the MDPDE for the GPD is established in the following theorem.

Theorem 1. *Let* (ξ_0, β_0) *be the target parameter in the GPD model. Suppose* $\xi_0 < (1+\alpha)/(2+\alpha)$ *for* $\alpha > 0$ *fixed, and that the integrability conditions (8)-(13) in Appendix I are satisfied. Then there exists a sequence of MDPD estimators* $\{(\hat{\xi}_{\alpha,n}, \hat{\beta}_{\alpha,n})\}$ *that is consistent for* (ξ_0, β_0) *as* $n \to \infty$*. Furthermore*

$$\sqrt{n}(\hat{\xi}_{\alpha,n} - \xi_0, \hat{\beta}_{\alpha,n} - \beta_0)^t \rightsquigarrow N((0,0)^t, V(\xi_0, \beta_0)), \tag{4}$$

where $V(\xi_0, \beta_0) = J_\alpha^{-1}(\xi_0, \beta_0) K_\alpha(\xi_0, \beta_0) J_\alpha^{-1}(\xi_0, \beta_0)$*.*

This result is proved in Juárez (2003). *Under model conditions*, meaning the true distribution g is GPD, the integrability conditions (8)-(13) are all satisfied and the matrices K_α and J_α have closed form expressions (Juárez, 2003). Note that when $\alpha \to 0$ and under model conditions we obtain the well known restriction $\xi_0 < 1/2$ for which the MLE of the GPD parameters is asymptotically normal (Smith, 1985). Also note that for any $\alpha > 0$ the MDPDE slightly enlarges the region of the parameter space where it is asymptotically normal as compared to the MLE. Furthermore, under model conditions, the integrability conditions in Appendix I do not imply $\alpha < 1/2$. In theory for certain ξ_0, the MDPDE can be implemented for any positive value of α, however, we will only be interested in small values of α (see next section).

The empirical versions of the matrices K_α and J_α are

$$\begin{aligned} \hat{K}_\alpha(\xi, \beta) &= \frac{1}{n}\sum_{i=1}^{n} S(X_i; \xi, \beta) S^t(X_i; \xi, \beta) f^{2\alpha}(X_i; \xi, \beta) \\ &\quad - \frac{1}{n^2}\left\{\sum_{i=1}^{n} S(X_i; \xi, \beta) f^\alpha(X_i; \xi, \beta)\right\}\left\{\sum_{i=1}^{n} S(X_i; \xi, \beta) f^\alpha(X_i; \xi, \beta)\right\}^t \end{aligned}$$

and

$$\begin{aligned} \hat{J}_\alpha(\xi, \beta) &= (1+\alpha)\int_{D(\xi,\beta)} S(x; \xi, \beta) S^t(x; \xi, \beta) f^{\alpha+1}(x; \xi, \beta)\mathrm{d}x \\ &\quad - \int_{D(\xi,\beta)} i(x; \xi, \beta) f^{\alpha+1}(x; \xi, \beta)\mathrm{d}x \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\{i(X_i; \xi, \beta) - \alpha S(X_i; \xi, \beta) S^t(X_i; \xi, \beta)\} f^\alpha(X_i; \xi, \beta). \end{aligned}$$

Substituting $(\hat{\xi}_\alpha, \hat{\beta}_\alpha)$ in these matrices, we obtain $n^{-1}\hat{J}_\alpha^{-1}(\hat{\xi}_\alpha, \hat{\beta}_\alpha)\hat{K}_\alpha(\hat{\xi}_\alpha, \hat{\beta}_\alpha)\hat{J}_\alpha^{-1}(\hat{\xi}_\alpha, \hat{\beta}_\alpha)$ as a consistent estimator of the variance-covariance matrix of the MDPDE for the GPD.

2.2 Influence Function and Efficiency

Under model conditions the influence function of the MDPDE with fixed α for the GPD is given by

$$\mathrm{IF}(x; \alpha) = J_\alpha^{-1}(\xi, \beta)[S(x; \xi, \beta) f^\alpha(x; \xi, \beta) - U_\alpha(\xi, \beta)]. \tag{5}$$

For $\xi < 0$, $S_\xi(x;\xi,\beta)$ and $S_\beta(x;\xi,\beta)$ are continuous on $[0,\infty)$ and the density vanishes as $x \to \infty$. Use of l'Hôpital's rule yields $\lim_{x\to\infty} S_\xi(x;\xi,\beta)f^\alpha(x;\xi,\beta) = 0$ and $\lim_{x\to\infty} S_\beta(x;\xi,\beta)f^\alpha(x;\xi,\beta) = 0$. Therefore, the influence function (5) is bounded. For $0 < \xi < (1+\alpha)/(2+\alpha)$ the boundedness of the influence function follows from the compactness of the support $[0,\beta/\xi]$. For $\xi > (1+\alpha)/(2+\alpha)$, J_α^{-1} does not exist, Juárez (2003), and so the influence function does not exist. When β is fixed, the influence function of ξ is

$$\mathrm{IF}_\xi(x;\alpha) = \frac{S_\xi(x;\xi,\beta)f^\alpha(x;\xi,\beta) - \int_{D(\xi,\beta)} S_\xi(x;\xi,\beta)f^{1+\alpha}(x;\xi,\beta)\mathrm{d}x}{\int_{D(\xi,\beta)} S_\xi^2(x;\xi,\beta)f^{1+\alpha}(x;\xi,\beta)\mathrm{d}x}. \tag{6}$$

For fixed ξ the influence function of β is

$$\mathrm{IF}_\beta(x;\alpha) = \frac{S_\beta(x;\xi,\beta)f^\alpha(x;\xi,\beta) - \int_{D(\xi,\beta)} S_\beta(x;\xi,\beta)f^{1+\alpha}(x;\xi,\beta)\mathrm{d}x}{\int_{D(\xi,\beta)} S_\beta^2(x;\xi,\beta)f^{1+\alpha}(x;\xi,\beta)\mathrm{d}x}. \tag{7}$$

In Figures 1 and 2 we see the unboundedness of the influence function of the MLE (the solid line). In particular, when the scale (shape) parameter is fixed, arbitrarily increasing a single observation would make the MLE of the shape (scale) parameter decrease (increase) with no lower (upper) bound. Also, the influence function of the MDPDE approaches a horizontal asymptote more slowly as α gets closer to zero.

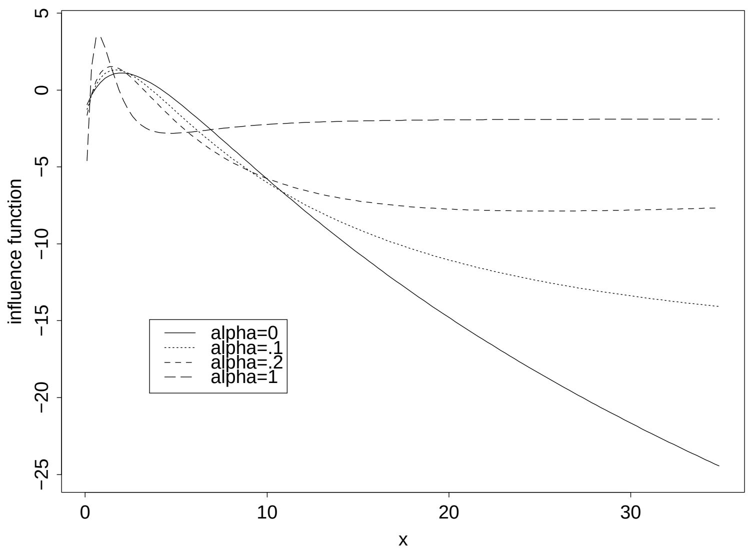

Figure 1: Influence functions for MDPDE of the shape parameter (6) for four values of α. The model is GPD($\xi = -.1, \beta = 1$). The solid line is the influence function of the MLE.

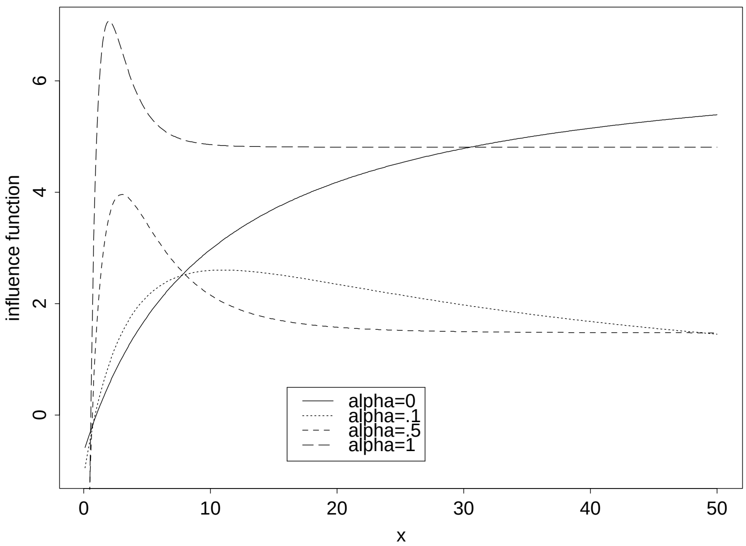

Figure 2: Influence functions for MDPDE of the scale parameter (7) for four values of α. The model is GPD($\xi = -.25, \beta = 1$). The solid line is the influence function of the MLE.

Table 1 displays some values of the ARE of the MDPDE with respect to the MLE for estimation of the shape parameter ξ with $\beta = 1$. These AREs are calculated from the 1-1 element of $V(\xi_0, \beta_0)$ in (4) and the asymptotic variance, $(1 - \xi_0)^2$, of the MLE of ξ. We see that the MDPDE has high relative efficiency for small values of α. However, for larger values of α the efficiency decreases to unacceptable levels. We also see that as the tail of the GPD gets heavier (ξ decreases) the ARE tends to decrease. Values of the tuning constant α close to zero retain a reasonable level of the ARE.

Table 1: Asymptotic relative efficiencies of the MDPDE of the shape parameter for different values of α with respect to the MLE for scale parameter $\beta = 1$ and six choices of the shape ξ.

	α										
ξ	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	1
0.45	0.97	0.92	0.79	0.67	0.56	0.46	0.39	0.32	0.27	0.22	0.19
0.10	0.98	0.93	0.79	0.64	0.51	0.41	0.33	0.27	0.22	0.19	0.16
−0.10	0.98	0.92	0.77	0.61	0.48	0.37	0.30	0.25	0.21	0.18	0.16
−0.25	0.98	0.92	0.75	0.58	0.45	0.35	0.29	0.24	0.20	0.18	0.16
−0.50	0.97	0.90	0.71	0.54	0.42	0.33	0.27	0.23	0.20	0.18	0.16
−1.00	0.96	0.86	0.65	0.48	0.37	0.30	0.26	0.22	0.19	0.17	0.15

2.3 The Effect of α in Finite Samples

We conducted a simulation experiment to investigate the effect of α under model conditions. For each pair (n, ξ), where $n \in \{50, 100, 200, 400\}$ and $\xi \in \{-1, -.5, -.25, .25, .45\}$, we generated 500 samples of size n from the GPD$(\xi, \beta = 1)$. Without loss of generality we fixed the parameter $\beta = 1$, because both the MLE and the MDPDE are equivariant under scale changes. For each value of α in { 0, .02, .04, .05, .06, .08, .1, .2, .3, .4, .5, .6, .7, .8, .9, 1} we evaluated the MDPDE $(\hat{\xi}_{\alpha,n}, \hat{\beta}_{\alpha,n})$ and computed their empirical variances, biases, and MSEs.

Table 2 reports the values of α yielding the smallest estimated MSEs for ξ and for β. The interesting feature is that small values of α (in general $\alpha < 0.2$) give empirical MSEs close to the empirical MSEs of MLE. However, this may not be surprising since under model conditions, the MLE is asymptotically efficient. We observed that the estimated MSEs increased drastically and the non-convergence rate of the MDPDE increased as the value of α gets closer to 1. There is good agreement among these minimizers for moderate to large n.

Table 2: The pairs (α_1, α_2) are the values of α_1 and α_2 yielding the smallest Monte Carlo MSEs for ξ and β, respectively.

	ξ				
n	-1	-0.5	-0.25	0.25	0.45
50	(0,.02)	(0,.1)	(0,.2)	(.2,.5)	(.2,.5)
100	(0,.04)	(0,.1)	(0,0)	(0,0)	(.2,.3)
200	(0,.04)	(0,.04)	(0,0)	(.08,0)	(.08,.1)
400	(0,0)	(.02,0)	(0,0)	(.06,.08)	(.1,.2)

3 Comparison to Other Estimators

To compare the performance of MLE, OBRE, Medians and MDPDE, we generated samples from the mixture contaminated model $(1-\epsilon)\text{GPD}(\xi_0, \beta_0) + \epsilon\text{GPD}(\xi_1, \beta_1)$, $\epsilon \in \{0, .1, .2\}$. This model is denoted by $(\xi_0, \beta_0, \epsilon, \xi_1, \beta_1)$. From each mixture model we generated 500 samples of size 100 and evaluated the four estimators. For the OBRE we used Dupuis' (1998) iterative algorithm with robustness constant $c = 4$. The S-Plus routines for computing the estimators are available at `http://www.smu.edu/statistics/faculty/SPlusFilesSergio.pdf`. Based on Tables 1 and 2, we decided to use a fixed $\alpha = 0.1$. We used several of the parameter values from Peng and

Welsh (2002). Table 3 presents the models in our simulations together with the frequency of failure to converge of each estimator. We see that Medians and OBRE had a very high non-convergence rate when $\xi_0 = 1/3$. Thus, to avoid selection bias, we did not include Medians and OBRE in the comparisons in Table 4 when $\xi_0 = 1/3$.

Table 3: Number of samples, out of 500, where there was no convergence. The models marked with * were also considered by Peng and Welsh (2002).

$(\xi_0, \beta_0, \epsilon, \xi_1, \beta_1)$	MLE	MDPDE	Medians	OBRE
$(-1, 1, \times, \times, \times)$ *	0	0	7	28
$(-1, 1, 0.1, -2, 1)$ *	0	0	9	25
$(-1, 1, 0.1, -1, 2)$ *	0	0	6	24
$(-1, 1, 0.2, -2, 1)$	0	0	0	18
$(-1, 1, 0.2, -1, 2)$	0	0	0	10
$(-1/4, 1, 0.1, -2, 1)$	6	6	52	30
$(1/3, 1, \times, \times, \times)$ *	0	0	187	307
$(1/3, 1, 0.1, 2/3, 1)$ *	0	0	196	274

As an empirical measure of relative efficiency, in Table 4 we display the estimated relative mean square errors (MSE) of the robust estimators with respect to the MLE. When the ratio is less than one, the MLE has smaller MSE by that factor. In parentheses are estimates of the standard errors of these ratios (see Appendix II for details of the computation of these standard errors). The magnitudes of these standard errors indicate that 500 Monte Carlo repetitions estimate the true ratios of MSEs with adequate precision.

For $\xi_0 = -1$ with no contamination, the MLE is the best for ξ but tied for β. MLE is better than the other estimators for ξ under contamination with 10% from a GPD with $\xi_1 = -2$. However, for estimation of β both the MDPDE and the OBRE perform slightly better than MLE. The MLE of ξ is also better than the other estimators under 10% contamination from a GPD with $\beta_1 = 2$. When we increased the amount of contamination from 10% to 20%, MDPDE and OBRE slightly improved their performance. Medians had a poor performance in all the models with $\xi_0 = -1$. In these cases OBRE performs slightly better than MDPDE. When $\xi_0 = -1/4$ and 10% of contamination with $\xi_1 = -2$, MLE was very badly affected. For estimation of ξ, MDPDE and OBRE halved their empirical MSE's with respect to the empirical MSE of the MLE (ratios > 2). Once again, Medians had a poor performance. When $\xi_0 = 1/3$, MDPDE had a slightly better performance than MLE.

Table 4: Estimated relative MSE of the robust estimation procedures with respect to the MLE. In parentheses are the estimated standard errors of these empirical relative MSEs. Information not available in the empty cells.

$(\xi_0, \beta_0, \epsilon, \xi_1, \beta_1)$	Parameter	MDPDE	Medians	OBRE
$(-1, 1, \times, \times, \times)$	ξ	0.84 (.03)	0.21 (.03)	0.87 (.03)
	β	0.98 (.02)	0.60 (.03)	0.99 (.02)
$(-1, 1, 0.1, -2, 1)$	ξ	0.99 (.04)	0.35 (.03)	0.99 (.03)
	β	1.02 (.02)	0.68 (.05)	1.02 (.01)
$(-1, 1, 0.1, -1, 2)$	ξ	0.87 (.03)	0.25 (.03)	0.93 (.01)
	β	1.03 (.02)	0.70 (.02)	1.07 (.05)
$(-1, 1, 0.2, -2, 1)$	ξ	1.11 (.04)	0.55 (.04)	1.11 (.04)
	β	0.99 (.02)	0.65 (.02)	1.02 (.02)
$(-1, 1, 0.2, -1, 2)$	ξ	0.86 (.03)	0.25 (.02)	0.84 (.03)
	β	1.02 (.02)	0.71 (.04)	1.02 (.03)
$(-1/4, 1, 0.1, -2, 1)$	ξ	2.12 (.08)	0.07 (.01)	2.60 (.03)
	β	1.23 (.04)	0.81 (.06)	1.21 (.05)
$(1/3, 1, \times, \times, \times)$	ξ	1.05 (.02)		
	β	1.05 (.01)		
$(1/3, 1, .1, 2/3, 1)$	ξ	1.02 (.02)		
	β	1.01 (.01)		

The high rate of non-convergence observed in our Medians and OBRE runs is troublesome. Peng and Welsh also considered very short tail scenarios ($\xi_0 = 2, 4$). In pilot Monte Carlo runs we observed high rates of non-convergence for *all* the estimators when $\xi_0 = 2, 4$. That Peng and Welsh do not report convergence problems is puzzling. The non-convergence of the MLE for the GPD when $\xi > 1$ is well documented in the literature; see for instance Hosking and Wallis (1987). While other authors may find ways to finesse the convergence problems that we encountered with Medians, we find the resulting estimates to be unacceptably erratic.

To get a better understanding of the sampling distributions of the MDPDE and MLE, we ran a larger more focused simulation experiment. For each of two models without contamination and two contaminated, $(-1/4, 1, \times, \times, \times)$, $(-2, 1, \times, \times, \times)$, $(-1/4, 1, .1, -1, 1)$ and $(-1/4, 1, 0.1, -2, 1)$, we generated 20,000 samples of size 100 and evaluated the MLE and the MDPDE.

Figure 3 displays nonparametric density estimates of the sampling distributions of the MLE and MDPDE of ξ and scatterplots of the pairs. From these graphs we see that under model conditions, the empirical sampling distribution of the MDPDE of ξ has a heavier lower tail than the MLE. Contamination with the heavier tails ($\xi_1 = -1$ and $\xi_1 = -2$) affects the MLE, causing a heavy tail to the left of the true value of the shape $\xi_0 = -1/4$. This tail is heavier than the tail of the sampling distribution of the MDPDE. In the scatterplots we see that under model conditions MLE tends to be closer to the true value of ξ than the MDPDE. However, for the contaminated models the MDPDE is consistently closer to the true value $\xi_0 = -1/4$ than the MLE. For estimation of β and under model conditions (we omit the plots for β), MLE and MDPDE had very similar empirical sampling distributions. Contamination of the shape with $\xi_1 = -1$ had no major effect on the estimation of β. However, the more severe contamination with $\xi_1 = -2$ affected the MLE and not the MDPDE. The MDPDE was consistently closer to the true value $\beta_0 = 1$ when the contamination was with $\xi_1 = -2$, and it was slightly closer than MLE under contamination with $\xi_1 = -1$. We examined Pitman closeness as well as MSE. The comparisons yield the same general conclusions as from Table 4.

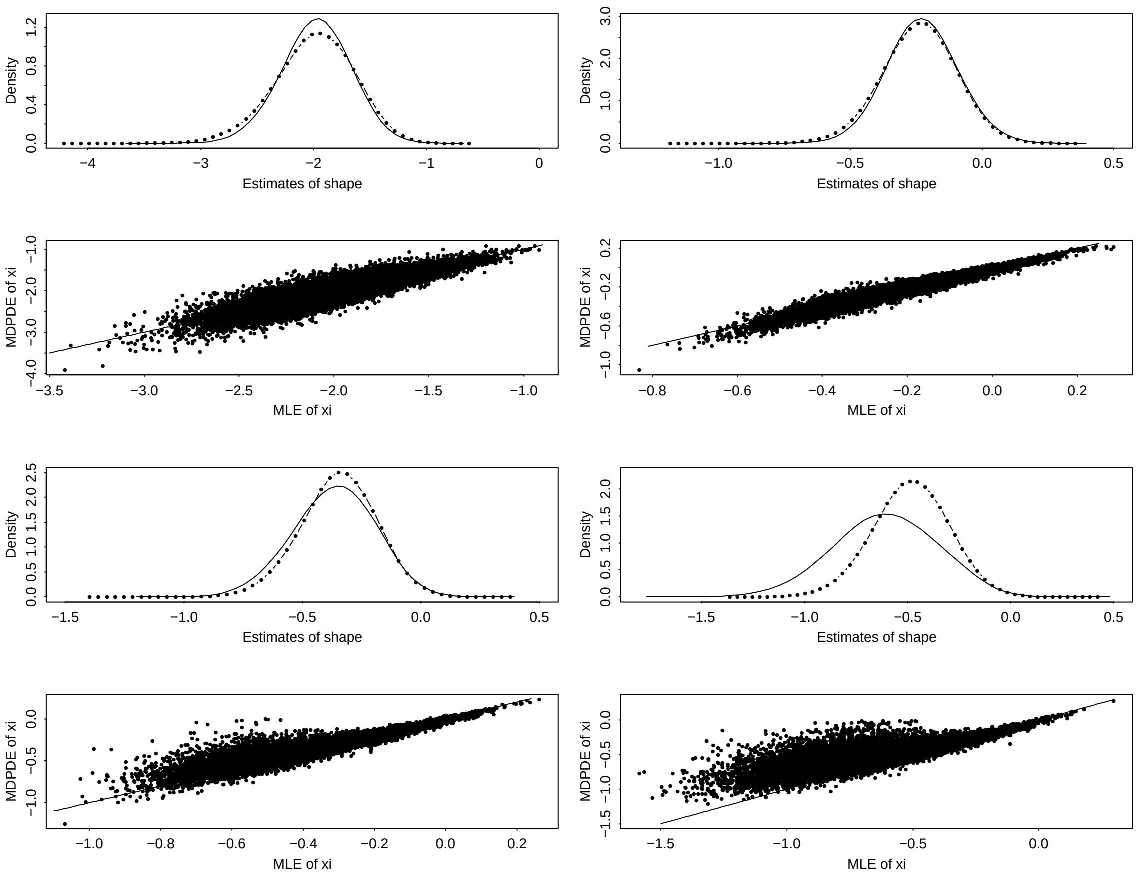

Figure 3: Empirical sampling distributions of the MDPDE (dotted line) and the MLE (solid line) of ξ based on 20,000 Monte Carlo samples of size 100 for each model. Scatterplots of MLE versus MDPDE, the line plotted is $y = x$. First and second rows: Left, $(-2, 1, \times, \times, \times)$; right, $(-1/4, 1, \times, \times, \times)$. Third and fourth rows: Left, $(-1/4, 1, 0.1, -1, 1)$; right, $(-1/4, 1, 0.1, -2, -1)$.

4 Example

Figure 4 displays the monthly accumulated rainfall in Xalapa, Ver., México, from 1904 to 2003. To see the differences between the estimators with a particular data set, we fit the GPD to the monthly rainfall excesses over 300 mm, see Table 6 in Appendix III. The resulting fits are presented in Table

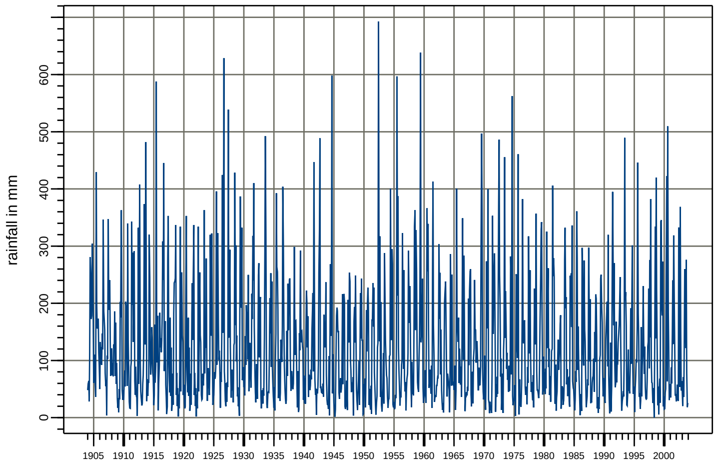

Figure 4: Monthly accumulated rainfall in Xalapa, Veracruz, México. Source: National Commission of Water, Management of the Center Gulf Region.

5. We see that all four estimators agree with a finite right end point for the excess distribution of rainfall. MLE and MDPDE with $\alpha = 0.1$ are in close agreement. OBRE produces a shape estimate that gives the largest right end point at 3338. Conversely, the Medians shape estimate is such that the estimated right end point is the smallest. In other words, OBRE has estimated the GPD with largest tail while Medians produced the GPD with the shortest tail. Both MLE and MDPDE yield similar GPDs.

Table 5: Estimated parameters.

	MDPDE	MLE	OBRE	Medians
$\hat{\xi}$	0.102	0.107	0.0286	0.156
$\hat{\beta}$	101.8	101.8	95.49	95.59
$\hat{\beta}/\hat{\xi}$	998	951	3338	612

5 Concluding Remarks

Maximum likelihood has been the preferred estimation procedure used for the GPD, and in extreme value analysis in general. Some of the arguments favoring maximum likelihood offered by the advocates are that: (1) These procedures can be generalized without major changes to more complicated data structures in which trends, covariates, or multivariate extremes can be included. (2) MLEs have convenient large-sample properties such as asymptotic efficiency and asymptotic normality. (3) It is relatively straightforward to produce standard errors using the likelihood function. In this article we have seen that the MDPDE has characteristics (2) and (3) plus the additional feature of robustness against gross errors. Furthermore, the MDPDE also has property (1), a general procedure that may be applied to problems involving complex structures in the data.

For example, suppose one wants to fit the GPD(ξ_t, β_t) to the excesses X_{it} of year t, $i = 1, \ldots, n_t$, $t = 1, \ldots, T$. Suppose one wants to include a yearly linear trend in the log-scale of the scale parameter and keep the shape parameter fixed. The MDPDE for this situation would be obtained by setting $\xi(t) = \xi$ and $\beta(t) = \exp(\gamma_0 + \gamma_1 t)$ in equation (2). More specifically, the MDPDE would be obtained by minimizing

$$
\begin{aligned}
H_\alpha(\xi, \gamma_0, \gamma_1) \;=\; & \sum_{t=1}^{T} \frac{1}{n_t} \sum_{i=1}^{n_t} \frac{1}{\{\exp(\gamma_0 + \gamma_1 t)\}^\alpha (1 + \alpha - \alpha\xi)} \\
& - \left(1 + \frac{1}{\alpha}\right) \sum_{t=1}^{T} \frac{1}{n_t} \sum_{i=1}^{n_t} \frac{1}{\{\exp(\gamma_0 + \gamma_1 t)\}^\alpha} \left(1 - \xi \frac{X_{it}}{\exp(\gamma_0 + \gamma_1 t)}\right)^{(\xi^{-1}-1)\alpha},
\end{aligned}
$$

over the parameter space

$$
\left\{ (\xi, \gamma_0, \gamma_1) \in \mathbb{R}^3 : \gamma_0, \gamma_1 \in \mathbb{R}, \max_{1 \le i \le n_t} \{X_{it}\}\xi < \exp(\gamma_0 + \gamma_1 t), -\infty < \xi < (1+\alpha)/\alpha, \xi \neq 0 \right\}.
$$

The relatively straightforward implementation of the MDPDE using S-Plus, the convergence problems of the OBRE with short tail data, and the poor performance of the Medians estimator, recommend the MDPDE for the GPD in practical situations. As the rainfall example suggests, when the MLE is reasonable the MDPDE can agree with it well. Our final conclusion is that MDPD estimation for the GPD offers a feasible and reliable robust alternative to maximum likelihood estimation.

Acknowledgements

The authors are grateful to the associate editor and a referee for helpful comments and suggestions. The authors also thank professor Ian R. Harris for his useful comments.

Appendix I: Integrability Conditions

The components of the score function and the information matrix,

$$S(x;\xi,\beta) = \begin{bmatrix} S_\xi(x;\xi,\beta) \\ S_\beta(x;\xi,\beta) \end{bmatrix} \text{ and } i(x;\xi,\beta) = \begin{bmatrix} i_\xi(x;\xi,\beta) & i_{\xi\beta}(x;\xi,\beta) \\ i_{\xi\beta}(x;\xi,\beta) & i_\beta(x;\xi,\beta) \end{bmatrix}$$

of the GPD, are given by the usual partial derivatives of $\log f(x;\xi,\beta)$, namely

$$\begin{aligned}
S_\xi(x;\xi,\beta) &= \frac{\partial}{\partial\xi}\log f(x;\xi,\beta) \\
&= -\frac{1}{\xi^2}\log\left(1-\xi\frac{x}{\beta}\right) + \frac{1}{\xi}\left(\frac{1}{\xi}-1\right)\left\{1-\left(1-\xi\frac{x}{\beta}\right)^{-1}\right\}, \\
S_\beta(x;\xi,\beta) &= \frac{\partial}{\partial\beta}\log f(x;\xi,\beta) = -\frac{1}{\xi\beta} + \frac{1}{\beta}\left(\frac{1}{\xi}-1\right)\left(1-\xi\frac{x}{\beta}\right)^{-1}, \\
i_\xi(x;\xi,\beta) &= -\frac{\partial^2}{\partial\xi^2}\log f(x;\xi,\beta) \\
&= -\frac{2}{\xi^3}\log\left(1-\xi\frac{x}{\beta}\right) + \frac{3-\xi}{\xi^3} - \frac{2(2-\xi)}{\xi^3}\left(1-\xi\frac{x}{\beta}\right)^{-1} \\
&\quad +\frac{1-\xi}{\xi^3}\left(1-\xi\frac{x}{\beta}\right)^{-2}, \\
i_\beta(x;\xi,\beta) &= -\frac{\partial^2}{\partial\beta^2}\log f(x;\xi,\beta) = -\frac{1}{\beta^2\xi} + \frac{1}{\beta^2}\left(\frac{1}{\xi}-1\right)\left(1-\xi\frac{x}{\beta}\right)^{-2}, \\
i_{\xi\beta}(x;\xi,\beta) &= -\frac{\partial^2}{\partial\xi\partial\beta}\log f(x;\xi,\beta) \\
&= -\frac{1}{\beta\xi^2} + \frac{2-\xi}{\beta\xi^2}\left(1-\xi\frac{x}{\beta}\right)^{-1} - \frac{1-\xi}{\beta\xi^2}\left(1-\xi\frac{x}{\beta}\right)^{-2}.
\end{aligned}$$

Let $[0,x^*]$, $0 < x^* \le \infty$, be the support of the unknown true excess density g. The consistency and asymptotic normality of the MDPDE of the GPD follow from the integrability conditions:

$$\int_0^{x^*} f^\alpha(x;\xi_0,\beta_0)g(x)\mathrm{d}x < \infty, \tag{8}$$

$$\int_0^{x^*} S_\xi^2(x;\xi_0,\beta_0)f^{2\alpha}(x;\xi_0,\beta_0)g(x)\mathrm{d}x < \infty, \tag{9}$$

$$\int_0^{x^*} S_\beta^2(x;\xi_0,\beta_0)f^{2\alpha}(x;\xi_0,\beta_0)g(x)\mathrm{d}x < \infty, \tag{10}$$

$$\int_0^{x^*} |i_\xi(x;\xi_0,\beta_0)f^{2\alpha}(x;\xi_0,\beta_0)|g(x)\mathrm{d}x < \infty, \tag{11}$$

$$\int_0^{x^*} |i_\beta(x;\xi_0,\beta_0)f^{2\alpha}(x;\xi_0,\beta_0)|g(x)\mathrm{d}x < \infty, \tag{12}$$

$$\int_0^{x^*} |i_{\xi\beta}(x;\xi_0,\beta_0)f^{2\alpha}(x;\xi_0,\beta_0)|g(x)\mathrm{d}x < \infty. \tag{13}$$

Appendix II: Variance of the Ratio of two Empirical MSEs

Suppose $\hat{\theta}$ and $\tilde{\theta}$ are estimators of θ. Furthermore, suppose one has N independent random samples and computes $\hat{\theta}_i$ and $\tilde{\theta}_i$ from each sample, $i = 1, \ldots, N$. According to equation (10.17) in Stuart and Ord (1994) page 351, the variance of the ratio of the empirical mean squared errors of $\hat{\theta}$ and $\tilde{\theta}$ is

$$\mathrm{Var}\left[\frac{\hat{\mathrm{MSE}}(\hat{\theta})}{\hat{\mathrm{MSE}}(\tilde{\theta})}\right] \approx \left\{\frac{\mathrm{E}[\hat{\mathrm{MSE}}(\hat{\theta})]}{\mathrm{E}[\hat{\mathrm{MSE}}(\tilde{\theta})]}\right\}^2 \left\{\frac{\mathrm{Var}[\hat{\mathrm{MSE}}(\hat{\theta})]}{\mathrm{E}^2[\hat{\mathrm{MSE}}(\hat{\theta})]} + \frac{\mathrm{Var}[\hat{\mathrm{MSE}}(\tilde{\theta})]}{\mathrm{E}^2[\hat{\mathrm{MSE}}(\tilde{\theta})]} - \frac{2\mathrm{E}[\hat{\mathrm{MSE}}(\hat{\theta})\hat{\mathrm{MSE}}(\tilde{\theta})]}{\mathrm{E}[\hat{\mathrm{MSE}}(\hat{\theta})]\mathrm{E}[\hat{\mathrm{MSE}}(\tilde{\theta})]} + 2\right\}. \quad (14)$$

A straightforward calculation shows that $\mathrm{E}[\hat{\mathrm{MSE}}(\hat{\theta})] = \mathrm{E}\hat{\theta}^2 - 2\theta\mathrm{E}\hat{\theta} + \theta^2$, $\mathrm{Var}[\hat{\mathrm{MSE}}(\hat{\theta})] = \mathrm{Var}[(\hat{\theta} - \theta)^2]/N$, and $\mathrm{E}[\hat{\mathrm{MSE}}(\hat{\theta})\hat{\mathrm{MSE}}(\tilde{\theta})] = \mathrm{E}[(\hat{\theta}-\theta)^2(\tilde{\theta}-\theta)^2]/N + (N-1)\mathrm{E}[(\hat{\theta}-\theta)^2]\mathrm{E}[(\tilde{\theta}-\theta)^2]/N$. Analogous expressions are obtained for $\mathrm{E}[\hat{\mathrm{MSE}}(\tilde{\theta})]$ and $\mathrm{Var}[\hat{\mathrm{MSE}}(\tilde{\theta})]$. So, if θ is known we have the following estimators

$$\begin{aligned}
\hat{\mathrm{E}}[\hat{\mathrm{MSE}}(\hat{\theta})] &= \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i^2 - 2\theta\frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i + \theta^2, \\
\hat{\mathrm{Var}}[\hat{\mathrm{MSE}}(\hat{\theta})] &= \frac{1}{N(N-1)}\sum_{i=1}^{N}\left\{(\hat{\theta}_i - \theta)^2 - \left[\frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2\right]\right\}^2, \\
\hat{\mathrm{E}}[\hat{\mathrm{MSE}}(\hat{\theta})\hat{\mathrm{MSE}}(\tilde{\theta})] &= \frac{1}{N^2}\sum_{i=1}^{N}\left\{(\hat{\theta}_i - \theta)^2(\tilde{\theta}_i - \theta)^2\right\} \\
&\quad + \frac{N-1}{N^3}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2\sum_{i=1}^{N}(\tilde{\theta}_i - \theta)^2.
\end{aligned}$$

Again, analogous expressions for $\hat{\mathrm{E}}[\hat{\mathrm{MSE}}(\tilde{\theta})]$ and $\hat{\mathrm{Var}}[\hat{\mathrm{MSE}}(\tilde{\theta})]$ follow. When these estimated quantities are plugged into (14), we obtain an estimator of the variance of the ratio of the empirical mean squared errors of $\hat{\theta}$ and $\tilde{\theta}$.

Appendix III: Rainfall Data

The following table displays the monthly accumulated rainfall excesses over 300 mm observed in Xalapa, Ver, México, from 1904 to 2003. The month and the year when the excess occurred is also included.

Table 6: Monthly accumulated rainfall excesses over 300 mm.

10/1904/4.2	06/1905/129.3	08/1906/46.3	06/1907/47.1	08/1909/62.8	09/1910/39.7
05/1911/42.9	06/1912/32.2	09/1912/107.9	06/1913/73.3	09/1913/181.9	04/1914/19.9
06/1915/287.9	07/1916/8.3	09/1916/145.5	06/1917/52.7	09/1918/36.8	06/1919/34.1
06/1920/52.7	09/1921/36.8	06/1922/34.1	06/1923/63.0	06/1924/19.8	09/1924/22.2
06/1925/95.6	09/1925/22.7	06/1926/124.5	09/1926/328.8	06/1927/238.7	07/1928/128.5
09/1928/1.2	06/1929/86.8	09/1929/32.0	07/1931/17.8	09/1931/110.3	08/1933/192.3
09/1933/13.3	06/1935/92.6	07/1936/103.8	09/1941/146.7	08/1942/ 30.0	09/1942/188.5
09/1944/298.3	06/1952/392.8	09/1952/17.0	06/1954/100.2	07/1955/296.7	09/1955/87.2
06/1956/22.8	06/1958/34.1	07/1958/63.0	09/1958/28.0	06/1959/338.3	07/1960/66.1
09/1960/38.7	07/1961/112.9	07/1962/3.6	06/1965/100.6	06/1966/48.8	08/1969/196.9
09/1969/57.4	09/1970/99.5	06/1971/53.1	06/1972/74.7	07/1972/186.1	06/1973/155.7
09/1974/262.5	09/1975/160.6	06/1976/82.2	06/1977/17.0	09/1978/56.9	07/1979/29.5
08/1979/41.9	06/1980/25.3	06/1981/105.9	07/1983/32.1	09/1984/36.0	06/1985/60.8
09/1990/19.8	06/1991/95.0	06/1993/189.4	09/1994/1.3	08/1995/146.1	10/1997/82.4
07/1998/33.9	09/1998/119.8	06/1999/43.1	07/1999/45.4	06/2000/122.2	08/2000/209.6
08/2001/18.7	06/2002/32.5	09/2002/68.6			

References

[1] Basu, A., Harris, I.R., Hjort, N.L., and Jones, M.C., Robust and efficient estimation by minimising a density power divergence, *Biometrika*, **85**(3), 549-559, (1998).

[2] Davison, A.C., and Smith, R.L., Models for exceedances over high threshold (with discussion), *J.R. Statistical Society*, B **52**, 393-442, (1990).

[3] Dupuis, D.J., Exceedances over high thresholds: A guide to threshold selection, *Extremes*, **1**(3), 251-261, (1998).

[4] Hosking, J.R.M., and Wallis, J.R., Parameter and quantile estimation for the Generalized Pareto distribution, *Technometrics*, **29**, 339-349, (1987).

[5] Juárez, S.F., *Robust and Efficient Estimation for the Generalized Pareto Distribution*, unpublished Ph.D. dissertation, Statistical Science Department, Southern Methodist University, (2003). Available at `http://www.smu.edu/statistics/faculty/SergioDiss1.pdf`.

[6] Peng, L., and Welsh, A., Robust estimation of the generalized Pareto distribution, *Extremes*, **4**(1), 53-65, (2002).

[7] Smith, R.L., Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, **72**, 67-90, (1985).

[8] Stuart, A., and Ord, K., *Kendall's Advanced Theory of Statistics. Sixth edition of Volume 1 Distribution Theory*, New York: Edward Arnold, (1994).