Estimating the Extremal Index

Author(s): Richard L. Smith and Ishay Weissman

Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, 1994, Vol. 56, No. 3 (1994), pp. 515-528

Published by: Wiley for the Royal Statistical Society

Stable URL: https://www.jstor.org/stable/2346124

J. R. Statist. Soc. B (1994)
56, *No.* 3, *pp.* 515–528

Estimating the Extremal Index

By RICHARD L. SMITH† and ISHAY WEISSMAN

University of North Carolina, Chapel Hill, USA *Technion, Haifa, Israel*

[Received November 1991. Final revision March 1993]

SUMMARY

The extremal index is an important parameter measuring the degree of clustering of extremes in a stationary process. If we consider the point process of exceedance times over a high threshold, then this can be shown to converge asymptotically to a clustered Poisson process. The extremal index, a parameter in the interval [0, 1], is the reciprocal of the mean cluster size. Apart from being of interest in its own right, it is a crucial parameter for determining the limiting distribution of extreme values from the process. In this paper we review current work on statistical estimation of the extremal index and consider an optimality criterion based on a bias–variance trade-off. Theoretical results are developed for a simple doubly stochastic process, and it is argued that the main formula obtained is valid for a much wider class of processes. The practical implications are examined through simulations and a real data example.

Keywords: BIAS–VARIANCE TRADE-OFF; EXTREME VALUES IN STATIONARY PROCESSES; THRESHOLDS; WAVE HEIGHTS

1. BACKGROUND

Suppose that we have n observations from a stationary process $\{X_i, i \geqslant 0\}$ with marginal distribution function F. For large n and u_n, typically

$$F_n(u_n) = \Pr\{\max(X_1, \ldots, X_n) \leqslant u_n\} \approx F^{n\theta}(u_n) \tag{1.1}$$

where $\theta \in [0, 1]$ is a constant for the process known as the *extremal index*. This concept originated in papers by Cartwright (1958), Newell (1964), Loynes (1965), O'Brien (1974) and Leadbetter (1983); see also Leadbetter *et al.* (1983), section 3.7, and Leadbetter and Rootzén (1988). The classical theory of extremes in independent, identically distributed (IID) sequences (Galambos, 1987; Leadbetter *et al.*, 1983) provides conditions for the existence of normalizing constants $a_n > 0$ and b_n such that

$$F^n(a_n x + b_n) \to H(x) \tag{1.2}$$

where $H(x)$ is one of the three classical extreme value types $\Phi_\alpha(x) = \exp(-x^{-\alpha})$, $x > 0$, $\Psi_\alpha(x) = \exp\{-(-x)^\alpha\}$, $x < 0$, or $\Lambda(x) = \exp\{-\exp(-x)\}$, $-\infty < x < \infty$. Combining expressions (1.1) and (1.2), it is seen that for a stationary sequence $\{X_n\}$ we have

$$F_n(a_n x + b_n) = \Pr\{\max(X_1, \ldots, X_n) \leqslant a_n x + b_n\} \to H^\theta(x). \tag{1.3}$$

Since H is computable from knowledge of the marginal distribution, it appears that θ is the key parameter for extending extreme value theory for IID random

†*Address for correspondence*: Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260, USA.

 0035–9246/94/56515

variables to stationary processes. From a *statistical* point of view, since there are already many publications on estimating extreme value distributions from IID data, it can be seen that estimating the extremal index is a key problem.

An alternative characterization of the extremal index (Hsing *et al.*, 1988) is that $1/\theta$ is the limiting mean cluster size in the point process of exceedance times over a high threshold. This suggests that a suitable way to estimate the extremal index is to identify clusters of high level exceedances, and to calculate the mean size of those clusters. This idea underlies the statistical proposals of Smith (1989) and Davison and Smith (1990), though in an informal way the concept has existed in the hydrology literature for at least 20 years. However, until recently there has been no discussion of any optimality criteria for selecting the clusters. This is the subject of the present paper.

For independent data, there has been much discussion of statistical methods based on exceedances of a high threshold u_n; references include Smith (1987, 1989), Davison and Smith (1990), Dekkers and de Haan (1989) and Dekkers *et al.* (1989). For our initial discussion, we shall assume that the same threshold is being used for the identification of clusters. Later discussion will include optimal choice of the threshold as well.

The discussion throughout is confined to stationary sequences. Although non-stationary sequences are of practical interest, an exact analogue of the extremal index may not exist in this case (Hüsler, 1986). In practice, in cases where the non-stationarity arises from such sources as seasonal variation in the data, a pragmatic approach is to break up the data into seasonal portions within which they may be assumed stationary, as in Smith (1989). In the following discussion we shall assume that such preliminary analysis has already been carried out.

Suppose, then, that we have n observations from a stationary series, and let N_n denote the number of observations which exceed a predetermined high threshold u_n. We consider two methods of defining clusters. The first, called the *blocks method*, divides the data into approximately k_n blocks of length r_n, where $n \approx k_n r_n$. Each block is treated as one cluster; thus if Z_n denotes the number of blocks in which there is at least one exceedance of the threshold u_n we consider N_n/Z_n to estimate the mean size of a cluster and hence estimate θ by

$$\hat{\theta}_n = Z_n/N_n. \tag{1.4}$$

The following argument leads to a refinement of this. We can consider Z_n/k_n to be an estimator of $1 - F_{r_n}(u_n)$ where, as in equation (1.1), F_r denotes the distribution function of $\max(X_1, \ldots, X_r)$. Also, N_n/n is an estimator of $1-F(u_n)$. If we relate these by the approximation $F_r \approx F^{r\theta}$, then we obtain the estimator

$$\tilde{\theta}_n = \frac{\log(1-Z_n/k_n)}{r_n \log(1-N_n/n)}. \tag{1.5}$$

Recalling that $n \approx k_n r_n$, it can be seen that equations (1.4) and (1.5) are asymptotically equivalent when the ratios Z_n/k_n and N_n/n are small, but it is conceivable that equation (1.5) has some advantages in terms of second-order asymptotic properties. It can easily be checked that, for numbers x and y between 0 and 1,

$$\frac{x}{y} < \frac{\log(1-x)}{\log(1-y)} \qquad \text{if and only if } y < x$$

and from this it follows that $\hat{\theta}_n < \tilde{\theta}_n$ provided that $N_n < r_n Z_n$, which will always be true except in the circumstance that every block with at least one exceedance consists entirely of exceedances.

Our second method is based on the idea of runs of observations below or above the threshold defining clusters. More precisely, suppose that we take any sequence of r_n consecutive observations *below* the threshold as separating two clusters. An equivalent and in some ways more attractive characterization is the following. Define $W_{n,i}$ to be 1 if the ith observation is above the threshold (i.e. $X_i > u_n$) and 0 otherwise. Let

$$\begin{aligned} N_n &= \sum_{i=1}^{n} W_{n,i}, \\ Z_n^* &= \sum_{i=1}^{n-r_n} W_{n,i}(1 - W_{n,i+1}) \ \ldots \ (1 - W_{n,i+r_n}). \end{aligned} \tag{1.6}$$

Then

$$\bar{\theta}_n = Z_n^*/N_n. \tag{1.7}$$

We call this the *runs estimator*. The definition of Z_n^* in equation (1.6) ensures that an exceedance in position i is counted if and only if the following r_n observations are all below the threshold u_n, i.e. if it is the rightmost member of a cluster according to the runs definition.

In principle we could consider a fourth estimator, formed by modifying $\bar{\theta}_n$ in the same way as $\tilde{\theta}_n$ modified $\hat{\theta}_n$, though this does not seem so natural since $\bar{\theta}_n$ is directly motivated from the way in which $F^{r\theta}$ approximates F_r. We do not consider this further.

The estimators $\hat{\theta}_n$ and $\bar{\theta}_n$ have appeared previously; references include Leadbetter *et al.* (1989), Hsing (1990, 1991) and Nandogopalan (1990). However, until now there has not been any discussion of the optimal choice of the key parameter r_n. This is the main focus of the present paper. In a meteorological or oceanographic setting it has the interpretation of an average storm length, and in some cases it is possible to argue what this should be directly from physical considerations; see Tawn (1988a). Our own point of view is more theoretical. In Section 2, we outline some optimality arguments based on bias–variance considerations but leave open the question of how the relevant quantities can be calculated for a specific stochastic process. In Section 3 we make detailed calculations for a specific process and use them to construct a general formula, and in Section 4 we present heuristic arguments in support of a general formula (equation (4.1)). The paper concludes in Section 5 with some simulations and real data analysis.

2. BIAS AND VARIANCE CALCULATIONS

The main results of this paper depend on the asymptotic properties of the point process of exceedance times of a sequence of high thresholds u_n. We begin by quoting one of the main results of Hsing *et al.* (1988).

Suppose that $\{u_n,\ n \geqslant 1\}$ is a sequence of thresholds such that $F(u_n) \to 1$ as $n \to \infty$, and let P_n denote the exceedance point process for the level u_n by $X_1, \ldots, X_n$, i.e., for any $0 \leqslant a < b \leqslant 1$, $P_n[a, b]$ denotes the number of exceedances of the level u_n by X_i, $na \leqslant i \leqslant nb$. Suppose that $\{X_n\}$ is stationary and for $i < j$ let $\mathscr{B}_i^j$ denote the σ-field generated by the events $\{X_t \leqslant u_n\}$, $i \leqslant t \leqslant j$. For $n \geqslant 1$ and $1 \leqslant l \leqslant n - 1$ let

$$\alpha_{n,l} = \max\{|\Pr(A \cap B) - \Pr(A)\Pr(B)|: A \in \mathscr{B}_1^k,\ B \in \mathscr{B}_{k+l}^n,\ 1 \leqslant k \leqslant n - l\}.$$

The process is said to satisfy the condition $\Delta(u_n)$ if $\alpha_{n,l_n} \to 0$ as $n \to \infty$ for some sequence $\{l_n\}$ with $l_n/n \to 0$. This is a condition of 'mixing in the tails', similar to, but somewhat stronger than, the condition $D(u_n)$ used extensively in the book by Leadbetter *et al.* (1983). Suppose also that we define sequences $\{k_n\}$ and $\{r_n\}$ where $k_n \to \infty$, $k_n l_n/n \to 0$ and $k_n \alpha_{n,l_n} \to 0$, while r_n is the integer part of n/k_n. Define the cluster size distribution π_n by

$$\pi_n(j) = \Pr\left\{ \sum_{i=1}^{r_n} W_{n,i} = j \,\middle|\, \sum_{i=1}^{r_n} W_{n,i} > 0 \right\}, \qquad j \geqslant 1,$$

where $W_{n,i}$ is the indicator of the event $\{X_i > u_n\}$. Thus π_n is the distribution of cluster size when the definition of a cluster is based on a block of length r_n.

Under these conditions, a combination of corollary 3.3, theorem 4.1 and theorem 4.2 of Hsing *et al.* (1988) leads to the following conclusion: if $F_n(u_n) \to \exp(-\lambda)$ and $\pi_n(j) \to \pi(j)$ as $n \to \infty$, where $0 < \lambda < \infty$ and π is a probability distribution on the positive integers, then the exceedance point process P_n converges to a compound Poisson process whose 'cluster centres' form a Poisson process of intensity λ on [0, 1], each cluster consisting of a random number of points following the distribution π, independently for each cluster. Moreover, under additional summability conditions on the $\pi_n(j)$, we also have $1/\theta = \Sigma j \pi(j)$, confirming the notion that the extremal index is the reciprocal of the limiting mean cluster size.

Let us now assume that these asymptotic results are an adequate description of the exceedance process, and consider the effect on estimation of θ. Our arguments are somewhat heuristic, but a precise formulation of the asymptotic results along the lines of Hsing (1991) or Nandogopalan (1990) involves considerable technicalities which we prefer to avoid. Let Z_n denote the number of clusters and N_n the total number of exceedances, and suppose that $1/\theta$ and σ^2 are the mean and variance of the cluster size distribution. For this calculation, we do not yet distinguish between the 'blocks' and 'runs' definitions of clusters (so Z_n will be replaced by Z_n^* in the latter case). From the obvious relationships $E\{N_n|Z_n\} = Z_n/\theta$ and $\operatorname{var}\{N_n|Z_n\} = Z_n\sigma^2$ it follows at once that $E(N_n) = \lambda/\theta$ and the asymptotic covariance matrix of (Z_n, N_n) is

$$\begin{pmatrix} \lambda & \lambda/\theta \\ \lambda/\theta & \lambda(\theta^{-2} + \sigma^2) \end{pmatrix}. \tag{2.1}$$

In practice, we may replace λ by $n\theta\bar{F}(u_n)$ where $\bar{F} = 1 - F$. Let $\hat{\theta}_n = Z_n/N_n$. It follows by the delta method (Taylor expansion of $f(Z_n, N_n)$ about $f\{E(Z_n), E(N_n)\}$, with $f(x, y) = x/y$) that asymptotically we have

$$E(\hat{\theta}_n) \approx \theta,$$
$$\operatorname{var}(\hat{\theta}_n) \approx \frac{\sigma^2\theta^4}{\lambda} \approx \frac{\sigma^2\theta^3}{n\overline{F}(u_n)}. \tag{2.2}$$

This result is essentially due to Hsing (1991), who established precise sufficient conditions for expressions (2.1) and (2.2) and the corresponding asymptotic normality results to hold. The asymptotics are a little different from those in Hsing *et al.* (1988), since in that paper it was sufficient to assume that λ was a constant, whereas here we need to allow $\lambda \to \infty$ to obtain consistent estimates. Moreover, if we replace $\hat{\theta}_n$ by $\tilde{\theta}_n$ defined by equation (1.5), it is clear that the same result will hold, to the first order of approximation, provided that k_n and n are both large compared with λ.

Now, however, let us use the same method to obtain a first-order approximation to the bias of $\hat{\theta}_n$ and $\tilde{\theta}_n$. By expression (2.1) and the delta method we obtain

$$E\left(\frac{Z_n}{N_n}\right) \approx \frac{E(Z_n)}{E(N_n)}\left\{1 + \frac{\operatorname{var}(N_n)}{E(N_n)^2} - \frac{\operatorname{cov}(Z_n, N_n)}{E(N_n)\,E(Z_n)}\right\}$$
$$\approx \frac{E(Z_n)}{E(N_n)}\left(1 + \frac{\sigma^2\theta^2}{\lambda}\right), \tag{2.3}$$

correct to $O(1/\lambda)$, where we have not replaced $E(Z_n)$ and $E(N_n)$ by their asymptotic values because we will also want to use approximation (2.3) in the situation where there are alternative definitions of Z_n and N_n which may not have the same means.

An extension of the same argument shows that

$$E\left\{\frac{\log(1 - Z_n/k_n)}{r_n \log(1 - N_n/n)}\right\} \approx \frac{E(Z_n)}{E(N_n)}\left(1 + \frac{\lambda}{2k_n} - \frac{\lambda}{2n\theta} + \frac{\sigma^2\theta^2}{\lambda}\right). \tag{2.4}$$

Again, we shall replace λ by $n\theta\overline{F}(u_n)$ in subsequent discussion. The relative sizes of the correction terms in approximation (2.4) depend on the relative magnitudes of λ, k_n and n; since these are unknown at present, no further simplification of equation (2.4) is attempted.

So far, we have not distinguished between the blocks and runs estimators. Now we do so, writing Z_n^* in place of Z_n in the latter case.

For integer $i \geqslant 2$ and threshold u, define

$$\theta(i, u) = \Pr\{X_2 \leqslant u,\ X_3 \leqslant u,\ \ldots,\ X_i \leqslant u \mid X_1 > u\} \tag{2.5}$$

and let $\theta(1, u) = 1$. O'Brien (1987) characterized the extremal index in the form

$$\theta = \lim_{n \to \infty} \{\theta(i_n, u_n)\} \tag{2.6}$$

for suitable sequences $i_n \to \infty$ and u_n such that $F(u_n) \to 1$. Essentially, the restriction on i_n and u_n is that i_n must not grow too rapidly relative to u_n, and in most practical cases (O'Brien, 1974; Smith, 1992a) it is possible to replace equation (2.6) by

$$\theta = \lim_{i \to \infty} \lim_{u: F(u) \to 1} \{\theta(i, u)\}. \tag{2.7}$$

Note, however, that the order of limits in equation (2.7) cannot be interchanged, the limit as $i \to \infty$ being 0 for each fixed u for which $F(u) < 1$.

We always have $E(N_n) = n\bar{F}(u_n)$. For the blocks method, we have

$$\begin{aligned} E(Z_n) &= k_n \bar{F}_{r_n}(u_n) \\ &= k_n \sum_{i=1}^{r_n} \Pr\{X_i > u_n,\ X_{i+1} \leqslant u_n,\ \ldots,\ X_{r_n} \leqslant u_n\} \\ &= k_n \bar{F}(u_n) \sum_{i=1}^{r_n} \theta(i, u_n). \end{aligned}$$

Hence from approximations (2.3) and (2.4), replacing λ by $n\theta\bar{F}(u_n)$, we have

$$E(\hat{\theta}_n) - \theta \approx \frac{1}{r_n} \sum_{i=1}^{r_n} \{\theta(i, u_n) - \theta\} + \frac{\sigma^2\theta^2}{n\bar{F}(u_n)} \tag{2.8}$$

and

$$E(\tilde{\theta}_n) - \theta \approx \frac{1}{r_n} \sum_{i=1}^{r_n} \{\theta(i, u_n) - \theta\} + \frac{r_n\theta^2\bar{F}(u_n)}{2} - \frac{\theta\bar{F}(u_n)}{2} + \frac{\sigma^2\theta^2}{n\bar{F}(u_n)}. \tag{2.9}$$

For the runs method, it follows immediately from equations (1.6) that

$$\begin{aligned} E(Z_n^*) &= (n - r_n) \Pr\{X_i > u_n,\ X_{i+1} \leqslant u_n,\ \ldots,\ X_{i+r_n} \leqslant u_n\} \\ &= (n - r_n)\bar{F}(u_n)\,\theta(r_n + 1,\ u_n) \end{aligned}$$

and hence, using approximations (2.3) with Z_n^* replacing Z_n, that

$$E(\bar{\theta}_n) - \theta \approx \{\theta(r_n + 1,\ u_n) - \theta\} - \frac{\theta}{k_n} + \frac{\sigma^2\theta^2}{n\bar{F}(u_n)}. \tag{2.10}$$

The comparison of approximations (2.8) and (2.10) already gives us a concrete reason to prefer the runs estimator: provided that r_n does not grow too fast it is reasonable to suppose that the dominant term in the bias will be the first term, and that $\theta(r_n + 1, u_n)$ will be closer to θ than the average of $\theta(i, u_n)$, $1 \leqslant i \leqslant r_n$, so that the bias in $\bar{\theta}_n$ will be smaller than that in $\hat{\theta}_n$. More detailed calculations in the following sections bear this out.

3. EXAMPLE: DOUBLY STOCHASTIC MODEL

We now consider a specific model, a form of doubly stochastic process, for which it is possible to make some explicit computations and comparisons. These will motivate a general result developed in detail in Section 4.

Let $\{\xi_i,\ i \geqslant 1\}$ be IID with distribution function G, and suppose that $Y_1 = \xi_1$, and for $i > 1$,

$$Y_i = \begin{cases} Y_{i-1} & \text{with probability } \psi, \\ \xi_i & \text{with probability } 1 - \psi, \end{cases}$$

the choice being made independently for each i. The doubly stochastic sequence $\{X_i\}$ is then defined by

$$X_i = \begin{cases} Y_i & \text{with probability } \eta, \\ 0 & \text{with probability } 1-\eta, \end{cases}$$

again independently of everything else. The idea behind this definition is that sequences of successive Ys define clusters, but the modification from Y_i to X_i means that high level exceedances within the same cluster are not necessarily consecutive.

Let $M_n = \max\{X_1, \ldots, X_n\}$ and define

$$\begin{aligned} P_n &= \Pr\{M_n \leqslant u \mid Y_1 \leqslant u\}, \\ Q_n &= \Pr\{M_n \leqslant u \mid Y_1 > u\} \end{aligned} \tag{3.1}$$

assuming that $u > 0$. Then $P_0 = Q_0 = 1$, $P_1 = 1$ and $Q_1 = 1 - \eta$. If $\epsilon = 1 - G(u) > 0$ we have

$$\begin{aligned} \Pr\{Y_2 \leqslant u \mid Y_1 \leqslant u\} &= 1 - \epsilon + \psi\epsilon, \\ \Pr\{Y_2 \leqslant u \mid Y_1 > u\} &= (1-\psi)(1-\epsilon) \end{aligned}$$

and hence for $n \geqslant 1$

$$\begin{aligned} P_n &= (1-\epsilon+\psi\epsilon)P_{n-1} + (1-\psi)\epsilon Q_{n-1}, \\ Q_n &= (1-\eta)\{(1-\psi)(1-\epsilon)P_{n-1} + (\psi+\epsilon-\psi\epsilon)Q_{n-1}\}. \end{aligned}$$

These difference equations have the solution

$$\begin{aligned} P_n &= C_1\lambda_1^n + C_2\lambda_2^n, \\ Q_n &= D_1\lambda_1^n + D_2\lambda_2^n \end{aligned} \tag{3.2}$$

where C_1, C_2, D_1 and D_2 are determined by the initial conditions to be

$$\left.\begin{aligned} C_1 &= (1-\lambda_2)/(\lambda_1-\lambda_2), \\ C_2 &= (\lambda_1-1)/(\lambda_1-\lambda_2), \\ D_1 &= (1-\lambda_2-\eta)/(\lambda_1-\lambda_2), \\ D_2 &= (\lambda_1+\eta-1)/(\lambda_1-\lambda_2), \end{aligned}\right\} \tag{3.3}$$

and λ_1 and λ_2 are the roots of the quadratic equation

$$\lambda^2 - (1+\psi-\psi\eta-\epsilon\eta+\epsilon\psi\eta)\lambda + (1-\eta)\psi = 0. \tag{3.4}$$

If λ_1 is the larger of the two roots then it is readily verified that, as $\epsilon \to 0$, $\lambda_1 = 1 - \epsilon\eta(1-\psi)/(1-\psi+\psi\eta) + O(\epsilon^2)$ while $\lambda_2 \to \beta$ where we define $\beta = \psi(1-\eta)$. Hence if $n \to \infty$, $\epsilon \to 0$ with $n\epsilon\eta \to \tau \in (0, \infty)$, we have $\Pr\{M_n \leqslant u\} = (1-\epsilon)P_n + \epsilon Q_n \sim P_n \sim \lambda_1^n \to \exp\{-\tau(1-\psi)/(1-\psi+\psi\eta)\}$ which confirms that the extremal index of $\{X_i\}$ is

$$\theta = \frac{1-\psi}{1-\psi+\psi\eta}. \tag{3.5}$$

Now consider

$$\theta(i,u) = (1-\psi)(1-\epsilon)P_{i-1} + (\psi+\epsilon-\psi\epsilon)Q_{i-1}.$$

The relevant asymptotics here are for $\epsilon \to 0$, $i \to \infty$ but $i\epsilon \to 0$. In particular, if we retain terms of $O(i\epsilon)$ and $O(\beta^i)$ and discard everything of smaller order, we find that

$$\theta(i, u) - \theta \approx -i\epsilon\theta^2\eta + (1-\theta)\beta^{i-1}. \tag{3.6}$$

Now let us consider some consequences of these calculations for the bias and variance approximations of Section 2. The cluster sizes have a geometric distribution with mean $1/\theta$ and hence variance $\sigma^2 = (1-\theta)/\theta^2$, so by approximations (2.2)

$$\operatorname{var}(\hat{\theta}_n) \approx \theta(1-\theta)/n\epsilon\eta, \tag{3.7}$$

which does not depend directly on r_n, so the initial choice of r_n can be made solely to minimize bias. Moreover, to the first order of accuracy approximation (3.7) holds equally well for $\tilde{\theta}_n$ and $\bar{\theta}_n$.

Now consider the bias, initially for $\bar{\theta}_n$. For our first calculation we base this entirely on the leading term of approximation (2.10). Note that u_n is related to ϵ by the formula $\bar{F}(u_n) = \epsilon\eta$, so to make the dependence on n explicit we write ϵ_n in place of ϵ. By approximation (3.6), we want to choose r_n to minimize $|-r_n\epsilon_n\theta^2\eta + (1-\theta)\beta^{r_n-1}|$. If we set

$$r_n = \frac{\log(1/\epsilon_n) + \log\{\log(1/\epsilon_n)\} + C_n}{\log\beta}$$

for bounded C_n, then both terms are of $O\{\epsilon_n \log(1/\epsilon_n)\}$. Because the two terms are of opposite sign, it might be possible to make the bias vanish completely by choosing C_n appropriately, but in general the requirement that r_n be an integer defeats this. This also makes it impossible to determine the exact constant of multiplicity.

Suppose that we are also free to choose u_n or equivalently ϵ_n to minimize the mean-squared error. The squared bias is of $O\{\epsilon_n^2 \log^2(1/\epsilon_n)\}$ and the variance is of $O(1/n\epsilon_n)$. For these to be the same order of magnitude we require $\epsilon_n = O\{n^{-1/3}(\log n)^{-2/3}\}$ leading to a mean-squared error of $O\{n^{-2/3}(\log n)^{2/3}\}$. It can also be checked that, for these choices of ϵ_n and $r_n = O\{\log(1/\epsilon_n)\} = O(\log n)$, the second and third terms in approximation (2.10) are of smaller order than the first, so justifying their neglect in these calculations.

It is a somewhat different matter, however, if we use the blocks estimator $\hat{\theta}_n$. In this case, if we use approximation (3.6) to evaluate the leading term in approximation (2.8), we find that the bias contains terms of $O(\epsilon_n r_n)$ and of $O(1/r_n)$, which are minimized by taking $r_n = O(\epsilon_n^{-1/2})$ with a bias of $O(\epsilon_n^{1/2})$. Then choosing ϵ_n to minimize the mean-squared error leads to an optimal ϵ_n of $O(n^{-1/2})$ and a resulting mean-squared error of $O(n^{-1/2})$. Calculations for $\tilde{\theta}_n$, starting from approximation (2.9), lead to bias and mean-squared error of the same order of magnitude, though with possibly different constants of proportionality.

Thus these calculations suggest that, if both the threshold and the r_n-parameter are chosen optimally, the runs estimator $\bar{\theta}_n$ is asymptotically better than either $\hat{\theta}_n$ or $\tilde{\theta}_n$, the comparison between the last two remaining an open question. This is a very limited conclusion because it is based only on one very artificial stochastic process, but it leads to a natural conjecture that the same qualitative result may hold much more generally. The remainder of the paper provides some justification for this point of view.

4. IMPLICATIONS FOR MORE GENERAL PROCESSES

The key formula of Section 3 was equation (3.6). This was used to examine in detail the bias–variance trade-off of various estimators. It is therefore of interest to consider whether such a formula might hold more generally.

In fact a mixture of heuristic reasoning and simulation results has suggested that a variant on this result should be valid for a wide class of processes. The claimed approximation is given by

$$\theta(i,u) - \theta \approx \theta\{F^{i\theta}(u) - 1\} + C\beta^i. \tag{4.1}$$

for some constants C (depending on u) and $\beta \in (0,1)$.

The heuristic reasoning suggesting approximation (4.1) is as follows. For small i, the main issue should be how quickly the process 'forgets' that it started in an extreme state, namely one for which $X_1 > u$. That in turn is related to how quickly the process starting from a point $x > u$ returns to its stationary distribution. For a wide class of processes known as geometrically ergodic, this is of order β^i for some $\beta < 1$. The constant β is independent of the starting state though the constant of proportionality will not be in general (Nummelin, 1984). Thus, for small i, we would expect the left-hand side of approximation (4.1) to be of order β^i.

For large i, the issue is how long it takes the process to return to the region $\{X_i > u\}$ after returning to its stationary distribution. We would expect the approximation $\theta(i,u) \approx \theta F^{i\theta}(u)$ to hold, where the first factor θ represents the probability of no short-term return to the region (u,∞) and the second factor $F^{i\theta}(u)$ is the probability that there is no return in the long term either. The claimed approximation (4.1) is of a form that allows both the 'small i' and 'large i' approximations to be valid.

For a geometrically ergodic Harris Markov chain, these concepts can be made precise by breaking up the chain into separate regeneration cycles. This is being developed elsewhere (Smith, 1992b).

In the case that $i\bar{F}(u) = O(1)$, we may also write $\theta\{F^{i\theta}(u) - 1\} \approx -i\theta^2\bar{F}(u)$ and with that modification approximation (4.1) is of the same structure as approximation (3.6). However, we believe that approximation (4.1) will work better than approximation (3.6) in most cases. It is easy to modify the arguments of Section 3 to show that approximation (3.6) may equally well be written in the form of approximation (4.1) for that example.

Although formula (4.1) does appear to be valid in some generality, it may not be appropriate for all processes. For example Hsing (1990) did a detailed study for the case of m-dependent processes. There is no obvious way to see that his calculations may be represented in the form of approximation (4.1).

As an example of the sort of process for which it seems to be worth examining these approximations in detail, consider the Markov chain in which the joint distribution of X_i and X_{i+1} is given by

$$\Pr\{X_i \leqslant x,\ X_{i+1} \leqslant y\} = \exp[-\{\exp(-Rx) + \exp(-Ry)\}^{1/R}], \qquad R \geqslant 1. \tag{4.2}$$

A theoretical calculation of the extremal index for this model was given by Smith (1992a). In the next section, we shall use this as a specific example to illustrate approximation (4.1) and of the associated estimators of the extremal index.

5. SIMULATIONS AND DATA ANALYSIS

Our first simulations are aimed at evaluating $\theta(i, u)$ based on approximately 10^5 replications. First, the doubly stochastic model of Section 3 was simulated. Numerous combinations of η and ψ were simulated, but we show only one of these, $\eta = 0.7$ and $\psi = 0.9$, which was chosen because in this case the two bias terms roughly balance leading to a range of positive and negative biases. Three thresholds were chosen corresponding to $\bar{F}(u) = 0.03$, 0.018 and 0.006 and the bias $\theta(i, u) - \theta$ plotted for different i in Fig. 1. Here and in all subsequent plots, the three thresholds are represented by dots, circles and + signs in that order, the + signs being for the highest threshold. It is natural to examine how well approximation (4.1) fits, and this was calculated from the data by a non-linear least squares fit of the parameters θ, C and β, fitted separately for each threshold (i.e. each data point from Fig. 1 was treated as an independent observation with common variance). The results were

$$\bar{F}(u) = 0.030: \quad \hat{\theta} = 0.128,\ \hat{C} = 3.22,\ \hat{\beta} = 0.270;$$

$$\bar{F}(u) = 0.018: \quad \hat{\theta} = 0.133,\ \hat{C} = 3.34,\ \hat{\beta} = 0.264;$$

$$\bar{F}(u) = 0.006: \quad \hat{\theta} = 0.141,\ \hat{C} = 3.35,\ \hat{\beta} = 0.263.$$

In this case the exact values computed by Section 3 are $\theta = 0.1370$, $C = 3.196$ and $\beta = 0.27$. It can be seen that the estimated values agree well with these. In Fig. 1 the three fitted curves are superimposed on the plot, which seems to confirm a good fit.

Next, we repeated the same exercise by using the Markov chain defined by equation (4.2) with $R = 2$. In this case $\theta \approx 0.328$ is known from Smith (1992a), but we know nothing about C and β.

Fig. 2 shows a plot of the simulated $\theta(i, u)$ together with fitted curves of the form (4.1). The three thresholds correspond to $\bar{F}(u) = 0.01$, 0.006 and 0.002. The fitted curves are

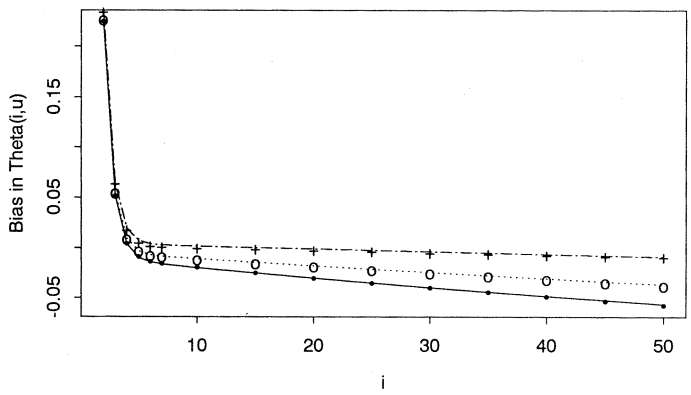

Fig. 1. Simulated values of $\theta(i, u) - \theta$ for the doubly stochastic model of Section 3, with $\eta = 0.7$ and $\psi = 0.9$ (•, ○, +, simulated values for three thresholds for which $\bar{F}(u)$ are respectively 0.03, 0.018 and 0.006): the curves are based on equation (4.1) fitted to the data separately for each threshold; it can be seen that the curves fit well through the data points

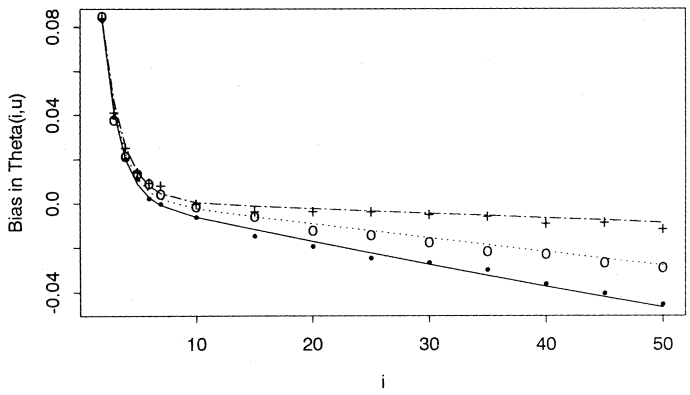

Fig. 2. Same as Fig. 1, but the simulated process is now a first-order Markov chain with bivariate distributions given by equation (4.2) with $R=2$: the three thresholds correspond to $\bar{F}(u)=0.01$ (•), $\bar{F}(u)=0.006$ (○) and $\bar{F}(u)=0.002$ (+), and the curves are again based on fitting approximation (4.1)

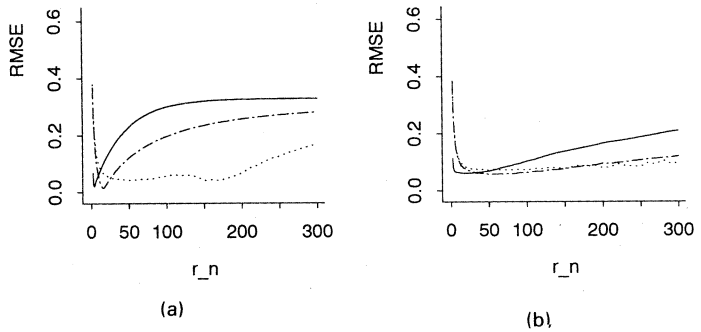

Fig. 3. Simulations of the square root of the mean-squared error for the three estimators $\hat{\theta}_n$ (–·–·–), $\tilde{\theta}_n$ (·····) and $\bar{\theta}_n$ (———), for $n=10000$ and r_n up to 300: both curves are based on 100 simulations of model (4.2) with $R=2$, with (a) $\bar{F}(u)=0.07$ and (b) $\bar{F}(u)=0.01$

$$\bar{F}(u) = 0.010: \quad \hat{\theta} = 0.332,\ \hat{C} = 0.319,\ \hat{\beta} = 0.501;$$
$$\bar{F}(u) = 0.006: \quad \hat{\theta} = 0.332,\ \hat{C} = 0.307,\ \hat{\beta} = 0.514;$$
$$\bar{F}(u) = 0.002: \quad \hat{\theta} = 0.330,\ \hat{C} = 0.270,\ \hat{\beta} = 0.550.$$

In this case all three estimated values of θ were very close to the true values, whereas the estimated C and β are reasonably close to each other over different thresholds. Again, the plot seems to confirm that approximation (4.1) is the correct functional shape.

A third set of simulations, in Fig. 3, is directly concerned with the three estimators when the true model is equation (4.2) with $R=2$. In each case the runs estimator $\bar{\theta}_n$ achieves its minimum mean-squared error at the smallest value of r_n, but after that the conclusions are not so clear cut—the fact that the other two estimators have smaller mean-squared error over a wide range of values of r_n may yet provide some grounds for preferring these estimators.

We now consider a real data example based on part of a large study that concerned heights of sea-waves sampled at intervals of 3 hours. It turns out that tidal effects are not important, but both common sense and previous studies such as Tawn (1988a) suggest that a 'storm length' in the region of 1–2 days ($r_n = 8$–16) should be appropriate. This series was picked out from a large number of similar series as one which caused particular difficulty in interpreting the estimates of the extremal index.

The full series was of length 29220, and three thresholds u_1, u_2 and u_3 were chosen over which there were respectively 2816, 1170 and 463 exceedances. Thus we estimate $\bar{F}(u_1) = 2816/29220 = 0.096$ and similarly $\bar{F}(u_2) = 0.040$, $\bar{F}(u_3) = 0.016$. The runs estimator $\bar{\theta}_n$ is calculated for each threshold and for a variety of values of r_n with results which are displayed in Fig. 4. It can be seen that the extremal index is difficult to specify, estimates from 0.044 to 0.194 being obtained. The standard errors of the estimates, obtained from equation (2.2), are respectively about 0.006, 0.01 and 0.02 for the three thresholds, and virtually independent of r_n. This implies that the differences among the calculated values are not merely the result of sampling variation.

For this example the non-linear least squares procedure based on approximation (4.1) led to the following results:

$$\bar{F}(u) = 0.096: \quad \hat{\theta} = 0.048,\ \hat{C} = 0.084,\ \hat{\beta} = 0.875;$$

$$\bar{F}(u) = 0.040: \quad \hat{\theta} = 0.091,\ \hat{C} = 0.087,\ \hat{\beta} = 0.884;$$

$$\bar{F}(u) = 0.016: \quad \hat{\theta} = 0.135,\ \hat{C} = 0.065,\ \hat{\beta} = 0.881.$$

From these results it can be seen that the estimates of β and C are quite consistent across thresholds, but not those of θ. This is rather difficult to interpret in the light of most of our results in this paper, but a possible interpretation is that we should be considering a different extremal index for each threshold. This idea is not new,

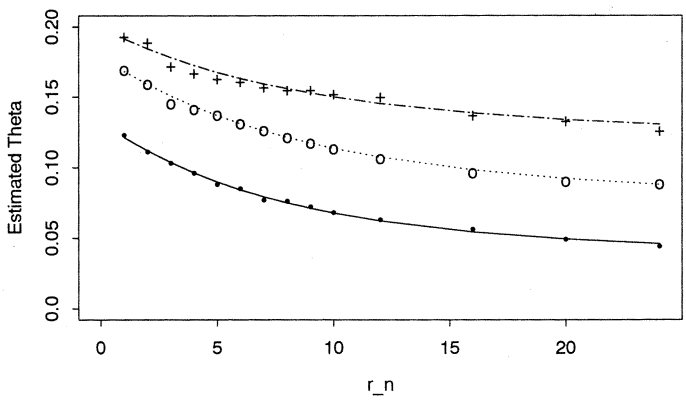

Fig. 4. Real data example based on wave heights: the three thresholds correspond to empirical $\bar{F}(u) = 0.096$ (•), $\bar{F}(u) = 0.04$ (○) and $\bar{F}(u) = 0.016$ (+), and the curves are again based on fitting approximation (4.1); again the curves fit well individually, but there is a very clear separation among the three thresholds, which unlike Figs 1 and 2 cannot be explained as the effects of bias; thus we conclude that the difference between the curves is real and represents the fact that different extremal indices are effective at different levels of the process

since Tawn (1990) has argued for it as a practical technique. Smith (1992b) has suggested also that there are theoretical reasons for considering this in connection with approximation (4.1).

Thus we conclude, for this example, that it is necessary to adopt a threshold-dependent extremal index, but that if we do so then the fit of model (4.1) is quite good and that it indeed helps to clarify what r_n we should use. An interesting side comment is that, for each of the three thresholds, the final estimate of θ is very close to $\bar{\theta}_n$ with $r_n = 20$, which is consistent with a storm length of about 2.5 days, so this part of the result has a nice physical intepretation.

A referee has pointed out that there could be difficulties in using a threshold-dependent extremal index to estimate quantiles of extreme value distributions. Thus, in solving for a quantile q_p for given tail probability p via the approximation

$$\Pr\{\max(X_1, \ldots, X_n) \leqslant q_p\} \approx F^{n\theta(q_p)}(q_p) = 1 - p,$$

we are implicitly constrained to values of q_p for which an estimate of $\theta(q_p)$ exists. In particular, q_p must lie within the range of the data. This difficulty could be avoided, at the cost of further model assumptions, by postulating some functional form for $\theta(u)$.

We may summarize the main results of the paper as follows.

(a) The theoretical analysis provides a firm basis for preferring the runs estimator over the blocks estimator and gives some asymptotic results about the optimum rate of convergence.

(b) A mixture of detailed calculations for the doubly stochastic model, heuristic arguments for more general stochastic processes and simulation results, all support model (4.1) as a basis for deriving the bias terms in approximations (2.8)–(2.10).

(c) It is possible to estimate the three parameters in approximation (4.1), and thus to obtain an estimate of θ which does not depend too sensitively on a particular value of r_n, but it may be necessary to consider a threshold-dependent extremal index.

ACKNOWLEDGEMENTS

This work has been supported by the Wolfson Foundation and the UK Science and Engineering Research Council, and by the Fund for the Promotion of Research at the Technion. The authors would like to thank Uri Cohen for assistance with the programming and graphics. They are also grateful to the referees for many helpful comments.

REFERENCES

Cartwright, D. E. (1958) On estimating the mean energy of sea waves from the highest waves in a record. *Proc. R. Soc. Lond.* A, **247**, 22–48.

Davison, A. C. and Smith, R. L. (1990) Models for exceedances over high thresholds (with discussion). *J. R. Statist. Soc.* B, **52**, 393–442.

Dekkers, A. L. M., Einmahl, J. H. J. and de Haan, L. (1989) A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, **17**, 1833–1855.

Dekkers, A. L. M. and de Haan, L. (1989) On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.*, **17**, 1795–1832.

Galambos, J. (1987) *The Asymptotic Theory of Extreme Order Statistics*, 2nd edn. Melbourne: Krieger.

Hsing, T. (1990) Estimating the extremal index under *M*-dependence and tail balancing. *Preprint*. Department of Statistics, Texas A&M University, College Station.

——— (1991) Estimating the parameters of rare events. *Stoch. Processes Appl.*, **37**, 117–139.

Hsing, T., Hüsler, J. and Leadbetter, M. R. (1988) On the exceedance point process for a stationary sequence. *Probab. Theory Reltd Flds*, **78**, 97–112.

Hüsler, J. (1986) Extreme values of non-stationary random sequences. *J. Appl. Probab.*, **23**, 937–950.

Leadbetter, M. R. (1983) Extremes and local dependence in stationary sequences. *Z. Wahrsch. Ver. Geb.*, **65**, 291–306.

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

Leadbetter, M. R. and Rootzén, H. (1988) Extremal theory for stochastic processes. *Ann. Probab.*, **16**, 431–478.

Leadbetter, M. R., Weissman, I., de Haan, L. and Rootzén, H. (1989) On clustering of high levels in statistically stationary series. In *Proc. 4th Int. Meet. Statistical Climatology* (ed. J. Sansom). Wellington: New Zealand Meteorological Service.

Loynes, R. M. (1965) Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Statist.*, **36**, 993–999.

Nandogopalan, S. (1990) Multivariate extremes and the estimation of the extremal index. *PhD Dissertation*. Department of Statistics, University of North Carolina, Chapel Hill.

Newell, G. F. (1964) Asymptotic extremes for *m*-dependent random variables. *Ann. Math. Statist.*, **35**, 1322–1325.

Nummelin, E. (1984) *General Irreducible Markov Chains and Non-negative Operators*, p. 120. Cambridge: Cambridge University Press.

O'Brien, G. L. (1974) The maximum term of uniformly mixing stationary processes. *Z. Wahrsch. Ver. Geb.*, **30**, 57–63.

——— (1987) Extreme values for stationary and Markov sequences. *Ann. Probab.*, **15**, 281–291.

Smith, R. L. (1987) Estimating tails of probability distributions. *Ann. Statist.*, **15**, 1174–1207.

——— (1989) Extreme value analysis of environmental time series: an example based on ozone data (with discussion). *Statist. Sci.*, **4**, 367–393.

——— (1992a) The extremal index for a Markov chain. *J. Appl. Probab.*, **29**, 37–45.

——— (1992b) On an approximation formula for the extremal index in a Markov chain. *Technical Report*. Department of Statistics, University of North Carolina, Chapel Hill.

Tawn, J. A. (1988a) An extreme value theory model for dependent observations. *J. Hydr.*, **101**, 227–250.

——— (1988b) Bivariate extreme value theory—models and estimation. *Biometrika*, **75**, 397–415.

——— (1990) Discussion on Models for exceedances over high thresholds (by A. C. Davison and R. L. Smith). *J. R. Statist. Soc.* B, **52**, 428–429.