A New and Efficient Estimation Method for the Generalized Pareto Distribution

Author(s): Jin Zhang and Michael A. Stephens

Source: *Technometrics*, August 2009, Vol. 51, No. 3 (August 2009), pp. 316-325

Published by: Taylor & Francis, Ltd. on behalf of American Statistical Association and American Society for Quality

Stable URL: https://www.jstor.org/stable/40586625

A New and Efficient Estimation Method for the Generalized Pareto Distribution

Jin Zhang

School of Mathematics and Statistics
Yunnan University
Kunming, Yunnan, 650091, China
(*jinzhang@ynu.edu.cn*)

Michael A. Stephens

Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, B.C. V5A 1S6, Canada

The generalized Pareto distribution (GPD) is widely used to model extreme values, for example, exceedences over thresholds, in modeling floods. Existing methods for estimating parameters have theoretical or computational defects. An efficient new estimator is proposed, which is computationally easy, free from the problems observed in traditional approaches, and performs well compared with existing estimators. A numerical example involving heights of waves is used to illustrate the various methods and tests of fit are performed to compare them.

KEY WORDS: Efficiency; Maximum likelihood estimation; Method of moment estimation; Probability-weighted moment estimation.

1. INTRODUCTION

The generalized Pareto distribution (GPD) named by Pickands (1975) is a two-parameter family of distributions with the probability distribution function

$$F_{\sigma,k}(x) = \begin{cases} 1-(1-kx/\sigma)^{1/k}, & \text{if } k \neq 0 \\ 1-e^{-x/\sigma}, & \text{if } k=0, \end{cases}$$

and the density function

$$f_{\sigma,k}(x) = \begin{cases} \sigma^{-1}(1-kx/\sigma)^{1/k-1}, & \text{if } k \neq 0 \\ \sigma^{-1}e^{-x/\sigma}, & \text{if } k=0, \end{cases}$$

where σ and k are scale and shape parameters ($\sigma > 0$). The domain of x is $(0,\infty)$ for $k \leq 0$ or $(0,\sigma/k)$ for $k > 0$. The GPD becomes the uniform distribution when $k=1$ and the exponential distribution when $k=0$ (taken as the limit). The GPD has mean $\sigma/(1+k)$ and variance $\sigma^2/[(1+k)^2(1+2k)]$; these exist only if $k > -1$ and $k > -1/2$, respectively.

1.1 Peaks Over Thresholds Application

An important application of the GPD is to model extreme values, such as exceedences $(X-t)$, where X is a sample value and t is a given threshold: examples are flood levels of rivers, heights of waves, and ozone levels in the upper atmosphere. See Davison (1984), Hosking and Wallis (1987), and Smith (1984, 1985, 1989).

An attractive feature of the GPD in these applications is its stability; it may easily be shown that if the distribution of X is GPD(σ,k), the conditional distribution of $X-t$, given that $X > t$, for any level t, is GPD$(\sigma-kt,k)$. Davison and Smith (1990), have therefore suggested a method of estimating the appropriate threshold, by raising the t-value until the distribution of the exceedences is the GPD. This method has been applied to waves in the Bay of Biscay (Castillo and Hadi 1997) and to flood levels of Canadian rivers (Choulakian and Stephens 2001). In the latter paper the Davison–Smith procedure is investigated in some detail; also, the versatility of the GPD to model other kinds of data is discussed. We return to the Bay of Biscay data in Section 6.

In addition, many important applications of the GPD involve situations where its parameters depend on some covariates (see, e.g., Méndez et al. 2006). The maximum goodness-of-fit estimators can be used to cope with this case (Luceño 2006). We do not consider this situation in this article.

1.2 Contents of This Paper

The contents of this paper are as follows. Various methods for estimating the GPD parameters will be discussed in Section 2; special attention is given to maximum likelihood (ML) in Section 3. In Section 4 a new estimator will be proposed which is easy to compute and free from the problems encountered in traditional approaches. In Section 5 we show that the new method performs well compared with others in terms of efficiency and bias. The wave data is discussed as a peaks over thresholds (POT) example in Section 6, and some final remarks are in Section 7.

2. ESTIMATION OF PARAMETERS

Given a random sample $X_1, X_2, \dots, X_n$ of a population X, we first discuss estimation of parameters assuming the sample is from the GPD(σ,k). Later, we examine tests of fit to the GPD.

The classical method of estimation (maximum likelihood) can give problems when used for the GPD. In the next section we will discuss these problems and adapt the ML method to resolve these problems. However, many authors have proposed alternative methods of estimation. Hosking and Wallis (1987) discussed the method of moment (MOM) estimators of σ and k:

$$\hat{\sigma}_{\text{MOM}} = \overline{X}(\overline{X}^2/s^2+1)/2, \qquad \hat{k}_{\text{MOM}} = (\overline{X}^2/s^2-1)/2,$$

TECHNOMETRICS, AUGUST 2009, VOL. 51, NO. 3
DOI 10.1198/TECH.2009.08017

where $\overline{X}$ and s^2 are the sample mean and the sample variance. They also considered probability-weighted moment (PWM) estimators of σ and k:

$$\hat{\sigma}_{\mathrm{PWM}} = 2\alpha\overline{X}/(\overline{X} - 2\alpha), \qquad \hat{k}_{\mathrm{PWM}} = \overline{X}/(\overline{X} - 2\alpha) - 2,$$

where one choice of α is $\alpha = n^{-1}\sum_{i=1}^{n}\frac{n-i}{n-1}X_{(i)}$ and $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are the order statistics of the sample. MOM estimators perform poorly for $k \le -1/2$ because $\mathrm{Var}(X)$ does not exist, and so do PWM estimators for $k \le -1$ where $E(X)$ does not exist. Also, they may give invalid estimates, that is, if $\hat{k} > 0$ and $\hat{\sigma}/\hat{k} < X_{(n)}$, some of the sample observations will fall outside the domain, $0 < x < \hat{\sigma}/\hat{k}$, suggested by the estimates (see, e.g., Dupuis and Tsao 1998). Finally, the MOM estimators have low efficiencies and the large-sample efficiencies of PWM estimators are normally low (see Zhang 2007 and Section 5 below).

Castillo and Hadi (1997) proposed an elemental percentile method (EPM) which essentially fits the GPD to all pairs of order statistics in the sample, and then choosing the median k and σ.

Zhang (2007) proposed an estimation procedure to solve the computational problems of the MLEs. The estimates are called likelihood moment estimators (LME). The LME of $\theta = k/\sigma$ is the solution $\hat{\theta}_{\mathrm{LME}}$ of

$$n^{-1}\sum_{i=1}^{n}(1-\theta X_i)^p - (1-r)^{-1} = 0, \qquad \theta < X_{(n)}^{-1}, \quad (1)$$

where $p = rn/\sum_{i=1}^{n}\log(1-\theta X_i)$ with $r < 1$ being a constant to be specified.

The (only) solution of (1) is easily computed, since the left-hand side is a smooth monotone function of θ. Then

$$\hat{k}_{\mathrm{LME}} = -n^{-1}\sum_{i=1}^{n}\log(1-\hat{\theta}_{\mathrm{LME}}X_i), \qquad \hat{\sigma}_{\mathrm{LME}} = \hat{k}_{\mathrm{LME}}/\hat{\theta}_{\mathrm{LME}}, \quad (2)$$

which are asymptotically efficient in the region $k < 1/2$ if $r = k$. However, r must be chosen without knowing k. Zhang shows that a workable and robust choice is $r = -1/2$, which is used in this paper. The Splus (Insightful Corporation, Seattle, USA) or R (R Development Core Team 2005) code for computing $\hat{\sigma}_{\mathrm{LME}}$ and $\hat{k}_{\mathrm{LME}}$ can be found in Zhang (2007).

3. MLE ESTIMATES

The maximum likelihood method of estimation is always an important method in statistics, so that the MLEs of the parameters for the GPD are preferred in the literature. In fact, Smith (1984) showed that when $k < 1/2$, the MLEs for the GPD are consistent, asymptotically normal, and efficient. The MLEs for the GPD have been studied by many authors, including Davison (1984), Smith (1985), Hosking and Wallis (1987), Grimshaw (1993), and Choulakian and Stephens (2001).

It is convenient to reparametrize (σ, k) of the GPD to (θ, k), where $\theta = k/\sigma$. Then, using any estimates of (θ, k), σ can be estimated by $\hat{\sigma} = \hat{k}/\hat{\theta}$. In terms of (θ, k), the log-likelihood for the sample is

$$l^*(\theta, k) = n\log(\theta/k) + (1/k - 1)\sum_{i=1}^{n}\log(1-\theta X_i). \quad (3)$$

It is easy to show that the estimating equations of the MLE for (θ, k) are equivalent to

$$1 - k = n\Big/\sum_{i=1}^{n}(1-\theta X_i)^{-1}; \qquad k = -n^{-1}\sum_{i=1}^{n}\log(1-\theta X_i).$$

By eliminating k from (3), it suffices to solve the equation for θ:

$$\begin{aligned} l(\theta) &= 1 - n\Big/\sum_{i=1}^{n}(1-\theta X_i)^{-1} + n^{-1}\sum_{i=1}^{n}\log(1-\theta X_i) \\ &= 0, \qquad \theta < 1/X_{(n)}. \end{aligned} \quad (4)$$

When $\hat{\theta}_{\mathrm{MLE}}$ is obtained, (σ, k) are estimated by

$$\hat{k}_{\mathrm{MLE}} = -n^{-1}\sum_{i=1}^{n}\log(1-\hat{\theta}_{\mathrm{MLE}}X_i), \qquad \hat{\sigma}_{\mathrm{MLE}} = \hat{k}_{\mathrm{MLE}}/\hat{\theta}_{\mathrm{MLE}}.$$

The numerical solution of θ in (4) can be complex. First, the profile log-likelihood function $l(\theta)$ may steadily decrease, as θ decreases, from an infinity at $\theta = 1/X_{(n)}$, so that no local maximum may be found. The situation of no maximum will occur with increasing probability as the true k increases towards and beyond 1. Second, if there is a local maximum, it may be extremely close to $\theta = 1/X_{(n)}$, for example, within 10^{-6}, for some datasets, and then the solution of (4) will easily be passed over or may give convergence problems (see, e.g., Hosking and Wallis 1987; Grimshaw 1993). We discuss the MLE further in Section 5.

4. NEW ESTIMATORS

In this section a new estimator of (σ, k) is proposed, in order to improve the MLE and avoid the computational problems. The procedure borrows from Bayesian methods.

An estimator of $\theta = k/\sigma$ will be defined by

$$\hat{\theta} = \int \theta\cdot\pi(\theta)L(\theta)\,d\theta \Big/ \int \pi(\theta)L(\theta)\,d\theta, \quad (5)$$

where $l(\theta) = n[\log(\theta/k) + k - 1]$ with $k = -n^{-1}\sum_{i=1}^{n}\log(1-\theta X_i)$ is the profile log-likelihood function, $L(\theta) = e^{l(\theta)}$ is the profile likelihood function, and $\pi(\cdot)$ is a data-driven "prior" density function for θ.

For most priors, the computation of the integral in (5) is not easy. Therefore, a simplified numerical version of (5) will be given. Let $X^* = X_{([n/4+0.5])}$ be the first quartile of the sample data and $m = 20 + [\sqrt{n}]$, where $[x]$ denotes the largest integer smaller than or equal to x. Then define

$$\theta_j = 1/X_{(n)} + \left[1 - \sqrt{\frac{m}{j-0.5}}\right]\Big/(3X^*)$$

for $j = 1, \ldots, m$ and $w(\theta_j) = L(\theta_j)/\sum_{t=1}^{m}L(\theta_t)$. A computationally easier expression is $w(\theta_j) = 1/\sum_{t=1}^{m}e^{l(\theta_t)-l(\theta_j)}$.

The estimator proposed is now

$$\hat{\theta}_{\mathrm{NEW}} = \sum_{j=1}^{m}\theta_j\cdot w(\theta_j). \quad (6)$$

Statistic $\hat{\theta}_{\mathrm{NEW}}$ is not sensitive to the choice of m, provided $m > 20$ and $m = O(n^{\alpha})$ for some $0 < \alpha < 1$ to save computer computing time and to give necessary precision. The m given above is chosen for simplicity.

The motivation for the formula for θ_j is to make θ_j ($j = 1, \ldots, m$) be the $\frac{j-0.5}{m}$th quantile of the proposed prior distribution. Then in order to gain efficiency and to reduce bias of the new estimators, the prior $\pi(\cdot)$ for θ should be well chosen. Extensive simulation reveals that the new estimators are sensitive to the shape of the prior distribution, but not sensitive to its scale. Let $y = 1/X_{(n)} - \theta$; we have the constraint that y must be greater than 0. Consider a prior density $\pi(\theta) = g(y)$ such that $g(\cdot)$ is a probability density function with support $(0, \infty)$ and parameters $shape = c$, $scale = d/X_{([n*p+0.5])}$, where c and d are constants, and $X_{([n*p+0.5])}$ is the pth quantile of the sample data.

A simple but good choice of $g(\cdot)$, based on a simulation study, is the GPD density function itself, with shape $k^* = -0.5$, scale $\sigma^* = 1/(6X^*)$. Throughout this paper, this density $g(\cdot)$ will be used to give the prior density $\pi(\theta)$. The corresponding distribution function for θ is then $1 - G(y)$, where $G(y) = 1 - [1 + 3X^*y]^{-2}$, $y > 0$, is the distribution function of $g(y)$. Since $\frac{j-0.5}{m} = 1 - G(1/X_{(n)} - \theta_j)$, θ_j takes the form given above.

Finally, the new estimators are

$$\hat{k}_{\text{NEW}} = -n^{-1}\sum_{i=1}^{n}\log(1 - \hat{\theta}_{\text{NEW}}X_i), \qquad \hat{\sigma}_{\text{NEW}} = \hat{k}_{\text{NEW}}/\hat{\theta}_{\text{NEW}}. \tag{7}$$

It is easy to see that the θ_j and $\hat{\theta}$ are both less than $1/X_{(n)}$, so that the new estimators will always give valid estimates.

The program for computing $(\hat{\sigma}_{\text{NEW}}, \hat{k}_{\text{NEW}})$ in Splus or R code is as follows.

```
# x is the sample data from the GPD
f  <- function(x) {n <- length(x);
      x <- sort(x)
lx <- function(b, x)
      {k <- -mean(log(1-b*x));
      log(b/k)+k-1}
m <- 20+floor(sqrt(n))
b <- w <- L <- 1/x[n]+(1-sqrt(m/(1:m-
      .5)))/3/x[floor(n/4+.5)]
      for (i in 1:m) L[i] <- n*lx(b[i],x)
      for (i in 1:m)
      w[i]<- 1/sum(exp(L-L[i]))
b <- sum(b*w);
k <- -mean(log(1-b*x)); sigma <- k/b
      list(sigma=sigma, k=k)}
```

5. BIAS AND EFFICIENCY FOR DIFFERENT ESTIMATORS

In this section tables and plots are given to show bias and efficiency for finite sample sizes. These are based on Monte Carlo simulations, with 100,000 replicates. Without loss of generality, the scale σ is taken as 1; values of shape k were taken in the range $-1 < k < 0.5$, which is the most common range for k.

It is well known that the likelihood function of the GPD has no local maximum in the region $k > 1$, and when $k = 1$ the GPD(σ, k) becomes a uniform distribution U($0, \sigma$) so that the MLE of σ is $X_{(n)}$. Therefore, the MLE of (σ, k) in our simulation is set to $(\hat{\sigma}, \hat{k}) = (X_{(n)}, 1)$ if the MLE iterations do not converge or the $\hat{k}$ value is greater than 1, as recommended by an associate editor.

Bias

The biases in $\hat{\sigma}$ and $\hat{k}$ are shown in Table 1 and are plotted versus k in Figure 1 for $n = 50$ and 500. These show that the existing four estimators all have generally positive biases for all k and n. The biases of $(\hat{\sigma}_{\text{NEW}}, \hat{k}_{\text{NEW}})$ are always small and are balanced between positive values if $k < -1/2$ or negative if $k > -1/4$. Statistics $(\hat{\sigma}_{\text{MOM}}, \hat{k}_{\text{MOM}})$ have the largest biases for $k < 0$, while $(\hat{\sigma}_{\text{PWM}}, \hat{k}_{\text{PWM}})$ have the smallest biases for $k > 0$. The LMEs are better than MLEs for significantly reducing biases.

Efficiency

(a) In measuring efficiency, the accepted yardstick is to compare the Cramér–Rao lower bound of the asymptotic variance with the asymptotic variance of the estimate. The efficiency of an estimator is here defined as the ratio of the Cramér–Rao lower bound of the variance to the mean squared error (MSE) of the estimate, since the MSE is an overall criterion for measuring the accuracy of estimation. Note that the Cramér–Rao lower bound will not be valid for certain values of the parameters since the support of the distribution family is not independent of its parameters.

(b) Smith (1984) showed that the asymptotic variance of the MLEs for the GPD achieve the bound

$$n\,\text{var}\begin{bmatrix}\hat{\sigma}_{\text{MLE}} \\ \hat{k}_{\text{MLE}}\end{bmatrix} \sim \begin{bmatrix}2\sigma^2(1-k) & \sigma(1-k) \\ \sigma(1-k) & (1-k)^2\end{bmatrix}$$

so that the MLEs are, as expected, asymptotically efficient for $k < 0.5$. Hosking and Wallis (1987) gave the asymptotic variances for the MOM and PWM estimates; the efficiencies depend on the shape parameter k only. Zhang (2007) showed that when k is 0 or -0.25, respectively, $(\hat{\sigma}_{\text{MOM}}, \hat{k}_{\text{MOM}})$, and $(\hat{\sigma}_{\text{PWM}}, \hat{k}_{\text{PWM}})$ are asymptotically efficient, but the asymptotic efficiency of $\hat{k}_{\text{PWM}}$ vanishes when k is close to -0.5. Zhang also derived the asymptotic efficiencies of $(\hat{\sigma}_{\text{LME}}, \hat{k}_{\text{LME}})$; these are generally high, especially when k is close to $r = -1/2$. Unfortunately, it seems difficult to obtain the asymptotic variances and asymptotic efficiencies of $(\hat{\sigma}_{\text{NEW}}, \hat{k}_{\text{NEW}})$.

Comments

Tables and plots of finite-sample efficiencies against k are given in Table 2 and Figure 2. Apart from near $k = 0$ and $k = -0.25$, respectively, the efficiencies of the MOM and PWM estimators are very low. The Monte Carlo results show the high efficiencies of the LME, and also show high efficiency for the NEW estimates, especially for small sample sizes; the efficiency of the NEW estimates exceeds those of other estimators for all cases unless k is near -0.25, where $\hat{k}_{\text{PWM}}$ may be slightly better (more efficient). Among the four existing estimators, $(\hat{\sigma}_{\text{MLE}}, \hat{k}_{\text{MLE}})$ and $(\hat{\sigma}_{\text{LME}}, \hat{k}_{\text{LME}})$ have the best efficiencies for all k, except for small samples, where the efficiency of the MLE is very low due to its heavy bias.

Overall, the new estimators of (σ, k) outperform the existing estimators in terms of efficiency and bias. They also have the advantage that they can always be found, and do not give invalid estimates.

Table 1. Biases of different estimators of (σ, k)

n	$\hat{\sigma}_{\mathrm{MOM}}$	$\hat{\sigma}_{\mathrm{PWM}}$	$\hat{\sigma}_{\mathrm{MLE}}$	$\hat{\sigma}_{\mathrm{LME}}$	$\hat{\sigma}_{\mathrm{NEW}}$	$\hat{k}_{\mathrm{MOM}}$	$\hat{k}_{\mathrm{PWM}}$	$\hat{k}_{\mathrm{MLE}}$	$\hat{k}_{\mathrm{LME}}$	$\hat{k}_{\mathrm{NEW}}$
					$\sigma = 1, k = -1$					
10	6.379	0.438	0.743	0.448	0.283	0.783	0.467	0.345	0.288	0.181
20	8.276	0.350	0.199	0.204	0.119	0.674	0.367	0.121	0.134	0.068
50	7.288	0.267	0.066	0.075	0.057	0.594	0.280	0.044	0.053	0.039
100	7.955	0.227	0.032	0.037	0.029	0.559	0.235	0.022	0.027	0.020
200	7.334	0.196	0.015	0.018	0.014	0.537	0.199	0.010	0.012	0.009
500	7.507	0.165	0.006	0.007	0.006	0.520	0.166	0.004	0.005	0.004
1,000	11.951	0.146	0.003	0.004	0.003	0.512	0.147	0.002	0.003	0.002
					$\sigma = 1, k = -1/2$					
10	0.700	0.160	0.802	0.331	0.125	0.433	0.213	0.428	0.255	0.083
20	0.529	0.094	0.192	0.147	0.034	0.314	0.132	0.137	0.115	0.010
50	0.394	0.049	0.059	0.054	0.015	0.223	0.072	0.046	0.043	0.006
100	0.326	0.031	0.028	0.027	0.007	0.179	0.045	0.022	0.022	0.001
200	0.276	0.019	0.014	0.013	0.003	0.147	0.028	0.011	0.011	0.001
500	0.225	0.009	0.005	0.005	0.001	0.118	0.014	0.004	0.004	0.000
1,000	0.196	0.006	0.003	0.002	0.001	0.102	0.008	0.002	0.002	0.000
					$\sigma = 1, k = -1/4$					
10	0.348	0.103	0.823	0.278	0.056	0.301	0.131	0.487	0.241	0.035
20	0.219	0.053	0.207	0.125	0.001	0.187	0.071	0.160	0.110	−0.012
50	0.125	0.022	0.060	0.046	−0.002	0.105	0.031	0.052	0.040	−0.007
100	0.082	0.011	0.028	0.022	−0.004	0.068	0.017	0.025	0.020	−0.006
200	0.054	0.005	0.013	0.010	−0.003	0.044	0.008	0.012	0.009	−0.004
500	0.031	0.002	0.005	0.004	−0.001	0.025	0.004	0.005	0.004	−0.001
1,000	0.020	0.001	0.003	0.002	0.000	0.016	0.002	0.002	0.002	−0.001
					$\sigma = 1, k = 1/4$					
10	0.164	0.080	0.662	0.206	−0.041	0.170	0.064	0.546	0.219	−0.053
20	0.071	0.033	0.269	0.087	−0.049	0.071	0.022	0.230	0.091	−0.059
50	0.027	0.012	0.073	0.030	−0.025	0.026	0.008	0.073	0.031	−0.027
100	0.012	0.005	0.035	0.013	−0.016	0.012	0.004	0.036	0.013	−0.016
200	0.007	0.003	0.019	0.006	−0.009	0.006	0.002	0.019	0.006	−0.008
500	0.002	0.001	0.008	0.002	−0.004	0.002	0.001	0.008	0.002	−0.004
1,000	0.001	0.000	0.004	0.001	−0.002	0.001	0.000	0.005	0.001	−0.002
					$\sigma = 1, k = 1/2$					
10	0.157	0.085	0.388	0.180	−0.076	0.167	0.061	0.436	0.212	−0.097
20	0.065	0.036	0.272	0.073	−0.065	0.066	0.022	0.262	0.084	−0.077
50	0.024	0.013	0.090	0.023	−0.029	0.024	0.007	0.093	0.026	−0.030
100	0.011	0.006	0.044	0.009	−0.016	0.011	0.003	0.046	0.010	−0.015
200	0.006	0.003	0.024	0.004	−0.007	0.005	0.001	0.025	0.003	−0.006
500	0.002	0.001	0.011	0.001	−0.002	0.002	0.001	0.012	0.000	−0.002
1,000	0.001	0.000	0.006	0.000	−0.001	0.001	0.000	0.007	−0.001	0.000

6. EXAMPLE

We now examine the dataset given by Castillo and Hadi (1997, table 3). The values are the zero-crossing hourly mean periods (in seconds) of the sea waves measured in a Bilbao bay in January 1997, Spain. Castillo and Hadi (1997) used the GPD to fit the data over known thresholds, based on their EPM statistic. In Table 3 the estimated parameters are listed for thresholds $t = 7.0, 7.5, 8.0, 8.5, 9.0$, and 9.5. We have not been able to replicate the EPM estimates of Castillo and Hadi (1997), so these are not included.

For thresholds $t = 7.0$ and 7.5, Figures 3 and 4 respectively display the empirical distribution function (EDF) given by the data, and the estimated distribution functions corresponding to the parameter estimates in Table 3. See also figures 1 and 2 in Luceño (2006), using the maximum goodness-of-fit estimates.

The figures suggest that the threshold $t = 7.0$ may not be large enough for the GPD to provide a good fit to the data, and MOM and PWM methods give nonsensical estimates of (σ, k). With threshold $t = 7.5$, the GPD fit in Figure 4 improves, particularly in the upper tail, which is important in POT analysis. Formal tests given below support these findings.

Figure 4 shows that even though the MOM or PWM fit is generally good in the whole distribution, the NEW fit is the best in both tails of the distribution.

Goodness of Fit

The fit of the GPD to the data, with some specified (σ_0, k_0) and for different thresholds, will be tested using the classical Cramér–von Mises W^2 statistic or the Anderson–Darling statistic A^2. Also we shall use Zhang's (2002) statistic Z_C. For

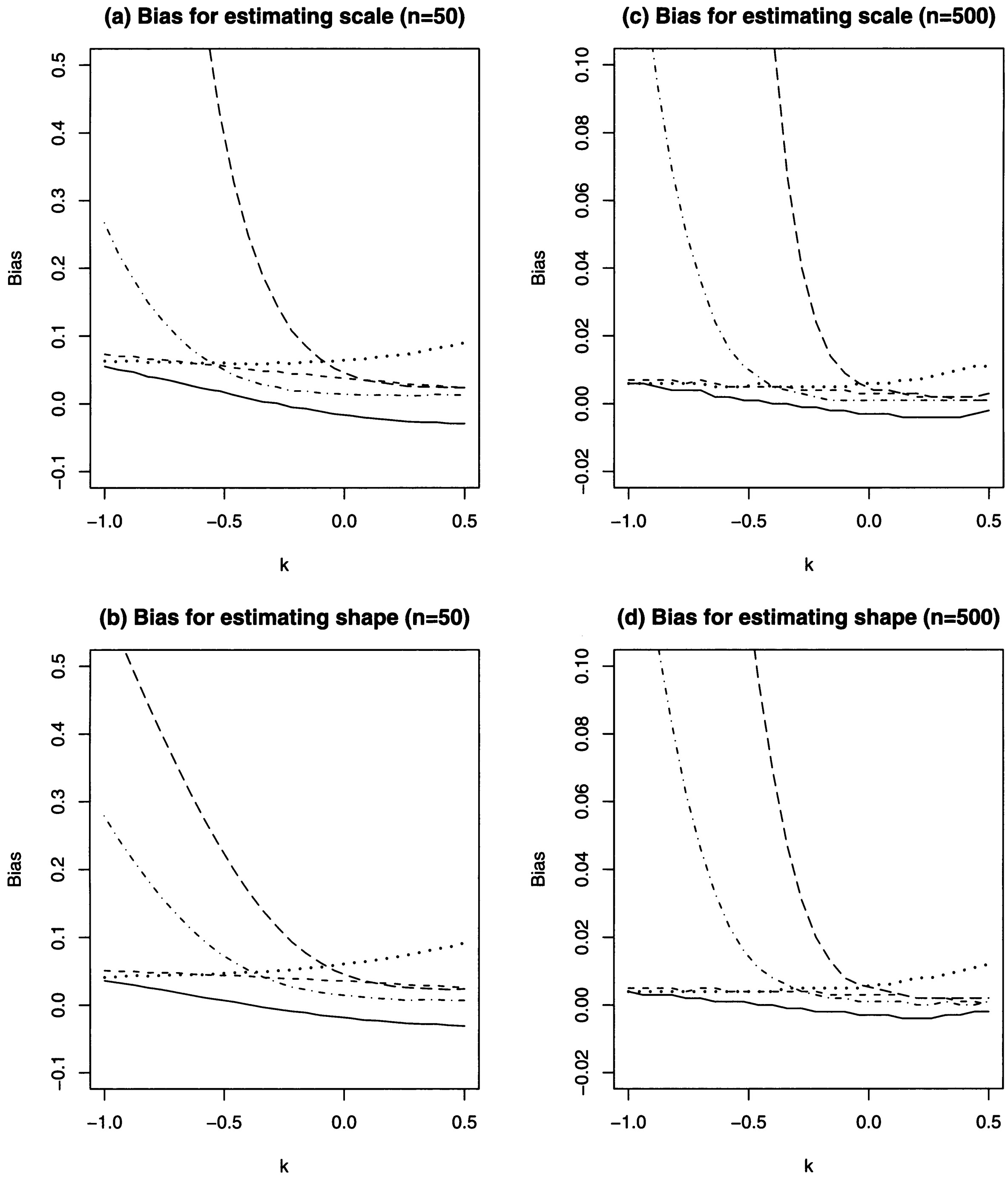

Figure 1. Bias of parameters. – – – – – MOM, – · – · – · – · – PWM, · · · · · · · · · MLE, - - - - - - - LME, ——— NEW.

basic definitions of W^2 and A^2, see Stephens (1986). For the present purpose, they are defined as follows. The Probability Integral Transformation (PIT) is: $u_{(i)} = F_{\hat{\sigma},\hat{k}}(x_{(i)})$, where $x_{(i)}, i = 1, \dots, n$, are the order statistics of the sample. Then

$$W^2 = \sum_{i=1}^{n} \left[u_{(i)} - (i - 0.5)/n\right]^2 + \frac{1}{12n}, \tag{8}$$

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) \times \left[\log\left(u_{(i)}\right) + \log\left(1 - u_{(n+1-i)}\right)\right]. \tag{9}$$

Statistic Z_C is defined by

$$Z_C = \sum_{i=1}^{n} \left[\log \frac{u_{(i)}^{-1} - 1}{n/(i - 0.5) - 1}\right]^2, \tag{10}$$

and large values of all three statistics indicate lack of fit.

We have used the parametric bootstrap to estimate p-values of the test statistics W^2, A^2, and Z_C. For each method of estimation (EPM excepted), samples of size m for the different t (given in Table 3) were generated from the GPD, using the parameter estimates for that threshold. For each run in the present study, 1,000 samples were generated. Finally the approximate p-value of each test statistic calculated from the original sample

Table 2. Efficiencies of different estimators of (σ, k) for the GPD

n	$\hat{\sigma}_{MOM}$	$\hat{\sigma}_{PWM}$	$\hat{\sigma}_{MLE}$	$\hat{\sigma}_{LME}$	$\hat{\sigma}_{NEW}$	$\hat{k}_{MOM}$	$\hat{k}_{PWM}$	$\hat{k}_{MLE}$	$\hat{k}_{LME}$	$\hat{k}_{NEW}$
					$\sigma = 1, k = -1$					
10	0	0.230	0.100	0.320	0.511	0.601	1.200	0.461	0.768	1.097
20	0	0.365	0.477	0.544	0.736	0.424	1.059	0.779	0.875	1.089
50	0	0.393	0.784	0.767	0.863	0.224	0.776	0.924	0.937	1.064
100	0	0.354	0.898	0.870	0.941	0.127	0.572	0.964	0.955	1.037
200	0	0.291	0.946	0.915	0.966	0.069	0.406	0.982	0.959	1.019
500	0	0.199	0.974	0.944	0.981	0.030	0.240	0.992	0.967	1.007
1,000	0	0.139	0.992	0.961	0.996	0.015	0.156	0.988	0.959	0.995
					$\sigma = 1, k = -1/2$					
10	0.055	0.657	0.101	0.420	0.737	0.805	1.122	0.290	0.634	0.939
20	0.141	0.839	0.428	0.654	0.915	0.850	1.278	0.601	0.806	0.975
50	0.123	0.946	0.788	0.848	0.974	0.710	1.234	0.843	0.925	1.024
100	0.101	0.956	0.887	0.915	0.986	0.562	1.103	0.918	0.960	1.014
200	0.082	0.954	0.954	0.967	1.005	0.421	0.952	0.956	0.978	1.007
500	0.083	0.900	0.987	0.993	1.007	0.267	0.772	0.987	0.996	1.007
1,000	0.058	0.852	0.999	1.002	1.010	0.183	0.661	0.992	0.996	1.002
					$\sigma = 1, k = -1/4$					
10	0.502	0.663	0.105	0.463	0.851	0.736	0.808	0.214	0.520	0.780
20	0.669	0.835	0.365	0.680	0.957	0.987	0.994	0.458	0.714	0.856
50	0.788	0.936	0.749	0.849	0.983	1.075	1.057	0.749	0.861	0.943
100	0.815	0.955	0.862	0.907	0.987	1.021	1.035	0.857	0.912	0.964
200	0.809	0.989	0.943	0.961	1.009	0.922	1.029	0.934	0.957	0.994
500	0.706	0.988	0.972	0.973	1.000	0.734	0.999	0.968	0.967	0.993
1,000	0.608	0.993	0.990	0.984	1.004	0.604	0.987	0.981	0.970	0.994
					$\sigma = 1, k = 1/4$					
10	0.370	0.418	0.169	0.431	0.835	0.189	0.204	0.117	0.226	0.346
20	0.577	0.533	0.238	0.595	0.824	0.341	0.273	0.164	0.348	0.403
50	0.704	0.607	0.577	0.738	0.857	0.442	0.313	0.381	0.485	0.537
100	0.730	0.617	0.717	0.778	0.867	0.475	0.323	0.521	0.554	0.624
200	0.754	0.633	0.807	0.810	0.898	0.496	0.332	0.636	0.602	0.713
500	0.766	0.638	0.889	0.831	0.934	0.507	0.335	0.763	0.641	0.812
1,000	0.779	0.647	0.934	0.847	0.962	0.517	0.340	0.840	0.664	0.876
					$\sigma = 1, k = 1/2$					
10	0.210	0.267	0.365	0.342	0.661	0.055	0.067	0.099	0.098	0.150
20	0.348	0.345	0.235	0.467	0.640	0.096	0.087	0.084	0.154	0.179
50	0.426	0.392	0.383	0.565	0.661	0.122	0.099	0.152	0.217	0.251
100	0.453	0.408	0.537	0.611	0.696	0.130	0.103	0.232	0.254	0.308
200	0.461	0.411	0.624	0.630	0.722	0.134	0.104	0.300	0.281	0.362
500	0.470	0.418	0.709	0.654	0.764	0.137	0.106	0.383	0.311	0.431
1,000	0.473	0.419	0.759	0.664	0.790	0.137	0.106	0.440	0.324	0.477

was found from the estimated distribution. The results are given in Table 4.

A diagram illustrating the relationship between the estimated and the exact distributions is described in Lockhart, O'Reilly, and Stephens (2007); also given are comparisons of bootstrap and exact p-values, which show the bootstrap to give good results. Many studies, including, for example, Konstantinides and Meintanis (2004), have assumed that this is normally the case, although the parametric bootstrap is based on an estimated distribution from which to draw the samples.

Comments

(a) When a method gave no solution for the original sample above the threshold (no local maximum for the MLE, or invalid $\hat{\theta}$ for MOM or PWM), as given in Table 4, the bootstrap resampling could obviously not be continued.

(b) The last column in Table 4 shows how many Monte Carlo samples gave good solutions (MLE estimates, or valid $\hat{\theta}$ for MOM or PWM estimators), on which the estimated p-values could be based. They were large enough to obtain approximate p-values, but the numbers show that there is a high probability of getting no estimate using these estimators, and this clearly weakens their appeal.

(c) We are then left with the LME and the NEW methods. In general, for thresholds $t = 7.5$ and above, all three test statistics suggest that the GPD is an adequate fit. There are some rather surprising differences in p-values given by the three statistics, which suggest that they do different jobs in testing the fit. This should be explored further.

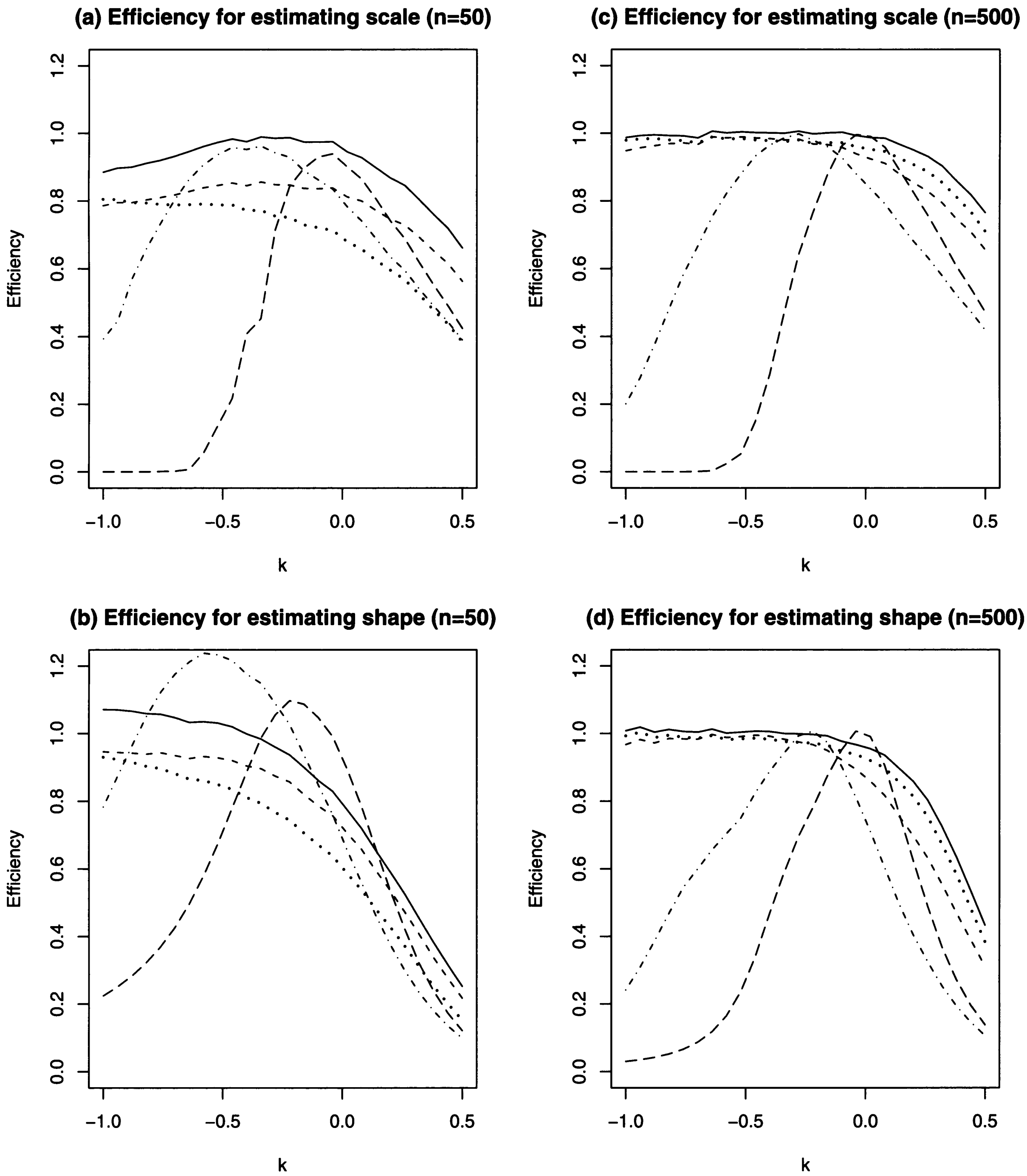

Figure 2. Efficiency of parameters. — — — — — MOM, — · — · — · — PWM, · · · · · · · · · MLE, - - - - - - - LME, ——— NEW.

Table 3. Different estimates of (σ, k) for the GPD, using Bilbao waves data

t	m	$\hat{\sigma}_{\text{MOM}}$	$\hat{\sigma}_{\text{PWM}}$	$\hat{\sigma}_{\text{MLE}}$	$\hat{\sigma}_{\text{LME}}$	$\hat{\sigma}_{\text{NEW}}$	$\hat{k}_{\text{MOM}}$	$\hat{k}_{\text{PWM}}$	$\hat{k}_{\text{MLE}}$	$\hat{k}_{\text{LME}}$	$\hat{k}_{\text{NEW}}$
7.0	179	2.750	2.780	2.500	2.450	2.380	1.050	1.070	0.861	0.838	0.808
7.5	154	1.620	1.620	1.860	1.670	1.750	0.606	0.602	0.768	0.651	0.706
8.0	106	1.380	1.370	1.650	1.510	1.460	0.647	0.630	0.864	0.727	0.768
8.5	69	1.130	1.110	No MLE	1.210	1.210	0.722	0.700	No MLE	0.833	0.833
9.0	41	0.814	0.809	No MLE	0.865	0.826	0.833	0.823	No MLE	0.938	0.878
9.5	17	0.626	0.601	No MLE	0.526	0.430	1.710	1.600	No MLE	1.310	1.010

NOTE: m is the number of exceedances over the threshold t, that is, the sample size.

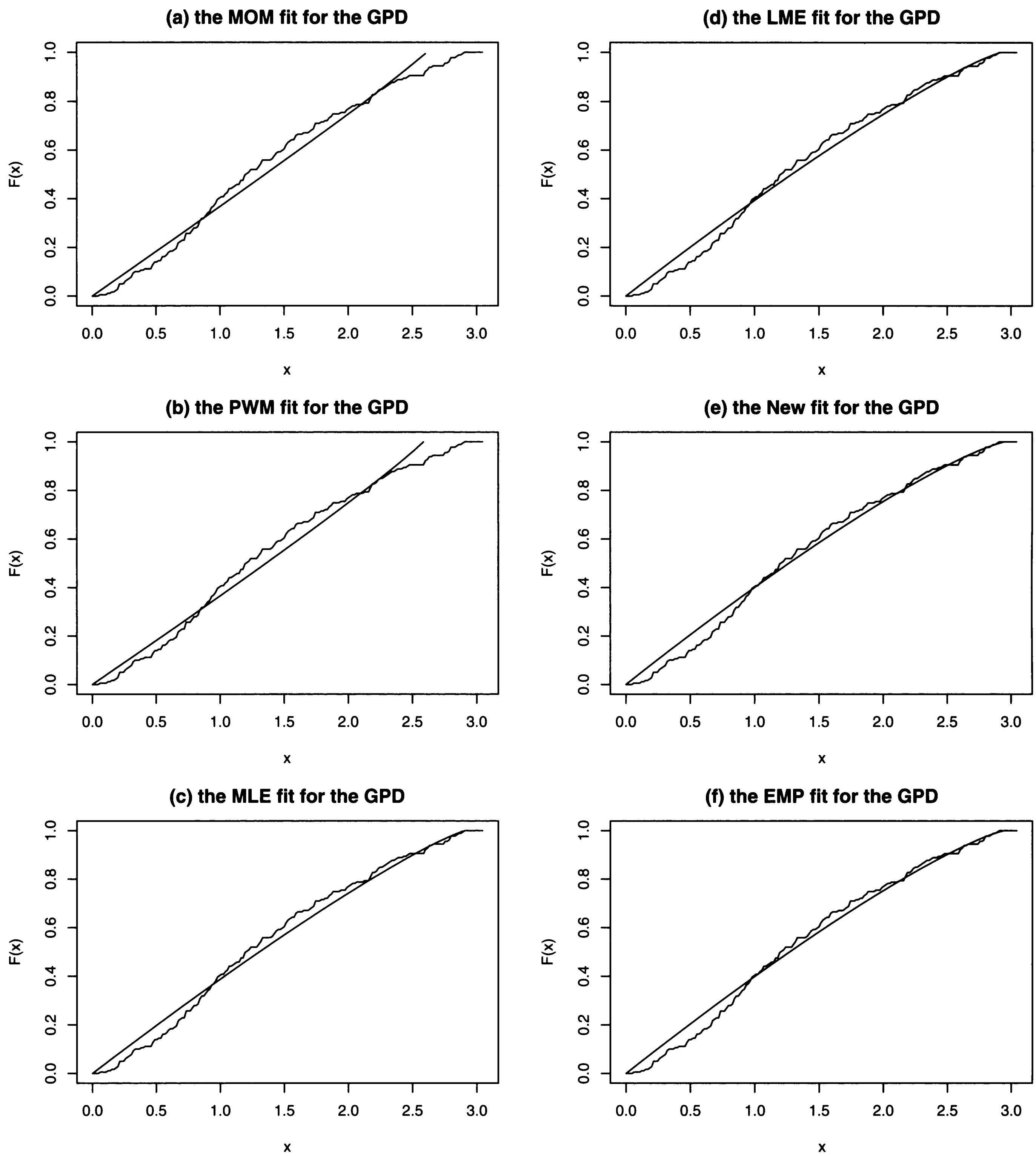

Figure 3. Quantile–quantile plots, $t = 7.0$.

7. FINAL REMARKS

A new estimating procedure has been introduced for the parameters of the GPD. It is based on MLE, but with a data-based prior which ensures that the estimates always exist. Studies of bias and efficiency show they compare very well with those of other estimates, some of which have serious problems in finding solutions. An example of fitting POT demonstrates these problems and shows the new procedure to be very effective.

This paper does not consider the situation where the parameters of the GPD depend on some covariates. For details about this aspect, refer to Méndez et al. (2006) and Luceño (2006).

ACKNOWLEDGEMENTS

The research is supported by the National Natural Science Foundation of China (NSFC, Grant No. 10871167), and by

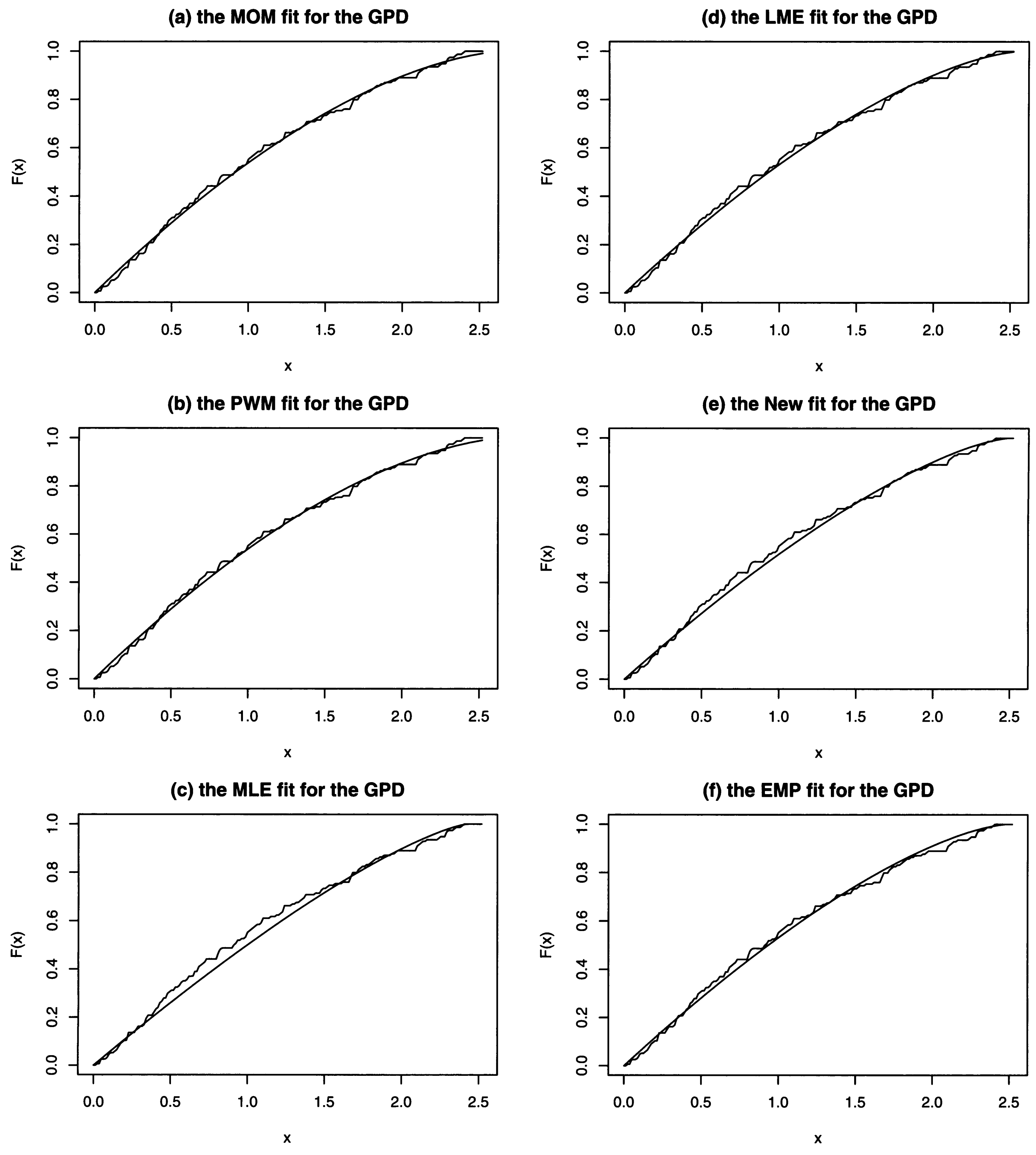

Figure 4. Quantile–quantile plots, $t = 7.5$.

the National Sciences and Engineering Research Council of Canada (NSERC). The authors are grateful to two referees, an Associate Editor, and the Editor for their helpful comments, corrections, and suggestions. The authors thank Professor Scott D. Grimshaw for providing us with his algorithm for computing the LMEs for the GPD.

[Received January 2008. Revised February 2009.]

REFERENCES

Castillo, E., and Hadi, A. S. (1997), "Fitting the Generalized Pareto Distribution to Data," *Journal of the American Statistical Association*, 92, 1609–1620.

Choulakian, V., and Stephens, M. A. (2001), "Goodness-of-Fit for the Generalized Pareto Distribution," *Technometrics*, 43, 478–484.

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

Table 4. Goodness-of-fit statistics and p-values for different estimators

Threshold	Z_C	p	W^2	p	A^2	p	Solutions
MLE							
7.0	28.12	0.27	0.228	0.26	1.605	0.24	533
7.5	10.43	0.45	0.168	0.14	0.896	0.18	910
8.0	9.82	0.54	0.133	0.26	0.745	0.30	539
8.5	No MLE						
9.0	No MLE						
9.5	No MLE						
NEW							
7.0	26.30	0.017	0.217	0.027	1.62	0.009	1,000
7.5	7.66	0.38	0.081	0.36	0.494	0.40	1,000
8.0	5.35	0.58	0.047	0.71	0.318	0.73	1,000
8.5	2.99	0.91	0.034	0.88	0.258	0.85	1,000
9.0	3.01	0.83	0.063	0.51	0.384	0.57	1,000
9.5	7.38	0.12	0.100	0.26	0.680	0.20	1,000
LME							
7.0	25.34	0.076	0.215	0.011	1.57	0.006	1,000
7.5	8.75	0.37	0.044	0.61	0.349	0.53	1,000
8.0	5.95	0.54	0.033	0.81	0.253	0.80	1,000
8.5	2.99	0.86	0.034	0.78	0.258	0.79	1,000
9.0	2.46	0.86	0.068	0.33	0.384	0.48	1,000
9.5	5.77	0.19	0.067	0.32	0.501	0.26	1,000
PWM							
7.0	$\hat{\theta}$ not valid						
7.5	11.46	0.33	0.033	0.58	0.323	0.49	712
8.0	10.11	0.35	0.024	0.95	0.253	0.79	697
8.5	6.63	0.54	0.033	0.88	0.291	0.78	684
9.0	4.56	0.66	0.059	0.44	0.398	0.29	675
9.5	$\hat{\theta}$ not valid						
MOM							
7.0	$\hat{\theta}$ not valid						
7.5	11.19	0.30	0.033	0.71	0.321	0.54	719
8.0	9.27	0.35	0.024	0.96	0.238	0.84	699
8.5	5.92	0.58	0.031	0.84	0.270	0.46	659
9.0	4.33	0.65	0.059	0.45	0.392	0.57	650
9.5	$\hat{\theta}$ not valid						

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

Dupuis, D. J., and Tsao, M. (1998), "A Hybrid Estimator for Generalized Pareto and Extreme-Value Distributions," *Communications in Statistics—Theory and Methods*, 27, 925–941.

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

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

Konstantinides, D., and Meintanis, S. G. (2004), "A Test of Fit for the Generalized Pareto Distribution Based on Transforms," in *Proceedings of the 3rd Conference in Actuarial Science and Finance in Samos, 2004*. Available at *http://www.stat.ucl.ac.be/Samos2004/proceedings2004/Meintanis.pdf*.

Lockhart, R. A., O'Reilly, F. J., and Stephens, M. A. (2007), "Use of the Gibbs Sampler to Obtain Conditional Tests, With Applications," *Biometrika*, 94, 992–998.

Luceño, A. (2006), "Fitting the Generalized Pareto Distribution to Data Using Maximum Goodness of Fit Estimators," *Computational Statistics & Data Analysis*, 51, 904–917.

Méndez, F. J., Menéndez, M., Luceño, A., and Losada, I. J. (2006), "Estimation of the Long-Term Variability of Extreme Significant Wave Height Using a Time-Dependent Peak Over Threshold (POT) Model," *Journal of Geophysical Research*, 111, C07024.

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

R Development Core Team (2005), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing, available at *http://www.R-project.org*.

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

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

——— (1989), "Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone," *Statistical Science*, 4, 367–393.

Stephens, M. A. (1986), "Tests Based on EDF Statistics," in *Goodness-of-Fit Techniques*, New York: Dekker, Chapter 4, pp. 112–191.

Zhang, J. (2002), "Powerful Goodness-of-Fit Tests Based on Likelihood Ratio," *Journal of the Royal Statistical Society*, Ser. B, 64, 281–294.

——— (2007), "Likelihood Moment Estimation for the Generalized Pareto Distribution," *Australian and New Zealand Journal of Statistics*, 49, 69–77.