# ESTIMATING THE EXTREMAL INDEX

by

**RICHARD L. SMITH**

**University of North Carolina, USA**

and

**ISHAY WEISSMAN**

**Technion, Haifa, Israel**

*Contact Address:* Richard L. Smith, Department of Statistics, University of North Carolina, Chapel Hill, N.C. 27599-3260, U.S.A.

*November 4 1991*

*Abstract*

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 between 0 and 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 presented for some simple stochastic processes, and the practical implications are examined through simulations and some real data analysis.

## 1. BACKGROUND

Suppose we have $n$ observations from a stationary process $X_i,\ i \geq 0$ with marginal distribution function $F$. For large $n$ and $u_n$, it is typically the case that

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

where $\theta \in [0, 1]$ is a constant for the process known as the *extremal index*. This concept originated in papers by Newell (1964), Loynes (1965) and O'Brien (1974), and was placed on a firm footing by Leadbetter (1983). It is explained further in Section 3.7 of the book by Leadbetter *et al.* (1983), and the review paper Leadbetter and Rootzén (1988) contains much more information. The classical theory of extremes in independent, identically distributed sequences (Galambos 1987, Leadbetter *et al.* 1983) provides conditions for the existence of normalising constants $a_n > 0$ and $b_n$ such that

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

where $H_n(x)$ 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 $\Lambda(x) = \exp(-e^{-x}),\ -\infty < x < \infty$. Combining (1.1) and (1.2), it is seen that for a stationary sequence $\{X_n\}$, we have

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

Thus it appears that $\theta$ is the key parameter for extending extreme value theory for i.i.d. random variables to dependent stochastic processes. From a *statistical* point of view, since there is already an extensive literature on estimating extreme value distributions from i.i.d. data, it can be seen that estimating the extremal index is a key problem.

An alternative characterisation of the extremal index (Hsing *et al.* 1988) is that $1/\theta$ is the mean cluster size in the point process of exceedance times over a high threshold. This suggest 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.

Let us assume that a high threshold $u_n$ has already been chosen. The choice of $u_n$ for the estimation of the tail of $F$ has been extensively studied from both a theoretical and practical point of view; references include Smith (1987, 1989), Davison and Smith (1990), Dekkers and de Haan (1989), Dekkers, Einmahl and de Haan (1989). There is no theoretical reason why the threshold for estimating $\theta$ has to be the same as the threshold for estimating the tail of $F$, and it is possible to consider procedures in which they are different. Initially, however, we treat the threshold as fixed and concentrate on the problem of identifying clusters. Later discussion will include optimal choice of the threshold as well.

The discussion throughout is confined to stationary sequences. Although nonstationary 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 nonstationarity 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. In the following discussion we shall assume that such preliminary analysis has already been carried out.

Suppose, then, 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 $\theta$ 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 (1.1), $F_r$ denotes the distribution function of $\max(X_1, ..., X_r)$. 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 get the estimator

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

Recalling that $n \approx k_n r_n$, it can be seen that (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 (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,

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

and from this it follows that $\hat{\theta}_n < \tilde{\theta}_n$ provided $N_n < r_n Z_n$, which will always be the case 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 we take any sequence of $r_n$ consecutive observations *below* the threshold as separating two clusters. An equivalent and in some ways more attractive characterisation is the following. Define $W_{n,i}$ to be 1 if the $i$'th observation is above the threshold (i.e. if $X_i > u_n$) and 0 otherwise. Let

$$N_n = \sum_{i=1}^{n} W_{n,i}, \quad Z_n^* = \sum_{i=1}^{n} W_{n,i}(1 - W_{n,i+1})...(1 - W_{n,i+r_n}). \tag{1.6}$$

Then

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

We call this the *runs estimator*. The definition of $Z_n^*$ in (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 $u_n$, in other words, if it is the rightmost member of a cluster according to the runs definition.

The theoretical properties of the blocks estimator (1.4) have been studied by Hsing (1990, 1991). It is a natural starting point for any statistical study because probabilistic analyses of the extremal index (Leadbetter 1983, Hsing *et al.* 1988) used the same definition. However, the runs estimator seems more natural from a statistical point of view. For example, Smith (1989) used precisely this method of separating clusters, calling $r_n$ the *cluster interval.* Nandagopalan (1990) and Leadbetter *et al.* (1989) have used a version of the runs estimator with $r_n = 1$ and Nandagopalan (1990) derived theoretical properties for it, but in general there seems no reason to confine ourselves to $r_n = 1$ and a more general estimator is desirable. Indeed a major theme of the present paper is the optimal choice of $r_n$ from bias and variance considerations.

## 2. BIAS AND VARIANCE CALCULATIONS

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

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

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

The process is said to satisfy the condition $\Delta(u_n)$ if $\alpha_{n,\ell_n} \to 0$ as $n \to \infty$ for some sequence $\{\ell_n\}$ with $\ell_n/n \to 0$. 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 we define sequences $\{k_n\}$ and $\{r_n\}$ where $k_n \to \infty$, $k_n\ell_n/n \to 0$ and $k_n\alpha_{n,\ell_n} \to 0$, while $r_n$ is the integer part of $n/k_n$. Define the cluster size distribution $\pi_n$ by

$$\pi_n(j) = \Pr\{\sum_{i=1}^{r_n} W_{n,i} = j | \sum_{i=1}^{r_n} W_{n,i} > 0\}, \quad j \geq 1$$

where $W_{n,i}$ is the indicator of the event $\{X_i > u_n\}$. Thus $\pi_n$ 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 $F_n(u_n) \to e^{-\lambda}$ and $\pi_n(j) \to \pi(j)$ as $n \to \infty$, where $0 < \lambda < \infty$ and $\pi$ is a probability distribution on the positive integers, then the exceedance point process $N_n$ converges to a compound Poisson process consisting of "cluster centres" forming a Poisson process of intensity $\lambda$ on $[0,1]$, and each cluster consisting of a random number of points which follows the distribution $\pi$, independently for each cluster. Moreover, under additional summability conditions on the $\pi_n(j)$'s, we also have $1/\theta = \sum j\pi_j$, confirming the notion that the extremal index is a reciprocal of 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 $\theta$. Our arguments are somewhat heuristic, but 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 $\mu$ and $\sigma^2$ are the mean and variance of the cluster size distribution. For the purpose of this calculation, we do not yet distinguish between the "blocks" and "runs" definitions of clusters. Also, $\theta = 1/\mu$. From the obvious relations $\mathrm{E}\{N_n|Z_n\} = Z_n\mu$, $\mathrm{var}\{N_n|Z_n\} = Z_n\sigma^2$ it follows at once that $\mathrm{E}N_n = \lambda\mu$ and the asymptotic covariance matrix of $(Z_n, N_n)$ is

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

In practice, we may replace $\lambda$ 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 $f(\mathrm{E}Z_n, \mathrm{E}N_n)$, with $f(x,y) = x/y$) that asymptotically we have

$$\mathrm{E}\,\hat{\theta}_n \approx \theta, \quad \mathrm{Var}\,\hat{\theta}_n \approx \frac{\sigma^2}{\lambda\mu^4} \approx \frac{\sigma^2}{n\bar{F}(u_n)\mu^3}. \tag{2.2}$$

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

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 (2.1) and the delta method we obtain

$$\begin{aligned} \mathrm{E}\left\{\frac{Z_n}{N_n}\right\} &\approx \frac{\mathrm{E}(Z_n)}{\mathrm{E}(N_n)}\left\{1 + \frac{\mathrm{Var}(N_n)}{(\mathrm{E}N_n)^2} - \frac{\mathrm{Cov}(Z_n, N_n)}{(\mathrm{E}N_n)(\mathrm{E}Z_n)}\right\} \\ &\approx \frac{\mathrm{E}(Z_n)}{\mathrm{E}(N_n)}\left(1 + \frac{\sigma^2}{\lambda\mu^2}\right), \end{aligned} \tag{2.3}$$

correct to $O(1/\lambda)$, where we have not replaced $\mathrm{E}Z_n$ and $\mathrm{E}N_n$ by their asymptotic values because we will also want to use (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

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

Again, we will replace $\lambda$ by $n\theta\bar{F}(u_n)$ in subsequent discussion. The relative sizes of the correction terms in (2.4) depend on the relative magnitudes of $\lambda$, $k_n$ and $n$; since these are unknown at present, no further simplification of (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 $i \geq 2$ and threshold $u$, define

$$\theta(i,u) = \Pr\{X_2 \leq u, X_3 \leq u, ..., X_i \leq u | X_1 > u\} \tag{2.5}$$

O'Brien (1987) characterised 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 1992) it is possible to replace (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 (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 $\mathrm{E}(N_n) = n\bar{F}(u_n)$. For the blocks method, we have

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

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

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

and

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

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

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

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

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

The comparison of (2.8) and (2.10) already gives us a concrete reason to prefer the runs estimator: provided $r_n$ does not grow too fast it is reasonable to suppose that $\theta(r_n+1, u_n)$ will be closer to $\theta$ than the average of $\theta(i, u_n)$, $1 \le i \le r_n$, 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: A 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 serve to motivate some conjectures about the general case.

Let $\{\xi_i,\ i \ge 1\}$ be i.i.d. with distribution function $G$, and suppose $Y_1 = \xi_1$, 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 $i$. Furthermore, let

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

again independently of everything else. In words, the $\{Y_i\}$ process consists of runs of observations which remain constant for a geometrically distributed length of time, while the $\{X_i\}$ process is a distorted version of $\{Y_i\}$ in which runs of high-level values are interrupted by randomly and independently distributed 0's. We assume $G(0) < 1$.

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

$$P_n = \Pr\{M_n \le u | Y_1 \le u\}, \quad Q_n = \Pr\{M_n \le u | Y_1 > u\} \tag{3.1}$$

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

$$\Pr\{Y_2 \le u | Y_1 \le u\} = 1 - \epsilon + \psi\epsilon, \quad \Pr\{Y_2 \le u | Y_1 > u\} = (1 - \psi)(1 - \epsilon)$$

and hence for $n \ge 1$

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

Defining

$$R_n = \begin{pmatrix} P_n \\ Q_n \end{pmatrix},$$

we have

$$R_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad R_n = AR_{n-1} \text{ where } A = \begin{pmatrix} (1 - \epsilon + \psi\epsilon) & (1 - \psi)\epsilon \\ (1 - \eta)(1 - \psi)(1 - \epsilon) & (1 - \eta)(\psi + \epsilon - \psi\epsilon) \end{pmatrix}. \tag{3.2}$$

These equations have the solution

$$P_n = C_1\lambda_1^n + C_2\lambda_2^n, \quad Q_n = D_1\lambda_1^n + D_2\lambda_2^n \tag{3.3}$$

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

$$C_1 = \frac{1 - \lambda_2}{\lambda_1 - \lambda_2}, \quad C_2 = \frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}, \quad D_1 = \frac{1 - \lambda_2 - \eta}{\lambda_1 - \lambda_2}, \quad D_2 = \frac{\lambda_1 + \eta - 1}{\lambda_1 - \lambda_2} \tag{3.4}$$

and $\lambda_1$, $\lambda_2$ are the eigenvalues of $A$, obtained by solving the quadratic equation

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

which reduces to

$$\lambda^2 - \lambda(1 + \psi - \psi\eta - \epsilon\eta + \psi\epsilon\eta) + (1 - \eta)\psi = 0$$

with roots

$$\lambda = \frac{1 + \psi - \psi\eta - \epsilon\eta(1 - \psi) \pm [\{1 + \psi - \psi\eta - \epsilon\eta(1 - \psi)\}^2 - 4\psi(1 - \eta)]^{1/2}}{2}. \tag{3.5}$$

As $\epsilon \to 0$ we have

$$\begin{aligned} \lambda_1 &= 1 - \frac{\epsilon\eta(1 - \psi)}{1 - \psi + \psi\eta} - \frac{\epsilon^2\psi\eta^2(1 - \psi)^2(1 - \eta)}{(1 - \psi + \psi\eta)^3} + O(\epsilon^3), \\ \lambda_2 &= \psi(1 - \eta) + \frac{\epsilon\eta\psi(1 - \eta)(1 - \psi)}{1 - \psi + \psi\eta} + \frac{\epsilon^2\psi\eta^2(1 - \psi)^2(1 - \eta)}{(1 - \psi + \psi\eta)^3} + O(\epsilon^3). \end{aligned} \tag{3.6}$$

We apply this first to the distribution of $M_n$. Since $\Pr\{Y_1 \le u\} = 1 - \epsilon$ we have from (3.1)

$$F_n(u) = \Pr\{M_n \le u\} = (1 - \epsilon)P_n + \epsilon Q_n$$

which can be determined exactly from (3.3)-(3.5). Now suppose $n \to \infty$, $\epsilon \to 0$ so that $n \Pr\{X_1 > u\} = n\epsilon\eta \to \tau$ where $\tau \in (0, \infty)$ is fixed. If we expand based on (3.6), ignoring the terms involving $\lambda_2^n$ because these are geometrically small compared with those involving $\lambda_1^n$, we find that

$$\begin{aligned} \Pr\{M_n \le u\} = &\exp\{-\tau(1 - \psi)/(1 - \psi + \psi\eta)\}. \\ &\cdot\left\{1 - \frac{\tau^2(1 - \psi)^2(1 + \psi - \psi\eta)}{2n(1 - \psi + \psi\eta)^3} - \frac{\tau\psi\eta}{n(1 - \psi + \psi\eta)^2} + O\left(\frac{1}{n^2}\right)\right\}. \end{aligned} \tag{3.7}$$

In particular, $\Pr\{M_n \le u\} \to e^{-\theta\tau}$ where

$$\theta = \frac{1 - \psi}{1 - \psi + \psi\eta} \tag{3.8}$$

is the extremal index for this problem.

Now let us turn our attention to the functions $\theta(i,u)$. In terms of (3.1) we have

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

Recall that $u$ and $\epsilon$ are related by $\epsilon = 1 - G(u)$. Taking a Taylor expansion in $\epsilon$ for fixed $i$ we find that

$$\begin{aligned}\theta(i,u) = &\left\{\frac{1-\psi}{1-\psi+\psi\eta} + \frac{\epsilon\eta(1-\psi)(1-\psi-\psi\eta)}{(1-\psi+\psi\eta)^3}\right\}\left\{1 - \frac{i\epsilon\eta(1-\psi)}{1-\psi+\psi\eta}\right\} \\ &+ \frac{\eta}{(1-\eta)(1-\psi+\psi\eta)}\left\{1 - \frac{\epsilon(1-\psi)(1-\eta)(1-\psi-\psi\eta)}{(1-\psi+\psi\eta)^2}\right\}\cdot \\ &\cdot\left\{1 + \frac{i\epsilon\eta(1-\psi)}{1-\psi+\psi\eta}\right\}\{\psi(1-\eta)\}^i + O(\epsilon^2).\end{aligned} \tag{3.9}$$

It follows at once from this that $\theta$ as defined by (2.7) is the same as that from (3.8). Let

$$\beta = \psi(1-\eta).$$

Our main interest is in cases in which $\epsilon$ is small, possibly very small, while $i$ is moderately large. Retaining the terms of $O(i\epsilon)$ and $O(\beta^i)$ while discarding everything else leads to the approximation

$$\theta(i,u) - \theta \approx -i\epsilon\theta^2\eta + (1-\theta)\beta^{i-1}. \tag{3.10}$$

Now let us consider some consequences of these calculations for the bias and variance approximations of Section 2. In practice $u$ will not be a constant but a sequence of thresholds $u_n$ so $\epsilon = \epsilon_n$ also depends on $n$. The cluster sizes have a geometric distribution: $\pi(j) = (1-\theta)^{j-1}\theta$ for $j \geq 1$. Hence $\mu = 1/\theta$ and $\sigma^2 = (1-\theta)/\theta^2$. By (2.2),

$$\text{Var}\ \hat{\theta}_n \approx \frac{\theta(1-\theta)}{n\epsilon_n} \tag{3.11}$$

which does not depend directly on $r_n$, so the initial choice of $r_n$ can be made solely to minimise bias. Moreover, to the first order of accuracy (3.11) holds equally well for $\tilde{\theta}_n$ and $\bar{\theta}_n$.

Now consider the bias, initially for $\bar{\theta}_n$. We use the approximation (2.10), ignoring the second term for the moment. Using the approximation (3.10), we want to choose $r_n$ to minimise $|-r_n\epsilon_n\theta^2\eta+(1-\theta)\beta^{r_n-1}|$. If we set $r_n=(\log(1/\epsilon_n)+\log\log(1/\epsilon_n)+C_n)/\log\beta$ for bounded $C_n$, then both terms are of $O(\epsilon_n\log(1/\epsilon_n))$. Note that, 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.

Thus, based on the approximation (3.10), we conclude that the best possible rate for $\theta(r_n,u)-\theta$ is of $O(\epsilon_n\log(1/\epsilon_n))$ which depends on $n$ only through $\epsilon_n$, i.e. the optimal $r_n$ depends only on the threshold and not the sample size.

Let us now combine the bias and variance terms, choosing first $u_n$ and then $r_n$ to minimise 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 $O(n^{-2/3}(\log n)^{2/3})$. The second term in (2.10) is of $O(1/n\epsilon_n)$ and hence negligible in comparison with the first, so at this point we have justified neglecting this term in the preceding calculations.

It is a somewhat different matter, however, if we use the blocks estimator $\hat{\theta}_n$. In this case, assuming (3.10), we have for $r$ large that

$$\frac{1}{r}\sum_{i=1}^{r}\{\theta(i,u)-\theta)\}\approx-\frac{\epsilon_n\theta^2\eta r}{2}+\frac{(1-\theta)}{r(1-\beta)}$$

so, considering the first term only in (2.8), we need to choose $r_n$ so that $r_n=O(\epsilon_n^{-1/2})$. That leads to a bias of $O(\epsilon_n^{1/2})$. Then choosing $\epsilon_n$ to minimise the mean squared error leads to an optimal $\epsilon_n$ of $O(n^{-1/2})$ and a resulting mean squared error of $O(n^{-1/2})$. From

this it can be seen that the blocks estimator leads to a worse optimal rate of convergence than the runs estimator.

Finally, let us consider the refinement that takes us from $\hat{\theta}_n$ into $\tilde{\theta}_n$, which results in two additional terms in (2.9) as compared with (2.8). The larger additional term is of $O(r_n\epsilon_n)$ which is of the same order as the first of the two terms resulting from (3.10). Thus, the correction leads to a final mean squared error of the same order of magnitude, but a different constant of proportionality. This suggests that the correction is worth considering, though we do not have any definitive results as to when it is better.

## 4. IMPLICATIONS FOR MORE GENERAL PROCESSES

In Section 2 we argued that the key feature in characterising the bias, of both the blocks and runs estimators, was the discrepancy between $\theta(i,u)$ and $\theta$ for finite $i$ and $u$. In Section 3, we derived the formula (3.10) for a specific but highly artificial model. This is the only case that we have calculated in detail. Nevertheless, we want to argue, by a mixture of heuristic reasoning and computer simulation, followed in the next section by a real data example, that this structure is considerably more general than that, and provides the basis for a practical method of selecting $r_n$ and hence estimating the extremal index.

The conjectured general result is given by rewriting equation (3.10) in the form

$$\theta(i,u) - \theta \approx -i\theta^2\bar{F}(u) + C\beta^i. \tag{4.1}$$

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

For a fixed threshold $u$ it is clear that $\theta(i,u)$ is decreasing to 0 as $i$ increases. Let $i_u$ be the index $i$ for which $\theta(i,u)$ is closest to $\theta$. Ideally, for given $u$ we would like to set $i$ equal to $i_u$. Suppose $i$ is much bigger than $i_u$. Then

$$
\begin{aligned}
&\theta(i_u, u) - \theta(i, u) \\
&= \Pr\{X_2 \le u, ..., X_{i_u} \le u,\ X_j > u \text{ for some } j,\ i_u < j \le i | X_1 > u\} \\
&= \Pr\{X_2 \le u, ..., X_{i_u} \le u | X_1 > u\}. \qquad (4.2) \\
&\quad . \Pr\{X_j > u \text{ for some } j,\ i_u < j \le i | X_2 \le u, ..., X_{i_u} \le u, X_1 > u\} \\
&\approx \theta\ .\ F^{i\theta}(u)
\end{aligned}
$$

the reasoning for the second factor being that, for very large $i$, the event whose probability is being computed is approximately independent of the conditioning event, by whatever form of mixing hypothesis is being assumed, and is therefore approximated by (1.1) applied with $n = i$.

For only moderately large $i$, the correct form of approximation is not so clear-cut, but for many processes satisfying a geometric ergodicity property, one would expect that probabilities of the form

$$\Pr\{X_2 \le u, ...., X_{i-1} \le u, X_i > u | X_1 > u\}$$

should decay exponentially fast as $i$ grows, implying an error of form $C\beta^i$. Putting the two together implies a general expression of form

$$\theta(i, u) - \theta \approx \theta\{F^{i\theta}(u) - 1\} + C\beta^i. \qquad (4.3)$$

On expanding $F^{i\theta}(u)$ in powers of $\bar{F}(u)$, (4.3) reduces to (4.1). However, if the reasoning leading to (4.3) is correct, then (4.3) should be a better approximation than (4.1), so we use (4.3) rather than (4.1) in subsequent calculation.

The above reasoning is very general and highly heuristic. Unfortunately, it seems too complicated at the moment to perform the calculations rigorously for specific classes of processes. One example where they might be possible is for Markov chains with continuous state space. Smith (1992) has shown that the calculation of the extremal index for such chains, under some conditions about the transition function, may be reduced to an equivalent calculation about random walks. The geometric ergodicity of random walks may

be used to justify the second term in (4.3), while the first term follows from the general argument given above. In this case, what we would expect is that the constants $C$ and $\beta$ will be independent of the threshold. This will be confirmed numerically in one example, defined by equation (4.4) below.

The only other case for which we are aware of detailed calculations being carried out is the paper of Hsing (1990) for $m$-dependent processes. However, the result in this case appears to be of a different form, arising from the fact that although functions of the process are independent over lags greater than $m$, up to lag $m$ they may be arbitrarily complicated, and this appears to preclude any direct approximation of the form we have been considering.

From this it can be seen that the proposed approximation is not universally applicable. The spirit in which we are proposing it, therefore, is as a first approach to be tried in a statistical procedure.

The remainder of this section gives some simulations. Each value of $\theta(i,u)$ is based on approximately $10^5$ replications. First, the doubly stochastic model of Section 3 was simulated. Numerous combinations of $\eta$ and $\psi$ 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 out 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 Figure 1. Here and in all subsequent plots, the three thresholds are represented by crosses, triangles and diamonds in that order, the diamonds being for the highest threshold. It is natural to examine how well the approximation (4.3) fits, and this was calculated from the data by a nonlinear least squares fit of the parameters $\theta, C$ and $\beta$, fitted separately for each threshold. 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 $\theta = 0.1370$, $C = 3.196$, $\beta = 0.27$. It can be seen that the estimated values agree well with these. In Figure 1 the three fitted curves are superimposed on the plot, which seems to confirm a good fit.

Now let us consider an example of a Markov chain. We use for this the logistic Gumbel distribution

$$\Pr\{X_1 \leq x_1, X_2 \leq x_2\} = \exp\{(e^{-rx_1} + e^{-rx_2})^{1/r}\}, \quad r \geq 1, \tag{4.4}$$

which has been shown (in particular by Tawn, 1988b) to be a realistic model for many applications of bivariate extreme value theory. The proposal here is that we should use the conditional distribution of $X_2$ given $X_1$, derived from (4.4), as a transition distribution in a Markov chain. We fixed $r = 2$ for illustration. In this case $\theta \approx 0.328$ is known from Smith (1992), but we know nothing about $C$ and $\beta$ including whether they exist at all.

Figure 2 shows a plot of the simulated $\theta(i, u)$ together with fitted curves of the form (4.3) fitted by nonlinear least squares to each threshold. The three thresholds correspond to $\bar{F}(u)$ =0.01, 0.006 and 0.002. The fitted curves are:

$$\bar{F}(u) = 0.010: \quad \hat{\theta} = 0.332, \hat{C} = 0.319, \hat{\beta} = 0.501,$$

$$\bar{F}(u) = 0.006: \quad \hat{\theta} = 0.332, \hat{C} = 0.307, \hat{\beta} = 0.514,$$

$$\bar{F}(u) = 0.002: \quad \hat{\theta} = 0.330, \hat{C} = 0.270, \hat{\beta} = 0.550.$$

In this case all three estimated values of $\theta$ were very close to the true values, while the estimated $C$ and $\beta$ are reasonably close to each other over different thresholds. Again, the plot seems to confirm that (4.3) is the right functional shape.

The foregoing simulations refer solely to the function $\theta(i,u)$ and not to the actual behaviour of the estimators. As an example of the latter, Figure 3 shows simulations of the root mean squared error for different values of $r_n$ of each of the three estimators $\hat{\theta}_n$, $\tilde{\theta}_n$ and $\bar{\theta}_n$ when the true model is (4.4) 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. It appears, however, that the blocks estimator $\hat{\theta}_n$ performs as well as $\bar{\theta}_n$ in terms of minimum mean squared error over all $r_n$. This partially contradicts the results of Section 3, but it may be that even larger sample sizes than $n = 10000$ are needed to demonstrate the effect, or it may be that we have not chosen the threshold here in a optimal way.

## 5. DATA ANALYSIS EXAMPLE

The example which follows is based on part of a large study that concerned heights of sea waves sampled at three-hourly intervals. 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 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 $\bar{F}(u_2) = 0.040$, $\bar{F}(u_3) = 0.016$. To give an example of the calculation, for threshold $u_1$ there were 2816 exceedances and hence 2815 intervals between exceedances, of which 2468 were of length 0 (i.e. consecutive exceedances), 35 of length 1, 21 of length 2, and so on. Counting the number of intervals of length $> r$ is equivalent to computing the runs estimator with the same $r = r_n$. For instance, with $r = 1$ there are $2815 - 2468 + 1 = 348$ clusters and hence estimated extremal index $348/2816 = 0.124$. With $r = 2$ the numbers are $2815 - 2468 - 35 + 1 = 313$ clusters, extremal index 0.111. The full data set, expressed as a frequency table for intervals between exceedances over three thresholds, is given in Table 1, and the extremal index estimates are calculated and

displayed in Figure 4. It can be seen that the extremal index is hard to specify, estimates from 0.044 to 0.194 being obtained. The standard errors of the estimates, obtained from (2.2), are respectively about 0.006, 0.01 and 0.02 for the three thresholds, and more or less independent of $r$. This implies that the differences among the calculated values are not merely the result of sampling variation.

For this example the procedure described in Section 4 was again applied, the model (4.3) ($i$ replaced by $r$) being fitted by nonlinear least squares to the estimated $\theta$'s for each threshold. The results in this case were:

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

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

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

We can now see what is going on much better. The estimates of $C$ and $\beta$ are quite consistent across different thresholds, but the values of $\theta$ are radically different. The plots in Figure 4 confirm this picture very strongly. We seem to have a case where the extremal index varies across different thresholds, though the model (4.3) fits very well across each of the three thresholds we have examined.

Strictly speaking, the case where the appropriate extremal index varies over different thresholds falls outside the scope of this paper, but it has been observed as a practical phenomenon in a number of different studies. See for example J. Tawn's remarks in the discussion of Davison and Smith (1990). It seems that a more detailed theory encompassing such cases is needed. In the context of the present paper, the message seems to be that direct application of (4.3) fails, but if we allow the extremal index to be dependent on the threshold, the rest of the analysis fits very well, even to the extent of the other parameters $C$ and $\beta$ being nearly constant across thresholds.

Another question to ask is the following: suppose, as in the original recipe, we did use the runs estimator with a single $r$ to estimate the extremal index. What value of $r$ would give the closest answer to our final estimate? The answer, surprisingly, is the same for all three thresholds: $r = 20$. Since this corresponds to a time period of 2.5 days, it seems that we need to take a slightly longer storm length than originally assumed. Nevertheless, it is pleasing that a consistent and readily interpretable storm length is obtained.

We may summarise the main results of the paper as follows:

1. 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.

2. Detailed calculations for the doubly stochastic model, supplemented by heuristic reasoning and simulation for other cases, gives support to the model (4.3) which is the main component of the bias terms derived in (2.8)-(2.10).

3. It is possible to estimate the three parameters in (4.3), and thus obtain an estimate of $\theta$ which does not depend too sensitively on a particular value of $r$, by a least-squares fit based on a range of values of $r$ or $i$. In the one data example we considered, the conclusion was that it was necessary to assume different values of $\theta$ for different thresholds, and since such a phenomenon has been suggested on a number of previous occasions, it would appear to be a genuine and practically important phenomenon.

## ACKNOWLEDGEMENTS

The bulk of the work was done while RLS was at Surrey University. Acknowledgement is made to the Wolfson Foundation for supporting RLS's research at Surrey, to Technion for supporting a visit by RLS to Technion, and to the SERC for a Visiting Fellowship which supported IW's visit to Surrey. The authors would like to thank Uri Cohen for assistance with the programming and graphics.

## REFERENCES

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

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.

Galambos, J. (1987), *The Asymptotic Theory of Extreme Order Statistics* (2nd. edn.). Krieger, Melbourne, Fl. (First edn. published 1978 by John Wiley, New York.)

Hsing, T., Hüsler, J. and Leadbetter, M.R. (1988), On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields* **78**, 97-112.

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

Hsing, T. (1991), Estimating the parameters of rare events. *Stoch. Proc. Appl.* **37**, 117-139.

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

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

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

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. *Proceedings of the Fourth International Meeting on Statistical Climatology*, ed. John Sansom. New Zealand Meteorological Service, P.O. Box 722, Wellington, New Zealand.

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. Ph.D. dissertation available as Technical Report Number 315, Center for Stochastic Processes, Department of Statistics, University of North Carolina.

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

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

O'Brien, G.L. (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.

Smith, R.L. (1989), Extreme value analysis of environmental time series: an example based on ozone data. *Statistical Science* **4**, 367-393.

Smith, R.L. (1992), The extremal index for a Markov chain. To appear, *J. Appl. Prob.*

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

Tawn, J.A. (1988), Bivariate extreme value theory - models and estimation. *Biometrika* **75**, 397-415.

TABLE 1: EXCEEDANCE DATA

The data represent exceedances over three thresholds denoted $u_1$, $u_2$ and $u_3$, based on a total number of 29220 observations. There were respectively 2816, 1170 and 463 exceedances over the three thresholds. In the table, the lengths of intervals between consecutive exceedances are given in the following form:

Col 1: $n$.
Col 2: Number of intervals of length $n$ between exceedances over $u_1$.
Col 3: Number of intervals of length $n$ between exceedances over $u_2$.
Col 4: Number of intervals of length $n$ between exceedances over $u_3$.

| | | | |
|---|---|---|---|
| 0 | 2468 | 971 | 372 |
| 1 | 35 | 12 | 2 |
| 2 | 21 | 16 | 8 |
| 3 | 20 | 5 | 2 |
| 4 | 22 | 5 | 2 |
| 5 | 10 | 7 | 1 |
| 6 | 21 | 4 | 2 |
| 7 | 3 | 6 | 1 |
| 8 | 12 | 5 | 0 |
| 9 | 11 | 5 | 1 |
| 10 | 7 | 4 | 1 |
| 11 | 7 | 4 | 0 |
| 12 | 4 | 4 | 0 |
| 13 | 7 | 4 | 3 |
| 14 | 4 | 0 | 0 |
| 15 | 4 | 4 | 3 |
| 16 | 7 | 0 | 2 |
| 17 | 4 | 2 | 0 |
| 18 | 7 | 3 | 0 |
| 19 | 3 | 1 | 0 |
| 20 | 6 | 1 | 0 |
| 21 | 1 | 0 | 1 |
| 22 | 3 | 1 | 1 |
| 23 | 4 | 1 | 1 |
| 24 | 2 | 2 | 0 |
| 25 | 3 | 1 | 1 |
| 26 | 2 | 3 | 0 |
| 27 | 4 | 2 | 1 |
| 28 | 0 | 0 | 0 |
| 29 | 1 | 2 | 2 |
| 30 | 1 | 0 | 0 |
| >30 | 111 | 94 | 55 |

## FIGURE CAPTIONS

*Figure 1.* Simulated values of $\theta(i,u) - \theta$ for the doubly stochastic model of Section 3, with $\eta = 0.7$, $\psi = 0.9$. The crosses, triangles and diamonds represent simulated values for three thresholds for which $1 - F(u)$ are respectively 0.03, 0.018 and 0.006. The curves are based on equation (4.3) fitted to the data separately for each threshold. It can be seen that the curves fit well through the data points.

*Figure 2.* Same as Figure 1, but the simulated process is now a first-order Markov chain with bivariate distributions given by (4.6) with $r = 2$. The three thresholds correspond to $1 - F(u) = 0.01$ (crosses), $1 - F(u) = 0.006$ (triangles), $1 - F(u) = 0.002$ (diamonds), and the curves are again based on fitting (4.3).

*Figure 3.* Simulations of the square root of mean squared error for the three estimators $\hat{\theta}_n$ (dot-dashed curve), $\tilde{\theta}_n$ (dashed curve) and $\bar{\theta}_n$ (solid curve), for $n = 10000$ and $r_n$ up to 300. All the curves are based on 100 simulations of the model (4.6) with $r = 2$, with $\bar{F}(u) = 0.07$ (Figure 3(a)), $\bar{F}(u) = 0.04$ (Figure 3(b)) and $\bar{F}(u) = 0.01$ (Figure 3(c)).

*Figure 4.* A real data example based on wave heights. The three thresholds correspond to empirical $1 - F(u) = 0.096$ (crosses), $1 - F(u) = 0.04$ (triangles), $1 - F(u) = 0.016$ (diamonds), and the curves are again based on fitting (4.3). Again the curves fit well individually, but there is a very clear separation among the three thresholds, which unlike Figures 1 and 2 cannot be explained away as the effects of bias. Thus we conclude that the difference between the curves is real and represents different extremal indices being effective at different levels of the process.

Figure 1

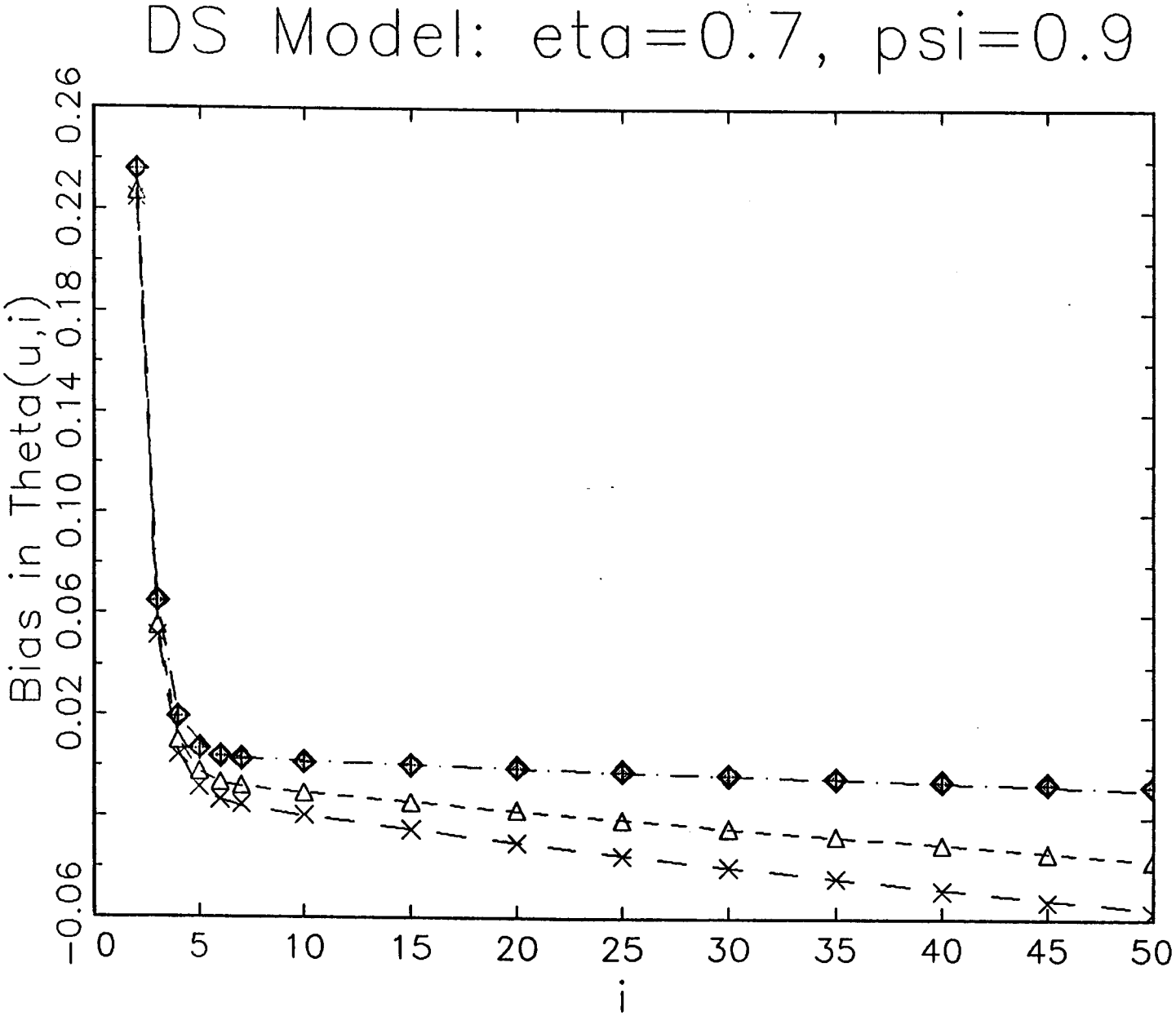

DS Model: eta=0.7, psi=0.9
Bias in Theta(u,i)
-0.06
0.02
0.06
0.10
0.14
0.18
0.22
0.26
0
5
10
15
20
25
30
35
40
45
50
i

Figure 2

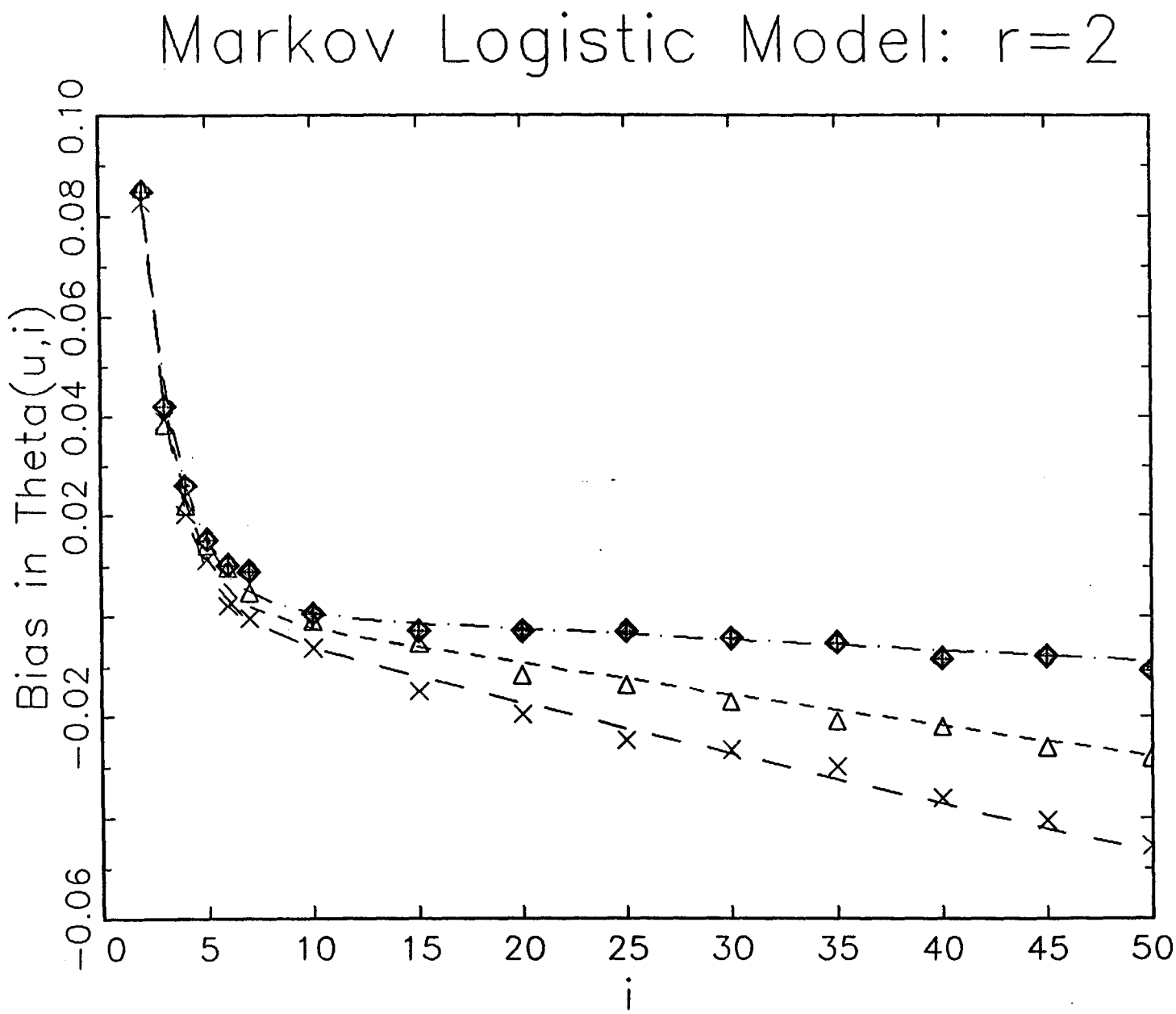

Markov Logistic Model: r=2
Bias in Theta(u,i)
0.10
0.08
0.06
0.04
0.02
−0.02
−0.06
0
5
10
15
20
25
30
35
40
45
50
i

Figure 3(a)

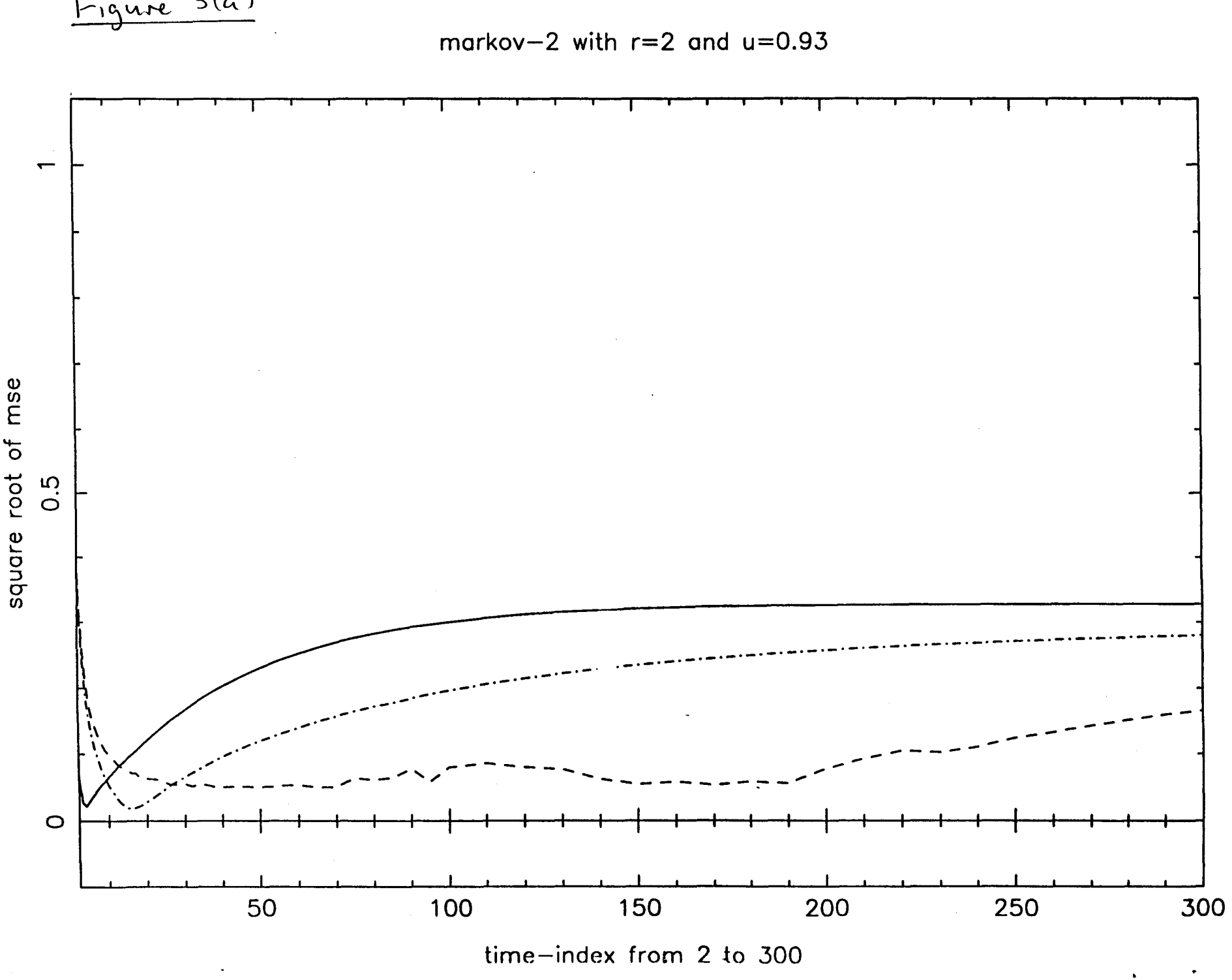

markov−2 with r=2 and u=0.93
square root of mse
0
0.5
1
50
100
150
200
250
300
time−index from 2 to 300

Figure 3(b)

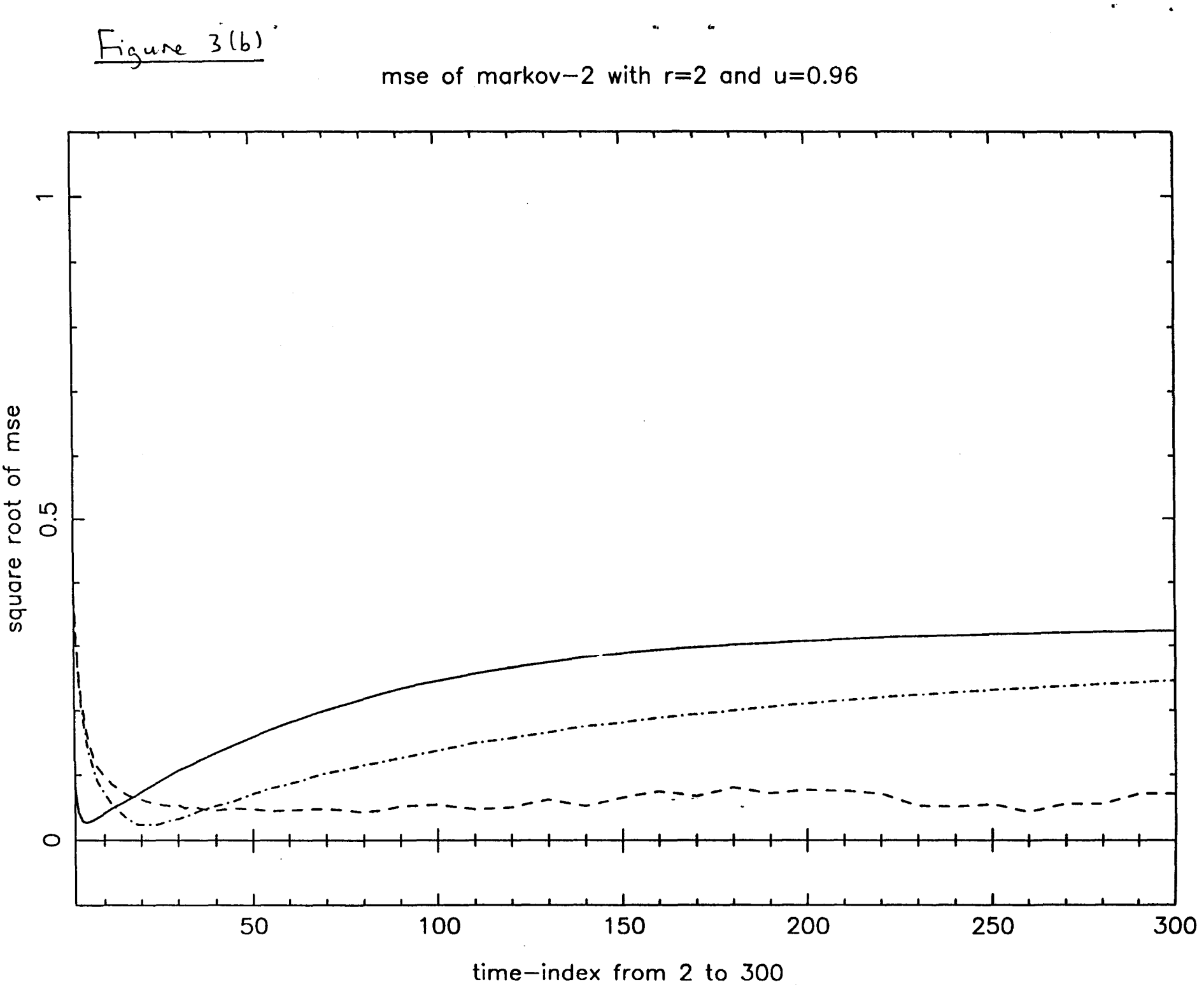

mse of markov−2 with r=2 and u=0.96
square root of mse
0
0.5
1
50
100
150
200
250
300
time−index from 2 to 300

Figure 3(c)

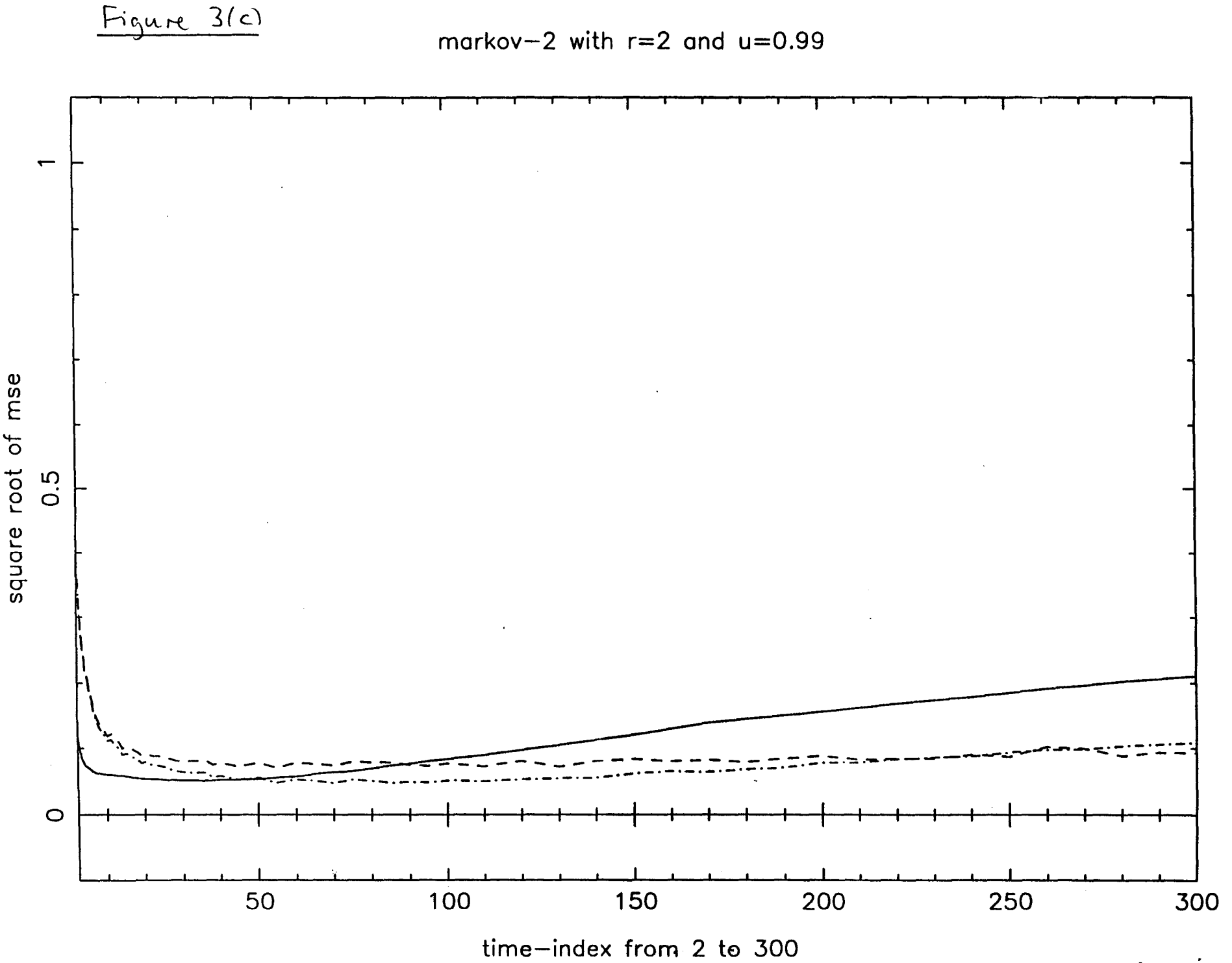
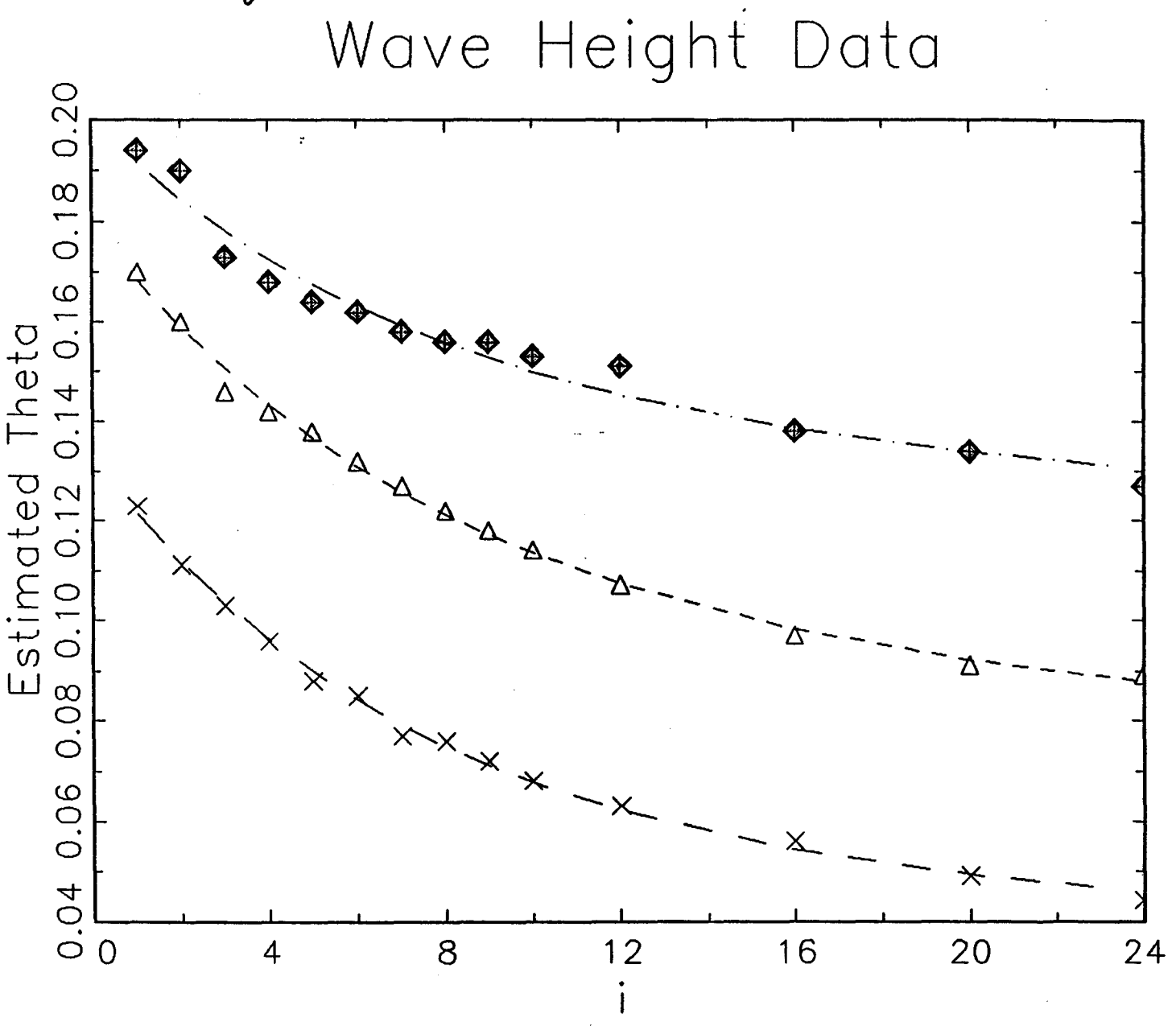

markov−2 with r=2 and u=0.99
square root of mse
0
0.5
1
50
100
150
200
250
300
time−index from 2 to 300

Figure 4


Wave Height Data
Estimated Theta
0.04 0.06 0.08 0.10 0.12 0.14 0.16 0.18 0.20
0 4 8 12 16 20 24
i