MOVING-MAXIMUM MODELS FOR EXTREMA OF TIME SERIES

Peter Hall[1] Liang Peng[1] Qiwei Yao[1,2]

ABSTRACT. We discuss moving-maximum models, based on weighted maxima of independent random variables, for extreme values from a time series. The models encompass a range of stochastic processes that are of interest in the context of extreme-value data. We show that a stationary stochastic process whose finite-dimensional distributions are extreme-value distributions may be approximated arbitrarily closely by a moving-maximum process with extreme-value marginals. It is demonstrated that bootstrap techniques, applied to moving-maximum models, may be used to construct confidence and prediction intervals from dependent extrema. Moreover, it is shown that bootstrapped moving-maximum models may be used to capture the dominant features of a range of processes that are not themselves moving maxima. Connections of moving-maximum models to more conventional, moving-average processes are addressed. In particular, it is demonstrated that a moving-maximum process with extreme-value distributed marginals may be approximated by powers of moving-average processes with stably-distributed marginals.

KEYWORDS. Autoregression, bootstrap, confidence interval, extreme value distribution, generalised Pareto distribution, moving average, Pareto distribution, percentile method, prediction interval.

SHORT TITLE. Moving-maximum models.

[1] Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia.

[2] Institute of Mathematics and Statistics, University of Kent at Canterbury, Canterbury, Kent CT2 7NF, UK

1. INTRODUCTION

We discuss moving-maximum models for exceedences of stationary time series over a threshold. It is shown that when the marginal distribution of the data is one of the three classical extreme-value distributions, the class of moving-maximum processes is dense in the class of stationary processes whose finite-dimensional distributions are extreme-value of a given type. We prove that a moving-maximum process with Type III extreme-value marginals may be expressed as the limit, as the exponent converges to 0, of a moving average of independent, stably-distributed disturbances. Bootstrap methods for moving-maximum models, leading to new approaches to constructing both confidence and prediction intervals for dependent extrema, are also described.

Tavares (1977, 1980, 1981), Deheuvels (1983, 1985), Sim (1986) and Lewis and McKenzie (1991) discussed models based on moving extrema of exponential or Weibull random variables. Processes of moving maxima are related to max-stable processes, introduced by de Haan (1984) and discussed by, for example, de Haan and Pickands (1986) and Balkema and de Haan (1988). Moving-maxima processes are closely related to max–ARMA and max–linear processes, introduced by Davis and Resnick (1989). Prediction methods for these and related processes have been discussed by Davis and Resnick (1989, 1993), where existence of consistent predictors was proved. Alternative approaches to inference for extrema of time series include both parametric and semi-parametric methods; see e.g. Tawn (1988), Coles and Tawn (1991), Ledford and Tawn (1997) and De Haan and Ronde (1998).

Our main results are outlined in section 2. In particular, theoretical properties are summarised in section 2.7, which also discusses asymptotes of the marginal distribution of a moving-maximum process, and large-sample properties of the distributions of estimators. Technical arguments are deferred to section 3.

2. METHODOLOGY AND PROPERTIES

2.1. Definition. Let Y_i, $-\infty < i < \infty$, denote independent and identically distributed random variables with distribution function F, and let $\{a_i\}$ and $\{b_i\}$ be sequences of constants, the former nonnegative. We define a multiplicative process of moving maxima, $\{X_j\}$, by

$$X_j = \sup\{a_{j-i} Y_i \, , \; -\infty < i < \infty\} \, , \tag{2.1}$$

and its additive version, $\{X_j\}$, by

$$X_j = \sup\{Y_i + b_{j-i}\,,\ -\infty < i < \infty\}\,. \tag{2.2}$$

The former is appropriate when the distribution F has support on the positive half-line and is in the domain of attraction of a Type III extreme-value distribution (i.e. the distribution $\exp(-x^{-\gamma})$, for $x > 0$, where $\gamma > 0$), and the latter, when F is attracted to a Type I extreme-value distribution (i.e. the distribution $\exp(-e^{-x})$, for $-\infty < x < \infty$).

To motivate the second, additive model, suppose Y_j represents the strength of the maximum windgust associated with the jth of a sequence of storms that are travelling across a region. Observations of windgust strengths are made as the storms progress through the point O, which they pass at a rate of one per unit time. The effects of the jth storm may be noticed some time before or after its peak, which arrives at O at time j. At time i units after the peak has passed, the observed windgust strength is less by an amount $-b_i$ than it was at the time at which the peak arrived; thus, we expect $b_0 = 0$ and $b_i \leq 0$ for each i. The maximum windgust strength, X_j say, observed at time j is therefore the maximum of the gust associated with the storm whose peak occurred at that time, along with the downweighted values of windgust strengths (downweighted as described in the previous sentence) from storms that were on their way or had recently departed at that time. Mathematically, the maximum windgust strength is therefore given by formula (2.2).

Of course, similar arguments apply in the contexts of other types of extrema, associated for example with events of an environmental or financial nature. And since there is a trivial connection between the two cases represented by (2.1) and (2.2), evident on taking logarithms of the multiplicative process (for strictly positive Y_i), then physical motivation for one of the models is readily obtained from that for the other. For the sake of brevity we shall confine attention to multiplicative moving maxima in the Type III case, noting that Type I and Type II settings admit similar treatments. Section 2.7 will discuss conditions that are sufficient for the process $\{X_j\}$ at (2.1) to be well defined.

Moving-average and autoregressive models are generally constructed so as to express the current value of a time series in terms of past and current values of independent disturbances, rather than future disturbances. Practical considerations suggest that this would usually be appropriate for moving-maximum processes too. Therefore, we would expect the weights a_i to vanish for all sufficiently large negative values of i. Without loss of generality, $a_i = 0$ for $i < 0$. We shall generally make this assumption when discussing models and methods for data analysis, although it will be clear that our procedures are easily modified to encompass more general contexts.

A moving-average model captures both the "shocks" from large, new independent variables, and the subsequent "decay" of those events. For example, if $a_i = 0$ for $i < 0$, if $a_0 = \max_i a_i$, and if a_i gradually decreases as i increases through positive integers, then the full force of the shock from the independent variable X_i is experienced at time i, after which its impact gradually decays at a rate determined by the a_i's.

We shall say that the process at (2.1) is of order m if, except for a consecutive sequence of length no more than $m+1$, we have $a_i = 0$. The process is of course stationary. More general definitions of moving-maxima processes, in terms of triangular arrays of constants $\{a_{ji}\}$ rather than the linear array $\{a_i\}$, or triangular arrays of possibly non-identically distributed random variables Y_{ji}, may be used to model nonstationarity.

The distribution of X_j may be expressed in terms of the distribution of the independent disturbances Y_i:

$$P(X_j \le x_j\,,\, 1 \le j \le k) = \prod_{-\infty<i<\infty} F\big[\min\big\{a_{j-i}^{-1} x_j\,,\, 1 \le j \le k\big\}\big]\,, \quad 1 \le j \le k\,. \tag{2.3}$$

To derive this formula, note that $X_j \le x_j$ for $1 \le j \le k$ if and only if $a_{j-i}Y_i \le x_j$ for all i and $1 \le j \le k$, or equivalently, if and only if $Y_i \le a_{j-i}^{-1}x_j$ for all i and $1 \le j \le k$. Since we do not constrain the a_i's to be strictly positive (they may take the value 0) then we should insist that each Y_i, and hence each X_j, be strictly positive with probability 1, and interpret $a_{j-i}^{-1}X_j$ as $+\infty$ if $a_{j-i} = 0$.

2.2. Relationship to multivariate extreme-value distributions. If F is a Type III extreme-value distribution, say $F(x) = \exp(-bx^{-\gamma})$ where $b, \gamma > 0$, then each X_j has distribution function $\exp(-bcx^{-\gamma})$ where $c = \sum_i a_i^\gamma$. Moreover, all finite-dimensional distributions of the process $\{X_j\}$ are multivariate extreme-value distributions of Type III, and the closure of the class of possible distributions of $\{X_j\}$ is the class $\mathcal{C}$ of stationary processes $\{\xi_i, -\infty < i < \infty\}$ all of whose finite-dimensional distributions are multivariate extreme-value distributions of Type III. Specifically, if $\{\xi_i\}$ is in $\mathcal{C}$ then, for each $\epsilon > 0$ and each positive integer k, there exists a moving-maximum process $\{X_j\}$ with the same marginal, univariate distribution as $\{\xi_i\}$, and such that

$$\sup_{-\infty < x_1, \ldots, x_k < \infty} \left| P(\xi_1 \le x_1, \ldots, \xi_k \le x_k) - P(X_1 \le x_1, \ldots, X_k \le x_k) \right| \le \epsilon \, . \quad (2.4)$$

In this sense, the class of moving-maximum processes with extreme-value marginals is indistinguishable from the class of stationary process with multivariate extreme-value distributions. Marshall and Olkin (1983) have discussed conditions for the weak convergence of extrema to limiting multivariate extreme-value distributions.

Result (2.4) may be proved from classical characterisations of multivariate extreme-value distributions (Geffroy, 1958; Tiago de Oliveira, 1958, 1984; Sibuya, 1960), and the fact that the maximum $\{\max_\ell X_j^{(\ell)}, \ -\infty < j < \infty\}$ of any finite number of independent moving-maximum processes $\{X_j^{(\ell)}, \ -\infty < j < \infty\}$ with Type III marginals may be approximated arbitrarily closely, in the sense of k-variate distributions for any k, by a single moving-maximum process with Type III marginals. The approximation is effected by superimposing translates of the corresponding weight sequences $\{a_i^{(\ell)}, \ -\infty < i < \infty\}$, where the translations are chosen so that large components of any one vector are added to negligibly small components of all the others. The class of superpositions of 'two point' dependence functions, introduced in the next section, is an example of such superimposed translates.

2.3. Bivariate distributions of moving-maximum processes. The circumstances under which many processes of extrema are recorded tend to relegate in favour of weak dependence. To appreciate why, note that if Z_j represents the j'th occasion on which a high level is exceeded by a time series, the time between successive exceedences is likely to be relatively large, and so consecutive Z_j's will tend to be only weakly associated. In such cases, much of the dependence of the process $\{Z_j\}$ will be captured by fitting models to bivariate distributions.

To address bivariate distributions of a moving-maximum process of Type III extremes, suppose the weight sequence $a = \{a_i\}$ has been standardised so that $\sum_i a_i^\gamma = 1$, and put $b_i = a_i^\gamma$. Then, the *dependence function* of the bivariate distribution (see e.g. Pickands 1981, 1989) is

$$A_a(u) = \sum_i \max\{b_{2-i} u, b_{1-i}(1-u)\} \, . \quad (2.5)$$

Note that A_a satisfies the characterising conditions of such a function: it is convex, it passes through the points $P_1 = (0,1)$ and $P_2 = (1,1)$, and it lies within the triangle $\mathcal{T}$ determined by P_1, P_2 and $P_3 = (\frac{1}{2}, \frac{1}{2})$.

Moreover, any function A satisfying these conditions may be obtained as the limit, as a varies, of a sequence of functions A_a defined at (2.5), where we may restrict attention to weight sequences a that are of finite order. In particular, the constant dependence function $A(u) \equiv 1$, representing total independence, is obtained with $a_{i_0} = 1$ for any i_0 and $a_i = 0$ otherwise, while $A(u) \equiv \min(u, 1-u)$, representing total dependence, is obtained as the limit as $m \to \infty$ of the case $a_i = m^{-\gamma}$ for $i_0 \le i \le i_0 + m$ and $a_i = 0$ otherwise, where i_0 is any integer.

Reflecting the case of total dependence, functions A that protrude lower into the triangle $\mathcal{T}$ require sequences a of relatively high order if we are to achieve a good approximation. The case where just two, adjacent a_i's are nonzero, i.e. $a = (\ldots, 0, t, 1-t, 0, \ldots)$ and $0 \le t \le 1$, gives the triangular dependence function $A_a(u) = \max\{t + (1-t)u, 1 - tu\}$, which always lies within the smaller triangle $\mathcal{T}'$ defined by P_1, P_2 and $P_3' = (\frac{1}{2}, \frac{3}{4})$. Any dependence function A lying wholly within $\mathcal{T}'$ may be obtained as the limit of superpositions of these two-point dependence functions. In particular, given $\epsilon > 0$ we may choose $m \ge 1$ and $t_1, \ldots, t_m \in [0,1]$ such that, with

$$a = m^{-1/\gamma}\big(\ldots, 0, 0, t_1^{1/\gamma}, (1-t_1)^{1/\gamma}, 0,\\ t_2^{1/\gamma}, (1-t_2)^{1/\gamma}, 0, \ldots, 0, t_m^{1/\gamma}, (1-t_m)^{1/\gamma}, 0, 0, \ldots\big),$$

we have $\sup_{0 \le u \le 1} |A(u) - A_a(u)| \le \epsilon$.

2.4. Relationship to moving averages and stable processes. Univariate extreme-value distributions of Type III may be derived as limits of stable distributions as the exponent converges to 0. Specifically, the characteristic function of a general stable law with exponent $\alpha \in (0,2)$ satisfying $\alpha \ne 1$, and with tail-balance parameter $\beta \in [0,1]$, after standardisation for location, is

$$\psi(t|\alpha, \beta, c) = \exp\big[-c\,|t|^\alpha \big\{1 - i\,\beta\, \tan\big(\tfrac{1}{2}\alpha\pi\big)\, \mathrm{sgn}\, t\big\}\big], \tag{2.6}$$

where $i = \sqrt{-1}$. If the random variable $S(\alpha, \beta, c)$ has this distribution then the limit, as $\alpha \to 0$, of the law of $|S(\alpha, \beta, c)|^{\alpha/\gamma}$ is Type III extreme value:

$$P\big\{|S(\alpha, \beta, c)|^{\alpha/\gamma} \le x\big\} \to F(x) = \exp\big(-cx^{-\gamma}\big), \quad 0 < x < \infty, \tag{2.7}$$

as $\alpha \to 0$. The convergence is uniform in β and x, for c fixed; this constant should be interpreted identically in (2.6) and (2.7).

The property that a sum of independent stably-distributed random variables is stably distributed implies that dependent sequences with stably-distributed marginals may be simulated very simply as moving averages. To appreciate how this may be done, note that if $\{W_i, -\infty < i < \infty\}$ are independent and identically distributed random variables with the distribution at (2.6), and if $\{b_i, -\infty < i < \infty\}$ are nonnegative constants satisfying $\sum_i b_i^\alpha = 1$, then $V_j = \sum_i b_{j-i} W_i$ defines a stable process with the same marginal distribution as W_i.

This result may in turn be used as the basis of an approximation to moving maximum processes by stable distributions. To appreciate how, note that the (α/γ)'th power of the absolute value of this moving average process converges, as $\alpha \to 0$, to a moving-maximum process with extreme-value marginals. Specifically, if $b_i = b_i(\alpha)$ varies with α in such a way that b_i^α converges to a_i^γ (not depending on α) as $\alpha \to 0$, in the strong sense that $\sum_i |b_i^\alpha - a_i^\gamma| \to 0$, then all finite-dimensional distributions of the process $\{|V_j|^{\alpha/\gamma}, -\infty < j < \infty\}$ converge to those of the process $\{X_j\}$ defined at (2.1), where Y_i should be taken to have distribution F at (2.7). An outline proof will be given in section 3.1.

2.5. Estimating F and the weights a_i. Let $\{X_j\}$ be a moving-maximum process, and assume a parametric model $F(\cdot|\theta)$ is available for the distribution of Y_i. Given a sample $\mathcal{X} = \{X_1, \ldots, X_n\}$ we may compute estimators $\hat\theta$ of θ and $\hat a = \{\hat a_i\}$ of a, such that low order, finite dimensional, conditional distributions of the process $\{X_j^*\}$, defined by

$$X_j^* = \sup\left\{\hat a_{j-i} Y_i^*\,,\ -\infty < i < \infty\right\}, \tag{2.8}$$

approximate the corresponding distributions of $\{X_j\}$. Section 2.7 will give appropriate regularity conditions, permitting $\{X_j\}$ to be more general than a moving-maximum process. In (2.8), $\{Y_i^*, -\infty < i < \infty\}$ is, conditional on $\mathcal{X}$, a sequence of independent random variables with distribution function $F(\cdot|\hat\theta)$.

Specifically, for all $k \ge 1$ we may construct $\hat\theta$ and $\hat a$ such that

$$\sup_{-\infty < x_1, \ldots, x_k < \infty} \left|P\left(X_1^* \le x_1, \ldots, X_k^* \le x_k \middle| \mathcal{X}\right) - P(X_1 \le x_1, \ldots, X_k \le x_k)\right| \to 0 \tag{2.9}$$

in probability as $n \to \infty$. One approach is maximum likelihood, exploiting independence of the variables Y_i through the representation at (2.3). Assuming the model is exact, this technique will produce estimates that reflect perfectly, with very high probability, some of the characteristics of the model. To appreciate why, consider attempting to fit a two-point moving-maximum model of the type discussed in Section 2.3. Then without loss of generality, $a_i = 0$ except for $i = 1$ or 2. It follows that $P(X_j/X_{j-1} = a_2/a_1) > 0$ for each j, and so with probability tending to 1 exponentially quickly the mode of the empirical distribution of X_j/X_{j-1} equals a_2/a_1. Therefore, using maximum likelihood we can in principle estimate a_2/a_1 with extraordinary accuracy. Similar remarks may be made about methods suggested by Davis and Resnick (1989, section 5).

An alternative approach, not tied so closely to features of the exact model, may be based on weighted least-squares approach and is described in the next paragraph. It employs iterative methods to fit k-variate distributions to an m'th order moving-maximum model, using the weight vector $a_{(m)} = (a_0, \ldots, a_m)$ and assuming $m \geq k$, and can achieve virtually root-n consistency in (2.9). (Theoretical results in section 2.7 will show that this is achievable even when k diverges with n.) We assume that scale is an unknown parameter of the distribution of Y_i, but we take the value of that parameter to equal 1, in effect incorporating scale into the weights a_i. Let $F(\cdot|\theta)$ denote the corresponding distribution function of Y_i. For example, in the extreme-value and generalised Pareto cases, θ is a scalar $\gamma > 0$, and $F(y|\gamma) = \exp(-y^{-\gamma})$ and $F(y|\gamma) = 1 - (1+y)^{-\gamma}$, respectively. In such cases it would usually be convenient to separately compute an estimator $\widehat{\gamma}$ of γ, using for example the method of Hill (1975), and fit $a_{(m)}$ only after replacing θ by $\widehat{\gamma}$ in the procedures below; but this is not strictly necessary.

Define $x = (x_1, \ldots, x_k)$,

$$\widehat{G}(x) = (n-k)^{-1} \sum_{i=1}^{n-k} I(X_{i+j-1} \leq x_j \text{ for } 1 \leq j \leq k)\,, \tag{2.10}$$

$$D_m(\theta, a_{(m)}) = \int \Bigg(\widehat{G}(x) - \prod_{i=2-m}^{k} F\big[\min\{a_{j-i}^{-1} x_j\,, \max(i,1) \leq j \leq \min(i+m,k)\} \,\big|\, \theta\big]\Bigg)^2 w(x)\, dx\,,$$

where w is a nonnegative weight function. (Thus, $\widehat{G}(x)$ represents an empirical approximation to the probability that $X_j \leq x_j$ for $1 \leq j \leq k$. The identity (2.3) motivates our definition of $D_m(\theta, a_{(m)})$.) For m given, choose $(\theta, a_{(m)})$ to minimise $D_m(\theta, a_{(m)})$. We suggest stopping at a relatively low value of m; see sections 2.7

and 3 for discussion. The resulting estimators of θ and a will be denoted by $\hat{\theta}$ and $\hat{a}$, respectively.

The order of a fitted moving-maximum model may be interpreted as a smoothing parameter. Increasing it tends to reduce the bias of bootstrap procedures but increase their variability. When the model is fitted to bivariate distributions (or to k-variate distributions for any fixed k), cross-validation may be used to select the appropriate value of m. Provided m is not allowed to increase too quickly with n, and the bivariate extreme-value distribution is smooth, it may be proved that cross-validation gives consistent estimation of the bivariate distribution function. An attractive practical procedure, valid when the marginal distribution is approximately Type III extreme value, is to estimate the shape parameter using the Hill estimator and fit a moving-maximum model, with relatively small m, to bivariate distributions under the assumption of Type III univariate distributions.

2.6. Bootstrap methods for confidence and prediction intervals. Let $\hat{\theta}$ and $\hat{a}$ denote the estimators of θ suggested in section 2.5, and define the process $\{X_j^*\}$ as at (2.8). Percentile-method confidence and prediction intervals may be constructed in the usual way. For example, to calculate an α-level confidence interval for the r'th component, θ_r say, of θ, first compute the version $\hat{\theta}^*$ of $\hat{\theta}$ for the sample $\mathcal{X}^* = \{X_1^*, \ldots, X_n^*\}$ rather than $\mathcal{X}$, let $\hat{\theta}_r^*$ be the r'th component of $\hat{\theta}^*$, and let $\hat{\lambda}_\alpha$, a function of $\mathcal{X}$, be the solution of $P(\hat{\theta}_r^* \leq \hat{\lambda}_\alpha | \mathcal{X}) = \alpha$. Then, $(-\infty, \hat{\lambda}_\alpha)$ is a nominal α-level confidence interval for θ_r. To compute the analogous prediction interval for the largest of the next n_0 values of the process $\{X_j\}$, i.e. for $X_{n+1}, \ldots, X_{n+n_0}$, define $\hat{q}_\alpha$ to be the solution of $P(\max_{1\leq j\leq n_0} X_{n+j}^* \geq \hat{q}_\alpha | \mathcal{X}) = \alpha$. Then, $[\hat{q}_\alpha, \infty)$ is a nominal α-level prediction interval for $\max_{1\leq j\leq n_0} X_{n+j}$. Both intervals may be calibrated using the double bootstrap, so as to improve their coverage accuracy.

2.7. Theoretical properties. Assume that (i) F has support on the positive half-line, and is in the domain of attraction of a Type III extreme-value distribution. That is, $1 - F$ is regularly varying at infinity with exponent $-\gamma$, where $\gamma > 0$. Suppose too that (ii) each a_i is nonnegative and, for some $\epsilon \in (0, \gamma)$, $0 < \sum_i a_i^{\gamma-\epsilon} < \infty$. Then the following results hold: (a) the random variable X_j defined at (2.1) is finite with probability 1 [compare Proposition 2.1 of Davis and Resnick (1989)], (b) the process $\{X_j\}$ satisfies the distributional mixing condition $D(u_n)$ (see Leadbetter, Lindgren and Rootzén, 1983, Chapter 3), (c) if G denotes the distribution function of X_j then $1-G$ is regularly varying at infinity with exponent $-\gamma$, (d)

$$\lim_{x\to\infty} \frac{1-G(x)}{1-F(x)} = \sum_i a_i^\gamma\,,$$

(e) if in addition to (i) and (ii), $p=p(n)\to\infty$ and $p/n\to 0$, then the Hill estimator

$$\widehat{\gamma}_p \equiv \left(p^{-1}\sum_{i=1}^{p} \log X_{n,n-i+1} - \log X_{n,n-p}\right)^{-1}$$

based on the largest p of the order statistics $X_{n1}\le\ldots\le X_{nn}$ of $\mathcal{X}$ is weakly consistent for γ, and (f) if $b_n=(1-F)^{-1}(1/n)$ then

$$P\Big(\max_{1\le j\le n} X_j \le b_n x\Big) \to \exp\big(-bx^{-\gamma}\big)\,,\quad 0\le x<\infty\,,$$

where b denotes the finite, positive constant defined by

$$b = \lim_{n\to\infty} n^{-1}\sum_i \max_{1\le j\le n} a_{i+j}^\gamma\,.$$

In particular, $b\le\sum_i a_i^\gamma$.

Property (f) asserts that to first order, the largest element of the dependent sample $\mathcal{X}$ behaves identically to the largest among the independent disturbances $Y_1,\ldots,Y_n$, except that scale is increased by a constant factor. However, this result does not extend to multivariate extremes. In particular, if $a_i = a_{i+1} = \ldots = a_{i+p-1} = \max_j a_j$ for some i, then the probability that the p largest values in $\mathcal{X}$ are equal to one another does not converge to 0 as $n\to\infty$.

The latter result, as well as (a), (c), (d) and (f) above, may be deduced by modifying standard theoretical methods used to prove limit theorems for extremes of independent data. Property (b) follows from the following result: under conditions (i) and (ii), for each $\epsilon>0$ we may choose the positive integer m so large that the probability that X_j equals $\sup\{a_{j-i}Y_i,\ |i|\le m\}$ exceeds $1-\epsilon$ for each j. Result (e) is derived in section 3.2. Methods in section 2 of Resnick and Starica (1997) may be used to show that if (i) holds, if $k\to\infty$ sufficiently slowly (the rate depending on F), and if $a_i = O(c^{|i|})$ as $|i|\to\infty$, for some $c\in(0,1)$, then $k^{1/2}(\widehat{\gamma}_k-\gamma)$ is asymptotically Normal $\mathrm{N}(0,\sigma^2)$, where

$$\sigma^2 = \gamma^2\Big[1 + A^{-1}\sum_{j_1}\sum_{j_2}\big\{a_{j_1}^\gamma + a_{j_1+j_2}^\gamma - \max(a_{j_1}^\gamma, a_{j_1+j_2}^\gamma)\big\}\Big]$$

and $A = \sum_i a_i^\gamma$.

To conclude, we give a theorem which confirms that bootstrapped moving-maximum processes can successfully approximate finite-dimensional distributions of processes that are actually more general than moving maxima. As a prelude to stating the theorem, assume that $Y_i^{(\ell)}$, for $-\infty < i < \infty$ and $\ell > 1$, are independent and identically distributed random variables with either a Type III extreme-value distribution, $F(x|\gamma) = \exp(-x^{-\gamma})$, or a Pareto distribution, $F(x|\gamma) = 1-(1+x)^{-\gamma}$. Let $a_i^{(\ell)}$ and c_ℓ, for $0 \le i < \infty$, be constants with the properties $0 \le a_i^{(\ell)} \le C_1\rho^i$ and $0 \le c_\ell \le C_1\rho^i$, where $0 < C_1 < \infty$ and $0 < \rho < 1$. Suppose too that $\sum_\ell c_\ell \sum_i a_i^{(\ell)} > 0$. Put

$$X_j = \sup_{\ell \ge 1} \left[c_\ell \sup \left\{ a_{j-i}^{(\ell)} Y_i^{(\ell)}, \ -\infty < i < \infty \right\} \right]. \tag{2.11}$$

In the algorithm for constructing estimators $\hat{\theta}_{(m)}$ and $\hat{a}_{(m)}$, assume that $m = m(n) \to \infty$ such that $m = O\{(\log n)^{C_2}\}$ for some $C_2 > 0$, and that ϵ is infinitesimal. Suppose too that the weight function w is the indicator of a rectangular prism $[0, C_3]^k$, where $0 < C_3 < \infty$.

Theorem. *Under these conditions,* (2.9) *holds for each* $k \ge 1$. *Moreover, if* $m \ge C_4(\log n)^2$ *for* C_4 *sufficiently large, then the rate of convergence in* (2.9) *is* $O_p(n^{-(1/2)+\delta})$ *for all* $\delta > 0$.

Note particularly that the sequence $\{X_j\}$ defined at (2.11) is not, in general, a moving-maximum process. In particular, it is not expressible in the form (2.1) if at least two of the c_ℓ's are nonzero. If the number of nonzero c_ℓ's and $a_i^{(\ell)}$'s are both finite then, despite the infinite process $\{X_j\}$ not being expressible as a moving maximum, all its finite-dimensional distributions are those of moving maximum processes. Nevertheless, if F is a Type III extreme-value distribution then all the finite-dimensional distributions of $\{X_j\}$ are multivariate extreme-value distributions. Therefore, the theorem confirms that moving-maximum process models can capture all the finite-dimensional distributions of processes that are not themselves of moving-maximum type.

Since, under the conditions of the theorem, (2.9) holds for all $k \ge 1$, then it remains true if $k = k(n) \to \infty$ sufficiently slowly. A longer proof than that given in section 3.@ shows that it is sufficient for k to increase no faster than a constant multiple of $\log n$.

3. TECHNICAL ARGUMENTS

3.1. Approximation by stable processes. Here we derive the result stated in the last paragraph of section 2.4. We use notation from that section, and note that if W (a generic W_i) has the distribution of which the characteristic function is given by (2.6), then the limit as $\alpha \to 0$ of the distribution of $|W|^{\alpha/\gamma}$ is a Type I extreme-value distribution with exponent γ:

$$P\big(|W|^{\alpha/\gamma} \le x\big) \to \exp\big(-c\,x^{-\gamma}\big) \tag{3.1}$$

as $\alpha \to 0$, for each $x > 0$. The convergence is uniform in all values of the tail-balance constant β in the stable law, and the result may be deduced from expansions of densities of stable laws; see Ibragimov and Linnik (1971, p. 54f) and Zolotarev (1986, p. 94f), but note that the argument of the sine function in Ibragimov and Linnik's Theorem 2.4.2 should be $\frac{1}{2}\pi n(2-\alpha)(\beta+1)$, and that in Zolotarev's formula (2.5.4), '$s \in$' should read 'sin'.

Observe too that if $b_i = b_i(\alpha)$ satisfies $\sum_i b_i^\alpha = 1$ and $\sum_i |b_i^\alpha - a_i^\gamma| \to 0$ as $\alpha \to 0$, where a_i does not depend on α, then $\sum_i a_i^\gamma = 1$, and as $\alpha \to 0$, for each j,

$$|V_j|^{\alpha/\gamma} - \sup\big\{a_{j-i}\,|W_i|^{\alpha/\gamma}, -\infty < i < \infty\big\} \to 0 \tag{3.2}$$

in probability. It follows from (3.1) and (3.2) that any finite collection of the variables $|V_j|^{\alpha/\gamma}$ converges in distribution to a vector of the same number of variables, with joint distribution equal to that of the X_j's at (2.1), where the distribution function of Y_i is that on the right-hand side of (3.1), or equivalently, on the the right-hand side of (2.7), and the Y_i's are independent.

3.2. Consistency of Hill estimator. We show that if conditions (i) and (ii) in section 2.7 hold then the Hill estimator is consistent for γ. Our argument adapts methods of Resnick and Starica (1995); see also Hsing (1991) and Rootzén, Leadbetter and de Haan (1998).

LIANG PENG TO PUT IN PROOF HERE, AND THEN RENUMBER SUBSEQUENT EQUATIONS.

3.3. Proof of Theorem. Let $C_5 > 0$ be so large that $m \le \frac{1}{2} C_5 \log n$ and $C_5 \log \rho < -2$, and let $N = N(n)$ denote the integer part of $C_5 \log n$. Put $a_{2N(\ell-1)+i} = c_\ell a_i^{(\ell)}$ if $0 \le i \le N-1$ and $1 \le \ell \le N$, and $a_i = 0$ for all values of i that are not covered by this prescription. Let $\{Y_i, -\infty < i < \infty\}$ be a sequence of independent random variables with the same distribution as $T_i^{(\ell)}$, and put $X_j' = \sup\{a_{j-i} Y_i, -\infty < i < \infty\}$. Then,

$$\sup_{-\infty<x_1,\ldots,x_k<\infty} \big|P(X_1' \le x_1,\ldots,X_k' \le x_k) - P(X_1 \le x_1,\ldots,X_k \le x_k)\big| = O\big(n^{-2+\delta}\big) \tag{3.@1}$$

for all $\delta > 0$. Therefore, it suffices to prove the version of (2.9) where X_j is replaced by X_j'. (We have approximated $\{X_j\}$ by a moving-maximum model $\{X_j'\}$, albeit one where the weights depend on n.)

Let $\widehat{G}(\cdot)$ be as at (2.10), put $G(y) = P(X_j \le x_j$ for $1 \le j \le k)$, and given a sequence $b_{(p)} = (b_0, \ldots, b_{p-1})$, define

$$d_p(\gamma, b_{(p)}) = \int \Big\{G(x) - \prod_i F\big(\min_j b_{j-i}^{-1} x_j \big| \gamma\big)\Big\}^2 w(x)\, dx\,. \tag{3.@2}$$

Now, $\sup_x |\widehat{G}(x) - G(x)| = O_p(n^{-1/2})$, and so

$$D_p(\gamma, b_{(p)})^{1/2} = d_p(\gamma, b_{(p)})^{1/2} + O_p\big(n^{-1/2}\big) \tag{3.@3}$$

uniformly in $\gamma > \gamma_0$, $2 \le p \le m$ and $b_{(p)}$, for any positive γ_0.

Let G' denote the version of G when X_j is replaced by X_j', and let d_p' be the version of d_p when G is replaced by G' on the right-hand side of (3.@2). The argument leading to (3.@1) may be employed to prove that $d_p(\gamma, b_{(p)})^{1/2} = d_p'(\gamma, b_{(p)})^{1/2} + O(n^{-1+\delta})$, and so by (3.@3), $D_p(\gamma, b_{(p)})^{1/2} = d_p'(\gamma, b_{(p)})^{1/2} + O_p(n^{-1/2})$, uniformly in $\gamma > \gamma_0$ and in $b_{(p)}$. If $p \ge 2(C_5 \log n)^2$ then γ and $b_{(p)}$ may be chosen so that $b_{(p)}'(\gamma, b_{(p)}) = 0$; and if $p = p(n) \to \infty$ then $b_{(p)}$ and γ may be chosen so that $d_p'(\gamma, b_{(p)}) \to 0$. The theorem now follows from a Taylor-expansion argument involving no more than $O\{(\log n)^{C_6}\}$ variables, for some $C_6 > 0$.

REFERENCES

BALKEMA, A.A. AND DE HAAN, L. (1988). Almost sure continuity of stable moving average processes with index less than one. *Ann. Probab.* **16**, 333–343.

COLES, S.G. AND TAWN, J.A. (1991). Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B* **53**, 377–392.

DAVIS, R.A. AND RESNICK, S.I. (1989). Basic properties and prediction of max-ARMA processes. *Adv. Appl. Probab.* **21**, 781–803.

DAVIS, R.A. AND RESNICK, S.I. (1993). Prediction of stationary max-stable processes. *Ann. Appl. Probab.* **3**, 497–525.

DE HAAN, L. (1984). A spectral representation for max-stable processes. *Ann. Probab.* **12**, 1194–1204.

DE HAAN, L. AND DE RONDE, J. (1998). Sea and wind: multivariate extremes at work. *Extremes 1*, 7–45.

DE HAAN, L. AND PICKANDS, J. III (1986). Stationary min-stable stochastic processes. *Probab. Theory Relat. Fields* **72**, 477–492.

DEHEUVELS, P. (1983). Point processes and multivariate extreme values. *J. Multivariate Anal.* **13**, 257–272.

DEHEUVELS, P. (1985). Point processes and multivariate extreme values (II). In: *Multivariate Analysis, VI*, Ed. P.R. Krishnaiah, pp. 145–164. North-Holland, Amsterdam.

GEFFROY, J. (1958). Contribution à la thórie des valeurs extrêmes. *Publ. Inst. Statist. Univ. Paris* **7**, 37–185.

HILL, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163–1174.

HSING, T. (1991). On tail estimation using dependent data. *Ann. Statist.* **19**, 1547–1569.

IBRAGIMOV, I.A. AND LINNIK, Yu.V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

LEADBETTER, M.R., LINDGREN, G. AND ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer, New York.

LEDFORD, A.W. AND TAWN, J.A. (1997). Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B* **59**, 475–499.

LEWIS, P.A.W. AND MCKENZIE, E. (1991). Minification processes and their transformations. *J. Appl. Probab.* **28**, 45–57.

MARSHALL, A.W. AND OLKIN, I. (1983). Domains of attraction of multivariate extreme value distributions. *Ann. Probab.* **11**, 168–177.

PICKANDS, J. (1981). Multivariate extreme value distributions. In: *Proc. 43rd Session ISI* (Buenos Aires), pp. 859–878.

PICKANDS, J. (1989). Multivariate negative exponential and extreme value distributions. In: *Extreme Value Theory*, Springer Lecture Notes in Statistics **51**, Eds. J. Hüsler and R.D. Reiss, pp. 859–878. Springer, Berlin.

RESNICK, S. AND STARICA, C. (1995). Consistency of Hill's estimator for dependent data. *J. Appl. Probab.* **32**, 139–167.

RESNICK, S. AND STARICA, C. (1997). Asymptotic behaviour of Hill's estimator for autoregressive data. *Comm. Statist. B—Stochastic Models.* **13**, 703–721.

ROOTZÉN, H., LEADBETTER, M. AND DE HAAN, L. (1998). Tail and quantile estimation for strong mixing stationary sequences. *Ann. Appl. Probab.* **8**, 868–885.

SIBUYA, M. (1960). Bivariate extreme statistics. *Ann. Inst. Statist. Math.* **11**, 195–210.

SIM, C.H. (1986). Simulation of Weibull and gamma autoregressive stationary process. *Comm. Statist. B—Simulation Comput.* **15**, 1141–1146.

TAVARES, L.V. (1977). The exact distribution of extremes of a non-Gaussian process. *Stochastic Processes Appl.* **5**, 151–156.

TAVARES, L.V. (1980a). An exponential Markovian stationary process. *J. Appl. Probab.* **17**, 1117–1120.

TAVARES, L.V. (1981). Letter to the editor: "An exponential Markovian stationary process". *J. Appl. Probab.* **18**, 957.

TAWN, J.A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika* **75**, 397–415.

TIAGO DE OLIVEIRA, J. (1958). Extremal distributions. *Rev. Fac. Cienc. Lisboa Ser. 2 A Mat.* **VII**, 215–227.

TIAGO DE OLIVEIRA, J. (1984) Bivariate models for extremes. In: *Statistical Extremes and Applications*, Ed. J. Tiago de Oliveira, pp 131–153. Reidel, Dordrecht.

ZOLOTAREV, V.M. (1986). *One-Dimensional Stable Distributions.* Amer. Math. Soc., Providence.