



Exceedances over High Thresholds: A Guide to Threshold Selection

D.J. DUPUIS

Department of Engineering Mathematics, DalTech, Dalhousie University, Halifax, Nova Scotia, Canada B3H 3J5, e-mail: dupuis@tuns.ca

[Received February 20, 1998; Revised August 24, 1998; Accepted October 19, 1998]

Abstract. In this paper, we consider the modeling of exceedances over high thresholds. The natural distribution for such exceedances, the generalized Pareto distribution (GPD), is used and the problematic issue of threshold selection is addressed. We fit the GPD robustly to the data using techniques based on optimal bias-robust estimates. The robust procedure will assign weights between 0 and 1 to each data point. These weights are used to assess the validity of the GPD model for exceedances of the proposed threshold and thus can guide threshold selection. That is, we can initially consider a low threshold and increase it (thus reducing the number of data points) until all weights are close to one. The new approach is used to analyze two of the NERC data sets.

Key words. generalized Pareto distribution, optimal bias robust estimators, weights

1. Introduction

Consider the problem of statistical inferences about the extreme values in a population or random process. Extreme-value distributions are often used to model natural phenomena such as sea levels, river heights, stream flows, and rainfall, amongst others, in order to obtain the distribution of annual maxima. Another approach is to use, instead of annual maxima or several largest order statistics, all exceedances over a high threshold. The latter threshold methods have been developed by hydrologists over the last 30 years. Early versions of the method assumed a non-homogeneous Poisson process to model the times of exceedances over the high thresholds in conjunction with independent exponentially distributed excesses. The first systematic developments are in Todorovic and Zelenhasic (1970) and Todorovic and Rousselle (1971). This approach was generalized in Davison and Smith (1990) where excesses are modeled as independent generalized Pareto random variables. The generalized Pareto distribution (GPD) has cumulative distribution function

$$H(x; \alpha, k) = 1 - \left(1 - \frac{kx}{\alpha}\right)^{1/k} \quad (1)$$

where $0 < x < \infty$ if $k \leq 0$, $0 < x < \alpha/k$ if $k > 0$, and $\alpha > 0$ but k is arbitrary. The exponential distribution is a special case of the GPD and is obtained from (1) by taking the limit as $k \rightarrow 0$. The use of (1) to model excesses is as natural as the GPD has an interpretation

as a limiting distribution similar to that which motivates the generalized extreme-value distribution. See Pickands (1975) for the original theoretical development and Davison and Smith (1990) (and references therein) for further developments and applications. The GPD's derivation suggests that it will be a practical family for statistical estimation provided that the threshold is taken sufficiently high.

In practice, the parameters α and possibly k must be estimated from the data. Present alternatives are Maximum Likelihood, Method of Moments and Probability-Weighted Moments. They have been reviewed by Hosking et al. (1987) where it is found that the latter two are more reliable (reduced bias and rms) in samples of size 500 or less. Note however that both these methods may yield nonfeasible estimates (Dupuis, 1996). The issue of feasibility is addressed in Dupuis and Tsao (1998) where hybrid estimators are examined.

In modeling the excess data, we choose the GPD as the model distribution but do not have any assurance that all the data are well modeled by this distribution. In fact, in many situations small deviations from the underlying model can greatly affect classical estimates such as maximum likelihood. Davison and Smith (1990) plot influence curves for ML and PWM estimators. They show that both estimators have unbounded influence functions and that, if anything, PWM estimators are even more sensitive than ML estimators. Thus, outliers can have a large effect when fitting (1).

Furthermore, in fitting the GPD, there is the issue of threshold selection. Did we choose a high enough threshold? The threshold should be high enough to justify the assumptions of the model, but low enough to capture a reasonable number of observations. In this paper, we seek to select the threshold based on robust estimation of the parameters. At a given threshold, estimates to the parameters in (1) are obtained using the optimal bias robust estimator (OBRE) which fits the bulk of the data well. This robust procedure also assigns a weight to each data point. We can use these weights to assess the validity of our model distribution as observations with low final weights are those which do not fit the model as determined by the robust parameter estimates. To ignore this information can put the validity of any inferences in jeopardy. Thus, if we have too low a threshold, the observed exceedances are not likely to have a GPD and this lack of fit will be evident through the low weights assigned to some of the observations. At this point, our options are

- a. base inferences on the exceedances which are well fit by the model and downweight exceedance(s) which are poorly fit,
- b. increase the threshold and reduce the number of data points until all weights are close to one, or
- c. leave the threshold unchanged but try to develop alternative models in which all the weights are close to one.

In this paper, we explore option (b) and use the OBRE to guide us in choosing a threshold appropriate for inferences based on the GPD.

The remainder of the paper is organized as follows. In Section 2, we give the basic ideas of the OBRE and discuss its implementation in the case of the GPD. In Section 3, we look

at two data sets from the *Flood Studies Report* (Natural Environment Research Council (NERC), 1975) and show how the OBRE is used to select a proper threshold for the analyses. Appendix I gives details of the OBRE; appendix II summarizes the fitting algorithm; and appendix III gives the details of the goodness of fit tests.

2. Optimal bias-robust estimators

As noted above, our aim is to develop robust estimates when the model distribution for the n exceedances over a threshold, $x_1, \dots, x_n$, is the GPD with density function

$$h_\theta(x) = \frac{1}{\alpha} \left(1 - k \frac{x}{\alpha}\right)^{(1-k)/k} \quad (2)$$

where the range of x is $0 \leq x \leq \infty$ for $k \leq 0$ and $0 \leq x \leq \alpha/k$ for $k > 0$. The distribution involves the two-dimensional parameter $\theta = [\alpha, k]^t$ where α is a scale parameter and k is a shape parameter. The corresponding score function ($s(x; \theta) = (\partial/\partial\theta) \ln h_\theta(x)$) is

$$s(x; \theta) = \begin{bmatrix} s_1(x; \theta) = -\frac{1}{\alpha} + \left(\frac{1-k}{k}\right) \frac{kx}{\alpha^2(1-k\frac{x}{\alpha})} \\ s_2(x; \theta) = -\frac{1}{k} \ln \left(1 - k \frac{x}{\alpha}\right) - \frac{1-k}{k^2} \ln \left(1 - k \frac{x}{\alpha}\right) - \frac{1-k}{k} \frac{x}{\alpha(1-k\frac{x}{\alpha})} \end{bmatrix}. \quad (3)$$

Let $H^{(n)}$ represent the empirical distribution of the data putting mass $1/n$ on each x . As estimators of θ , we consider statistics $T_n = T_n(x_1, \dots, x_n)$ which can be represented, at least asymptotically, as functionals of the empirical distribution function, i.e. $T_n(x_1, \dots, x_n) = T(H^{(n)})$. The robustness of the estimator T is assessed by means of the influence function (*IF*) (Hampel, 1968 & 1974). The *IF* describes the effect of a small contamination ($\varepsilon\Delta_x$) at the point x on the estimate (standardized by the mass of the contamination). A desirable robustness property for an estimator is that it have a bounded *IF*. Such an estimator is called B-robust (Bias-robust). By bounding the influence function, we are able to ensure that small deviations from the model distribution do not cause large changes in the estimates. The OBRE are constructed to have bounded influence and provide the weights to indicate which observations, if any, are not well fit by the model. On the other hand, the *IF* of a ML estimator (MLE) is proportional to its score function. Unfortunately, the MLE for (2) has an unbounded score function (3) in x , and therefore an unbounded *IF*. PWM can also be shown to have an unbounded *IF*. In fact, Davison and Smith (1990) plot influence curves for both estimators and point out that, if anything, PWM are even more sensitive than MLE.

The OBRE, defined in Hampel et al. (1986), belongs to the class of M-estimators which are a generalization of the MLE (Huber, 1964). The estimator is described for a p -dimensional parameter θ of an arbitrary density function $h_\theta(x)$ in Victoria-Feser and Ronchetti (1994) and appears in appendix I. The OBRE for the parameters of (2) must be found iteratively. We have adapted the fitting algorithm of Victoria-Feser and Ronchetti

(1994) to the GPD and it appears in appendix II. The OBRE algorithm is convergent provided that the starting point is near the solution. One may argue that the use of PWM or MLE as starting points may not be robust enough. If problems are encountered, one may choose to compute the OBRE with a larger robustness constant and then use these estimates as starting values for the more robust estimates. The OBRE procedure yields weights which will always be less than or equal 1. They will be less than 1 when the norm of the appropriately standardized score function exceeds the robustness constant. This downweights observations which are not close to the working values of the parameters in this score function metric. For observations which are consistent with the current values of θ , the weights are one and the score function is that of the maximum likelihood estimate. This ensures that the OBRE is efficient since it is as similar as possible to the MLE for the bulk of the data. The downweighted exceedances are likely to be those at either extreme of the data. Since the extremes can be very informative in predicting return levels, it is important to know whether our model is consistent with the extreme data. The weight function gives us precisely the information needed to make this assessment. If a potentially informative observation is downweighted, we know that inferences based on (2) are likely to be flawed.

The constant c is the bound on the IF and acts as a regulator between robustness and efficiency. We found $c = 4$ to be a good compromise.

In order to properly evaluate the weights, it is important to note that the OBRE may in fact downweight points for data generated from a GPD. Thus, for each sample size and shape parameter of interest, we run a small simulation. Note that the OBRE is invariant under linear transformation so we fix α to 1 in all simulations. The mean and the standard deviation of the weights will be examined. This provides us with a calibration for determining when the downweighting indicates a serious lack of fit. Unfortunately no simple rule, such as treat any weight below a certain quantity as suspicious, can effectively be adopted as there is considerable variability in the resulting weights (as can be seen in tables 2 and 5). One must consider each value of n and k individually to maximize the benefits of the approach.

In the following section we use OBRE to help us determine the appropriate threshold in the analysis of two data sets from the *Flood Studies Report* (NERC, 1975).

3. Application to NERC data sets

3.1. Example 1

The data consists of 141 exceedances of the threshold level $95 \text{ m}^3 \text{ s}^{-1}$ by the River Derwent at Longbridge from June 7, 1935 to March 31, 1969. The source is pp. 258–259 of the *Flood Studies Report* (NERC, 1975), volume 4. The data already satisfy some independence criteria (NERC, 1975, volume 4, p. 13). The usual approach is to fit the GPD to the exceedances. However, the same problems always arise: *i*) where does one set the threshold for the analysis; and *ii*) at a given threshold, does the GPD give a reasonable fit to the resulting exceedances. In the past, one would deal with (*i*) by trying many thresholds

and choosing one where little difference in the results (e.g. estimated quantiles) would ensue if we would have chosen a slightly higher or lower threshold value (see e.g. Davison and Smith (1990)). As there was previously no goodness of fit test for the GPD, (ii) was sometimes dealt with by using the renormalized Kolmogorov-Smirnov and Anderson-Darling statistics for testing an exponential distribution with unknown mean (see e.g. Davison and Smith (1990)). Now, a goodness of fit test has been established (Choulakian and Stephens, 1998) and this option is open to us for (ii). Here, we use the OBRE (and the weights that it provides) to deal with both (i) and (ii). Goodness of fit results are added for completeness.

Table 1 shows the estimated parameters of the GPD and the weight of the down-weighted observations at various thresholds. Table 2 shows the results of the calibrating simulations. We compare the weights in table 1 to those for the GPD in Table 2. Also included in Table 1 is the probability (according to the empirical distribution of the weights in the simulation) of seeing a weight that small under the null hypothesis that the GPD model is correct. At thresholds 95, 100, 105, and 110, we can see that the larger exceedances are being downweighted significantly. Three options are available to us (as listed in Section 1). Here, we chose option (b) and increase the threshold until all weights are acceptable. At thresholds 115, 120, and 125, weights assigned to all exceedances are in line with the weights from a GPD and we can be quite confident of any inference drawn

Table 1. Results of OBRE fits to River Derwent Data. OBRE are computed using $c = 4$. The quantile estimate $\hat{q}_p = u + \frac{\hat{\alpha}}{\hat{k}}(1 - p^k)$, where u is the threshold, is also included for $p = 0.02$. P-value is the probability (according to the empirical distribution of the weights in the simulation) of seeing a weight that small under the null hypothesis that the GPD model is correct.

Thres.	# of exc.	$\hat{\alpha}$	$\hat{k}$	$\hat{q}_{0.02}$	Weights $\neq 1$							
95	141	31.77	0.01	216.89	rank	136	137	138	139	140	141	
					wght	0.79	0.33	0.12	0.10	0.07	0.02	
					p-value	0.11	0.03	0.03	0.04	0.05	0.05	
100	123	30.09	-0.03	227.72	rank	118	119	120	121	122	123	
					wght	0.99	0.44	0.17	0.15	0.11	0.04	
					p-value	0.08	0.04	0.03	0.04	0.06	0.07	
105	110	26.15	-0.16	247.19	rank	106	107	108	109	110		
					wght	0.66	0.32	0.29	0.23	0.12		
					p-value	0.03	0.02	0.05	0.11	0.17		
110	92	26.43	-0.19	263.41	rank	88	89	90	91	92		
					wght	0.95	0.47	0.43	0.33	0.17		
					p-value	0.03	0.01	0.11	0.18	0.26		
115	80	23.92	-0.35	315.40	rank	77	78	79	80			
					wght	0.87	0.81	0.65	0.38			
					p-value	0.04	0.21	0.36	0.45			
120	62	31.25	-0.23	318.24	rank	61	62					
					wght	0.81	0.38					
					p-value	0.42	0.40					
125	56	28.45	-0.31	341.83	rank	55	56					
					wght	0.94	0.48					
					p-value	0.45	0.47					

Table 2. Mean and standard deviation of weights for OBRE ($c = 4$) for GPD data, $\alpha = 1$. Based on 200 simulations.

n	k	Weights $\neq 1$							
141	0.01	rank	135	136	137	138	139	140	141
		mean	0.99	0.95	0.86	0.73	0.57	0.42	0.30
		stdv	0.05	0.13	0.19	0.23	0.25	0.24	0.24
123	-0.03	rank	118	119	120	121	122	123	
		mean	0.98	0.92	0.80	0.64	0.48	0.31	
		stdv	0.07	0.16	0.24	0.24	0.25	0.22	
110	-0.15	rank	106	107	108	109	110		
		mean	0.97	0.88	0.74	0.55	0.35		
		stdv	0.09	0.18	0.23	0.25	0.22		
92	-0.20	rank	89	90	91	92			
		mean	0.96	0.81	0.63	0.39			
		stdv	0.11	0.24	0.27	0.27			
80	-0.35	rank	77	78	79	80			
		mean	0.99	0.91	0.73	0.45			
		stdv	0.06	0.16	0.26	0.25			
62	-0.25	rank	60	61	62				
		mean	0.96	0.78	0.52				
		stdv	0.13	0.27	0.30				
56	-0.30	rank	54	55	56				
		mean	0.98	0.82	0.54				
		stdv	0.09	0.26	0.30				

based on a GPD fit to these data. Of course, one would choose an analysis based on a threshold of $115 \text{ m}^3 \text{ s}^{-1}$ since the latter would include the largest number of observations.

Details of the goodness of fit tests can be found in appendix III and results appear in Table 3. Conclusions are mixed for both the Cramer-von Mises and Anderson-Darling statistics, and are somewhat difficult to interpret. We note that both goodness of fit tests are based on ML estimators. Since ML estimators are biased in the presence of

Table 3. Results of Goodness of Fit tests to River Derwent Data. Relevant upper-tail 5% critical points for the Cramer-von Mises statistic $W_{n,k}^2$ and the Anderson-Darling statistic $A_{n,k}^2$ are $W_{100,-0.2}^2 = 0.135$, $W_{100,-0.3}^2 = 0.132$, $W_{100,-0.4}^2 = 0.128$, $W_{50,-0.3}^2 = 0.128$, and $A_{100,-0.2}^2 = 0.894$, $A_{100,-0.3}^2 = 0.895$, $A_{100,-0.4}^2 = 0.848$, and $A_{50,-0.3}^2 = 0.895$ respectively (from Choulakian and Stephens (1998)). Test statistics are compared to the critical points for the closest k and n available. Significant tests are denoted by *.

Thres.	# of exc.	$\hat{\alpha}_{MLE}$	$\hat{k}_{MLE}$	$\hat{q}_{0.02,MLE}$	Cramer-von Mises	Anderson-Darling
95	141	29.25	-0.19	264.78	0.126	0.978*
100	123	28.13	-0.22	274.49	0.106	0.896*
105	110	25.13	-0.29	287.81	0.088	1.26*
110	92	25.87	-0.30	302.61	0.098	0.818
115	80	23.99	-0.37	325.87	0.170*	2.46*
120	62	31.30	-0.27	337.43	0.142*	0.946*
125	56	28.50	-0.33	352.68	0.118*	0.781

outliers or when we are away from the model, it is more difficult to use this approach to deal with (i) and (ii). We believe the weights from the OBRE to be much more effective. We note that results at a threshold of 125 are consistent with our previous conclusions however.

3.2. Example 2

The data consists of 131 exceedances of the threshold level $59 \text{ m}^3 \text{ s}^{-1}$ by the River Wye at Rhayader from November 9, 1937 to October 6, 1969. The source is pp. 380–381 of the *Flood Studies Report* (NERC, 1975), volume 4.

Table 4 shows the estimated parameters of the GPD and the weight of the down-weighted observations at various thresholds. Table 5 shows the results of the calibrating simulations. In this case, the threshold used to analyze the data is shown adequate and should be chosen for the analysis. We note that once an adequate threshold has been reached, higher thresholds are likely to suffice as well. This is well illustrated in Tables 4 and 5. Results for the goodness of fit tests appear in Table 6. In this case, the data for all the considered thresholds pass both goodness of fit tests.

In both of the above examples it was not necessary to let the α and k parameters depend on covariates to explain any trend, however the method presented could be extended to include such models. The use of the OBRE on these more general models in the context of an extreme value analysis based on the r largest annual events is illustrated in Dupuis (1997).

Table 4. Results of OBRE fits to River Wye Data. OBRE are computed using $c = 4$.

Thres.	# of exc.	$\hat{\alpha}$	$\hat{k}$	$\hat{q}_{0.02}$	Weights $\neq 1$				
59	131	24.88	− 0.32	253.13	rank	128	129	130	131
					wght	0.84	0.68	0.63	0.62
					p-value	0.35	0.46	0.69	0.93
63	114	24.33	− 0.38	282.09	rank	112	113	114	
					wght	0.83	0.76	0.76	
					p-value	0.50	0.75	0.98	
67	97	26.22	− 0.37	297.47	rank	95	96	97	
					wght	0.90	0.82	0.82	
					p-value	0.44	0.67	0.92	
71	83	28.53	− 0.35	310.02	rank	82	83		
					wght	0.90	0.90		
					p-value	0.69	0.95		

Table 5. Mean and standard deviation of weights for OBRE ($c = 4$) for GPD data, $\alpha = 1$. Based on 200 simulations.

n	k	Weights $\neq 1$						
131	− 0.30	rank	126	127	128	129	130	131
		mean	0.99	0.95	0.85	0.70	0.54	0.35
		stdv	0.05	0.12	0.18	0.22	0.23	0.19
114	− 0.40	rank	110	111	112	113	114	
		mean	0.98	0.91	0.77	0.60	0.39	
		stdv	0.07	0.16	0.23	0.23	0.20	
97	− 0.35	rank	94	95	96	97		
		mean	0.94	0.83	0.65	0.43		
		stdv	0.14	0.22	0.26	0.25		
83	− 0.35	rank	80	81	82	83		
		mean	0.98	0.88	0.71	0.46		
		stdv	0.07	0.19	0.25	0.26		

Appendix I: Optimal bias-robust estimators

The OBRE are the solution T_n of the (system) of equation(s)

$$\sum_{i=1}^n \Psi(t_i; T_n) = 0$$

for some function $\Psi : T \times R^p \rightarrow R^p$. The IF of an M -estimator Ψ at H_θ is given by

$$IF(t; \Psi, H_\theta) = M(\Psi, H_\theta)^{-1} \Psi(t, \theta)$$

where

$$M(\Psi, H_\theta) = - \int \frac{\partial}{\partial \theta} \Psi(t, \theta) dH_\theta(t). \quad (\text{A.1})$$

Table 6. Results of Goodness of Fit tests to River Wye Data. Relevant upper-tail 5% critical points for the Cramer-von Mises statistic $W_{n,k}^2$ and the Anderson-Darling statistic $A_{n,k}^2$ are $W_{100,-.3}^2 = 0.132$ and $A_{50,-.3}^2 = 0.895$ respectively (from Choulakian and Stephens (1998)). Test statistics are compared to the critical points for the closest k and n available. All tests were not significant at 5%.

Thres.	# of exc.	$\hat{\theta}_{MLE}$	$\hat{k}_{MLE}$	$\hat{q}_{0.02,MLE}$	Cramer-von Mises	Anderson-Darling
59	131	25.58	− 0.27	236.69	0.04	0.416
63	114	25.17	− 0.30	250.40	0.04	0.349
67	97	27.13	− 0.28	259.84	0.052	0.556
71	83	29.56	− 0.25	267.18	0.072	0.462

The OBRE is optimal in the sense that it is the M -estimator which minimizes the trace of the asymptotic covariance matrix under the constraint that it has a bounded IF . There are several versions of the OBRE: each different in the way they choose to bound the influence function. Here, we use the standardized OBRE. For a given bound c on the IF ($c \geq p^{1/2}$), these estimators are defined implicitly by

$$\sum_{i=1}^n \Psi(t_i; \boldsymbol{\theta}) = \sum_{i=1}^n \{s(t_i; \boldsymbol{\theta}) - \mathbf{a}(\boldsymbol{\theta})\} W_c(t_i; \boldsymbol{\theta}) \quad (\text{A.2})$$

where $s(t; \boldsymbol{\theta})$ is as in (3) and $W_c(t; \boldsymbol{\theta})$ is the weight function

$$W_c(t; \boldsymbol{\theta}) = \min \left\{ 1; \frac{c}{\|\mathbf{A}(\boldsymbol{\theta})\{s(t, \boldsymbol{\theta}) - \mathbf{a}(\boldsymbol{\theta})\}\|} \right\} \quad (\text{A.3})$$

where $\|\cdot\|$ denotes the Euclidean norm. The $p \times p$ matrix $\mathbf{A}(\boldsymbol{\theta})$ and the $p \times 1$ vector $\mathbf{a}(\boldsymbol{\theta})$ are defined implicitly by

$$\begin{aligned} E\{\Psi(t; \boldsymbol{\theta})\Psi(t; \boldsymbol{\theta})^T\} &= \{\mathbf{A}(\boldsymbol{\theta})^T \mathbf{A}(\boldsymbol{\theta})\}^{-1}, \\ E\Psi(t; \boldsymbol{\theta}) &= 0. \end{aligned} \quad (\text{A.4})$$

Appendix II: Summary of OBRE fitting algorithm

The details for an observed sample $t_1, \dots, t_n$ are as follows:

- Step 1. Fix an initial value for the parameter $\boldsymbol{\theta} = [\alpha, k]^T$. That is, compute initial estimates (starting values) of the parameters α and k .
- Step 2. Fix a precision threshold η , the robustness constant c , $\mathbf{a} = \mathbf{0}$ and $\mathbf{A} = \mathbf{J}^{1/2}(\boldsymbol{\theta})^{-T}$ where

$$\mathbf{J}(\boldsymbol{\theta}) = \int s(t, \boldsymbol{\theta}), s(t, \boldsymbol{\theta})^T dH_{\boldsymbol{\theta}}(t)$$

is the Fisher information matrix.

- Step 3. Solve the following equations for $\mathbf{a}$ and $\mathbf{A}$:

$$\mathbf{A}^T \mathbf{A} = \mathbf{M}_2^{-1}$$

where

$$\mathbf{a} = \frac{\int s(t, \boldsymbol{\theta}) W_c(t, \boldsymbol{\theta}) dH_{\boldsymbol{\theta}}(t)}{\int W_c(t, \boldsymbol{\theta}) dH_{\boldsymbol{\theta}}(t)},$$

and

$$\mathbf{M}_k = \int \{s(t, \boldsymbol{\theta}) - \mathbf{a}\} \{s(t, \boldsymbol{\theta}) - \mathbf{a}\}^t W_c(t; \boldsymbol{\theta})^k dH_{\boldsymbol{\theta}}(t) \quad k = 1, 2.$$

The current values of $\boldsymbol{\theta}$, $\mathbf{a}$ and $\mathbf{A}$ are used as starting values to solve the above equations.

- Step 4. Compute $\mathbf{M}_1$ using $\mathbf{a}$ and $\mathbf{A}$ from Step 3 and

$$\Delta \boldsymbol{\theta} = \mathbf{M}_1^{-1} \left[\frac{1}{n} \sum_{i=1}^n \{s(t_i, \boldsymbol{\theta}) - \mathbf{a}\} W_c(t_i, \boldsymbol{\theta}) \right].$$

- Step 5. If $\max_j \left| \frac{\Delta \theta_j}{\theta_j} \right| > \eta$ ($j = 1, 2$) then $\boldsymbol{\theta} \rightarrow \boldsymbol{\theta} + \Delta \boldsymbol{\theta}$ and return to Step 3, else stop.

Note that integration can be avoided in Step 2 and in the computation of $\mathbf{M}_k$ (Steps 3 and 4) by replacing $H_{\boldsymbol{\theta}}$ with its empirical distribution function. However numerical integration must be used to compute $\mathbf{a}$ otherwise (A.2) will be satisfied by all estimates.

Appendix III: Goodness of fit tests

The null hypothesis is

$$H_0 : \text{the random sample } x_1, x_2, \dots, x_n \text{ comes from (1)}$$

Let $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ be the order statistics. Then, the goodness of fit test procedure is as follows:

- Find estimators $\hat{\alpha}_{MLE}$ and $\hat{k}_{MLE}$ for the unknown parameters and make the transformation

$$z_{(i)} = H\left(x_{(i)}; \hat{\alpha}_{MLE}, \hat{k}_{MLE}\right), \quad \text{for } i = 1, \dots, n. \quad (\text{A.5})$$

b. The Anderson-Darling statistic A^2 and the Cramer-von Mises statistic W^2 are

$$A^2 = -n - (1/n) \sum_{i=1}^n (2i-1) [\ln\{z_{(i)}\} + \ln\{1 - z_{(n+1-i)}\}]$$

$$W^2 = \sum_{i=1}^n \{z_{(i)} - (2i-1)/(2n)\}^2 + 1/(12n). \quad (\text{A.6})$$

Critical values for both statistics are found in Choulakian and Stephens (1998).

Acknowledgment

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to the author. The author wishes to thank the Editor and a referee for helpful comments.

References

- [1] Choulakian, V. and Stephens, M.A., "Goodness of fit tests for the generalized Pareto distribution," submitted to *Technometrics*, (1998).
- [2] Davison, A.C. and Smith, R.L., "Models for exceedances over high thresholds," *J. R. Statist. Soc. B* 52(3), 393–442, (1990).
- [3] Dupuis, D.J., "Estimating the probability of obtaining nonfeasible parameter estimates of the generalized Pareto distribution," *J. Statist. Comput. Simul.* 54, 197–209, (1996).
- [4] Dupuis, D.J., "Extreme-value theory based on the r largest annual events: a robust approach," *Journal of Hydrology* 200, 295–306, (1997).
- [5] Dupuis, D.J. and Tsao, M., "A hybrid estimator for Generalized Pareto and Extreme-Value Distributions," *Comm. in Stat.: Theory and Methods* 27(4), 925–941, (1998).
- [6] Hampel, F.R., *Contributions to the theory of robust estimation*, Ph.D. Thesis, University of California, Berkeley, 1968.
- [7] Hampel, F.R., "The influence curve and its role in robust estimation," *J. Amer. Statist. Assoc.* 69, 383–393, (1974).
- [8] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A., *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York, 1986.
- [9] Hosking, J.R.M. and Wallis, J.R., "Parameter and quantile estimation of the generalized Pareto distribution," *Technometrics* 29(3), 339–349, (1987).
- [10] Huber, P.J., "Robust estimation of a location parameter," *Ann. Math. Statist.* 35, 73–101, (1964).
- [11] Natural Environment Research Council, Flood Studies Report, London: Natural Environment Research Council, (1975).
- [12] Pickands, J., "Statistical inference using extreme order statistics," *Ann. Statist.* 3, 119–131, (1975).
- [13] Todorovic, P. and Rousselle, J., "Some problems of flood analysis," *Wat. Resour. Res.* 7, 1144–1150, (1971).
- [14] Todorovic, P. and Zelenhasic, E., "A stochastic model for flood analysis," *Wat. Resour. Res.* 6, 1641–1648, (1970).
- [15] Victoria-Feser, M. and Ronchetti, E.M., "Robust methods for personal income models," *Canad. J. Statist.* 22(2), 247–258, (1994).