Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Fitting the Generalized Pareto Distribution to Data

Enrique Castillo & Ali S. Hadi

Fitting the Generalized Pareto Distribution to Data

Enrique CASTILLO and Ali S. HADI

The generalized Pareto distribution (GPD) was introduced by Pickands to model exceedances over a threshold. It has since been used by many authors to model data in several fields. The GPD has a scale parameter ($\sigma > 0$) and a shape parameter ($-\infty < k < \infty$). The estimation of these parameters is not generally an easy problem. When $k > 1$, the maximum likelihood estimates do not exist, and when k is between $\frac{1}{2}$ and 1, they may have problems. Furthermore, for $k \leq -\frac{1}{2}$, second and higher moments do not exist, and hence both the method-of-moments (MOM) and the probability-weighted moments (PWM) estimates do not exist. Another and perhaps more serious problem with the MOM and PWM methods is that they can produce nonsensical estimates (i.e., estimates inconsistent with the observed data). In this article we propose a method for estimating the parameters and quantiles of the GPD. The estimators are well defined for all parameter values. They are also easy to compute. Some asymptotic results are provided. A simulation study is carried out to evaluate the performance of the proposed methods and to compare them with other methods suggested in the literature. The simulation results indicate that although no method is uniformly best for all the parameter values, the proposed method performs well compared to existing methods. The methods are applied to real-life data. Specific recommendations are also given.

KEY WORDS: Elemental percentile method; Generalized extreme value distribution; Maximum likelihood; Method of moments; Order statistics; Probability-weighted moments; Quantile estimation.

1. INTRODUCTION

The distribution of the largest or smallest values of certain natural phenomena (e.g., waves, winds, temperatures, earthquakes, floods) is of interest in many practical applications. For example, the distribution of high waves is important in the design of sea structures, the distribution of large floods is important in designing dams, and so on. This interest has given rise to a rapid development of extreme value theory in recent years (see, e.g., Castillo 1988, Galambos 1978, Leadbetter, Lindgren, and Rootzén 1983, and Resnick 1987). The traditional approach to the analysis of extreme values in a given population is based on the family of generalized extreme value distributions (GEVD) (see Galambos 1981, 1984).

The GEVD is appropriate when the data consist of a set of maxima. However, there has been some criticism of this approach, because using only maxima leads to the loss of information contained in other large-sample values in a given period. This problem is remedied by considering several of the largest order statistics instead of just the largest one; that is, considering all values larger than a given threshold. The differences between these values and a given threshold are called exceedances over the threshold. These exceedances are typically assumed to have a generalized Pareto distribution GPD(k, σ), whose cdf is defined by

$$F(x; k, \sigma) = \begin{cases} 1 - (1 - kx/\sigma)^{1/k}; & k \neq 0, \quad \sigma > 0 \\ 1 - \exp(-x/\sigma); & k = 0, \quad \sigma > 0, \end{cases} \tag{1}$$

where σ and k are scale and shape parameters. The range of x is $x > 0$ for $k \leq 0$ and $0 < x < \sigma/k$ for $k > 0$. The case $k = 0$, which is the exponential distribution, is the limiting distribution as $k \to 0$.

The GPD is a generalization of the Pareto distribution (PD). The PD was studied extensively by Arnold (1983), and the problem of estimation in the PD was considered by Arnold and Press (1989). Here we consider the GPD introduced by Pickands (1975) to model exceedances over a threshold. Since then, it has been used to model data arising in several fields. For example, Hosking and Wallis (1987) used the GPD to model the annual maximum flood of the River Nidd at Hunsingore, England, and Grimshaw (1993) used it to model tensile strength data from a random sample of nylon carpet fibers. Other examples and uses of the GPD in extreme value analysis have been discussed by Castillo (1994). (Other relevant references include Davison 1984, DuMouchel 1983, Joe 1987, and van Monfort and Witter 1985, 1986.)

One particularly interesting and useful property is that the GPD is stable with respect to excesses over threshold operations. In fact we have

$$\begin{aligned} \text{Prob}[X - u \leq x | X > u] &= \frac{\text{Prob}[u < X \leq x + u]}{\text{Prob}[X > u]} \\ &= \frac{F(x + u; k, \sigma) - F(u; k, \sigma)}{1 - F(u; k, \sigma)} \\ &= \frac{(1 - ku/\sigma)^{1/k} - (1 - k(x + u)/\sigma)^{1/k}}{(1 - ku/\sigma)^{1/k}} \\ &= -(1 - kx/(\sigma - ku))^{1/k}, \end{aligned} \tag{2}$$

which means that if x is a GPD(k, σ), then $x - u$, given that $x > u$ for any u is a GPD$(k, \sigma - ku)$. This property implies that if the model is consistent with a set of data for a given threshold, then it must be consistent with the data for all higher thresholds.

The following values of the parameter k are of particular interest:

- When $k = 0$, the GPD reduces to the exponential distribution with mean σ.

Enrique Castillo is Professor of Applied Mathematics, Department of Applied Mathematics and Computing Sciences, University of Cantabria, 39005 Santander, Spain. Ali S. Hadi is Professor of Statistics, Cornell University, Ithaca, NY 14853-3901.

Journal of the American Statistical Association
December 1997, Vol. 92, No. 440, Theory and Methods

- When $k = 1$, the GPD becomes a uniform $U[0, \sigma]$ distribution.
- When $k \le -\frac{1}{2}$, $\text{var}(X) = \infty$. In fact, the rth central moment exists only if $k > -1/r$.
- When $k < 0$, the GPD reduces to the Pareto distribution.

Other properties of the GPD, including examples with graphs for different values of the parameters, have been given by Hosking and Wallis (1987).

Several methods have been proposed for estimating the parameters of the GPD. Pickands (1975, 1984, 1993) suggested some methods for estimating the parameters that are combined with procedures for choosing the threshold value defining the exceedances. These include nonparametric, Bayesian, order statistics, and other approaches.

The method-of-moments (MOM) estimates of the parameters are given by

$$\hat{k}_{\text{MOM}} = (\bar{x}^2/s^2 - 1)/2$$

and

$$\hat{\sigma}_{\text{MOM}} = \bar{x}(\bar{x}^2/s^2 + 1)/2, \tag{3}$$

where $\bar{x}$ and s^2 are the sample mean and the sample variance.

The maximum likelihood (ML) method has been considered by many authors, including Davison (1984), DuMouchel (1983), and Hosking and Wallis (1987). Grimshaw (1993) gave an algorithm for computing the maximum likelihood estimates (MLEs) over a restricted range of the parameter k $(k \le 1)$.

Hosking and Wallis (1987) also considered the MOM and the method of probability-weighted moments (PWM). The PWM estimates of the parameters are

$$\hat{k}_{\text{PWM}} = \bar{x}/(\bar{x} - 2t) - 2$$

and

$$\hat{\sigma}_{\text{PWM}} = 2\bar{x}t/(\bar{x} - 2t), \tag{4}$$

where

$$t = n^{-1}\sum_{i=1}^{n}(1 - p_{i:n})x_{i:n},$$

$p_{i:n} = (i - .35)/n$, and $x_{i:n}$ is the ith order statistic in a sample of size n.

Unfortunately, these estimation methods are not without problems. For $k > 1$, the likelihood function can be made infinite, and hence the MLEs do not exist. Furthermore, Cramer's regularity conditions do not hold for some values of k. Additionally, because for $k \le -\frac{1}{2}$, the $\text{var}[X] = \infty$, both MOM and PWM estimates do not exist. Even when the MOM and PWM estimates exist, a serious problem with the MOM and PWM estimates is that they may not be consistent with the observed sample values; that is, some of the sample values may fall outside the range suggested by the estimated parameter values. This situation occurs whenever $\hat{\sigma}/\hat{k} < x_{n:n}$, where $x_{n:n}$ is the largest order statistic in a sample of size n. As we discuss in Section 5, this event has been observed in numerous simulated datasets.

Hosking and Wallis (1987) compared the ML, MOM, and PWM estimates by a well-designed simulation study, but they considered only the cases where $-\frac{1}{2} < k < \frac{1}{2}$. Their main conclusions are as follows:

1. The ML does not display its asymptotic efficiency even in samples as large as 500.
2. The MOM is not reliable for $k < -.2$.
3. The PWM is good when $-\frac{1}{2} < k < 0$.

Hosking and Wallis (1987) gave the following reasons for restricting the values of k to $-\frac{1}{2} < k < \frac{1}{2}$:

1. This range of values is commonly observed in practical applications (Hosking, Wallis, and Wood 1985).
2. When $k > \frac{1}{2}$, the GPD has finite end points with $f(x; k, \sigma) > 0$, and for $k > 1, f(x; k, \sigma)$ is increasing in the right tail; "such shapes rarely occur in statistical applications."
3. The rth moment of X exists only if $k > -1/r$. Thus, for example, for the variance to exist, k must be larger than $-\frac{1}{2}$.
4. The exponential distribution, which is commonly used in similar situations, is a special case of the GPD $(k = 0)$.

Although this line of argument has some merits, one can still provide practical and theoretical reasons for imposing no restrictions on the parameters. First, there are values of $k > \frac{1}{2}$ that are encountered in practical applications (see, e.g., Walshaw 1990). Also, the uniform distribution, which is encountered in many statistical applications, is a special case of the GPD $(k = 1)$. Other examples of real-life data where the estimates of k fall outside the interval $(-.5, .5)$ are commonly observed in heavy-tailed data and in truncated data (e.g., data obtained using truncation procedures). In Section 4 we give two examples of a real-life data where estimates of k exceed .5.

Second, $F(x; k, \sigma)$ in (1) is a bonafide probability distribution function for all values of $\sigma > 0$ and $-\infty < k < \infty$. It thus is of interest, at least from a theoretical viewpoint, to estimate the distribution for all possible values of its parameters.

When $k \ne 0$, it is sometimes more convenient to work with a reparameterized version of the GPD in (1). In this case the GPD can be written as

$$F(x) = 1 - (1 - x/\delta)^{1/k}; \qquad k \ne 0, \qquad \delta k > 0, \tag{5}$$

where $\delta = \sigma/k$. Thus the range of x is $x > 0$ for $k \le 0$ and $0 < x < \delta$ for $k > 0$. This reparameterization is useful in obtaining simple algorithms for the estimators proposed in this article. Another way to reparameterize the GPD is to let $\theta = k/\sigma = 1/\delta$, as noted by Davison (1984) and used by Grimshaw (1993) to obtain an algorithm that involves only one- instead of two-dimensional numerical searches for the MLE when they exist.

In Section 2 we give an alternative method for fitting the GPD to data (estimating the parameters and quantiles of the GPD). We use the elemental percentile method (EPM) to estimate the parameters and quantiles of the GPD. The EPM is a two-stage procedure (see Castillo and Hadi 1995a). In the first stage, preliminary initial estimates of the parameters are calculated. These initial estimates are combined in the second stage to give overall estimates of the parameters. These final estimates are then substituted in the quantile function to obtain estimates of all desired quantiles. The proposed method works for all possible values of the parameters and is the only viable option when other estimators either do not exist or are inconsistent with the observed evidence. The EPM has been successfully applied to other problems (see Castillo and Hadi 1995b).

The rest of the article is organized as follows. The proposed estimators are derived in Section 2. In Section 3 some asymptotic results for the initial estimates are given and it is shown that the estimates obtained in the second stage are better than the initial estimates. In Section 4 the methods are applied to real-life data. In Section 5 a simulation study is carried out to evaluate the performance of the proposed estimators and to compare them with other estimators suggested in the literature. The simulation study indicates that although no method is uniformly best for all of the parameter values, the proposed method performs well compared to other methods. A summary and specific recommendations are given in Section 6.

2. ESTIMATION

When $k = 0$, the GPD reduces to an exponential distribution with mean σ. The estimation of σ in this case is easy, because the MLE is efficient. Henceforth, we then concentrate on the more difficult case where $k \neq 0$.

The idea of the EPM method is to make full use of the information contained in the order statistics by first equating the cdf evaluated at the observed order statistics to their corresponding percentile values and then using the resulting equations as a basis for obtaining initial estimates of the parameters (Sec. 2.1). These estimates are then combined in a suitable way to obtain final estimates of the parameters (Sec. 2.2).

2.1 Initial Estimates

Let $x_{i:n}$ and $x_{j:n}$ be two distinct order statistics in a random sample of size n from $F(x; k, \sigma)$. Then, equating the cdf evaluated at the observed order statistics to their corresponding percentile values, we obtain

$$F(x_{i:n}; k, \sigma) = p_{i:n}$$

and

$$F(x_{j:n}; k, \sigma) = p_{j:n}, \tag{6}$$

where

$$p_{i:n} = \frac{i-\gamma}{n+\beta} \tag{7}$$

is a suitable plotting position. In the simulation study of Section 5, we tried several values of γ and β and found that $\gamma = 0$ and $\beta = 1$ give better results for the proposed method. Substituting (5) in (6) and taking the log, we obtain

$$\ln(1 - x_{i:n}/\delta) = kC_i$$

and

$$\ln(1 - x_{j:n}/\delta) = kC_j, \tag{8}$$

where $C_i = \ln(1 - p_{i:n}) < 0$. It can be seen that (8) is a system of two equations in two unknowns, δ and k. Eliminating k, we obtain

$$C_i \ln(1 - x_{j:n}/\delta) = C_j \ln(1 - x_{i:n}/\delta), \tag{9}$$

whereas eliminating δ, we get

$$x_{i:n}[1 - (1 - p_{j:n})^k] = x_{j:n}[1 - (1 - p_{i:n})^k]. \tag{10}$$

Each of the Equations (9) and (10) is a function of only one variable; hence they can be solved easily using the bisection method as outlined in Theorem 1 and Algorithm 1. Thus using Algorithm 1, one can solve (9), for example, for δ and obtain an estimator of $\delta, \hat{\delta}(i,j)$, say. This estimator is then substituted in one of the two equations in (8) to obtain a corresponding estimator of $k, \hat{k}(i,j)$, which is given by

$$\hat{k}(i,j) = \ln(1 - x_{i:n}/\hat{\delta}(i,j))/C_i. \tag{11}$$

An estimator of σ can then be computed as

$$\hat{\sigma}(i,j) = \hat{k}(i,j)\hat{\delta}(i,j). \tag{12}$$

Equivalently, one can solve (10) for k to obtain $\hat{k}(i,j)$ and substitute $\hat{k}(i,j)$ in (8) to obtain $\hat{\delta}(i,j)$.

The estimators proposed by Pickands (1975, p. 121), are obtained by setting $i = n/2$ and $j = 3n/4$. Thus they are a special case of the proposed estimators. For these values of i and j, the system of equations in (8) has a closed-form solution, which is obtained as follows. Substituting these values of i and j in (8), we get

$$\ln(1 - x_{n/2:n}/\delta) = k \ln\left(1 - \frac{1}{2}\right)$$

and

$$\ln(1 - x_{3n/4:n}/\delta) = k \ln\left(1 - \frac{3}{4}\right),$$

which leads to

$$\ln(1 - x_{n/2:n}/\delta) - \ln(1 - x_{3n/4:n}/\delta) = k \ln(2).$$

From this equation, we can express k as a function of δ,

$$k = \frac{1}{\ln(2)} \ln\left(\frac{\delta - x_{n/2:n}}{\delta - x_{3n/4:n}}\right). \tag{13}$$

To obtain an expression for δ, write

$$2\ln(1 - x_{n/2:n}/\delta) = \ln(1 - x_{3n/4:n}/\delta),$$

which implies

$$[1 - x_{n/2:n}/\delta]^2 = 1 - x_{3n/4:n}/\delta,$$

from which it follows that

$$\hat{\delta} = \frac{x_{n/2:n}^2}{2x_{n/2:n} - x_{3n/4:n}}, \tag{14}$$

which is the Pickands' estimator of δ. Substituting (14) in (13) gives

$$\hat{k} = \frac{1}{\ln(2)} \ln\left(\frac{x_{n/2:n}}{x_{3n/4:n} - x_{n/2:n}}\right), \tag{15}$$

which are the Pickands estimator of k.

We now show that the foregoing estimators are well defined; that is, (9) has one more solution in addition to the trivial solutions $\delta = \pm\infty$ or, equivalently, (10) has one more solution in addition to the trivial solution $k = 0$.

Theorem 1. Equation (9) has a finite solution, in the interval $(\delta_0, 0)$ if $x_{i:n} < C_i x_{j:n}/C_j$, or in the interval $(x_{j:n}, \delta_0)$ if $x_{i:n} > C_i x_{j:n}/C_j$, where

$$\delta_0 = \frac{x_{i:n} x_{j:n}(C_j - C_i)}{C_j x_{i:n} - C_i x_{j:n}}. \tag{16}$$

The proof of Theorem 1, which is given in the Appendix, is constructive for it gives rise to the following algorithm for solving (9) for δ.

Algorithm 1

1. Select any two distinct order statistics, $x_{i:n} < x_{j:n}$, and compute C_i and C_j. Let $d = C_j x_{i:n} - C_i x_{j:n}$.
2. If $d = 0$, then let $\hat{\delta}(i,j) = \pm\infty, \hat{k}(i,j) = 0$, and go to Step 5; otherwise, go to Step 3.
3. Compute $\delta_0 = x_{i:n} x_{j:n}(C_j - C_i)/d$. If $\delta_0 > 0$, then $\delta_0 > x_{j:n}$ (see Figs. 1–3 and the proof of Theorem 1). Thus use the bisection method on the interval $[x_{j:n}, \delta_0]$ to obtain a solution $\hat{\delta}(i,j)$ of (9) and go to Step 5; otherwise, go to Step 4.
4. Use the bisection method on the interval $[\delta_0, 0]$ to solve (9) and obtain $\hat{\delta}(i,j)$.
5. Use $\hat{\delta}(i,j)$ to compute $\hat{k}(i,j)$ and $\hat{\sigma}(i,j)$ using (11) and (12).

2.2 Final Estimates

The estimates found by Algorithm 1 are based on only two order statistics $\{x_{i:n}, x_{j:n}\}$, and thus they do not utilize the information contained in other order statistics. More statistically efficient estimators are obtained using the following algorithm.

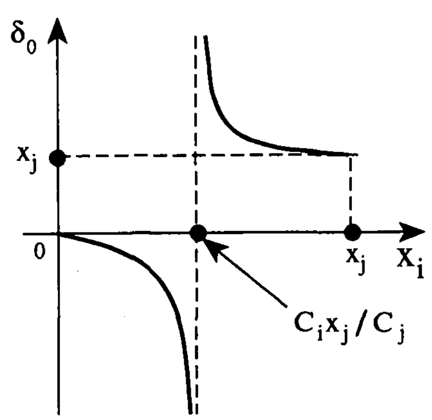

Figure 1. Plot of δ_0 Versus x_i for a Fixed Value of $x_j \geq x_i$.

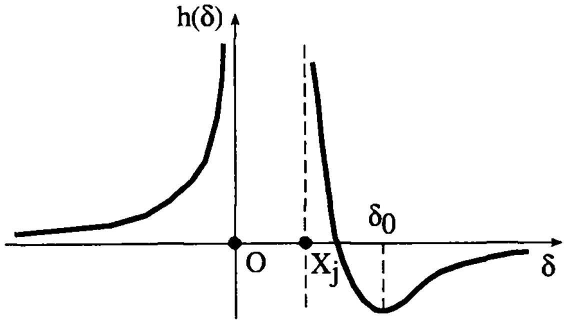

Figure 2. Plot of $h(\delta)$ Versus δ for the Case $x_i > C_i x_j/C_j$.

Algorithm 2

1. Use Algorithm 1 to compute $\hat{k}(i,j)$ and $\hat{\sigma}(i,j)$ for all distinct pairs $x_{i:n} < x_{j:n}$.
2. Use the median of each of the foregoing sets of estimators to obtain a corresponding overall estimators of k, and σ; that is,

$$\hat{k}_{\text{EPM}} = \text{median}(\hat{k}(1,2), \hat{k}(1,3), \ldots, \hat{k}(n-1,n))$$

and

$$\hat{\sigma}_{\text{EPM}} = \text{median}(\hat{\sigma}(1,2), \hat{\sigma}(1,3), \ldots, \hat{\sigma}(n-1,n)), \tag{17}$$

where $\text{median}(y_1, y_2, \ldots, y_n)$ is the median of $\{y_1, y_2, \ldots, y_n\}$.

When n is large, the number of all distinct pairs of order statistics (and hence the number of all possible initial estimators) becomes very large. In this case, step 1 of Algorithm 2 can be modified to reduce the number of calculations as follows. Instead of computing all possible initial estimates, one can compute only a subset of them. We have tried several possibilities for selecting this subset:

1. Random sampling. Select M pairs of distinct order statistics at random with replacement and compute the corresponding initial estimates, where M is a number to be specified by the data analyst. This suggestion is similar to that used for computing the minimum volume ellipsoid estimators in multivariate data (see Rousseeuw and van Zomeron 1990).

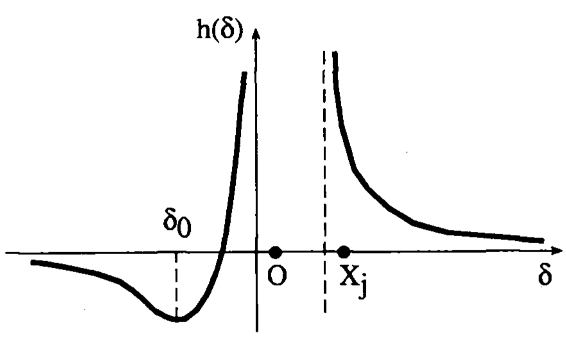

Figure 3. Plot of $h(\delta)$ Versus δ for the Case $x_i < C_i x_j/C_j$.

Table 1. Optimal Choices for p and q for the σ and k Estimators and Associated Variances

	Initial estimators						Pickands' estimator	
	$\hat{\sigma}$			$\hat{k}$			$\hat{\sigma}$	$\hat{k}$
k	p	q	n × var	p	q	n × var	n × var	n × var
−2.0	.29	.96	7.693	.18	.92	13.172	14.780	30.527
−.4	.48	.98	3.789	.32	.98	2.944	6.814	14.622
.4	.60	.98	2.380	.60	.98	1.386	4.874	12.204
1.0	.70	.98	1.837	.70	.98	2.146	3.947	12.488
2.0	.78	.98	1.580	.78	.98	6.350	3.028	16.651
5.0	.80	.98	1.544	.80	.98	38.606	2.206	55.554

2. Systematic sampling. Choose pairs of order statistics at least r steps apart, where r is a number to be specified by the data analyst. (We have tried $r = n/2$.)

3. Setting $j = n$ and selecting the pairs $x_{i:n}$ and $x_{n:n}$, for $i = 1, 2, \ldots, n-1$. The reason for setting $j = n$ is to force all the initial parameter estimates to be consistent with the observed data. A potential drawback of this is that initial estimates are influenced by the largest order statistics, which can be outliers is some datasets.

We experimented with all three possibilities and observed that the obtained estimates are similar. Thus for space-saving purposes, we report only the results of the third alternative in Section 5.

2.3 Confidence Intervals

Because the estimates exist for any combination of parameter values, the use of sampling-based methods such as the bootstrap methods (Diaconis and Efron 1983; Efron 1979) to obtain variances and confidence intervals is justified.

The bootstrap sampling can be performed in two ways: The samples can be drawn directly from the data or they can be drawn parametrically from $F(x; \hat{k}, \hat{\sigma})$. Efron and Tibshirani (1993) discussed and compared parametric and nonparametric bootstrap. For comparison purposes, one should use the first way, because the various methods would be applied to the same samples. On the other hand, when the bootstrap samples are taken from $F(x; \hat{k}, \hat{\sigma})$, one would obtain different samples. In practice, however, one may prefer using the parametric bootstrap to obtain the variance of the estimates of a particular method.

Table 2. Asymptotic Variances of the Estimators for the Optimal Initial Estimators and the Proposed Estimators for n = 200

	Proposed estimators		Optimal initial estimator	
k	$\hat{k}$	$\hat{\sigma}$	$\hat{k}$	$\hat{\sigma}$
−2.0	.0803	.0925	.0659	.0385
−.4	.0097	.0166	.0147	.0189
.4	.0023	.0118	.0069	.0119
1.0	.0065	.0065	.0107	.0092
2.0	.0273	.0068	.0317	.0080
5.0	.1628	.0065	.1930	.0077

3. ASYMPTOTIC RESULTS

In this section we derive some asymptotic results. We start with the initial estimators and then analyze the final estimators.

3.1 Asymptotic Results for the Initial Estimates

Let $i = [np]$ and $j = [nq]$, where $[\cdot]$ denotes the integer part and p and q are fixed constants in the interval (0, 1), such that $p < q$. Then $\sqrt{n}[(X_{(i)}, X_{(j)}) - (F^{-1}(p), F^{-1}(q))]$ is asymptotically normally distributed with (0, 0) expectation and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \frac{p(1-p)}{f(F^{-1}(p))^2} & \frac{p(1-q)}{f(F^{-1}(p))f(F^{-1}(q))} \\ \frac{p(1-q)}{f(F^{-1}(p))f(F^{-1}(q))} & \frac{q(1-q)}{f(F^{-1}(q))^2} \end{pmatrix}, \quad (18)$$

where f and F are the pdf and the cdf of the GPD. It is a straightforward to show that

$$\boldsymbol{\Sigma} = \begin{pmatrix} \frac{\sigma^2 p(1-p)^{2k-1}}{n} & \frac{\sigma^2 p(1-p)^{k-1}(1-q)^k}{n} \\ \frac{\sigma^2 p(1-p)^{k-1}(1-q)^k}{n} & \frac{\sigma^2 q(1-q)^{2k-1}}{n} \end{pmatrix}. \quad (19)$$

Using the delta method, the asymptotic distribution (as $n \to \infty$) of the initial estimators $(\hat{k}(i,j), \hat{\sigma}(i,j))$ is shown to be normal with mean (σ, k) and covariance matrix $\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T$, where

$$\mathbf{C} = \frac{1}{\gamma_{pq}(k)} \times \begin{pmatrix} \frac{-[x(q) + \log(1-q)\sigma(1-q)^k]}{(1-q)^k - 1} & \frac{[x(p) + \log(1-p)\sigma(1-p)^k]}{1-(1-p)^k} \end{pmatrix}, \quad (20)$$

$$\gamma_{pq}(k) = x(q)\log(1-p)(1-p)^k - x(p)\log(1-q)(1-q)^k, \quad (21)$$

and

$$x(p) = \sigma[1 - (1-p)^k]/k \quad (22)$$

is the pth quantile, $0 \le p \le 1$. From (19)–(21), we see that the estimates are asymptotically consistent.

The optimal (in the sense of minimum asymptotic variance) asymptotic choice for p and q depends strongly on the value of the parameter k. Table 1 shows the optimal choices of p and q for the σ and k estimators. We note that the optimal values do not coincide for the σ and k estimators for negative values of k. On the other hand, without

Table 3. *The Bilbao Waves Data: Zero-Crossing Hourly Mean Periods (in Seconds) of the Sea Waves Measured in a Bilbao Buoy in January 1997*

7.05	7.12	7.15	7.18	7.19	7.20	7.20	7.20	7.20	7.25
7.26	7.27	7.28	7.30	7.31	7.31	7.32	7.33	7.37	7.40
7.46	7.46	7.47	7.48	7.48	7.52	7.54	7.55	7.55	7.58
7.59	7.59	7.61	7.63	7.65	7.66	7.66	7.67	7.67	7.68
7.69	7.72	7.72	7.72	7.72	7.72	7.77	7.77	7.79	7.79
7.82	7.83	7.83	7.83	7.84	7.85	7.85	7.88	7.88	7.90
7.90	7.91	7.93	7.93	7.93	7.94	7.95	7.95	7.97	7.97
7.97	7.99	8.00	8.03	8.03	8.05	8.06	8.06	8.07	8.10
8.11	8.12	8.15	8.15	8.15	8.18	8.18	8.18	8.19	8.20
8.21	8.23	8.23	8.30	8.30	8.31	8.31	8.32	8.32	8.33
8.40	8.41	8.42	8.43	8.43	8.45	8.48	8.49	8.50	8.50
8.51	8.52	8.53	8.54	8.56	8.58	8.59	8.59	8.60	8.65
8.69	8.71	8.72	8.74	8.74	8.74	8.74	8.79	8.81	8.84
8.85	8.86	8.88	8.88	8.94	8.98	8.98	8.99	9.01	9.03
9.06	9.12	9.16	9.17	9.17	9.18	9.18	9.18	9.21	9.22
9.23	9.24	9.27	9.29	9.30	9.32	9.33	9.36	9.38	9.43
9.46	9.47	9.59	9.59	9.60	9.61	9.62	9.63	9.66	9.74
9.75	9.78	9.79	9.79	9.80	9.84	9.85	9.89	9.90	

NOTE: Only periods above 7 seconds are listed.

the knowledge of σ and k, the optimum value cannot be selected. Thus a single initial estimator cannot be optimum for all cases. In addition, the efficiency of the initial estimators is sensitive to the selection of the p and q values. For example, for $k = -2.0$ the asymptotic variance of the optimal initial estimate of k, which is $13.172/n$ for $p = .18$ and $q = .92$, jumps to $46.12/n$ for $p = .8$ and $q = .999$, which is almost four times the optimum.

Table 1 also shows the asymptotic variances of Pickands's estimates, which are a special case of the asymptotic variances of the initial estimators. Note that these variances are much higher than those corresponding to the optimal initial estimators, as would be expected.

3.2 Asymptotic Results for the Final Estimates

The foregoing results allow one to compare the initial and final estimators given in Section 2. Table 2 shows a comparison of the asymptotic variances of the estimators for the optimal asymptotic initial estimators and the proposed method for $n = 200$. Because the theoretical asymptotic variances for the proposed estimator are very difficult to obtain, they have been obtained by simulation (1,000 replications). With the only exception the case $k = -2.0$, the proposed method gives smaller variances. This shows that the proposed method is better than the optimum initial estimates for that range of k values. In the case of large negative values of k, the optimal value p is very sensitive to the parameter k; thus optimum initial estimates cannot be used.

4. TWO EXAMPLES

As a first example, we fit the GPD to the Bilbao waves data. The data are the zero-crossing hourly mean periods (in seconds) of the sea waves measured in a Bilbao buoy in January 1997 as reported by the Maritime Climate Program of CEDEX, Spain. One purpose of the data is to study the influence of periods on beach morphodynamics and other problems related to the right tail. Thus only data above 7 seconds are shown in Table 3. We use the GPD model

Table 4. *The Bilbao Waves Data: Estimated Parameters for Three Estimation Methods*

		$\hat{k}$			$\hat{\sigma}$		
u	*m*	*MOM*	*PWM*	*EPM*	*MOM*	*PWM*	*EPM*
7.0	179	1.052	1.075	.815	2.75	2.78	2.40
		(.158)	(.149)	(.073)	(.304)	(.289)	(.202)
7.5	154	.606	.606	.682	1.62	1.62	1.69
		(.114)	(.121)	(.079)	(.186)	(.190)	(.174)
8.0	106	.647	.635	.743	1.38	1.37	1.46
		(.158)	(.168)	(.111)	(.209)	(.215)	(.194)
8.5	69	.723	.707	.814	1.13	1.12	1.18
		(.229)	(.240)	(.155)	(.228)	(.232)	(.198)
9.0	41	.834	.834	.912	.81	.81	.85
		(.299)	(.288)	(.231)	(.209)	(.203)	(.190)
9.5	17	1.709	1.584	1.271	.63	.60	.52
		(.835)	(.572)	(.420)	(.272)	(.203)	(.152)

NOTE: *m* is the number of exceedances over the threshold value *u*; standard errors in parentheses.

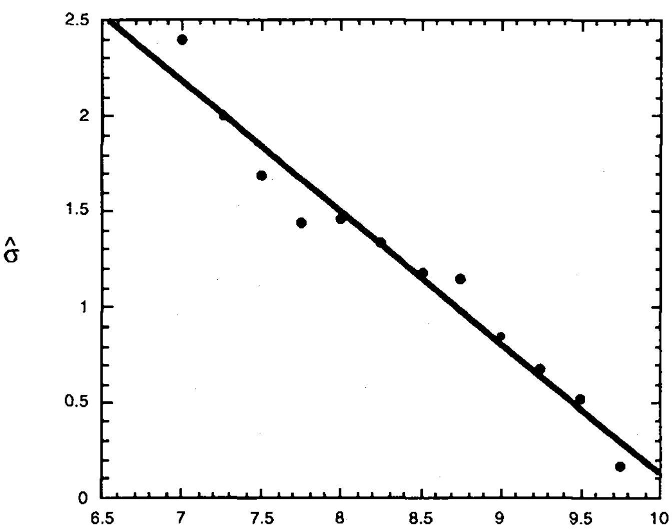

Figure 4. *The Bilbao Waves Data: Scatterplot of $\hat{\sigma}$ as a Function of Threshold Value u. The Corresponding Least Squares Regression Line has a Slope of −.687 and an Intercept of 6.994.*

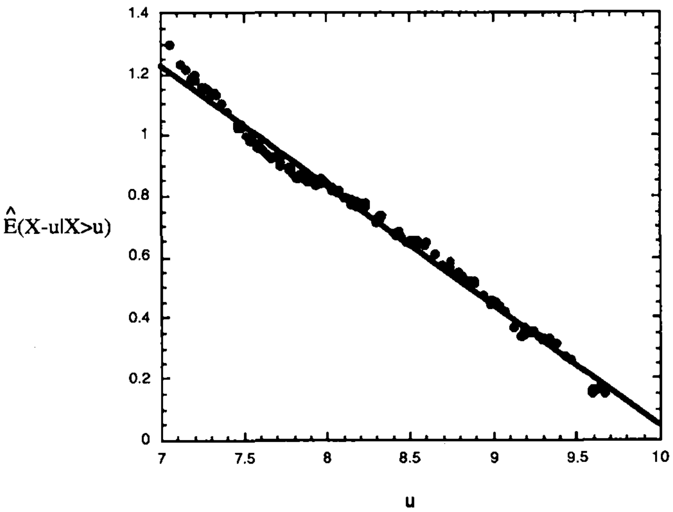

Figure 5. The Bilbao Waves Data: Scatterplot of $\hat{E}(X - u|X > u)$ as a Function of Threshold Value u. The Corresponding Least Squares Regression Line has a Slope of −.395 and an Intercept of 3.999.

Table 5. The Bilbao Waves Data: Average Scaled Absolute Errors (ASAE) for Three Estimation Methods

u	*m*	*MOM*	*PWM*	*EPM*
7.0	179	.0378	.0394	.0314
7.5	154	.0200	.0206	.0230
8.0	106	.0242	.0250	.0267
8.5	69	.0314	.0324	.0320
9.0	41	.0458	.0469	.0454
9.5	17	.0909	.0932	.0842

NOTE: *m* is the number of exceedances over the threshold value *u*.

to fit these data and to predict exceedances over known thresholds. To study the sensitivity of the estimates to the specification of the threshold, we repeat the calculations for several thresholds and monitor the effect of changing the threshold on the obtained results. The estimated parameters and their standard errors (in parentheses) for three estimation methods are given in Table 4 for several threshold values u. The standard errors are computed based on 1,000 bootstrap samples. It can be seen from Table 4 that all estimation methods lead to values of k outside the range $(-.5, .5)$. For example, for the threshold value $u = 7$, the EPM estimate of k is .815 with a standard error of .073. The standard errors in Table 4 are small compared to the estimated parameter values, indicating that k is significantly different from 0. In fact, some of the results indicate that the k parameter is significantly larger than .5.

It can also be seen from Table 4 that the standard errors of the EPM estimators of both κ and σ are smaller than the corresponding standard errors of the MOM and PWM methods for *all* values of u.

To further assess the quality of the GPD model and the EPM estimates, we note the following:

1. The estimates of σ are plotted versus the threshold values u in Figure 4. Note that the theoretical value of σ as a function of u is (see expression 2):

$$\sigma(u) = \sigma_0 - ku, \tag{23}$$

where σ_0 is the value of σ associated with $u = 0$. Thus, as expected, the points in Figure 4 show a linear trend ($R^2 = .96$) with slope $-.687$ and intercept 6.994, which, from (23), leads to global estimates $\hat{k} = .687$ and $\hat{\sigma}_0 = 6.994$.

2. Provided that $k > -1, u > 0$, and $\sigma - ku > 0$, we have (Davison and Smith 1990; Hall and Wellner 1981; Yang 1978):

$$E(X - u|X > u) = \frac{\sigma - ku}{1 + k}. \tag{24}$$

Accordingly, if the GPD is appropriate, then the scatterplot of the mean observed excess over u versus u should resemble a straight line with a slope of $-k/(1+k)$ and an intercept of $\sigma/(1 + k)$. This scatterplot is shown in Figure 5. The graph shows a strong linear relationship ($R^2 = .99$) with with slope $-.395$ and intercept 3.999. Therefore, the GPD assumption seems reasonable. The associated estimates are $\hat{k} = .654$ and $\hat{\sigma}_0 = 6.614$, which are similar to the ones obtained earlier.

To judge the overall goodness of fit, we use the average scaled absolute error (ASAE),

$$\text{ASAE } = n^{-1} \sum_{i=1}^{n} |x_{i:n} - \hat{x}_{i:n}|/(x_{n:n} - x_{1:n}), \tag{25}$$

where

$$\hat{x}_{i:n} = \hat{\sigma}[1 - (1 - p_{i:n})^{\hat{k}}]/\hat{k}. \tag{26}$$

Note that we divided by a scale factor $(x_{n:n} - x_{1:n})$ to account for the fact that the range of the random variable changes with the parameters σ and k. The ASAEs for the three methods are given in Table 5. The MOM has smaller ASAEs than the PWM. The ASAEs produced by the MOM and EPM methods are smaller than those produced by PWM. The MOM and EPM methods have comparable ASAEs.

As a second example, we fit the GPD to the fatigue data for the Kevlar/Epoxy strand lifetime (in hours) at a stress level equal to 70% (Barlow, Toland, and Freeman 1984). The dataset, as reported by Andrews and Herzberg (1985, p. 184) is shown in Table 6.

In this example we are interested in the left tail. However, the GPD is applicable to lower extremes by taking

Table 6. Fatigue Data: Lifetime Data in Hours for the Kevlar/Epoxy Strand at 70% Stress Level

1,051	1,337	1,389	1,921	1,942	2,322	3,629	4,006	4,012	4,063
4,921	5,445	5,620	5,817	5,905	5,956	6,068	6,121	6,473	7,501
7,886	8,108	8,546	8,666	8,831	9,106	9,711	9,806	10,205	10,396
10,861	11,026	11,214	11,362	11,604	11,608	11,745	11,762	11,895	12,044
13,520	13,670	14,110	14,496	15,395	16,179	17,092	17,568	17,568	

Table 7. Fatigue Data: Estimated Parameters for Three Estimation Methods

		$\hat{k}$			$\hat{\sigma}$		
u	*m*	*MOM*	*PWM*	*EPM*	*MOM*	*PWM*	*EPM*
−18,000	49	1.538	1.489	1.118	23,337	22,887	19,187
		(.510)	(.413)	(.181)	(6,091)	(5,106)	(2,926)
−16,000	45	1.493	1.441	1.061	19,776	19,359	16,141
		(.418)	(.342)	(.189)	(4,456)	(3,801)	(2,698)
−14,000	42	1.0184	1.0134	.8772	12,924	12,892	11,795
		(.331)	(.301)	(.201)	(3,080)	(2,858)	(2,389)
−12,000	39	.472	.4218	.6541	7,103	6,860	7,866
		(.250)	(.282)	(.229)	(1,976)	(2,097)	(2,083)
−10,000	28	.833	.821	.853	8,108	8,052	8,088
		(.432)	(.417)	(.281)	(2,780)	(2,674)	(2,246)
−8,000	21	.927	.933	.921	6,845	6,866	6,798
		(.472)	(.432)	(.354)	(2,499)	(2,295)	(2,205)

NOTE: *m* is the number of exceedances over the threshold value *u*; standard errors in parentheses.

$Y = -X$, which transforms the lower tail to the upper tail, and then fitting the GPD to $Y - u|Y > u$. With this transformation, we use the GPD model to fit these data and to predict exceedances over known thresholds. As in the foregoing example, we study the sensitivity of the estimates to the specification of the threshold. The estimated parameters and their standard errors (in parentheses) for the three estimation methods are given in Table 7 for several threshold values u. The standard errors are computed based on 1,000 bootstrap samples. It can be seen from Table 7 that the EPM estimates of k are outside the range $(-.5, .5)$. For example, for the threshold value $u = -18{,}000$, the EPM estimate of k is 1.118 with a standard error of .181. The standard errors in Table 7 are small compared to the estimated parameter values, indicating that k is significantly different from 0. In fact, foregoing results indicate that the k parameter is significantly larger than .5.

As in the previous example, the estimates of σ are plotted versus the threshold values u in Figure 6. As expected, the points in Figure 6 show a linear trend ($R^2 = .937$) with slope -1.248 and intercept $-4{,}461$, which, from (23), leads to global estimates $\hat{k} = 1.248$ and $\hat{\sigma}_0 = 4{,}461$.

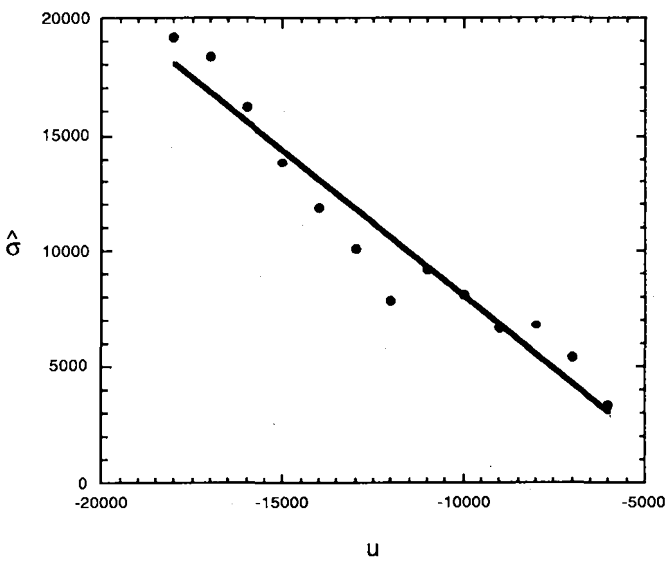

Figure 6. Fatigue Data: Scatterplot of $\hat{\sigma}$ as a Function of Threshold Value u. The corresponding least squares regression line has a slope of −1.248 and an intercept of −4,461.

Similarly, the scatterplot of the mean observed excesses over u versus u is shown in Figure 7. The graph shows a strong linear relationship ($R^2 = .973$) with slope $-.550$ and intercept 1,066. Therefore, the GPD assumption seems reasonable. The associated estimates are $\hat{k} = 1.223$ and $\hat{\sigma}_0 = 2{,}370$.

Finally, the ASAEs for the three methods are given in Table 8. The MOM has smaller ASAEs than the PWM. The ASAEs produced by the EPM are smaller than those produced by the MOM and PWM.

5. SIMULATIONS

In this section we carry out a simulation study to evaluate the performance of the proposed method and to compare it with other methods suggested in the literature. Our simulation parallels that of Hosking and Wallis (1987), but we consider values of k both inside and outside the range $-\frac{1}{2} < k < \frac{1}{2}$.

The results are invariant with respect to the scale parameter σ; we thus set $\sigma = 1$, following Hosking and Wallis (1987). We consider the following values of

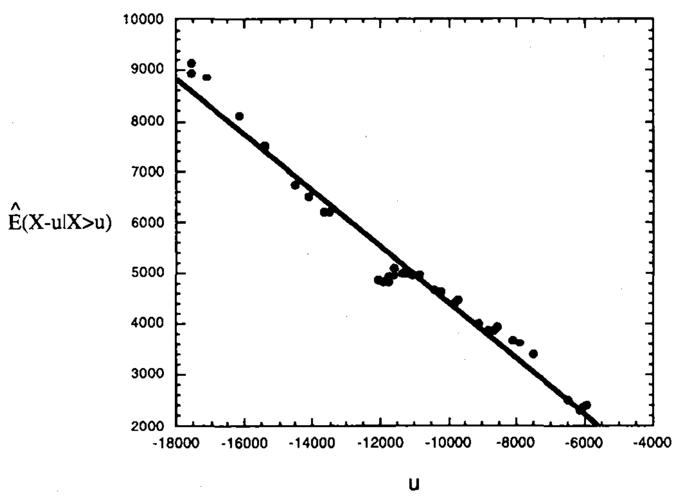

Figure 7. Fatigue Data: Scatterplot of $\hat{E}(X - u|X>u)$ as a Function of Threshold Value u. The corresponding least squares regression line has a slope of −.550 and an intercept of 1,066.

Table 8. Fatigue Data: Average Scaled Absolute Errors (ASAE) for Estimation Methods

u	m	MOM	PWM	EPM
−18,000	49	.0514	.0519	.0475
−16,000	45	.0576	.0583	.0490
−14,000	42	.0446	.0459	.0417
−12,000	39	.0492	.0534	.0504
−10,000	28	.0526	.0542	.0514
−8,000	21	.0712	.0734	.0664

NOTE: *m* is the number of exceedances over the threshold value *u*.

k: $\{-2, -.4, -.2, 0, .2, .4, 1, 2\}$. The values of $k \leq -.5$ represent cases where the random variable X has an infinite variance, whereas $k > 1$ represent cases where the MLE do not exist. We consider sample sizes $n = \{15, 50, 100\}$. The results are based on 1,000 simulation runs.

For each combination of k and n, the parameters k and σ are estimated using the following methods:

1. The MOM estimates given by (3).
2. The PWM estimates given by (4).
3. The proposed estimators $\hat{k}_{\text{EPM}}$ given by (17).

Because it is efficient for large samples, the MLE should be used when the sample size is large, provided that its algorithm does not run into convergence problems. The simulation results of Hosking and Wallis (1987) indicate that the ML method does not display its asymptotic efficiency even in samples as larger as 500. Because the scope of this study is for samples not larger than 100, and for these sample sizes the ML has been shown (see Hosking and Wallis 1987) to give worse performance than the MOM and PWM methods, the ML is not considered here.

We should also note that we have also estimated the quantiles $x(p)$ for $p = \{.01, .02, .98, .99\}$. The results are not included for space-saving purposes, but they lead to the same conclusions as those based on the parameter estimates.

As we pointed out in Section 1, the MOM and PWM can give parameter estimates inconsistent with the observed data. The frequency of this event in 1,000 simulated datasets is reported in Table 9. For $\hat{k} < 0$, the condition $\hat{\sigma}/\hat{k} < x_{n:n}$ is always satisfied. This is the reason why the frequencies in Table 9 are smaller for $k < 0$ than for $k > 0$. For $k > 0$, the relative frequency of inconsistent results is disturbingly large. Also, Table 9 indicates that when $k < 0$, the number of positive estimates of k decreases with the sample size as it should. On the other hand, Table 9 seems to indicate that for $k > 0$, the number of inconsistent results increases with the sample size.

Table 9. Number of Datasets (in 1,000 Simulated Datasets) Where the MOM and PWM Parameter Estimates are Inconsistent With the Observed Data

	n = 15		n = 50		n = 100	
k	MOM	PWM	MOM	PWM	MOM	PWM
−2.0	0	0	0	0	0	0
−1.0	1	1	1	1	0	0
−.4	7	14	0	1	0	0
−.2	10	28	0	3	0	0
0	27	52	6	29	0	15
.2	69	104	49	98	26	83
.4	141	167	144	199	144	197
1.0	330	291	400	382	425	419
2.0	425	297	450	381	489	446

Next, we compare the estimation methods, first using the bias and the root mean squared error (RMSE) of the parameter estimates, then using the ASAE as an overall goodness-of-fit measure.

5.1 Properties of the Parameter Estimators

The bias and the RMSE for the k and σ estimators are given in Tables 10 and 11. The results can be summarized as follows:

1. For all methods, the bias and RMSE substantially increase as k deviates from 0 for $n = 15$. Elsewhere, the trend is that as k increases from -2, the bias decreases.

2. For all methods, the bias and RMSE decrease as the sample size increases, an indication that all of the estimators are consistent.

Table 10. Bias of k and σ Estimators

			k								
Estimate	n	Method	−2.00	−1.00	−.40	−.20	0	.20	.40	1.00	2.00
$\hat{k}$	15	MOM	−1.60	−.72	−.31	−0.21	−.14	−.10	−.10	−.16	−.43
		PWM	−1.22	−.47	−.19	−0.14	−.11	−.08	−.08	−.07	−.07
		EPM	.52	.25	.16	0.12	.09	.06	.03	−.03	−.06
	50	MOM	−1.53	−.60	−.17	−0.08	−.05	−.03	−.02	−.05	−.07
		PWM	−1.08	−.30	−.08	−0.04	−.04	−.02	−.02	−.03	0
		EPM	.39	.19	.09	0.08	.05	.03	.03	−.01	−.01
	100	MOM	−1.52	−.56	−.13	−0.05	−.02	−.01	−.01	−.02	−.04
		PWM	−1.05	−.25	−.04	−0.02	−.02	−.01	−.01	−.01	0
		EPM	.28	.14	.08	0.07	.04	.03	.01	0	0
$\hat{\sigma}$	15	MOM	−19494.03	−2.86	−0.43	−0.25	−.15	−.10	−0.10	−0.13	−.19
		PWM	−1793.72	−0.68	−0.20	−0.15	−.12	−.08	−0.09	−0.08	−.06
		EPM	−0.12	−0.03	−0.04	−0.04	−.04	−.03	−0.04	−0.05	−.03
	50	MOM	−13769.68	−5.44	−0.24	−0.09	−.05	−.03	−0.02	−0.04	−.03
		PWM	−393.43	−0.45	−0.06	−0.03	−.04	−.03	−0.02	−0.03	−.01
		EPM	0.05	0.03	0.02	0.02	.00	−.01	0.01	−0.01	−.01
	100	MOM	−19104.02	−3.89	−0.20	−0.06	−.02	−.01	−0.01	−0.01	−.02
		PWM	−276.65	−0.30	−0.04	−0.02	−.02	0	−0.01	−0.01	−.01
		EPM	0.06	0.03	0.01	0.02	.01	.01	0	0	0

Table 11. RMSE of k and σ Estimators

Estimate	n	Method	k = −2.00	−1.00	−.40	−.20	0	.20	.40	1.00	2.00
$\hat{k}$	15	MOM	1.60	.73	.40	.33	.32	.33	.40	.83	1.89
		PWM	1.23	.53	.38	.35	.35	.36	.42	.65	1.07
		EPM	1.39	.76	.55	.47	.42	.38	.37	.44	.68
	50	MOM	1.53	.60	.22	.16	.15	.15	.18	.35	.72
		PWM	1.08	.34	.19	.16	.17	.18	.20	.33	.58
		EPM	.90	.53	.29	.24	.19	.16	.15	.19	.35
	100	MOM	1.52	.56	.16	.11	.10	.10	.12	.22	.45
		PWM	1.05	.27	.14	.12	.12	.12	.14	.21	.38
		EPM	.73	.40	.23	.18	.14	.11	.09	.12	.23
$\hat{\sigma}$	15	MOM	272,410.20	6.16	.70	.51	.44	.43	.46	.61	.82
		PWM	25,003.08	1.16	.55	.50	.46	.44	.46	.49	.50
		EPM	1.11	.61	.51	.48	.44	.42	.41	.37	.33
	50	MOM	192,525.66	22.97	.36	.23	.21	.21	.21	.26	.31
		PWM	5,370.76	.85	.26	.23	.23	.22	.22	.24	.26
		EPM	.48	.33	.26	.23	.22	.21	.19	.18	.17
	100	MOM	198,966.18	10.46	.27	.16	.15	.14	.14	.16	.19
		PWM	2,846.29	.41	.18	.16	.16	.15	.15	.16	.17
		EPM	.36	.25	.18	.16	.15	.14	.13	.12	.11

3. For $k = -2$ and $k = -1$, the bias and RMSE of the MOM and PWM estimators are extremely large. This is due to the fact that the moments of the random variable do not exist in these two cases.

4. Confirming the conclusion of Hosking and Wallis (1987), the performance of the methods depends on the parameter k. Generally, the PWM performs the best for $-.4 \leq k \leq 0$, the MOM performs the best for $0 < k \leq .4$, and the EPM performs the best for $k < -.4$ and $k > .4$. Although, the EPM is not the best for the range $-.4 \leq k \leq .4$, it competes well with the best method in this range.

5.2 Goodness of Fit

Because we shall include cases with infinite variances, in this section we use the ASAE as an overall goodness-of-fit measure to compare the methods. The ASAEs and their standard errors are given in Table 12. The following can be inferred from the results:

1. For all sample sizes, the EPM gives the smallest ASAE for small $(k < -.4)$ and large $(k > .4)$ values of k. Thus the EPM performs better than the other two methods for extreme values of k.

Table 12. Simulated Data: Average of Scaled Absolute Errors (ASAE) and Their Standard Errors (SE) for Various Values of n and k

		MOM		PWM		EPM	
n	k	ASAE	SE (ASAE)	ASAE	SE (ASAE)	ASAE	SE (ASAE)
15	−2.00	.0864	.0170	.0585	.0167	.0233	.0229
15	−1.00	.0735	.0182	.0484	.0201	.0314	.0205
15	−.40	.0602	.0162	.0456	.0212	.0375	.0184
15	−.20	.0561	.0145	.0471	.0202	.0396	.0174
15	0	.0537	.0143	.0501	.0197	.0432	.0172
15	.20	.0518	.0136	.0518	.0178	.0445	.0163
15	.40	.0527	.0166	.0556	.0204	.0481	.0183
15	1.00	.0574	.0226	.0607	.0227	.0530	.0201
15	2.00	.0671	.0321	.0642	.0264	.0561	.0239
50	−2.00	.0369	.0091	.0242	.0069	.0086	.0090
50	−1.00	.0346	.0091	.0210	.0099	.0145	.0099
50	−.40	.0263	.0077	.0208	.0100	.0198	.0091
50	−.20	.0245	.0068	.0223	.0092	.0221	.0089
50	0	.0234	.0064	.0233	.0086	.0234	.0088
50	.20	.0237	.0063	.0249	.0081	.0248	.0087
50	.40	.0248	.0071	.0264	.0085	.0260	.0090
50	1.00	.0294	.0103	.0301	.0104	.0289	.0093
50	2.00	.0341	.0137	.0317	.0115	.0301	.0108
100	−2.00	.0222	.0071	.0143	.0053	.0053	.0062
100	−1.00	.0223	.0069	.0127	.0070	.0093	.0069
100	−.40	.0161	.0053	.0126	.0062	.0131	.0061
100	−.20	.0146	.0042	.0134	.0054	.0146	.0058
100	0	.0147	.0039	.0147	.0050	.0161	.0060
100	.20	.0156	.0043	.0164	.0053	.0174	.0061
100	.40	.0171	.0048	.0180	.0056	.0187	.0062
100	1.00	.0204	.0066	.0207	.0066	.0204	.0064
100	2.00	.0236	.0093	.0219	.0078	.0207	.0069

2. For small sample sizes, the EPM gives the smallest ASAE for all values of k. Thus the EPM performs better than the other two methods for small sample sizes.

3. For $n = 100$, the ASAE depends on k. The PWM is best for $-.4 \le k \le 0$, the MOM is best for $.2 \le k \le .4$, and the EPM is best for other values of k.

4. For large values of k, the standard errors for the EPM are smaller than the corresponding standard errors for the other two methods.

6. SUMMARY AND CONCLUDING REMARKS

One of the main problems in dealing with the estimation of multiparameter families is the need to optimize complicated functions or solve nonlinear equations. Thus it is important to demonstrate that the associated functions or systems of equations have only one optimal solution. In addition, an efficient method for finding the optimal solution must be given.

In this article we have proposed a method for estimating the parameters of the GPD. Because the estimates are obtained by solving only one equation in one unknown, and the existence of a unique solution for any possible combination of parameter values is proved, they are also easy to compute using numerical methods. Unlike the existing estimation methods, the proposed estimators are defined for all possible values of the parameters. The PWM and MOM have closed-form expressions for large-sample confidence intervals, but confidence intervals for the proposed final estimators must be computed by bootstrap methods.

We applied the methodology to real-life data, and implemented a simulation study has been implemented to evaluate the performances of the proposed estimators and to compare them with the existing estimators. The results seem to indicate that the proposed method works well over a wide range of parameter values. Although no method is uniformly best, the results indicate that the proposed method performs well compared to existing methods.

Based on our simulation results and on those obtained by Hosking and Wallis (1987), we make the following specific recommendations as to which estimation method to use in practice:

1. If the sample size is large (e.g., $n > 500$) and it is believed that $-\frac{1}{2} < k < \frac{1}{2}$, then the MLE is to be preferred because of its efficiency in this case. However, as Hosking and Wallis (1987) observed, algorithms for computing the MLE can have convergence problems even when $n > 500$.

2. If the sample size is not large and there is a reason to believe that $-.5 \le k \le 0$, then use the PWM.

3. In all other cases, use the proposed EPM.

4. In all cases, if the MLE has convergence problems or if the PWM gives nonsensical estimates, then use the EPM.

APPENDIX: PROOF OF THEOREM 1

Assume without loss of generality that $i < j$, which implies $C_i > C_j$, and consider the following function of δ:

$$h(\delta) = C_i \ln(1 - x_{j:n}/\delta) - C_j \ln(1 - x_{i:n}/\delta), \qquad \text{(A.1)}$$

which is defined in the set $\{(-\infty, 0) \cup (x_{j:n}, \infty)\}$. Then (9) can be written as

$$h(\delta) = 0; \qquad \text{(A.2)}$$

that is, the solutions of (9) are the 0s of (A.1). Clearly, two 0s of $h(\delta)$ are $\delta = \pm\infty$. We now show that there exists a finite solution to (A.2).

The function $h(\delta)$ has the following properties:

$$h(-\infty) = 0; \quad h(-0) = \infty; \quad h(x_j) = \infty; \quad h(\infty) = 0. \qquad \text{(A.3)}$$

Additionally, it has no relative maximum and only one relative minimum, which is given by

$$\frac{dh(\delta)}{d\delta} = \frac{1}{\delta}\left[\frac{C_i x_{j:n}}{\delta - x_{j:n}} - \frac{C_j x_{i:n}}{\delta - x_{i:n}}\right] = 0. \qquad \text{(A.4)}$$

The solutions of (A.4) are $\delta_0 = \pm\infty$ and

$$\delta_0 = \frac{x_{i:n} x_{j:n}(C_j - C_i)}{C_j x_{i:n} - C_i x_{j:n}}. \qquad \text{(A.5)}$$

Thus we have (see Fig. 1)

$$\begin{array}{lll} \delta_0 = x_{j:n} & \text{if} & x_{i:n} = x_{j:n}, \\ \delta_0 > 0 & \text{if} & x_{i:n} > C_i x_{j:n}/C_j, \\ \delta_0 \to \pm\infty & \text{if} & x_{i:n} \to C_i x_{j:n}/C_j, \end{array}$$

and

$$\delta_0 < 0 \quad \text{if} \quad x_{i:n} < C_i x_{j:n}/C_j. \qquad \text{(A.6)}$$

The continuity of $h(\delta)$ together with (A.3) and the existence of a relative minimum imply that $h(\delta)$ has only one finite 0 if $C_j x_{i:n} \ne C_i x_{j:n}$. This 0 is in the interval $(x_{j:n}, \delta_0)$ if $x_{i:n} > C_i x_{j:n}/C_j$ (see Fig. 2) or in the interval $(\delta_0, 0)$ if $x_{i:n} < C_i x_{j:n}/C_j$ (see Fig. 3). Thus we can use the bisection method to determine the solution of (9), or the 0 of $h(\delta)$. This completes the proof.

[Received August 1995. Revised April 1997.]

REFERENCES

Andrews, D. F., and Herzberg, A. M. (1985), *Data. A Collection of Problems From Many Fields for the Student and Research Worker*, New York: Springer-Verlag.

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

Arnold, B., and Press, S. J. (1989), "Bayesian Estimation and Prediction for Pareto Data," *Journal of the American Statistical Association*, 84, 1079–1084.

Barlow, R. E., Toland, R. H., and Freeman, T. (1984), "A Bayesian Analysis of Stress-Rupture Life of Kevlar/Epoxy Spherical Pressure Vessels," in *Proceedings of the Canadian Conference in Applied Statistics*, ed. T. D. Dwivedi, New York: Marcel Dekker.

Castillo, E. (1988), *Extreme Value Theory in Engineering*, New York: Academic Press.

——— (1994), "Extremes in Engineering Applications," in *Proceedings of the Conference on Extreme Value Theory and its Applications*, eds. J. Galambos et al., Gaithersburg, MD: Kluwer, pp. 15–42.

Castillo, E., and Hadi, A. S. (1995a), "A Method for Estimating Parameters and Quantiles of Distributions of Continuous Random Variables," *Computational Statistics and Data Analysis*, 20, 421–439.

——— (1995b), "Modeling Lifetime Data With Applications to Fatigue Models," *Journal of the American Statistical Association*, 90, 1041–1054.

Davison, A. C. (1984), "Modelling Excesses Over High Thresholds, With an Application," in *Statistical Extremes and Applications*, ed. J. Tiago de Oliveira, NATO ASI Series, Dordrecht: Reidel, pp. 461–482.

Davison, A. C., and Smith, R. L. (1990), "Models for Exceedances Over High Thresholds," *Journal of the Royal Statistical Society*, Ser. B, 52, 393–442.

Diaconis, P., and Efron, B. (1983), "Computer Intensive Methods in Statistics," *Scientific American*, 248, 116–130.

DuMouchel, W. (1983), "Estimating the Stable Index σ in Order to Measure Tail Thickness," *The Annals of Statistics*, 11, 1019–1036.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7, 1–26.

Efron, B., and Tibshirani, R. (1993), *An Introduction to the Bootstrap*, New York: Chapman and Hall.

Galambos, J. (1978), *The Asymptotic Theory of Extreme Order Statistics*, Malabar, FL: Krieger.

——— (1981), "Extreme Value Theory in Applied Probability," *Mathematical Scientist*, 6, 13–26.

——— (1984), "Introduction, Order Statistics, Exceedances. Law of Large Numbers," in *Statistical Extremes and Applications*, NATO ASI Series, Dordrecht: Reidel.

Grimshaw, S. D. (1993), "Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution," *Technometrics*, 35, 185–191.

Hall, W. J., and Wellner, J. (1981), "Mean Residual Life," in *Statistics and Related Topics*, eds. M. Csörgo, D. A. Dawson, J. N. K. Rao, and A. K. Md. E. Saleh, Amsterdam: North-Holland, pp. 169–184.

Hosking, J. R. M., and Wallis, J. R. (1987), "Parameter and Quantile Estimation for the Generalized Pareto Distribution," *Technometrics*, 29, No. 3, 339–349.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985), "Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments," *Technometrics*, 27, 251–261.

Joe, H. (1987), "Estimation of Quantiles of the Maximum of N Observations," *Biometrika*, 74, 347–354.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Series*, New York: Springer.

Pickands, J. (1975), "Statistical Inference Using Extreme Order Statistics," *The Annals of Statistics*, 3, 119–131.

——— (1984), "Spline and Isotonic Estimation of the Pareto Function," in *Statistical Extremes and Applications*, ed. J. Tiago de Oliveira, NATO ASI Series, Dordrecht: Reidel, pp. 285–296.

——— (1993), "Bayes Quantile Estimation and Threshold Selection for the Generalized Pareto Family," in *Proceedings of the Conference on Extreme Value Theory and its Applications*, eds. J. Galambos et al., Gaithersburg, MD: Kluwer, pp. 123–138.

Resnick, S. (1987), *Extreme Values, Point Processes and Regular Variation*, New York: Springer-Verlag.

Rousseeuw, P. J., and van Zomeren, B. C. (1990), "Unmasking Multivariate Outliers and Leverage Points" (with discussion), *Journal of the American Statistical Association*, 85, 633–651.

van Montfort, M. A. J., and Witter, J. V. (1985), "Testing Exponentially Against Generalized Pareto Distributions," *Journal of Hydrology*, 78, 305–315.

——— (1986), "The Generalized Pareto Distribution Applied to Rainfall Depths," *Hydrology Science Journal*, 31, 151–162.

Walshaw, D. (1990), Discussion of "Models for Exceedances Over High Thresholds," by A. C. Davison and R. L. Smith, *Journal of the Royal Statistical Society*, Ser. B, 52, 393–442.

Yang, G. L. (1978), "Estimation of a Biometric Function," *The Annals of Statistics*, 6, 112–116.