Parameter and Quantile Estimation for the Generalized Pareto Distribution
Author(s): J. R. M. Hosking and J. R. Wallis
Reviewed work(s):
Source: *Technometrics*, Vol. 29, No. 3 (Aug., 1987), pp. 339-349
Published by: American Statistical Association and American Society for Quality
Stable URL: http://www.jstor.org/stable/1269343
Accessed: 29/02/2012 08:56

TECHNOMETRICS, AUGUST 1987, VOL. 29, NO. 3

Parameter and Quantile Estimation for the Generalized Pareto Distribution

J. R. M. Hosking

T. J. Watson Research Center
IBM Corporation
Yorktown Heights, NY 10598

Institute of Hydrology
Wallingford, Oxon OX10 8BB
England

J. R. Wallis

T. J. Watson Research Center
IBM Corporation
Yorktown Heights, NY 10598

The generalized Pareto distribution is a two-parameter distribution that contains uniform, exponential, and Pareto distributions as special cases. It has applications in a number of fields, including reliability studies and the analysis of environmental extreme events. Maximum likelihood estimation of the generalized Pareto distribution has previously been considered in the literature, but we show, using computer simulation, that, unless the sample size is 500 or more, estimators derived by the method of moments or the method of probability-weighted moments are more reliable. We also use computer simulation to assess the accuracy of confidence intervals for the parameters and quantiles of the generalized Pareto distribution.

KEY WORDS: Maximum likelihood; Method of moments; Probability-weighted moments.

1. THE GENERALIZED PARETO DISTRIBUTION

The generalized Pareto distribution is the distribution of a random variable X defined by $X = \alpha(1 - e^{-kY})/k$, where Y is a random variable with the standard exponential distribution. The generalized Pareto distribution has distribution function

$$F(x) = 1 - (1 - kx/\alpha)^{1/k}, \qquad k \neq 0$$
$$= 1 - \exp(-x/\alpha), \qquad k = 0, \qquad (1)$$

and density function

$$f(x) = \alpha^{-1}(1 - kx/\alpha)^{1/k-1}, \qquad k \neq 0$$
$$= \alpha^{-1} \exp(-x/\alpha), \qquad k = 0; \qquad (2)$$

the range of x is $0 \le x < \infty$ for $k \le 0$ and $0 \le x \le \alpha/k$ for $k > 0$. The parameters of the distribution are α, the scale parameter, and k, the shape parameter. The special cases $k = 0$ and $k = 1$ yield, respectively, the exponential distribution with mean α and the uniform distribution on $[0, \alpha]$; Pareto distributions are obtained when $k < 0$. The shapes of generalized Pareto distributions for different values of k are illustrated in Figure 1.

The generalized Pareto distribution was introduced by Pickands (1975), and interest in it was shown by Davison (1984), Smith (1984, 1985), and van Montfort and Witter (1985). Its applications include use in the analysis of extreme events, in the modeling of large insurance claims, as a failure-time distribution in reliability studies, and in any situation in which the exponential distribution might be used but in which some robustness is required against heavier tailed or lighter tailed alternatives.

Some elementary but important properties of the generalized Pareto distribution are as follows:

1. The failure rate $r(x) = f(x)/\{1 - F(x)\}$ is given by $r(x) = 1/(\alpha - kx)$ and is monotonic in x, decreasing if $k < 0$, constant if $k = 0$, and increasing if $k > 0$.
2. If the random variable X has a generalized Pareto distribution, then the conditional distribution of $X - t$ given $X \ge t$ is also generalized Pareto, with the same value of k.
3. Let $Z = \max(0, X_1, \ldots, X_N)$, where the X_i are independent and identically distributed as (1) and N has a Poisson distribution. Then Z has, essentially, a generalized extreme value (GEV) distribution as defined by Jenkinson (1955); that is, there exist quantities β, γ, and δ, independent of z, such that

$$F_Z(z) = \Pr(Z \le z)$$
$$= \exp[-\{1 - \delta(z - \gamma)/\beta\}^{1/\delta}], \qquad z \ge 0;$$

furthermore, $\delta = k$; that is, the shape parameters of

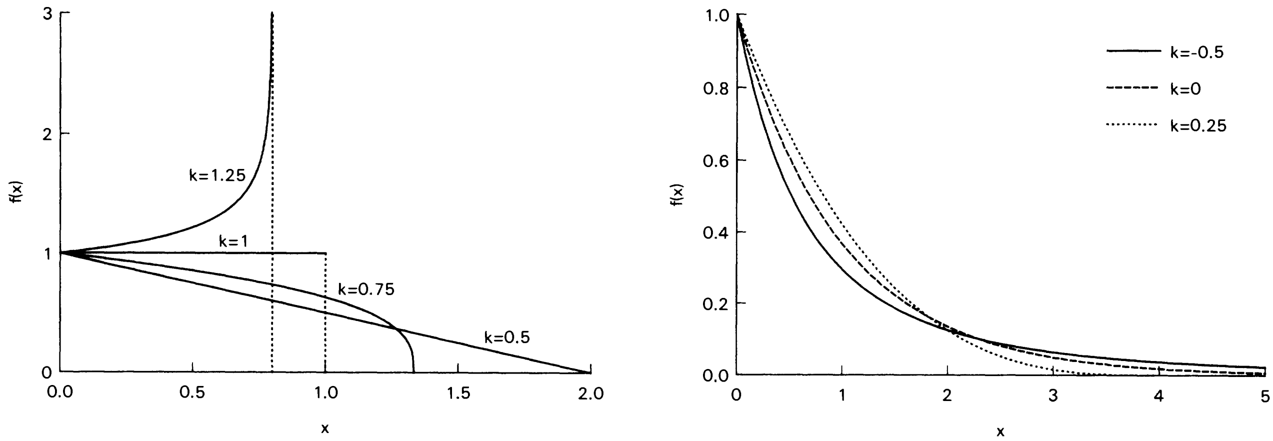

Figure 1. Probability Density Function of the Generalized Pareto Distribution for Different Values of the Shape Parameter k.

the GEV and the generalized Pareto distributions are equal.

In property 3, Z has only "essentially" a GEV distribution, because Z cannot take negative values; we have $\Pr(Z < 0) = 0$ and $\Pr(Z = 0) = e^{-\lambda}$, and only for $z \geq 0$ can the cumulative distribution function $F_Z(z)$ be matched by that of a GEV distribution. Properties 2 and 3 are relevant to the analysis of extreme events. The GEV distribution is widely used in the United Kingdom to describe the annual maximum floods at river gauging stations, and each annual maximum flood may be regarded as the maximum of a number of floods arising from separate storm events. If it is reasonable to assume that successive floods arrive according to a Poisson process and have independent magnitudes, then properties 2 and 3 make the generalized Pareto distribution the logical choice for modeling those flood magnitudes that exceed any fixed threshold.

2. SUMMARY

In this article we consider the problems of estimating the parameters and quantiles of the generalized Pareto distributions. We restrict attention to the case $-\frac{1}{2} < k < \frac{1}{2}$, for both practical and theoretical reasons. Property 3 of Section 1 implies a close connection between generalized Pareto and GEV distributions with equal values for their shape parameters, and, as Hosking, Wallis, and Wood (1985) remarked, applications of the GEV distributions, particularly in hydrology, usually involve the case $-\frac{1}{2} < k < \frac{1}{2}$. When the generalized Pareto distribution is used as an alternative to the exponential distribution, it is likely that values of k near 0 will be of the greatest interest, because the exponential distribution is a generalized Pareto distribution with $k = 0$. Generalized Pareto distributions with $k > \frac{1}{2}$ have finite endpoints with $f(x) > 0$ at each endpoint (see Fig. 1), and such shapes rarely occur in statistical applications. Generalized Pareto distributions with $k \geq \frac{1}{2}$ have infinite variance, and this too is unusual in statistical applications.

Maximum likelihood estimation of generalized Pareto parameters was discussed by Davison (1984) and Smith (1984, 1985). In Section 3, we derive estimators of parameters and quantiles by the method of moments and the method of probability-weighted moments (PWM's) and discuss the three methods and some of their large-sample properties. In Section 4, we compare the performances of the estimators for samples of size 15–500, using computer simulation. Our main conclusions are that maximum likelihood estimation, although asymptotically the most efficient method, does not clearly display its efficiency even in samples as large as 500, that the method of moments is generally reliable except when $k < -.2$, and that PWM estimation may be recommended if it seems likely that $k < 0$, particularly if it is important that estimated extreme quantiles should have low bias or that asymptotic theory should give a good approximation to the standard errors of the estimates.

In Section 5, we apply our results to the hydrological problem of estimation of extreme floods, using as an example a series of flood peaks for the River Nidd at Hunsingore, England.

3. ESTIMATION METHODS FOR THE GENERALIZED PARETO DISTRIBUTION

3.1 Method of Maximum Likelihood

The log-likelihood function for a sample $\mathbf{x} = \{x_1, \ldots, x_n\}$ is

$$\log L(\mathbf{x}; \alpha, k) = -n \log \alpha - (1 - k) \sum_{i=1}^{n} y_i,$$

$$y_i = -k^{-1} \log(1 - kx_i/\alpha).$$

The log-likelihood may be made arbitrarily large by taking $k > 1$ and α/k arbitrarily close to $\max(x_i)$, so the maximum likelihood estimators (MLE's) are taken to be the values $\hat{\alpha}$ and $\hat{k}$, which yield a local maximum of $\log L$ [an alternative approach might be to maximize a grouped-data likelihood function, as done by Giesbrecht and Kempthorne (1976)]. To find the local maximum of log L requires numerical methods; we used a procedure based on Newton–Raphson iteration, with the same structure as Hosking's (1985) algorithm for the GEV distribution.

Smith (1984) obtained the information matrix and gave the asymptotic variance of the MLE's:

$$n \operatorname{var}\begin{bmatrix}\hat{\alpha}\\ \hat{k}\end{bmatrix} \sim \begin{bmatrix}2\alpha^2(1-k) & \alpha(1-k)\\ \alpha(1-k) & (1-k)^2\end{bmatrix}, \qquad k < \frac{1}{2}. \quad (3)$$

When $k < \frac{1}{2}$, the estimators have their familiar properties of consistency, asymptotic normality, and asymptotic efficiency. Smith (1984) also discussed the nonregular case $k > \frac{1}{2}$, but this does not concern us here.

3.2 Method of Moments

Moments of the generalized Pareto distribution are obtained by noting that $E(1 - kX/\alpha)^r = 1/(1 + rk)$ if $1 + rk > 0$. The rth moment of X exists if $k > -1/r$. Provided that they exist, the mean, variance, skewness, and kurtosis are, respectively,

$$\mu = \alpha/(1 + k),$$

$$\sigma^2 = \alpha^2/(1 + k)^2(1 + 2k),$$

$$\gamma = 2(1 - k)(1 + 2k)^{1/2}/(1 + 3k),$$

and

$$\kappa = \frac{3(1 + 2k)(3 - k + 2k^2)}{(1 + 3k)(1 + 4k)} - 3.$$

The moment estimators of α and k are, therefore,

$$\hat{\alpha} = \tfrac{1}{2}\bar{x}(\bar{x}^2/s^2 + 1), \qquad \hat{k} = \tfrac{1}{2}(\bar{x}^2/s^2 - 1),$$

where $\bar{x}$ and s^2 are the sample mean and variance, respectively. Provided that $k > -\frac{1}{4}$, we can show by standard methods (e.g., Rao 1973, sec. 6h) that $\hat{\alpha}$ and $\hat{k}$ are asymptotically normally distributed with

$$n \operatorname{var}\begin{bmatrix}\hat{\alpha}\\ \hat{k}\end{bmatrix} \sim \frac{(1 + k)^2}{(1 + 2k)(1 + 3k)(1 + 4k)} \times \begin{bmatrix}2\alpha^2(1+6k+12k^2) & \alpha(1+2k)(1+4k+12k^2)\\ \alpha(1+2k)(1+4k+12k^2) & (1+2k)^2(1+k+6k^2)\end{bmatrix}. \quad (4)$$

When $k \le -\frac{1}{4}$, the variance of s^2 is infinite and the variances of $\hat{\alpha}$ and $\hat{k}$ are not of asymptotic order n^{-1}. When $k = 0$, (3) and (4) are identical, so the moment estimators are asymptotically 100% efficient.

3.3 Method of Probability-Weighted Moments

The PWM's of a continuous random variable X with distribution function F are the quantities

$$M_{p,r,s} = E[X^p\{F(X)\}^r\{(1 - F(X))\}^s]$$

for real p, r, and s (Greenwood, Landwehr, Matalas, and Wallis 1979). Greenwood et al. (1979) exhibited several distributions for which the relationship between the parameters of the distribution and the PWM's $M_{1,r,s}$ is simpler than the relationship between the parameters and the conventional moments $M_{p,0,0}$. Hosking et al. (1985) showed that efficient estimators of parameters and quantiles of the GEV distribution can be obtained using PWM's. Hosking (1986) gave a general exposition of the theory of PWM's. For the generalized Pareto distribution, it is convenient to work with the quantities

$$\alpha_s = M_{1,0,s} = E[X\{1 - F(X)\}^s] = \frac{\alpha}{(s+1)(s+1+k)},$$

which exist provided that $k > -1$ and in terms of which the parameters are given by

$$\alpha = \frac{2\alpha_0\alpha_1}{\alpha_0 - 2\alpha_1}, \qquad k = \frac{\alpha_0}{\alpha_0 - 2\alpha_1} - 2. \quad (5)$$

The PWM estimators $\hat{\alpha}$ and $\hat{k}$ are obtained by replacing α_0 and α_1 in (5) by estimators based on an observed sample of size n. Two possibilities are

$$a_r = n^{-1}\sum_{j=1}^{n} \frac{(n-j)(n-j-1)\cdots(n-j-r+1)}{(n-1)(n-2)\cdots(n-r)} x_{j:n}$$

and

$$\tilde{\alpha}_r = n^{-1}\sum_{j=1}^{n}(1 - p_{j:n})^r x_{j:n},$$

where $x_{1:n} \le \cdots \le x_{n:n}$ is the ordered sample and $p_{j:n} = (j + \gamma)/(n + \delta)$, where γ and δ are suitable constants. The estimator a_r is unbiased (Landwehr, Matalas, and Wallis 1979a), but $\tilde{\alpha}_r$ is merely consistent. The use of $\tilde{\alpha}_r$ with $\gamma = -.35$ and $\delta = 0$ was recommended by Landwehr et al. (1979b) for the Wakeby distribution, of which the generalized Pareto distribution is a special case, and by Hosking et al. (1985) for the GEV distribution. Whichever variant is used, the estimators of α_r, α, and k are asymptotically equivalent. The methods of Hosking et al. (1985) may be used to show that, provided $k > -\frac{1}{2}$, the $\{a_r\}$ are asymptotically normally distributed with variance given by

$$n \operatorname{cov}(a_r, a_s) \sim \frac{\alpha^2}{(r+1+k)(s+1+k)(r+s+1+2k)}$$

and that the PWM estimators $\hat{\alpha}$ and $\hat{k}$ are asymptotically normally distributed with

$$n \operatorname{var}\begin{bmatrix}\hat{\alpha}\\ \hat{k}\end{bmatrix} \sim \frac{1}{(1 + 2k)(3 + 2k)} \times \begin{bmatrix}\alpha^2(7+18k+11k^2+2k^3) & \alpha(2+k)(2+6k+7k^2+2k^3)\\ \alpha(2+k)(2+6k+7k^2+2k^3) & (1+k)(2+k)^2(1+k+2k^2)\end{bmatrix}. \quad (6)$$

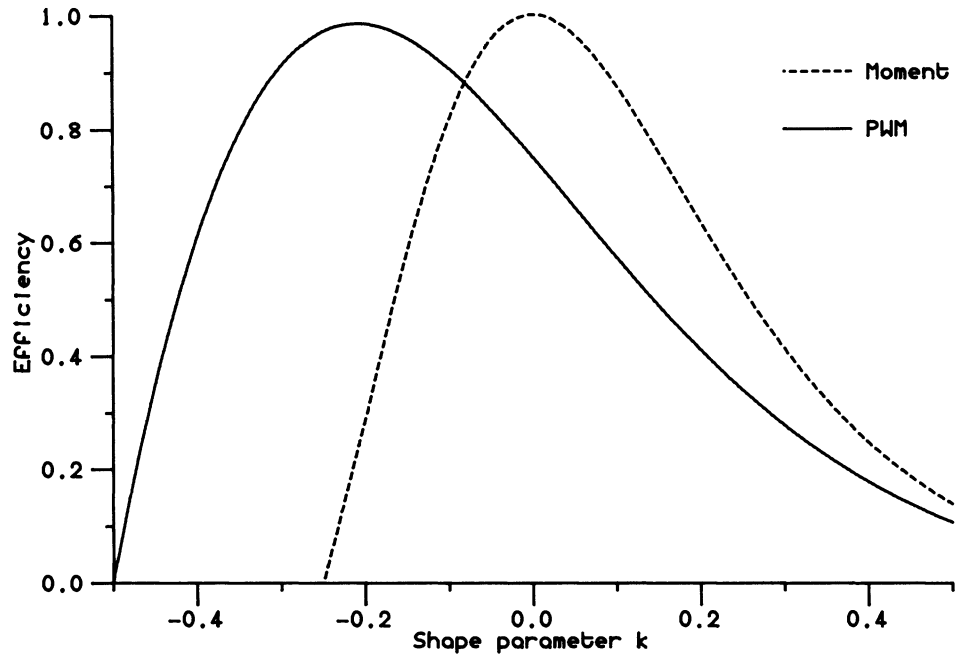

Figure 2. Asymptotic Efficiency, Relative to the Maximum Likelihood Estimator, of Moment and PWM Estimators of the Shape Parameter of the Generalized Pareto Distribution.

3.4 Estimation of Quantiles

Quantiles of the generalized Pareto distribution are given in terms of the parameters by

$$x(F) = \alpha\{1 - (1 - F)^k\}/k, \qquad k \neq 0$$
$$= -\alpha \log(1 - F), \qquad k = 0. \qquad (7)$$

A quantile estimator $\hat{x}(F)$ is defined by substituting estimators $\hat{\alpha}$ and $\hat{k}$ for the parameters in (7). The variance of $\hat{x}(F)$ is given asymptotically by

$$\text{var } \hat{x}(F) \sim \{s(k)\}^2 \text{var } \hat{\alpha} + 2\alpha s(k)s'(k)\text{cov}(\hat{\alpha}, \hat{k}) + \alpha^2\{s'(k)\}^2 \text{var } \hat{k}, \qquad (8)$$

where

$$s(k) = \{1 - (1 - F)^k\}/k,$$
$$s'(k) = \{-s(k) + (1 - F)^k \log(1 - F)\}/k.$$

We have $x(F) = \alpha F\{1 + \frac{1}{2}(1 - k)F + O(F^2)\}$ as $F \to 0$, so the accuracy of $\hat{x}(F)$ for small F is effectively determined by the accuracy of $\hat{\alpha}$.

3.5 Confidence Intervals

Approximate confidence intervals for parameters and quantiles can be obtained from the asymptotic distribution theory of Sections 3.1–3.4. If $\hat{\theta} = (\hat{\alpha}\ \hat{k})^T$ is any of the previously described estimators of $\theta = (\alpha\ k)^T$ and if $h(\theta)$ is a continuous function of the parameters whose estimator $h(\hat{\theta})$ is asymptotically normally distributed with n var $h(\hat{\theta}) \sim v(\theta)$ as $n \to \infty$, then a confidence interval for $h(\theta)$ with asymptotic confidence level t is

$$h(\hat{\theta}) + \{n^{-1}v(\hat{\theta})\}^{1/2} z_{(1-t)/2} \leq h(\theta) \leq h(\hat{\theta}) + \{n^{-1}v(\hat{\theta})\}^{1/2} z_{(1+t)/2}, \qquad (9)$$

where z_t is the tth quantile of the standard normal distribution.

If $\hat{\theta}$ is the MLE of θ, then n var $\hat{\theta}$ is given by (3), which is the inverse of the expected information matrix for θ. An alternative is to use the observed information matrix: in this case n var $\hat{\theta}$ is estimated by

$$\left[-n^{-1} \left. \frac{\partial^2 \log L(\mathbf{x}; \theta)}{\partial\theta\,\partial\theta^T} \right|_{\theta = \hat{\theta}} \right]^{-1},$$

and confidence intervals can be constructed analogously to (9). The use of observed information for construction of confidence intervals was proposed by Efron and Hinkley (1978) and was recommended for the GEV distribution in preference to the use of expected information by Prescott and Walden (1983).

3.6 Comparison of Estimators

We can calculate the first-order asymptotic biases and variances of estimators of generalized Pareto parameters and quantiles for each of the candidate methods of estimation, as Hosking et al. (1985) did for the GEV distribution. The results are of less interest for the generalized Pareto distribution, because it appears from our simulations that very large sample sizes are needed for asymptotic theory to give a useful approximation in finite samples for all of the estimation methods. Figure 2, however, in which the asymptotic efficiencies are plotted, relative to the

Table 1. Failure Rate of Maximum Likelihood Estimation for the Generalized Pareto Distribution

	k				
n	−.4	−.2	.0	.2	.4
15	3.6	4.8	12.2	22.7	41.7
25	.2	.3	1.5	4.7	14.6
50	.0	.0	.0	.0	.8
100	.0	.0	.0	.0	.0

NOTE: Tabulated values are the number of failures to converge of Newton–Raphson iteration per 100 simulated samples.

MLE, of the moment and PWM estimators of k, shows a pattern the implications of which are largely fulfilled in our simulations: that PWM estimators perform well when $k < 0$, and particularly well when $k \approx -.2$, but that moment estimators have high efficiency when k is near 0, and they outperform PWM estimators when $k \geq 0$.

4. FINITE-SAMPLE PROPERTIES OF THE ESTIMATORS

A computer simulation experiment was run to compare different estimation methods for the generalized Pareto distribution. Simulations were performed for sample sizes n = 15, 25, 50, 100, 200, 500 with the shape parameter taking the values $k = -.4$, $-.3$, …, .4. The scale parameter α was set to 1 throughout. All of the estimation methods considered are equivariant under scale changes of the data, so setting $\alpha = 1$ involves no loss of generality. For each combination of values of n and k, 50,000 random samples were generated from the generalized Pareto distribution, and for each sample the parameters α and k and the quantiles $x(F)$, F = .001, .01, .1, .2, .5, .8, .9, .98, .99, .998, .999, were estimated by each of the methods described in Section 3.

As was mentioned in Section 3.1, maximum likelihood estimates sometimes cannot be obtained for the generalized Pareto distribution. The frequency with which our estimation algorithm failed to converge is given in Table 1. We investigated in detail 100 of the simulated samples for which maximum likelihood estimation failed. The Newton–Raphson iteration was restarted from a variety of starting values of α and k, some based on the sample moments or PWM's, some chosen at random. For 91 of the 100 samples the iteration failed to converge for any choice of starting values. We conclude that the vast majority of failures of the algorithms are caused by the nonexistence of a local maximum of the likelihood function rather than by failure of our algorithm to find a local maximum when one exists. Failure of the algorithm to converge occurred exclusively for samples for which the other estimation methods gave large positive estimates of k, often with $\hat{k} > .5$; therefore, to ignore such failures would artificially increase the bias and reduce the variability of maximum likelihood quantile estimators. For this reason, (k, n) combinations that gave

Table 2. Bias of Estimators of Generalized Pareto Parameters

		k					k				
n	Method	−.4	−.2	.0	.2	.4	−.4	−.2	.0	.2	.4
				Bias ($\hat{\alpha}$)					Bias ($\hat{k}$)		
15	ML	.22	.20	.16*	.11*	.02*	.16	.15	.14*	.11*	.05*
	MOM	.41	.23	.14	.10	.09	.30	.20	.14	.10	.09
	PWM	.18	.13	.10	.08	.08	.18	.13	.10	.08	.07
25	ML	.13	.13	.13	.12	.09*	.10	.11	.12	.12	.10*
	MOM	.33	.16	.08	.06	.05	.23	.14	.09	.06	.05
	PWM	.11	.08	.06	.05	.04	.12	.08	.06	.05	.04
50	ML	.06	.06	.06	.07	.08	.05	.05	.06	.07	.08
	MOM	.25	.10	.04	.03	.02	.17	.09	.04	.03	.02
	PWM	.06	.04	.03	.02	.02	.07	.04	.03	.02	.02
100	ML	.02	.03	.03	.03	.04	.02	.02	.03	.04	.04
	MOM	.19	.06	.02	.01	.01	.13	.05	.02	.01	.01
	PWM	.03	.02	.01	.01	.01	.04	.02	.01	.01	.01
200	ML	.01	.01	.02	.02	.02	.01	.01	.01	.02	.02
	MOM	.15	.04	.01	.01	.01	.10	.03	.01	.01	.01
	PWM	.02	.01	.01	.01	.01	.02	.01	.01	.01	.00
500	ML	.01	.01	.01	.01	.01	.00	.00	.01	.01	.01
	MOM	.11	.02	.01	.00	.00	.07	.02	.00	.00	.00
	PWM	.01	.00	.00	.00	.00	.01	.00	.00	.00	.00

NOTE: ML is maximum likelihood, MOM is method of moments, and PWM is probability-weighted moments.
*Values are unreliable for reasons discussed in the text.

Table 3. RMSE of Estimators of Generalized Pareto Parameters

		k					k				
n	*Method*	−.4	−.2	.0	.2	.4	−.4	−.2	.0	.2	.4
				RMSE ($\hat{\alpha}$)					RMSE ($\hat{k}$)		
15	ML	.62	.56	.49*	.40*	.31*	.46	.41	.36*	.30*	24*
	MOM	.72	.49	.44	.43	.45	.38	.33	.32	.34	.40
	PWM	.52	.48	.46	.45	.45	.36	.35	.35	.37	.41
25	ML	.43	.41	.38	.34	.29*	.34	.32	.29	.26	.23*
	MOM	.53	.35	.31	.31	.32	.29	.24	.22	.24	.28
	PWM	.37	.34	.33	.33	.33	.27	.25	.25	.27	.30
50	ML	.27	.25	.23	.23	.22	.22	.20	.17	.17	.17
	MOM	.37	.24	.21	.21	.21	.21	.16	.15	.15	.18
	PWM	.25	.23	.23	.22	.22	.19	.17	.17	.18	.21
100	ML	.18	.17	.16	.15	.14	.15	.13	.12	.105	.102
	MOM	.27	.16	.14	.14	.15	.16	.12	.100	.104	.12
	PWM	.17	.16	.16	.16	.16	.14	.12	.12	.13	.14
200	ML	.12	.11	.106	.098	.092	.101	.088	.077	.068	.064
	MOM	.20	.12	.101	.099	.102	.12	.086	.070	.072	.086
	PWM	.12	.11	.109	.109	.109	.099	.084	.082	.089	.101
500	ML	.076	.070	.065	.060	.055	.063	.054	.046	.040	.036
	MOM	.14	.077	.063	.062	.064	.091	.059	.044	.045	.054
	PWM	.076	.070	.068	.068	.069	.066	.054	.051	.056	.064
∞	ML	1.67	1.55	1.41	1.26	1.10	1.40	1.20	1.00	.80	.60
	MOM	—	2.73	1.41	1.38	1.42	—	2.23	1.00	1.00	1.21
	PWM	1.80	1.57	1.53	1.52	1.53	1.79	1.21	1.15	1.25	1.42

NOTE: ML is maximum likelihood, MOM is method of moments, and PWM is probability-weighted moments. The ∞ row gives the theoretical asymptotic limits as $n \to \infty$ of $n^{1/2} \times$ RMSE.

*Values are unreliable for reasons discussed in the text.

more than a 10% proportion of failures of the maximum likelihood procedure have been identified in our tables of results.

In our simulations, we used PWM estimators calculated using both a_r and $\tilde{a}_r$ with $\gamma = -.35$ and $\delta = 0$ as estimators of α_r. The biased estimators based on $\tilde{a}_r$ gave the better overall performance, so it is this variant whose results are given in the tables following.

Our simulation results are summarized in Tables 2–5, which give the bias and root mean squared error (RMSE) of estimators of the parameters α and k and of the upper-tail quantiles $x(.9)$, $x(.99)$, and $x(.999)$. Biases and RMSE's of quantile estimators have been scaled by the true value of the quantile being estimated. Quantiles in the center of the distribution ($.1 < F < .9$) are estimated equally well by all three methods and, as noted in Section 3.4, lower-tail quantiles are essentially scalar multiples of the parameter α.

The biases of parameter estimators are all positive but are generally not severe for samples of size 100 or more. Overall, the smallest bias is achieved by the PWM estimators. Moment estimators of both parameters have large positive biases when $k < -.2$, and this bias decays only slowly, certainly not as fast as n^{-1}, as the sample size n increases.

The parameter estimators with the smallest RMSE are generally the moment estimators when $k > 0$ and the PWM estimators when $k < -.2$. For $-.2 \le k < 0$, there is little to choose between the moment and PWM estimators, although the PWM estimators might be preferred in practice because of their lower bias. MLE's are shown to good advantage only when n is large and k is large and positive, and it is barely evident that they are asymptotically the most efficient estimators from the simulation results even for $n = 500$.

The three estimators of the $F = .9$ quantile have very similar properties, except that the moment estimator has a relatively high RMSE in small samples when $k < -.2$. For more extreme quantiles, all estimators have a high RMSE when $1 - F < 1/n$, as might be expected, but the moment estimators generally have the lowest RMSE. Moment estimators of quantiles have large negative biases, however, when $k \le -.2$, so the PWM estimators, which, when $k \le -.2$ have the smallest bias, again seem preferable

Table 4. Bias of Estimators of Generalized Pareto Quantiles

		$k = -.4$			$k = -.2$			$k = 0$			$k = .2$			$k = .4$		
	$F =$	.9	.99	.999	.9	.99	.999	.9	.99	.999	.9	.99	.999	.9	.99	.999
n	Method $x(F) =$	3.78	13.27	37.12	2.92	7.56	14.91	2.30	4.61	6.91	1.85	3.01	3.74	1.50	2.10	2.34
15	ML	.01	.39	—[a]	−.02	.11	—[a]	−.03[b]	−.01[b]	—[a]	−.04[b]	−.05[b]	—[a]	−.06[b]	−.05[b]	.02[b]
	MOM	.01	−.26	−.44	−.03	−.17	−.25	−.04	−.09	−.09	−.03	−.04	.00	−.03	−.01	.05
	PWM	−.05	−.12	−.01	−.04	−.06	.09	−.04	−.02	.12	−.03	.00	.13	−.03	.02	.14
25	ML	.00	.14	—[a]	−.02	.01	—[a]	−.03	−.06	.04	−.03	−.07	−.06	−.03[b]	−.06[b]	.07[b]
	MOM	.01	−.22	−.39	−.02	−.13	−.19	−.02	−.06	−.06	−.03	−.03	.00	−.02	−.01	.03
	PWM	−.04	−.08	.01	−.03	−.03	.08	−.02	−.01	.08	−.02	.00	.09	−.02	.02	.09
50	ML	.00	.05	.32	−.01	−.01	.10	−.01	−.04	−.02	−.01	−.05	−.06	−.01	−.05	−.06
	MOM	.02	−.17	−.32	−.01	−.09	−.13	−.01	−.03	−.03	−.01	−.01	.00	−.01	.00	.02
	PWM	−.02	−.05	.02	−.01	−.02	.05	−.01	.00	.05	−.01	.00	.05	−.01	.01	.04
100	ML	.00	.02	.13	−.01	.00	.04	−.01	−.02	−.02	−.01	−.03	−.04	−.01	−.03	−.04
	MOM	.02	−.14	−.27	−.01	−.06	−.09	−.01	−.02	−.02	−.01	−.01	.00	.00	.00	.01
	PWM	−.01	−.02	.02	−.01	−.01	.03	−.01	.00	.03	−.01	.00	.02	.00	.01	.02
200	ML	.00	.01	.06	.00	.00	.02	.00	.01	−.01	.00	−.01	−.02	.00	−.02	−.02
	MOM	.02	−.11	−.22	.00	−.04	−.05	.00	.01	−.01	.00	.00	.00	.00	.00	.01
	PWM	−.01	−.01	.01	.00	.00	.02	.00	.00	.01	.00	.00	.01	.00	.00	.01
500	ML	.00	.01	.03	.00	.00	.01	.00	.00	.00	.00	−.01	−.01	.00	−.01	−.01
	MOM	.02	−.08	−.17	.00	−.02	−.03	.00	.00	.00	.00	.00	.00	.00	.00	.00
	PWM	.00	−.01	.01	.00	.00	.01	.00	.00	.01	.00	.00	.01	.00	.00	.01

NOTE: ML is maximum likelihood, MOM is method of moments, and PWM is probability-weighted moments. Tabulated biases are for the ratio $\hat{x}(F)/x(F)$ rather than for the estimator $\hat{x}(F)$ itself.

[a] Values were too large to be estimated reliably.

[b] Values are unreliable for reasons discussed in the text.

under these circumstances. Maximum likelihood quantile estimators are very unreliable in small samples, particularly when $k < 0$.

Standard deviations obtained by simulation of parameter and quantile estimators can be compared with the values obtained from the asymptotic approximations of Section 3, Equations (3), (4), (6), and (8), which are also given in Tables 3 and 5. For sample sizes smaller than 200, the moment estimators, both of parameters and of quantiles, tend to perform much better, and the MLE's much worse, than asymptotic theory suggests. The theoretical and simulated standard deviations are closest for the PWM estimators, being no more than 10% different provided that $n \geq 50$ and $k > -.4$.

We also included in our simulations the empirical coverage probabilities of confidence intervals with nominal coverage probabilities .8, .9, .95, .98, and .99, for parameters and quantiles, as described in Section 3.5. In assessing the results, we shall consider the accuracy of a method of constructing confidence intervals to be the degree of closeness between the nominal and empirical coverage probabilities. In this sense, confidence intervals based on the MLE's were almost uniformly more accurate when using observed information than when using expected information, the exceptional cases being some confidence intervals for extreme upper quantiles for which both methods performed poorly. Apart from this, it is difficult to draw general conclusions from the simulation results, for the results depend on the numerical values of the shape parameter k, the sample size n, and the nominal coverage probability of the confidence interval, and on whether the confidence interval is for k, α, a quantile near the center of the distribution, or a quantile in the upper tail of the distribution. Results for confidence intervals for α and for quantiles in the lower tail of the distribution are, however, very similar, because of the equivalence $x(F) \sim \alpha F$ as $F \rightarrow 0$, noted in Section 3.4. Table 6 contains some representative results and illustrates some of the following conclusions, drawn from the whole range of our simulations. Confidence intervals for α and for quantiles $x(F)$ with $F < .8$ are reasonably accurate for sample sizes $n \geq 50$, except for confidence intervals based on method-of-moment estimators when $k \leq 0$; these tend to be too long. Confidence intervals for k and for quantiles $x(F)$ with $F \geq .8$ sometimes require very large sample sizes, 200 or more, before acceptable accuracy is obtained. This is particularly true of confidence intervals with nominal coverage probabilities greater than .95. Inaccuracy in confidence intervals for upper-tail quantiles generally takes the form that the intervals are too short. Although none of the methods of constructing confidence intervals for quantiles is uniformly accurate, the use of PWM-

Table 5. RMSE of Estimators of Generalized Pareto Quantiles

		$k = -.4$			$k = -.2$			$k = 0$			$k = .2$			$k = .4$		
	$F =$	.9	.99	.999	.9	.99	.999	.9	.99	.999	.9	.99	.999	.9	.99	.999
n	Method $x(F) =$	3.78	13.27	37.12	2.92	7.56	14.91	2.30	4.61	6.91	1.85	3.01	3.74	1.50	2.10	2.34
15	ML	.45	—[a]	—[a]	.34	—[a]	—[a]	.26[a]	—[a]	—[a]	.20[b]	.99[b]	—[a]	.16[b]	.24[b]	.98[b]
	MOM	.54	.64	.77	.33	.46	.63	.26	.36	.53	.21	.28	.43	.17	.24	.37
	PWM	.39	.70	1.45	.32	.56	1.19	.26	.43	.88	.21	.35	.70	.17	.29	.57
25	ML	.34	1.41	—[a]	.26	.70	—[a]	.20	.40	1.47	.16	.25	.60	.12[b]	.15[b]	.27[b]
	MOM	.41	.52	.66	.26	.38	.55	.20	.29	.44	.16	.22	.34	.13	.18	.28
	PWM	.30	.57	1.19	.25	.45	.93	.20	.35	.66	.16	.27	.51	.13	.23	.40
50	ML	.23	.61	1.80	.18	.40	.91	.14	.25	.47	.11	.16	.25	.088	.102	.14
	MOM	.29	.40	.55	.18	.29	.44	.14	.21	.34	.11	.16	.24	.089	.13	.19
	PWM	.21	.44	.92	.18	.33	.65	.14	.25	.45	.11	.20	.34	.092	.16	.26
100	ML	.16	.38	.83	.13	.26	.51	.101	.17	.29	.079	.11	[illegible]	.062	.066	.088
	MOM	.20	.31	.45	.13	.22	.36	.101	.16	.25	.080	.11	.17	063	.091	.13
	PWM	.16	.34	.69	.13	.25	.45	.101	.18	.31	.081	.14	.23	.065	.11	.17
200	ML	.11	.26	.49	.091	.18	.33	.072	.12	.20	.056	.074	.11	.043	.043	.057
	MOM	.15	.24	.38	.091	.17	.29	.072	.11	.18	.056	.080	.12	.044	.065	.094
	PWM	.11	.26	.52	.090	.17	.31	.072	.13	.22	.057	.098	.16	.046	.080	.12
500	ML	.072	.16	.29	.057	.11	.20	.045	.074	.12	.035	.045	.067	.027	.025	.032
	MOM	.091	.17	.29	.058	.12	.21	.045	.072	.12	.036	.051	.078	.028	.041	.059
	PWM	.074	.17	.34	.057	.11	.19	.045	.080	.14	.036	.062	.099	.029	.051	.074
∞	ML	1.59	3.43	5.97	1.28	2.48	4.24	1.01	1.64	2.65	.79	.96	1.38	.61	.45	.50
	MOM	—	—	—	1.28	3.70	7.11	1.01	1.64	2.65	.79	1.14	1.75	.62	.92	1.32
	PWM	1.79	4.34	7.65	1.29	2.49	4.26	1.02	1.81	3.00	.81	1.40	2.21	.64	1.13	1.64

NOTE: ML is maximum likelihood, MOM is method of moments, and PWM is probability-weighted moments. The ∞ row gives the theoretical asymptotic limits as $n \to \infty$ of $n^{1/2} \times$ RMSE. Tabulated RMSE's are for the ratio $\hat{x}(F)/x(F)$ rather than for the estimator $\hat{x}(F)$ itself.

[a] Values were too large to be estimated reliably.

[b] Values are unreliable for reasons discussed in the text.

based estimators in (9) gives the best overall results; for example, provided that $n \geq 50$, $k \geq -.3$, and $F \leq .99$, the 80% confidence interval for $x(F)$ based on PWM's has empirical coverage probability between 75% and 85%.

To deduce from the simulation results a recommendation as to which estimation method to use in practice requires some further specification of the problem, particularly of the range of values of k that is likely to be encountered. For k near 0 and for $k > 0$, moment estimators have the best overall performance. If there is a strong possibility that k is substantially less than 0, and particularly if k might be less than $-.2$, then PWM estimators will probably be preferred because of their low bias. The PWM method is also recommended if inferences concerning the variability of estimated parameters or quantiles are to be based on approximate variances derived from asymptotic theory. Maximum likelihood estimation, with the additional computational burden that it involves, appears to be justified only for very large samples when $k > .2$.

5. EXAMPLE

Hosking et al. (1985) fitted the GEV distribution to the annual maximum floods of the River Nidd at Hunsingore, England. Some hydrologists feel, however, that extreme upper quantiles of the distribution of annual maximum floods are better estimated by analyzing data for all flood peaks exceeding some threshold. The threshold can be chosen so as to give a larger sample than is available from the annual maximum series, while excluding the smallest annual maxima, which may have arisen from meteorological conditions untypical of those that cause the largest floods. Accordingly we now examine the flood peak data for the Nidd at Hunsingore. The data, taken from Natural Environment Research Council (NERC) (1975, pp. 235–236), consist of the 154 flood peaks between October 1, 1934, and September 30, 1969, which exceeded 65 m^3 s^{-1} and satisfied the independency criteria of NERC (1975, p. 13). In Figure 3 the ordered data are plotted against the

Table 6. Empirical Noncoverage Probabilities, Percent, of Nominal 90% Confidence Intervals for Parameters and Quantiles of the Generalized Pareto Distribution

Interval for	n	k = −.2 ML(E)	k = −.2 ML(O)	k = −.2 MOM	k = −.2 PWM	k = .2 ML(E)	k = .2 ML(O)	k = .2 MOM	k = .2 PWM
k	15	30.3	17.8	15.5	7.8	36.0	1.4	0.0	1.0
	25	25.9	18.8	16.3	9.0	37.4	16.6	2.5	4.7
	50	18.6	15.2	15.2	8.7	29.5	16.6	5.2	7.4
	100	14.8	13.1	13.5	8.1	22.9	14.6	7.9	8.8
	200	12.3	11.5	11.3	7.4	18.1	12.7	8.9	9.7
	500	10.9	10.5	8.5	8.9	14.2	11.2	9.5	9.7
α	15	18.0	8.4	2.5	7.0	20.3	7.2	6.4	9.9
	25	16.4	10.3	3.5	7.4	22.4	8.2	7.7	9.8
	50	12.8	10.4	4.5	7.9	17.5	10.8	8.9	9.8
	100	11.3	10.2	5.1	8.7	14.6	10.6	9.5	9.8
	200	11.0	10.4	5.1	9.6	13.0	10.4	10.1	10.1
	500	10.3	10.1	4.9	9.7	11.6	10.4	9.9	9.9
$x(.5)$	15	16.7	9.5	6.6	11.2	18.8	9.7	9.7	11.9
	25	14.9	10.2	6.0	10.7	19.7	9.0	10.0	11.2
	50	12.1	10.3	5.7	10.1	15.6	10.4	10.0	10.5
	100	10.9	10.2	5.6	9.9	13.2	10.3	10.0	10.2
	200	10.8	10.4	5.2	10.3	12.0	10.4	10.3	10.4
	500	10.3	10.1	4.7	10.0	11.1	10.2	10.0	10.0
$x(.99)$	15	34.9	35.9	34.2	28.1	31.6	28.0	21.8	19.9
	25	31.5	31.9	28.9	23.1	33.4	31.5	17.8	16.4
	50	23.7	24.1	23.3	18.3	30.6	26.8	14.2	13.3
	100	17.4	18.2	18.6	14.6	23.7	20.5	11.8	11.1
	200	13.5	14.2	14.7	11.9	18.2	16.1	10.7	10.2
	500	11.3	11.4	11.0	10.0	14.2	12.7	10.0	10.0

NOTE: Tabulated values are probabilities that the true value of the quantity of interest lies outside the interval constructed for it and should be close to 10% if the construction of the confidence interval is accurate. The methods are: ML(E)—maximum likelihood, expected information; ML(O)—maximum likelihood, observed information; MOM—method of moments; and PWM—probability weighted moments.

expected order statistics of the standard exponential distribution.

We shall assume that successive flood peaks have independent magnitudes and that the number of peaks per year exceeding threshold t has a Poisson distribution with mean $\lambda(t)$. The relationship between the quantiles $z(\cdot)$ of the distribution of annual maximum floods and the quantiles $x_t(\cdot)$ of the distribution of excesses over the threshold t is then given by

$$z(F) = t + x_t[1 + \{\lambda(t)\}^{-1} \log F], \qquad e^{-\lambda(t)} < F < 1$$

$$z(F) \le t, \qquad 0 < F \le e^{-\lambda(t)}$$

In any given year there is probability $e^{-\lambda(t)}$ that no peak exceeds the threshold, so if $F \le e^{-\lambda(t)}$, then $z(F)$ is not uniquely defined.

As remarked in Section 1, these assumptions, together with the assumption of a GEV distribution for the annual maximum floods, suggest the use of the generalized Pareto distribution for modeling excesses over a threshold. We use the method of probability-weighted moments to fit generalized Pareto distributions to the Nidd data. This is because we have a priori reason to believe that the shape parameter k might be markedly negative (negative values of k are about twice as common as positive values when fitting GEV distributions to British annual maximum flood series, and, as remarked in Sec. 1, we expect the same to be true for peak-over-threshold series) and because we shall make use of asymptotic theory to approximate finite-sample standard errors of estimated quantiles. In Table 7 we give the estimated parameters of generalized Pareto distributions fitted to the excesses of the Nidd peak floods over four different thresholds and the corresponding estimates of the .9, .99, and .999 quantiles of the distribution of annual maxima. Results of fitting a GEV distribution to the annual maximum series itself are also included. Table 7 also contains significance levels of the χ^2 test of the hypothesis that the number of peaks per year has a Poisson distribution. The χ^2 statistic was calculated from a histogram of the number of peaks per year. All years with $[2\hat{\lambda}(t)]$ or more peaks, where $[x]$

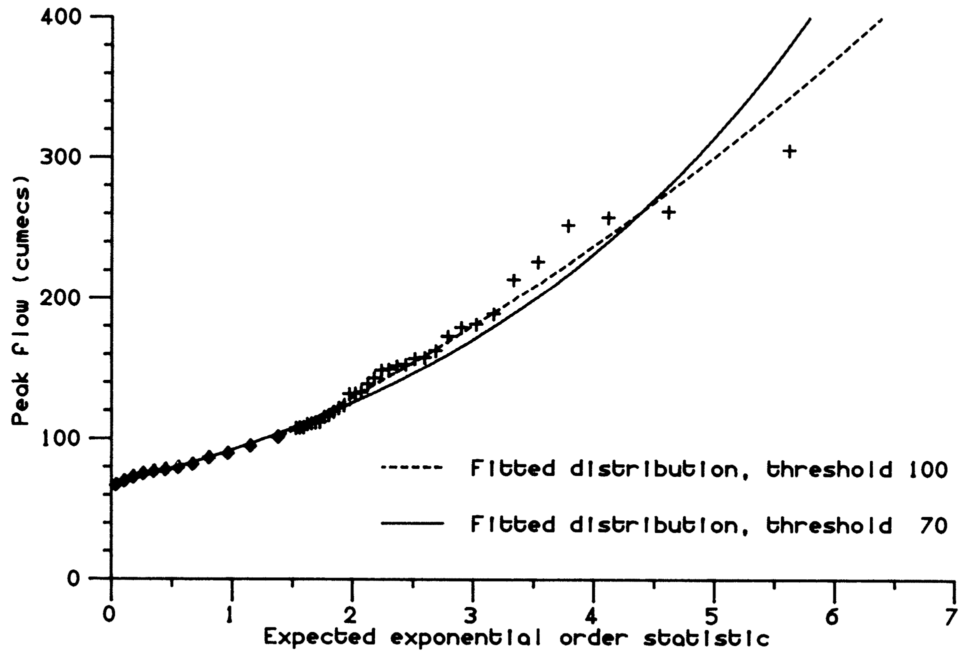

Figure 3. Flood Peaks of the River Nidd at Hunsingore, 1934–1969, and the Generalized Pareto Distributions Fitted to Those Peaks Exceeding the Thresholds 70 m^3 s^{-1} (solid line) and 100 m^3 s^{-1} (dashed line). To avoid crowding of the smaller peaks, the ◆ is used to indicate the average position of 10 plotted points.

denotes the largest integer not exceeding x, were combined into a single histogram class. Thus the histogram had $[2\hat{\lambda}(t)] + 1$ classes and the χ^2 test had $[2\hat{\lambda}(t)]$ df. The Nidd data are consistent with the Poisson hypothesis for all of the thresholds considered.

If all of the assumptions made in this analysis were valid, we would expect all of the estimates $\hat{k}$ in Table 7 to be approximately equal. In fact $\hat{k}$ tends to decrease as the threshold is reduced, and this suggests that there are more relatively small flood peaks than the generalized Pareto assumption implies. The fitted generalized Pareto distributions for thresholds 70 and 100 are plotted on Figure 3, and it seems that the distribution for threshold 100 gives a better fit to the larger observed peaks, particularly those with magnitudes in the range from 100 to 200 m^3 s^{-1}. On the whole, therefore, it seems best to use a fairly high threshold for the peak-over-threshold analysis; this also gives better agreement with the results obtained from analyzing the annual maximum series. The differences between estimated quantiles in Table 7, however, are not large compared with their standard errors. For example, the values of $\hat{z}(.99)$ in Table 7 differ by up to 22%; standard errors of $\hat{z}(.99)$ derived from asymptotic theory assuming that $k = -.2$ are

Table 7. Results of Fitting Generalized Pareto Distributions to Excesses Over a Threshold of Flood Peaks of the River Nidd at Hunsingore

			Estimated parameters			Estimated quantiles of distribution of annual maximum floods		
Threshold, t	No. of peaks	P	$\hat{\lambda}(t)$	$\hat{\alpha}$	$\hat{k}$	$\hat{z}(.9)$	$\hat{z}(.99)$	$\hat{z}(.999)$
AM	35			42.6	−.13	217	372	577
100	39	.54	1.11	45.5	−.10	222	377	571
90	57	.76	1.63	32.3	−.25	218	425	793
80	86	.26	2.46	24.8	−.32	216	454	938
70	138	.16	3.94	22.3	−.30	214	437	880

NOTE: P is the significance level of the χ^2 test for a Poisson distribution of the number of peaks per year. The AM row contains the results of fitting a GEV distribution to the annual maximum series.

$\pm 41\%$ for the estimate based on the annual maximum series and between $\pm 30\%$ and $\pm 41\%$ for estimates based on peaks over a threshold.

[*Received April 1986. Revised December 1986.*]

REFERENCES

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

Efron, B., and Hinkley, D. V. (1978), "Assessing the Accuracy of the Maximum-Likelihood Estimators: Observed Versus Expected Fisher Information," *Biometrika*, 65, 457–487.

Giesbrecht, F., and Kempthorne, O. (1976), "Maximum Likelihood Estimation in the 3-Parameter Lognormal Distribution," *Journal of the Royal Statistical Society*, Ser. B, 38, 257–264.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979), "Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form," *Water Resources Research*, 15, 1049–1054.

Hosking, J. R. M. (1985), "Algorithm AS215: Maximum-Likelihood Estimation of the Parameters of the Generalized Extreme-Value Distribution," *Applied Statistics*, 34, 301–310.

——— (1986), "The Theory of Probability Weighted Moments," Research Report RC12210, IBM Thomas J. Watson Research Center, Yorktown Heights, NY.

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.

Jenkinson, A. F. (1955), "The Frequency Distribution of the Annual Maximum (or Minimum) of Meteorological Elements," *Quarterly Journal of the Royal Meteorological Society*, 81, 158–171.

Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979a), "Probability Weighted Moments Compared With Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles," *Water Resources Research*, 15, 1055–1064.

——— (1979b), "Estimation of Parameters and Quantiles of Wakeby Distributions," *Water Resources Research*, 15, 1361–1379.

Natural Environment Research Council (NERC) (1975), *Flood Studies Report* (Vol. 4), London: Author.

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

Prescott, P., and Walden, A. T. (1983), "Maximum-Likelihood Estimation of the Parameters of the Three-Parameter Generalized Extreme-Value Distribution From Censored Samples," *Journal of Statistical Computation and Simulation*, 16, 241–250.

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications* (2nd ed.), New York: John Wiley.

Smith, R. L. (1984), "Threshold Methods for Sample Extremes," in *Statistical Extremes and Applications*, ed. J. Tiago de Oliveira, Dordrecht: D. Reidel, pp. 621–638.

——— (1985), "Maximum Likelihood Estimation in a Class of Nonregular Cases," *Biometrika*, 72, 67–90.

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