Kybernetika

Marta Ferreira

On the extremal behavior of a Pareto process: an alternative for ARMAX modeling

Kybernetika, Vol. 48 (2012), No. 1, 31--49

Persistent URL: `http://dml.cz/dmlcz/142062`

KYBERNETIKA — VOLUME *48* (2012), NUMBER 1, PAGES 31–49

ON THE EXTREMAL BEHAVIOR OF A PARETO PROCESS: AN ALTERNATIVE FOR ARMAX MODELING

Marta Ferreira

In what concerns extreme values modeling, heavy tailed autoregressive processes defined with the minimum or maximum operator have proved to be good alternatives to classical linear ARMA with heavy tailed marginals (Davis and Resnick [8], Ferreira and Canto e Castro [13]). In this paper we present a complete characterization of the tail behavior of the autoregressive Pareto process known as *Yeh–Arnold–Robertson Pareto(III)* (Yeh et al. [32]). We shall see that it is quite similar to the first order max-autoregressive ARMAX, but has a more robust parameter estimation procedure, being therefore more attractive for modeling purposes. Consistency and asymptotic normality of the presented estimators will also be stated.

Keywords: extreme value theory, Markov chains, autoregressive processes, tail dependence

Classification: 60G70, 60J20

1. INTRODUCTION

The main objective of an extreme value analysis is to estimate the probability of events that are more extreme than any that have already been observed. By way of example, suppose that a sea-wall projection requires a coastal defense from all sea-levels, for the next 100 years. Extremal models are a precious tool that enables extrapolations of this type. The central result in classical Extreme Value Theory (EVT) states that, for an i.i.d. sequence, $\{X_n\}_{n\geq 1}$, having marginal cumulative distribution function (cdf) F, if there are real constants $a_n > 0$ and b_n such that,

$$P(\max(X_1,\ldots,X_n) \leq a_n x + b_n) \longrightarrow_{n\to\infty} G_\gamma(x)\,, \tag{1}$$

for some non-degenerate function G_γ, then it must be the Generalized Extreme Value function (GEV),

$$G_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}), \quad 1+\gamma x > 0,\, \gamma \in \mathbb{R},$$

($G_0(x) = \exp(-e^{-x})$) and we say that F belongs to the max-domain of attraction of G_γ, in short, $F \in \mathcal{D}(G_\gamma)$. The parameter γ, known as the tail index, is a shape parameter as it determines the tail behavior of F, being so a crucial issue in EVT. If $\gamma > 0$ we have a heavy tail (Fréchet max-domain of attraction), $\gamma = 0$ means an exponential tail

(Gumbel max-domain of attraction) and $\gamma < 0$ indicates a short tail (Weibull max-domain of attraction). Here we will be interested in heavy tails. Considering the tail quantile function (q.f.), $F^{-1}(1-t) = \inf\{x : F(x) \geq 1-t\}$, we have, $F \in \mathcal{D}(G_\gamma)$ for $\gamma > 0$, if and only if

$$F^{-1}(1-tx) \sim x^{-\gamma} F^{-1}(1-t) \text{ , as } t \to \infty, \tag{2}$$

which is also equivalent to a $-1/\gamma$-regularly varying tail at ∞, i. e.,

$$1 - F(x) = x^{-1/\gamma} L_F(x), \tag{3}$$

where L_F is a slowly varying function at ∞ (i. e., $L(tx)/L(t) \sim 1$, as $t \to \infty$). The form (3) is also called a Pareto-type tail.

The first results in EVT were developed considering independent observations but, more recently, models for extreme values have been constructed under the more realistic assumption of temporal dependence. Among these models, stationary Markov chains are very interesting, specially because they may have a somewhat simple treatment in what concerns extremal properties. The max-autoregressive moving average processes MARMA (Davis and Resnick [8]), and also the particular case MAR(1) or ARMAX, given by,

$$X_i = \max(c\, X_{i-1}, W_i), \tag{4}$$

with $0 < c < 1$ and $\{W_i\}_{i\geq 1}$ i.i.d. (Alpuim [1]; Canto e Castro [5]; Lebedev [23]) are some examples. Heavy tailed MARMA, in particular ARMAX, and classical linear ARMA can be good choices for modeling time series data with sudden large peaks, although the former processes are more convenient for analysis as their finite-dimensional distributions can easily be written explicitly (Davis and Resnick [8]). Actually, the well-known MARMA processes and their generalizations have been applied to various phenomena, e. g., a solar thermal energy storage system (Daley and Haslett [9]), the water density in a sill fjord (Helland and Nielsen [15]) or financial series (Zhang and Smith [33]). Heavy tailed power max-autoregressive processes have also been developed with successful application to financial time series modeling (Ferreira and Canto e Castro [13]).

Here we shall focus on the not so well-known autoregressive Pareto processes. Any stochastic process whose marginal distributions are of the Pareto or generalized Pareto form is called a Pareto process. As Vito Pareto [26] observed, many economic variables have heavy tailed distributions not well modeled by the normal curve. Instead, he proposed a model, subsequently called, in his honor, the Pareto distribution. The defining feature of this distribution is that its survival function $P(X > x)$ decreases at a negative power of x as $x \to \infty$, i. e.,

$$P(X > x) \sim cx^{-\alpha}, \text{ as } x \to \infty. \tag{5}$$

Generalizations of Pareto's distribution have been proposed for modeling economic variables (a survey can be seen in Arnold [2]).

The classical Pareto distribution has a survival function of the form

$$\overline{F}_X(x) = (x/\sigma)^{-\alpha}, \; x > \sigma, \tag{6}$$

where $\sigma > 0$ is a scale parameter and $\alpha > 0$ is a shape (or inequality) parameter. If X has distribution (6) we will write $X \frown \mathscr{P}(I)(\sigma, \alpha)$.

A minor modification of (6) is obtained by introducing a location parameter μ, i. e.,

$$\overline{F}_X(x) = \left[1 + \frac{x-\mu}{\sigma}\right]^{-\alpha},\ x > \mu. \tag{7}$$

If X has distribution (7) we will write $X \frown \mathscr{P}(II)(\mu, \sigma, \alpha)$.

A third variant of Pareto's distribution has as its survival function

$$\overline{F}_X(x) = \left[1 + \left(\frac{x-\mu}{\sigma}\right)^{\alpha}\right]^{-1},\ x > \mu. \tag{8}$$

and if X has distribution (8) we will write $X \frown \mathscr{P}(III)(\mu, \sigma, \alpha)$.

Clearly all three of the Pareto distributions (6) – (8) exhibit the tail behavior (5) postulated by Pareto, i. e., an heavy tail. In practice, it is difficult to discriminate between models (7) and (8) and the choice may be justifiably made on the basis of which model is mathematically more tractable.

The classical normal autoregressive processes have proved to be flexible and useful modeling tools. The Pareto processes can be expected to better model time series with heavy tailed marginals as well. We will focus on autoregressive Pareto(III) processes, more precisely, the Yeh–Arnold–Robertson Pareto(III) (Yeh et al. [32]). We shall characterize the right tail behavior and conclude that it is similar to the process ARMAX (Sections 2.1 and 2.2). Therefore, the above mentioned phenomena usually modeled by ARMAX processes can also be modeled through a Yeh–Arnold–Robertson Pareto(III) if we are interested in the tail. We will see that the parameter estimation is more robust in Yeh–Arnold–Robertson Pareto(III), which makes this process more attractive for modeling purposes (Section 2.3). We will also state consistency and asymptotic normality of the presented estimators (Section 3).

2. THE YEH–ARNOLD–ROBERTSON PARETO(III) PROCESS

Consider an innovations sequence $\{\varepsilon_n\}_{n\geq 1}$ of i.i.d. random variables (r.v.'s) Pareto(III)(0,σ,α), with $\sigma, \alpha > 0$, and sequence $\{U_n\}_{n\geq 1}$ of i.i.d. r.v.'s Bernoulli(p) (independent of the ε's). The process $\{X_n\}_{n\geq 1}$ is a first order Yeh–Arnold–Robertson Pareto(III), in short YARP(III)(1), if it has the form

$$X_n = \min\left(p^{-1/\alpha} X_{n-1}, \frac{1}{1-U_n}\varepsilon_n\right), \tag{9}$$

where $1/0$ is interpreted as $+\infty$. By conditioning on U_n, it is readily verified that the YARP(III)(1) process has a Pareto(III)(0,σ,α) stationary distribution and will be a completely stationary process if $X_0 \frown \mathscr{P}(III)(0, \sigma, \alpha)$ (Arnold [3]). This process presents sudden large peaks as can be seen in Figure 1, a similar behavior as the above mentioned max-autoregressive processes.

Fluctuation probabilities are given by

$$P(X_{n-1} < X_n) = \frac{1+p}{2} \tag{10}$$

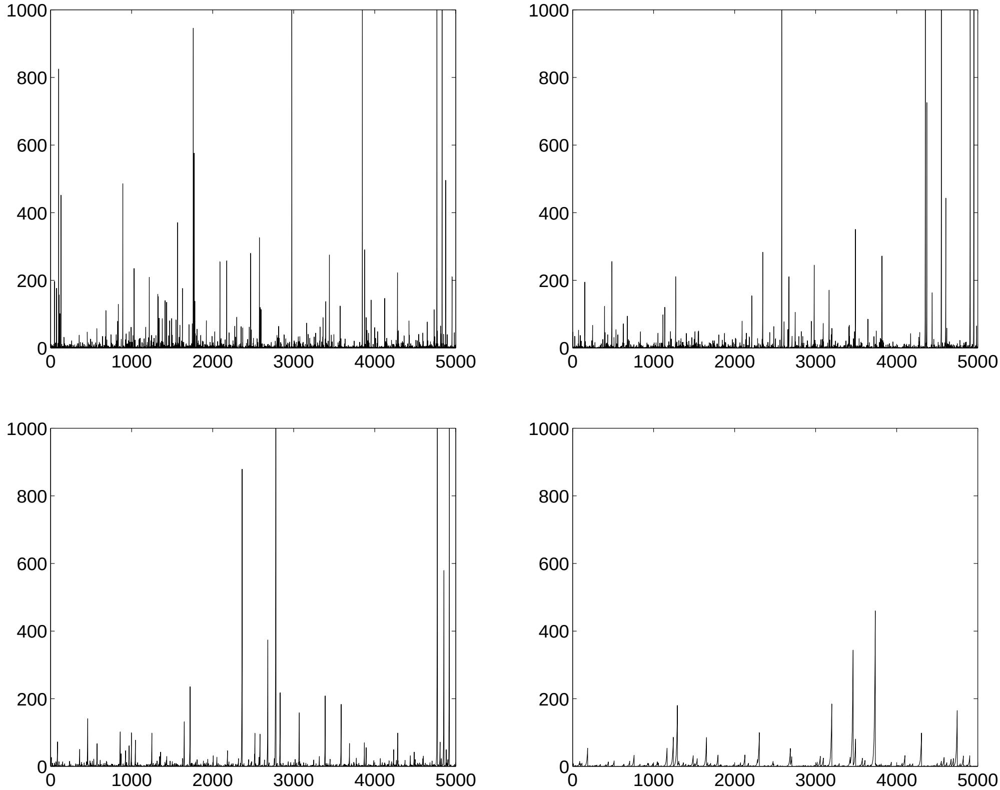

Fig. 1. Simulated sample paths of YARP(III)(1) processes with marginals $\mathscr{P}(III)(0,1,1)$ for $p = 0.3$ (top-left), $p = 0.5$ (top-right), $p = 0.7$ (bottom-left), $p = 0.9$ (bottom-right).

and this can be used to develop a simple consistent estimate of p based on an observed sample path from the process. Estimation of α and σ can be accomplished via the method of moments provided they exist. Note that, for the first and second moments we must have $\alpha > 1$ and $\alpha > 2$, respectively. Yeh et al. [32] proposed a logarithmic transformation to avoid moment assumptions. Since α is a tail index, it can be estimated through tail index estimators. We shall see that the Hill estimator (Hill [16]) has good properties such as consistency and asymptotic normality.

Another interesting feature is its well behaved extreme values. Consider

$$T_n = \min_{0 \le i \le n} X_i$$

and

$$M_n = \max_{0 \le i \le n} X_i.$$

It is readily seen that $T_n \overset{d}{=} \min_{i\leq N} \varepsilon_i$, where ε_i, $i \geq 1$, are i.i.d. Pareto(III)$(0,\sigma,\alpha)$ and N, independent of ε_i is such that $N-1 \frown Binomial(n,1-p)$ (Arnold [3]). It follows by conditioning on N that

$$P(T_n > t) = [1+(\tfrac{t}{\sigma})^\alpha]^{-1}([1+p(\tfrac{t}{\sigma})^\alpha]/[1+(\tfrac{t}{\sigma})^\alpha])^n\,, \quad t \geq 0$$

and the asymptotic behavior of T_n is given by $n(1-p)^{1/\alpha}T_n/\sigma \xrightarrow{d} Weibull(\alpha)$.

To determine de distribution of M_n it is convenient to consider a family of level crossing processes $\{Z_n(t)\}$ indexed by $t \in \mathbb{R}^+$, defined by

$$Z_n(t) = \begin{cases} 1 & \text{if } X_n > t \\ 0 & \text{if } X_n \leq t. \end{cases}$$

These two processes are themselves Markov chains with corresponding transition matrices given by

$$P = \left(1+\left(\tfrac{t}{\sigma}\right)^\alpha\right)^{-1} \begin{bmatrix} p+\left(\frac{t}{\sigma}\right)^\alpha & 1-p \\ (1-p)\left(\frac{t}{\sigma}\right)^\alpha & 1+p\left(\frac{t}{\sigma}\right)^\alpha \end{bmatrix}$$

Hence, for $t \geq 0$, we have

$$\begin{aligned} F_{M_n}(t) &= P(M_n \leq t) = P(Z_0(t)=0, Z_1(t)=0,\ldots,Z_n(t)=0) \\ &= P(X_0 \leq t)P(Z_i(t)=0|Z_{i-1}(t)=0)^n = \frac{\left(\frac{t}{\sigma}\right)^\alpha}{1+\left(\frac{t}{\sigma}\right)^\alpha}\left(\frac{p+\left(\frac{t}{\sigma}\right)^\alpha}{1+\left(\frac{t}{\sigma}\right)^\alpha}\right)^n \end{aligned} \tag{11}$$

and $\frac{n^{-1/\alpha}}{\sigma}M_n \xrightarrow{d} Fréchet(0,(1-p)^{-1},\alpha)$.

We also point out that these processes are closed for geometric minima and maxima, i. e., $T = \min_{0\leq i\leq N} X_i$ and $M = \max_{0\leq i\leq N} X_i$ where $N \frown Geometric(p)$, have also Pareto(III) distribution. Further details can be seen in Arnold, [3].

Before going any further, we determine the transition probability function (tpf) of the YARP(III)(1) process, as it will be a fundamental tool in the proofs of the results in the next sections. We start to compute the 1-step tpf:

$$\begin{aligned} Q(x,]0,y]) &= P(X_{n+1} \leq y|X_n = x) = P\big(\min(p^{-1/\alpha}x, \tfrac{\varepsilon_n}{1-U_n}) \leq y\big) \\ &= \begin{cases} 1-P(\frac{\varepsilon_n}{1-U_n} > y), & x > yp^{1/\alpha} \\ 1, & x \leq yp^{1/\alpha} \end{cases} \\ &= \begin{cases} (1-p)F_\varepsilon(y), & x > yp^{1/\alpha} \\ 1, & x \leq yp^{1/\alpha}. \end{cases} \end{aligned} \tag{12}$$

Similarly, and after some calculations, we derive the m-step tpf:

$$\begin{aligned} Q^m(x,]0,y]) = &\ P\big(X_{n+m} \leq y|X_n = x\big) \\ = &\begin{cases} 1-\prod_{j=0}^{m-1}[\overline{F}_\varepsilon(p^{j/\alpha}y)(1-p)+p], & x > yp^{m/\alpha} \\ 1, & x \leq yp^{m/\alpha}. \end{cases} \end{aligned} \tag{13}$$

Based on the this function, we compute the m-step fluctuation probabilities $f_m := P(X_{n-m} < X_n)$, for any positive integer m:

$$\begin{aligned} f_m := P(X_{n-m} < X_n) &= \int_0^\infty P(X_n > x | X_{n-m} = x)\, \mathrm{d}F_X(x) \\ &= \int_0^\infty (1 - Q^m(x,]0, x])\, \mathrm{d}F_X(x) \\ &= \int_0^\infty \prod_{j=0}^{m-1} \left[\overline{F}_\varepsilon(p^{j/\alpha} x)(1-p) + p\right] \mathrm{d}F_X(x) \\ &= \frac{1}{2}(1 + p^m), \end{aligned} \tag{14}$$

where the last step is due to the fact that $F_\varepsilon(x) = F_X(x)$ and can be derived if we take first $m = 1$, then $m = 2$, and so on. Observe that f_1 was already introduced in (10). We will use the fluctuation probabilities to estimate the process parameter, p, in Section 3.

2.1. The dependence structure and the tail dependence

We shall focus on the dependence conditions that will allow a characterization of the process tail behavior.

We start with the β-mixing condition. A stationary sequence $\{X_i\}_{i\geq 1}$ is said to be β-mixing if

$$\beta(l) := \sup_{p\in\mathbb{N}} E\Big(\sup_{B\in\mathcal{F}(X_{p+l+1},\ldots)} |P(B|\mathcal{F}(X_1,\ldots,X_p)) - P(B)| \Big) \underset{l\to\infty}{\longrightarrow} 0, \tag{15}$$

with $\mathcal{F}(.)$ denoting the $\sigma-$field generated by the indicated random variables.

We will show that YARP(III)(1) is regenerative and aperiodic, sufficient conditions to derive a β-mixing structure (Asmussen, 1987).

Proposition 2.1. The YARP(III)(1) process is regenerative and aperiodic.

Proof. For regeneration we must prove that it has a regeneration set R, i. e., a recurrent set R such that, for some $m \in \mathbb{N}$, a distribution λ and $\epsilon \in (0,1)$, we have

$$Q^m(x, B) \geq \epsilon\lambda(B),\ x \in \mathbb{R} \tag{16}$$

for all real borelian B. In what concerns aperiodicity, we must prove that, for any regeneration set R and any real borelian B, we have

$$Q^{m+1}(x, B) \geq \epsilon_1\lambda(B) \ \text{ and } \ Q^m(x, B) \geq \epsilon_2\lambda(B),\ \forall x \in \mathbb{R}, \tag{17}$$

for some $m \in \mathbb{N}$ and $\epsilon_1, \epsilon_2 \in (0,1)$. The proof runs along the same steps as in Ferreira and Canto e Castro [13]).

Consider $R =]r, \infty[\subset]0, \infty[$ (which is recurrent because it is in the support of the process) and B a real borelian set. Let $x \in R$ and $S =]0, r]$. Observe that

$$Q(x, B) = \int_B \mathrm{d}Q(x, z) \geq \int_{B\cap S} \mathrm{d}Q(x, z)$$

and, for all $x \in R$, $x > r > rp^{1/\alpha}$. Hence, by (12),

$$Q(x, B) \geq \int_{B \cap S} \mathrm{d}Q(x, z) = \int_{B \cap S} (1 - p)\, \mathrm{d}F_\varepsilon(z) = \epsilon\lambda(B), \tag{18}$$

where $\epsilon = (1-p)P(\varepsilon \in S)$ and $\lambda(\cdot) = P(\varepsilon \in \cdot \cap S)/P(\varepsilon \in S)$. If $B \cap S =$ development (18) still holds. Hence condition (16) holds. Observe now that

$$Q^2(x, B) = \int P(X_{n+2} \in B | X_{n+1} = z)\, \mathrm{d}Q(x, z) \geq \int_S Q(z, B)\, \mathrm{d}Q(x, z)$$

and by (12),

$$Q^2(x, B) \geq \textstyle\int_S (1-p)P(\varepsilon \in B)\, \mathrm{d}Q(x, z) \geq (1-p)P(\varepsilon \in B \cap S)Q(x, S)$$
$$= (1-p)P(\varepsilon \in B \cap S)P(\varepsilon \in S)(1-p) = \epsilon_1\lambda(B),$$

with $\epsilon_1 = \epsilon(1-p)P(\varepsilon \in S)$. Hence, condition (17) holds by taking, in addition, $\epsilon_2 = \epsilon$ and $m = 1$. □

The β-mixing condition ensures that the local dependence condition $D(u_n)$ of Leadbetter [20] holds for any real sequence $\{u_n\}_{n\geq 1}$. This latter is a condition like mixing but only required to hold for events of the form $\{X_i \leq u_n\}$ or their intersections. The D condition leads to the appearance of a dependence parameter, the extremal index $\theta \in [0, 1]$, associated with the tendency of clustering of high levels. Whenever $\theta = 1$ we have a similar behavior of an i.i.d. sequence, i. e., large values occur isolated and no clustering takes place. By a result in Chernick [6], if for each $\tau > 0$ there is a real sequence $\{u_n\}_{n\geq 1}$ satisfying

$$n(1 - F_X(u_n)) \to \tau,\ n \to \infty, \tag{19}$$

and $\{X_n\}_{n\geq 1}$ satisfies $D(u_n)$, then $P(M_n \leq u_n) \to e^{-\theta\tau}$ as $n \to \infty$, with θ independent of τ.

Proposition 2.2. The YARP(III)(1) process has extremal index $\theta = 1 - p$.

P r o o f. First observe that the quantile function is given by

$$F_X^{-1}(t) = \sigma((1-t)^{-1} - 1)^{1/\alpha} \tag{20}$$

and that real levels $\{u_n\}_{n\geq 1}$ satisfying (19) are of the form $\sigma(n/\tau - 1)^{1/\alpha}$. Hence, applying (11) and after some calculations, we have

$$P(M_n \leq u_n) = P(M_n \leq \sigma(n/\tau - 1)^{1/\alpha}) = (1 - \tfrac{\tau}{n}(1-p))^n \to e^{-\tau(1-p)}.$$

□

As $\theta < 1$ we have clustering of high values. We can also conclude that the local dependence condition $D'(u_n)$ of Leadbetter et al. [21] doesn't hold since this latter inhibits high levels clustering. As can be seen in the definition below, condition $\mathrm{D}'(u_n)$ bounds the probability of more than one exceedance of u_n on a time-interval of $r_n = [n/k_n]$ integers, with $k_n \to \infty$, as $n \to \infty$.

Definition 2.3. Condition $\mathrm{D}'(u_n)$ will be said to hold for $\{X_i\}_{i\in\mathbb{Z}}$ if for some sequence $\{k_n\}$ such that $k_n \to \infty$, as $n \to \infty$, we have

$$\limsup_{n\to\infty} n \sum_{j=2}^{r_n} P\big(X_1 > u_n, X_j > u_n\big) = 0.$$

Several local dependence conditions provide formulas for the computation of θ. For instance, the local dependence condition $\Delta(u_n)$ introduced in Hsing et al. [18], which is weaker than β-mixing as it only requires (15) to hold for σ-fields $\mathcal{F} = \sigma(X_I, \dots, X_L)$ generated by events $\{X_I \le u_n, \dots, X_L \le u_n\}$, allows to derive θ as the arithmetic inverse of the mean cluster size (Hsing et al. [18]).
Also the family of conditions $D^{(k)}(u_n)$, for $k \ge 1$ (Chernick et al. [7]) are sufficient to

$$\theta = \lim_{n\to\infty} P(M_{2,k} \le u_n | X_1 > u_n)$$

when the limit exists, where $M_{i,j} = \max(X_i, \dots, X_j)$ for $i \le j$ and $M_{i,j} = -\infty$ for $i > j$. The condition $D^{(k)}(u_n)$ holds for $\{X_i\}_{i\ge 1}$ when

$$nP(X_1 > u_n \ge M_{2,k}, M_{k+1,r_n} > u_n) \underset{n\to\infty}{\to} 0$$

with $r_n = [n/k_n]$ and sequence $\{k_n\}$ such that $k_n \to \infty$ as $n \to \infty$ satisfying some specific conditions. In particular, $D^{(1)}(u_n) \equiv D'(u_n)$ and $D^{(2)}(u_n)$ is implied by condition $D''(u_n)$ of Leadbetter and Nandagopalan [22], defined below.

Definition 2.4. Condition $D''(u_n)$ will be said to hold for $\{X_i\}_{i\in\mathbb{Z}}$ if condition $D(u_n)$ also holds and, considering a real sequence $\{k_n\}$ such that

$$k_n \underset{n\to\infty}{\longrightarrow} \infty \ , \ k_n \alpha_{n,l_n} \underset{n\to\infty}{\longrightarrow} 0 \ , \ k_n l_n / n \underset{n\to\infty}{\longrightarrow} 0 \ , \tag{21}$$

$k_n(1 - F(u_n)) \underset{n\to\infty}{\longrightarrow} 0$ we have

$$\limsup_{n\to\infty} n \sum_{j=2}^{r_n - 1} P\big(X_1 > u_n, X_j \le u_n < X_{j+1}\big) = 0.$$

Proposition 2.5. Condition $D''(u_n)$ holds for process YARP(III)(1), for levels u_n satisfying (19).

Proof. Observe that

$$\begin{aligned}
&P(X_1 > u_n, X_j \le u_n < X_{j+1}) \\
&= P\Big(X_1 > u_n, \min\big(p^{-\frac{j-1}{\alpha}} X_1, \tfrac{p^{-\frac{j-2}{\alpha}} \varepsilon_2}{1-U_2}, \dots, \tfrac{p^{-\frac{1}{\alpha}} \varepsilon_{j-1}}{1-U_{j-1}}, \tfrac{\varepsilon_j}{1-U_j}\big) \le u_n, \\
&\qquad p^{-\frac{j}{\alpha}} X_1 > u_n, \tfrac{p^{-\frac{j-1}{\alpha}} \varepsilon_2}{1-U_2} > u_n, \dots, \tfrac{p^{-\frac{1}{\alpha}} \varepsilon_j}{1-U_j}, \tfrac{\varepsilon_{j+1}}{1-U_{j+1}}\big) > u_n\Big) \\
&\le P\Big(X_1 > u_n, \tfrac{p^{-\frac{j-k}{\alpha}} \varepsilon_k}{1-U_k} \le u_n, p^{-\frac{j}{\alpha}} X_1 > u_n, \tfrac{p^{-\frac{j-1}{\alpha}} \varepsilon_2}{1-U_2} > u_n, \dots, \\
&\qquad \tfrac{\varepsilon_{j+1}}{1-U_{j+1}}\big) > u_n\Big), \text{ for some } k = 2, \dots, j \\
&= P(X_1 > u_n)[F_\varepsilon(p^{\frac{j-k}{\alpha}} u_n) - F_\varepsilon(p^{\frac{j-k+1}{\alpha}} u_n)] \prod_{\substack{i=0 \\ i \ne j-k+1}}^{j-1} [\overline{F}_\varepsilon(p^{\frac{i}{\alpha}} u_n)(1-p) + p]
\end{aligned} \tag{22}$$

where we have considered $\min(p^{-\frac{j-1}{\alpha}}X_1, \frac{p^{-\frac{j-2}{\alpha}}\varepsilon_2}{1-U_2}, \ldots, \frac{\varepsilon_j}{1-U_j}) \neq p^{-\frac{j-1}{\alpha}}X_1$ and $U_k \neq 1$, otherwise the probability will be immediately null. Now observe that, for some constant a,

$$\frac{1-F_\varepsilon(au_n)}{1-F_\varepsilon(u_n)} = \frac{[1+(au_n/\sigma)^\alpha)]^{-1}}{[1+(u_n/\sigma)^\alpha)]^{-1}} = \frac{\sigma^\alpha + u_n^\alpha}{\sigma^\alpha + (au_n)^\alpha} \underset{n\to\infty}{\sim} \frac{1}{a^\alpha}$$

and that levels u_n satisfying (19) also satisfy $n(1-F_\varepsilon(u_n)) \to \tau$, as $n \to \infty$, since $F_X(\cdot) = F_\varepsilon(\cdot)$. Considering (22) and as $F_\varepsilon(p^{\frac{j-k}{\alpha}}u_n) \leq F_\varepsilon(u_n)$, we have

$$\begin{aligned}
& n\textstyle\sum_{j=2}^{[n/k_n]-1} P(X_1 > u_n, X_j \leq u_n < X_{j+1}) \\
& \leq n\tfrac{n}{k_n} P(X_1 > u_n)[1 - F_\varepsilon(p^{\frac{j-k+1}{\alpha}}u_n) - (1-F_\varepsilon(u_n))] \cdot \\
& \qquad\qquad \cdot \prod_{\substack{i=0 \\ i\neq j-k+1}}^{j-1} [\overline{F}_\varepsilon(p^{\frac{i}{\alpha}}u_n)(1-p)+p] \\
& = \tfrac{1}{k_n} nP(X_1 > u_n)\Big[n(1-F_\varepsilon(u_n))\Big(\tfrac{1-F_\varepsilon(p^{\frac{j-k+1}{\alpha}}u_n)}{1-F_\varepsilon(u_n)} - 1\Big)\Big] \cdot \\
& \qquad\qquad \cdot \prod_{\substack{i=0 \\ i\neq j-k+1}}^{j-1} [\overline{F}_\varepsilon(p^{\frac{i}{\alpha}}u_n)(1-p)+p] \\
& \underset{n\to\infty}{\sim} \tfrac{1}{k_n}\,\tau\,[\tau(\tfrac{1}{p^{j-k+1}} - 1)]p^j \underset{n\to\infty}{\sim} 0.
\end{aligned} \tag{23}$$

□

A generalization of condition $D''(u_n)$ is obtained by replacing exceedances with upcrossings in $D^{(k)}(u_n)$ and this new family of local conditions, slightly stronger than $D^{(k)}(u_n)$, is denoted $\widetilde{D}^{(k)}(u_n)$ (cf. Ferreira [12]).

Consider notation $\widetilde{N}_n(B) = \sum_{i=1}^n \mathbf{1}_{\{X_i \leq u_n < X_{i+1}\}}\delta_{i/n}(B)$, with $B \subset [0,1]$, and let $\widetilde{N}_n[i/n, j/n] \equiv \widetilde{N}_{i,j}$.

Definition 2.6. For any $k \geq 2$, $\{X_i\}_{i\in\mathbb{Z}}$ satisfies condition $\widetilde{D}^{(k)}(u_n)$ if condition $\Delta(u_n)$ holds and

$$nP\big(X_1 \leq u_n < X_2, \widetilde{N}_{3,k} = 0, \widetilde{N}_{k+1,r_n} > 0\big) \underset{n\to\infty}{\to} 0,$$

for some sequence $r_n = [n/k_n]$ with $\{k_n\}$ satisfying (21).

Condition $\widetilde{D}^{(2)}(u_n)$ is also implied by $D''(u_n)$. Analogous to the extremal index as a measure of clustering of exceedances of high levels, Ferreira [12] introduces the upcrossings index, $\vartheta \in [0,1]$, a measure for clustering of upcrossings of high levels. The family $\widetilde{D}^{(k)}(u_n)$, for $k \geq 1$, provide a way to compute ϑ too. More precisely, under conditions $\Delta(u_n)$ and $\widetilde{D}^{(k)}(u_n^{(\varsigma)})$ for some $k \geq 2$ and for each $\varsigma > 0$, then the upcrossings index of $\{X_i\}_{i\in\mathbb{Z}}$ exists and is equal to ϑ if and only if

$$\vartheta = \lim_{n\to\infty} P(\widetilde{N}_{3,k}(\widetilde{u}_n^{(\varsigma)}) = 0 | X_1 \leq \widetilde{u}_n^{(\varsigma)} < X_2),$$

for each $\varsigma > 0$ (Corollary 3.1 in Ferreira [12]). We also have the following relation between the upcrossings index and the extremal index:

$$\theta = \frac{\varsigma}{\tau}\vartheta \tag{24}$$

Proposition 2.7. The YARP(III)(1) process has unit upcrossings index.

Proof. Observe that condition $D''(u_n)$ holds for levels satisfying (19) which imply $nP(X_1 \leq u_n < X_2) \to \varsigma$, with $\varsigma = \tau(1-p)$:

$$\begin{aligned} nP(X_1 \leq u_n < X_2) &= n[P(X_2 > u_n) - P(X_1 > u_n, X_2 > u_n)] \\ &= n[P(X_2 > u_n) - P(X_1 > u_n, p^{-1/\alpha}X_1 > u_n, \tfrac{\varepsilon_2}{1-U_2} > u_n)] \\ &= nP(X_2 > u_n) - nP(X_1 > u_n)[\overline{F}_\varepsilon(u_n)(1-p) + p] \underset{n\to\infty}{\to} \tau(1-p). \end{aligned} \tag{25}$$

Therefore, by (24) we obtain $\vartheta = 1$. □

An unit upcrossings index means that no clustering of upcrossings of high levels takes place.

2.2. Coefficients of tail dependence

Several dependence coefficients have been introduced in the literature to measure the dependence of a random pair (X, Y) occurring in the tail. Ferreira and Ferreira [14] have defined tail dependence coefficients for random pairs (X_1, X_{1+m}), i. e., for observations separated in time by a lag m, $m \in \mathbb{N}$ (Ferreira and Ferreira [14]). This formulation is interesting for model diagnosis purposes, similar to the role of the autocorrelation function in linear models.

Considering marginal uniform normalization, we have the lag-m *tail dependence coefficient*,

$$\lambda_m = \lim_{t\downarrow 0} P(X_{1+m} > F_X^{-1}(1-t)|X_1 > F_X^{-1}(1-t)). \tag{26}$$

Loosely stated, λ_m is the probability of X_{1+m} being extreme given that X_1 is extreme. According to Ferreira and Ferreira [14] (Proposition 3.2), we can relate the extremal index θ with coefficients λ_m, for $m \in \mathbb{N}$. In the case $\lambda_m = 0$, the r.v.'s X_1 and X_{1+m} are said to be asymptotically independent, and if $0 < \lambda_m \leq 1$ they are asymptotically dependent. Observe that the boundary cases of total dependence and total independence correspond to $\lambda_m = 1$ and $\lambda_m \sim P(X_{1+m} > F_X^{-1}(1-t))$, respectively.

Whenever $\lambda_m = 0$ and one assumes independence, and hence calculates the probability of a jointly extreme event $p = P(X_1 > F_X^{-1}(1-t), X_{1+m} > F_X^{-1}(1-t))$ as $P(X_1 > F_X^{-1}(1-t))P(X_{1+m} > F_X^{-1}(1-t))$, then one may underestimates p in the case of positive dependence (i. e., $P(X_1 > F_X^{-1}(1-t), X_{1+m} > F_X^{-1}(1-t)) \geq P(X_1 > F_X^{-1}(1-t))P(X_{1+m} > F_X^{-1}(1-t)))$ or may overestimates p in case of negative dependence (i. e., $P(X_1 > F_X^{-1}(1-t), X_{1+m} > F_X^{-1}(1-t)) \leq P(X_1 > F_X^{-1}(1-t))P(X_{1+m} > F_X^{-1}(1-t)))$.

In order to overcome this problem, Ledford and Tawn [24, 25] proposed a model to specify the "speed" of convergence of $P(X_1 > F_X^{-1}(1-t), X_{1+m} > F_X^{-1}(1-t))$ towards zero. Based on this approach, Ferreira and Ferreira introduced the lag-m *Ledford and Tawn coefficient*, η_m, such that

$$P(X_1 > F_X^{-1}(1-t), X_{1+m} > F_X^{-1}(1-t)) \sim t^{1/\eta_m} L_{\eta_m}(t), \tag{27}$$

as $t \downarrow 0$, where $\eta_m \in (0,1]$ and $L_{\eta_m}(t)$ is a slowly varying function at 0. Parameter η_m regulates the "speed" of convergence in (27) and $L_{\eta_m}(t)$ gives the relative "strength" of dependence within a particular value of η_m. The case $\eta_m = 1$ and $L_{\eta_m}(t) \to a > 0$, as $t \downarrow 0$, corresponds to asymptotic dependence (total dependence if $L_{\eta_m}(t) = 1$), otherwise X_1 and X_{1+m} are asymptotic independent. We also have positive dependence if $\eta_m > 1/2$, negative dependence if $\eta_m < 1/2$ and (almost) independence if $\eta_m = 1/2$ (perfect if $L_{\eta_m}(t) = 1$).

Observe that all these measures concern tail dependence based on extremal events of the type $\{X_1 > x\}$ for large x, i. e. an exceedance of a high level x. This is an extremal event widely used and applicable in literature but an adverse situation may also occur with other type of extremal events. For instance, suppose that a sea-wall projection requires a coastal defense from all sea-levels. An estimation of the probability of upcrossing the sea-wall height may be of interest. Thus, in an analogous way to λ_m and η_m, Ferreira and Ferreira [14] introduced coefficients μ_m and ν_m based on upcrossings events $\{X_1 \le x < X_2\}$ and $\{X_{2+m} \le x < X_{3+m}\}$. More precisely, we have the lag-m *upcrossings tail dependence coefficient*

$$\mu_m = \lim_{t\downarrow 0} P(X_{2+m} \le F_X^{-1}(1-t) < X_{3+m} | X_1 \le F_X^{-1}(1-t) < X_2), \tag{28}$$

and the ν_m-*upcrossings coefficient*

$$\begin{aligned} P(X_1 \le F_X^{-1}(1-t) < X_2, X_{2+m} \le F_X^{-1}(1-t) < X_{3+m}) \\ \sim P(X_1 \le F_X^{-1}(1-t) < X_2)^{1/\nu_m} L_{\nu_m}(t), \end{aligned} \tag{29}$$

as $t \downarrow 0$, with function $L_{\nu_m}(t)$ slowly varying at 0. The coefficient ν_m regulates the "speed" of convergence in (29) and $L_{\nu_m}(t)$ its relative "strength" given a particular value of ν_m. Analogous conclusions of the Ledford and Tawn coefficient η_m are derived, now concerning upcrossings. More precisely, when $\nu_m = 1$ and $L_{\nu_m}(t) \not\to 0$, we have asymptotic dependence of the upcrossings between random pairs (X_1,X_2) and (X_{2+m},X_{3+m}) (total dependence if $L_{\nu_m}(x) = 1$) and asymptotic independence otherwise. The cases $\nu_m > 1/2$ and $\nu_m < 1/2$ correspond to, respectively, positive and negative dependence, and $\nu_m = 1/2$ an (almost) independence (perfect if $L_{\nu_m}(t) = 1$).
According to Ferreira and Ferreira [14] (Proposition 3.3), we can also relate the upcrossings index ϑ with coefficients μ_m, for $m \in \mathbb{N}$.

We compute these measures for YARP(III)(1) process. In the following consider notation $a_t = F_X^{-1}(1-t)$.

Proposition 2.8. The YARP(III)(1) process has lag-m tail dependence coefficient $\lambda_m = p^m$.

Proof. We have

$$\begin{aligned}P(X_1 > a_t, X_{1+m} > a_t) &= \int_{a_t}^{\infty} P(X_{1+m} > a_t | X_1 = u)\,\mathrm{d}F_X(u)\\ &= \int_{a_t}^{\infty} [1 - Q^m(u,]0, a_t])]\,\mathrm{d}F_X(u)\\ &= \int_{a_t}^{\infty} \prod_{j=0}^{m-1} [\overline{F}_\varepsilon(p^{j/\alpha} a_t)(1-p) + p]\,\mathrm{d}F_X(u)\end{aligned}$$

where in last step we have applied (13). Considering the quantile function given in (20), we have

$$\begin{aligned}&1 - Q^m(u,]0, a_t]) = \prod_{j=0}^{m-1} [\overline{F}_\varepsilon(p^{j/\alpha} a_t)(1-p) + p]\\ &= \prod_{j=0}^{m-1} \left[\tfrac{t(1-p)}{t+p^j(1-t)} + p\right] = t + p^m(1-t).\end{aligned} \tag{30}$$

Thus being we obtain

$$\begin{aligned}&P(X_1 > a_t, X_{1+m} > a_t) = t + p^m(1-t) \int_{a_t}^{\infty} \mathrm{d}F_X(u)\\ &= t[t + p^m(1-t)] = t^2(1-p^m) + tp^m \sim tp^m.\end{aligned} \tag{31}$$

The result follows from (26). □

For curiosity, observe the power decay of λ_m, similar to the auto-correlation function of AR(1) processes.

Proposition 2.9. The YARP(III)(1) process has unit lag-m Ledford and Tawn coefficient, i. e., $\eta_m = 1$, for all positive integer m.

Proof. Straightforward from calculations of Proposition 2.8 and (27). □

This result is expected since our process is tail dependent, i. e., $\lambda_m = p^m > 0$ by Proposition 2.8.

In the next two propositions we compute, respectively, μ_m and ν_m for process YARP(III)(1).

Proposition 2.10. The YARP(III)(1) process has null lag-m upcrossings tail dependence coefficient, i. e., $\mu_m = 0$, for all positive integer m.

Proof. Applying the reasoning of (25) and (31) we obtain, respectively,

$$P(X_1 \le a_t < X_2) = t - t[t + p(1-t)] \tag{32}$$

and

$$P(X_2 > a_t, X_{3+m} > a_t) = t[t + p^{m+1}(1-t)].$$

Now observe that

$$\begin{aligned}&P(X_1 \le a_t < X_2, X_{2+m} \le a_t < X_{3+m})\\ &= P(X_2 > a_t, X_{3+m} > a_t) - P(X_1 > a_t, X_2 > a_t, X_{3+m} > a_t)\\ &-P(X_2 > a_t, X_{2+m} > a_t, X_{3+m} > a_t)\\ &+P(X_1 > a_t, X_2 > a_t, X_{2+m} > a_t, X_{3+m} > a_t).\end{aligned}$$

We have successively

$$\begin{aligned}
&P(X_1 > a_t, X_2 > a_t, X_{3+m} > a_t)\\
&= \int_{a_t}^{\infty} P(X_{3+m} > a_t, X_2 > a_t | X_1 = u_1)\, \mathrm{d}F_X(u_1)\\
&= \int_{a_t}^{\infty} \int_{a_t}^{\infty} P(X_{3+m} > a_t | X_2 = u_2) Q(u_1, du_2)\, \mathrm{d}F_X(u_1)\\
&= \int_{a_t}^{\infty} \int_{a_t}^{\infty} [1 - Q^{m+1}(u_2,]0, a_t])] Q(u_1, du_2)\, \mathrm{d}F_X(u_1).
\end{aligned}$$

Applying (30), we obtain

$$\begin{aligned}
&P(X_1 > a_t, X_2 > a_t, X_{3+m} > a_t)\\
&= t[t + p^{m+1}(1-t)] \int_{a_t}^{\infty} \int_{a_t}^{\infty} Q(u_1, du_2)\, \mathrm{d}F_X(u_1)\\
&= t[t + p^{m+1}(1-t)] \int_{a_t}^{\infty} [1 - Q(u_1,]0, a_t])]\, \mathrm{d}F_X(u_1)\\
&= [t + p^{m+1}(1-t)][t + p(1-t)]t.
\end{aligned}$$

A similar reasoning lead us to

$$P(X_2 > a_t, X_{2+m} > a_t, X_{3+m} > a_t) = [t + p^m(1-t)][t + p(1-t)]t$$

and

$$\begin{aligned}
&P(X_1 > a_t, X_2 > a_t, X_{2+m} > a_t, X_{3+m} > a_t)\\
&= [t + p^m(1-t)][t + p(1-t)]^2 t.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&P(X_1 \le a_t < X_2, X_{2+m} \le a_t < X_{3+m})\\
&= t[t + p^{m+1}(1-t)] - [t + p^{m+1}(1-t)][t + p(1-t)]t\\
&-[t + p^m(1-t)][t + p(1-t)]t + [t + p^m(1-t)][t + p(1-t)]^2 t\\
&= (1-p)^2(1-p^m)(1-t)^2 t^2.
\end{aligned} \tag{33}$$

By (32) and (33) we obtain

$$\frac{P(X_1 \le a_t < X_2, X_{2+m} \le a_t < X_{3+m})}{P(X_1 \le a_t < X_2)} = (1-p)(1-p^m)(1-t)t$$

and taking $t \downarrow 0$, the upcrossings tail dependence coefficient μ_m given in (28) is null. □

Proposition 2.11. The YARP(III)(1) process has lag-m coefficient $\nu_m = 1/2$, for all positive integer m.

P r o o f. By (33), we have

$$P(X_1 \leq a_t < X_2, X_{2+m} \leq a_t < X_{3+m}) \sim t^2(1-p)^2(1-p^m), \text{ as } t \downarrow 0.$$

and the result follows from (29), corresponding to tail upcrossings (almost) independence. □

2.3. YARP(III)(1) and ARMAX

The ARMAX process given in (4), with marginals $Fréchet(0,\sigma,\alpha)$, i. e., $F_X(x) = \exp(-(x/\sigma)^{-\alpha})$, and i.i.d. innovations $\{Z_i\}_{i\geq 1}$ with cdf, $F_Z(x) = F_X(x)/F_X(x/c) = \exp(-(x/\sigma)^{-\alpha}(1-c^\alpha))$, have a right tail behavior similar to the YARP(III)(1) process. More precisely, they are heavy-tailed processes belonging to the Fréchet(α) max-domain of attraction, with $\theta = 1 - c^\alpha$ and presenting the same mixing structure and local dependence conditions studied above (Alpuim [1], Canto e Castro [5]). They have also $\lambda_m = c^{m\alpha}$ (i. e., a power decay similar to the one obtained for YARP(III)(1)), $\eta_m = 1$, $\mu_m = 0$ and $\nu_m = 1/2$, for all $m \in \mathbb{N}$ (Ferreira and Ferreira [14]).Therefore, phenomena presenting an ARMAX tail behavior can also be modeled through a YARP(III)(1) process.

Moreover, based on the m-step transition probability function in ARMAX which is given by, $Q^m(x,]0, y]) = \prod_{i=0}^{m-1} F_Z(y/c^i)\mathbf{1}_{\{x<y/c^m\}}$, where $\mathbf{1}_{\{.\}}$ is the indicator function, we have

$$\begin{aligned} P(X_{n-1} < X_n) &= \int_0^\infty P(X_n > X_{n-1} | X_{n-1} = x)\, \mathrm{d}F_X(x) \\ &= \int_0^\infty (1 - Q(x,]0, x])\, \mathrm{d}F_X(x) = \int_0^\infty (1 - F_Z(x))\, \mathrm{d}F_X(x) = \frac{1 - c^{1/\alpha}}{2 - c^{1/\alpha}}. \end{aligned} \tag{34}$$

Now observe that every mentioned coefficients involving parameter c depends on α as well, including the fluctuation probabilities in (34). Hence, if we want to estimate the ARMAX parameter c we have also to estimate the tail index α, a drawback when compared with the YARP(III)(1) process (see (10)). Alternatively, we can consider the unit Fréchet ARMAX, i. e., with marginal cdf $F_X(x) = \exp(-1/x)$, by normalizing the values so that they have the standard Fréchet distribution (Lebedev [23]). This is achieved through the transformation, $-1/\log F_X(X)$, for which we still must have to estimate the parameters σ and α of F_X, and hence, once again an inclusion of error components in advance.

Therefore, YARP(III)(1) processes are more advantageous than ARMAX regarding data modeling.

3. ESTIMATION OF PROCESS PARAMETERS P AND α

We consider the estimation of the m-step fluctuation probabilities f_m in (14). There exist simple estimates for these probabilities:

$$\widehat{f}_m = \frac{1}{n-m} \sum_{j=m+1}^{n} \mathbf{1}_{\{X_{j-m} < X_j\}},\ m \geq 1. \tag{35}$$

The next result states consistency and asymptotic normality for estimators $\widehat{p^m}$ obtained from equation (14) by plugging in the empirical estimates $\widehat{f}_m$. Observe that $\widehat{p} \equiv \widehat{p^1}$ estimates the model parameter p.

Proposition 3.1. Let $\{X_i\}_{i\geq 1}$ be a stationary YARP(III)(1). Then, for each positive integer m,

$$n^{1/2}(\widehat{p^m} - p^m) \xrightarrow{D} N(0, 4\sigma_m^2) \tag{36}$$

where

$$\sigma_m^2 = f_m(1 - f_m)(1 - 2f_m + \chi_m)/(1 - \chi_m), \tag{37}$$

with f_m given in (14) and

$$\chi_m = \frac{P(X_{j-m} < X_j, X_{j-m-1} < X_{j-1})}{f_m}. \tag{38}$$

P r o o f. Observe that $\widehat{f}_m$ is the mean of Bernoulli trials with Markov dependence. From Klotz [19], we have that $n^{1/2}(\widehat{f}_m - f_m) \xrightarrow{D} N(0, \sigma_m^2)$ holds for σ_m^2 given in (37), where $\chi_m = P(X_{j-m} < X_j | X_{j-m-1} < X_{j-1})$ with $\max(0, (2f_m - 1)/f_m) \leq \chi_m \leq 1$. Hence, the result (36) is straightforward by the Delta Method. □

Note that $f_m \in [1/2, 1]$ and no definite results can be obtained for $\widehat{f_m} < 1/2$. However, the probability of such events goes to zero as $n \to \infty$ and hence, this may be an indication of an inconsistency in our choice of the model. In what concerns the lag m, it can be chosen in order to obtain the smallest variance (σ_m^2) provided that the estimate, $\widehat{f_m}$, takes value in $[1/2, 1]$.

Now we focus on process parameter, α, which can be estimated as the tail index of the marginal distribution of the YARP(III)(1) process. There are several estimators in literature such as, Hill estimator [16], Pickands' estimator [27], maximum likelihood estimator (Smith [31]), moments estimator (Dekkers et al. [10]), generalized weighted moments (Hosking and Wallis [17]), among others. Their properties have been derived in an i.i.d. framework, but there are some studies considering a stationary context (see, for instance, Rootzén et al. [30], Resnick and Stărică [28, 29], Drees [11]). The Hill estimator is the most used in heavy tails or Pareto-type tails which is our case. The β-mixing structure of YARP(III)(1) process, stated in Proposition 2.1, allows to conclude consistency and asymptotic normality of Hill estimator (Rootzén et al. [30]). In the example considered below, the sample paths of Hill estimator for parameter α can be seen in Figure 2.

3.1. An illustrative example

An illustration of the estimation procedure is now presented. We consider 5000 realizations from YARP(III)(1), for cases $p = 0.3, 0.5, 0.7, 0.9$, with marginal distribution Pareto(III)(0,1,1).

In order to obtain an estimate for the variance, we can replace in (37), f_m by $\widehat{f_m}$ stated in (35) and χ_m by the empirical counterpart

$$\widehat{\chi_m} = \frac{1}{n-m-1} \sum_{j=m+2}^{n} \mathbf{1}_{\{X_{j-m}<X_j, X_{j-m-1}<X_{j-1}\}} \Big/ \widehat{f_m} \tag{39}$$

or alternatively, use the estimator proposed by Klotz [19],

$$\widetilde{\chi_m} = \frac{r-\widehat{q_m}(2s-t)+(n-1)\widehat{f_m}+((r-\widehat{q_m}(2s-t)+(n-1)\widehat{f_m})^2+4r(1-2\widehat{f_m})(n-1)\widehat{f_m})^{1/2}}{2(n-1)(1-\widehat{f_m})} \tag{40}$$

where $\widehat{q_m} = 1-\widehat{f_m}$, $r = \sum_{i=2}^{n} x_i x_{i-1}$, $s = \sum_{i=1}^{n} x_i$ and $t = x_1 + x_n$, which is asymptotically equivalent to the maximum likelihood estimator. Again by Klotz [19], we have that $\widetilde{\chi_m}$ is consistent, more precisely, $\sqrt{n}(\chi_m - \widetilde{\chi_m}) \xrightarrow{D} N(0, \chi_m(1-\chi_m)/f_m)$. Results of estimation are summarized in Table 1.

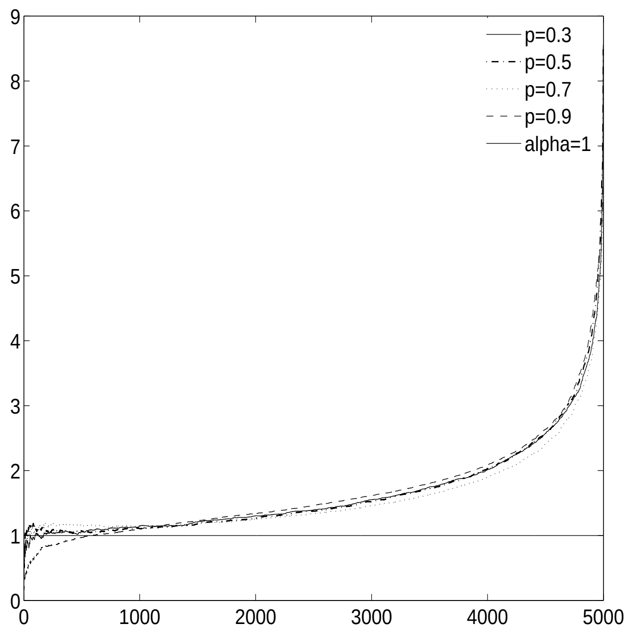

Fig. 2. Hill sample paths of YARP(III)(1) process, with marginal Pareto(III)(0,1,1), for $p = 0.3$, $p = 0.5$, $p = 0.7$ and $p = 0.9$.

	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$
p^m	0.3	0.09	0.027	0.0081	0.00243
$\widehat{p^m}$	0.2975	0.1012	0.02622	0.0052	—
IC($\widehat{\lambda}$)	(0.2886, 0.3063)	(0.0790, 0.123516)	(-0.0033, 0.0558)	(-0.0273,0.0377)	—
IC($\widetilde{\lambda}$)	(0.2886, 0.3063)	(0.0790, 0.1235)	(-0.0033,0.0558)	(0.0377,0.0052)	—
$\widehat{p}$ ($p=0.3$)	0.297459	0.3182	0.2971	0.2686	—
p^m	0.5	0.25	0.125	0.0625	0.03125
$\widehat{p^m}$	0.5011	0.2646	0.1255	0.0625	0.015015
IC($\widehat{\lambda}$)	(0.4926, 0.5096)	(0.2384, 0.2906)	(0.0879, 0.1630)	(0.0141, 0.1108)	(-0.0368,0.0669)
IC($\widetilde{\lambda}$)	(0.4925, 0.5097)	(0.2384, 0.2906)	(0.0879, 0.1631)	(0.0140, 0.1109)	(-0.0369, 0.0669)
$\widehat{p}$ ($p=0.5$)	0.5011	0.5143	0.5006	0.4999	0.4318
p^m	0.7	0.49	0.343	0.2401	0.16807
$\widehat{p^m}$	0.6903	0.4746	0.3212	0.2266	0.1572
IC($\widehat{\lambda}$)	(0.6839, 0.6968)	(0.4491, 0.5002)	(0.2764, 0.3660)	(0.1667,0.2865)	(0.0853, 0.2290)
IC($\widetilde{\lambda}$)	(0.6839, 0.6968)	(0.4490, 0.5002)	(0.2763, 0.3661)	(0.1666, 0.2866)	(0.0852, 0.2291)
$\widehat{p}$ ($p=0.7$)	0.6903	0.6889	0.6848	0.6899	0.6907
p^m	0.9	0.81	0.729	0.6561	0.59049
$\widehat{p^m}$	0.9084	0.8251	0.7519	0.6841	0.6244
IC($\widehat{\lambda}$)	(0.9060, 0.9108)	(0.8125, 0.8377)	(0.7242, 0.7795)	(0.6375, 0.7308)	(0.5581, 0.6908)
IC($\widetilde{\lambda}$)	(0.9060, 0.9108)	(0.8125,0.8378)	(0.7241, 0.7797)	(0.6373,0.7310)	(0.5578, 0.6911)
$\widehat{p}$ ($p=0.9$)	0.9084	0.9084	0.9093	0.9095	0.9101

Tab. 1. True values of p^m and of parameter p and respective estimates, considering $n = 5000$ realizations of the YARP(III)(1) process, with marginal Pareto(III)(0,1,1), for cases $p = 0.3, 0.5, 0.7, 0.9$; IC($\widehat{\lambda}$) and IC($\widetilde{\lambda}$) are 95% confidence intervals obtained, respectively, with σ^2 estimated using $\widehat{\lambda}$ given in (39) and $\widetilde{\lambda}$ given in (40); non filled cells mean that a $\widehat{f_m}$ less than 0.5 was obtained.

ACKNOWLEDGEMENT

This work was partially supported by the Research Centre of Mathematics of the University of Minho through the FCT Pluriannual Funding Program and by PTDC/FEDER grants from Portugal.

(Received November 2, 2010)

REFERENCES

[1] M. T. Alpuim: An extremal markovian sequence. J. Appl. Probab. *26* (1989), 219–232.

[2] B. C. Arnold: Pareto Distributions. International Cooperative Publishing House, Fairland 1983.

[3] B. C. Arnold: Pareto processes. In: Handbook of Statistics (D. N. Shanbhag and C. R. Rao, eds.), Elsevier Science B.V. 2001, Vol. 19.

[4] S. Asmussen: Applied Probability and Queues. John Wiley & Sons, Chichester 1987.

[5] L. Canto e Castro: Sobre a Teoria Assintótica de Extremos. Ph.D. Thesis, FCUL 1992.

[6] M. R. Chernick: A limit theorem for the maximum of autoregressive processes with uniform marginal distribution. Ann. Probab. *9* (1981), 145–149.

[7] M. R. Chernick, T. Hsing, and W. P. McCormick: Calculating the extremal index for a class of stationary sequences. Adv. Probab. *23* (1991), 835–850.

[8] R. Davis and S. Resnick: Basic properties and prediction of max-ARMA processes. Adv. Appl. Probab. *21* (1989), 781–803.

[9] D. J. Daley and J. Haslett: A thermal energy storage process with controlled input. Adv. Appl. Probab. *14* (1982), 257–271.

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

[11] H. Drees: Extreme quantile estimation for dependent data with applications to finance. Bernoulli *9* (2003), 617–657.

[12] H. Ferreira: The upcrossings index and the extremal index. J. Appl. Probab. *43(4)* (2006), 927–937.

[13] M. Ferreira and L. Canto e Castro: Modeling rare events through a pRARMAX process. J. Statist. Plann. Inference *140* (2010), 11, 3552–3566.

[14] M. Ferreira and H. Ferreira: On extremal dependence: some contributions. (In press).

[15] I. S. Helland and T. S. Nilsen: On a general random exchange model. J. Appl. Probab. *13* (1976), 781–790.

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

[17] J. R. M. Hosking and J. R. Wallis: Parameter and quantile estimation for the generalized Pareto distribution. Technometrics *29* (1987), 339–349.

[18] T. Hsing, J. Hüsler, and M. R. Leadbetter: On the exceedance point process for a stationary sequence. Probab. Theory Related Fields *78* (1988), 97–112.

[19] J. Klotz: Statistical inference in Bernoulli trials with dependence. Ann. Statist. *1* (1973), 373–379.

[20] M. R. Leadbetter: On extreme values in stationary sequences. Z. Wahrsch. verw. Gebiete *28* (1974), 289–303.

[21] M. R. Leadbetter, G. Lindgren, and H. Rootzén: Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag, New York 1983.

[22] M. R. Leadbetter and S. Nandagopalan: On exceedance point processes for stationary sequences under mild oscillation restrictions. In: Extreme Value Theory (J. Hüsler and R.-D. Reiss, eds.), Springer-Verlag 1989, pp. 69–80.

[23] A. V. Lebedev: Statistical analysis of first-order MARMA processes. Mat. Zametki *83* (2008), 4, 552–558.

[24] A. Ledford and J. A. Tawn: Statistics for near independence in multivariate extreme values. Biometrika *83* (1996), 169–187.

[25] A. Ledford and J. A. Tawn: Modelling dependence within joint tail regions. J. Royal Statist. Soc. Ser. B *59* (1997), 475–499.

[26] V. Pareto: Cours d'economie Politique. F. Rouge, Lausanne Vol. II., 1897.

[27] J. Pickands III: Statistical inference using extreme order statistics. Ann. Statist. *3* (1975), 119–131.

[28] S. Resnick and C. Stărică: Consistency of Hill's estimator for dependent data. J. Appl. Probab. *32* (1995), 139–167.

[29] S. Resnick and C. Stărică: Tail index estimation for dependent data. Ann. Appl. Probab. *8* (1998), 4, 1156–1183.

[30] H. Rootzén, M. R. Leadbetter, and L. de Haan: Tail and Quantile Estimation for Strongly Mixing Stationary Sequences. Technical Report, UNC Center for Stochastic Processes, 1990.

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

[32] H. C. Yeh, B. C. Arnold, and C. A. Robertson: Pareto processes. J. Appl. Probab. *25* (1988), 291–301.

[33] Z. Zhang and R. L. Smith: Modelling Financial Time Series Data as Moving Maxima Processes. Technical Report Dept. Stat. (Univ. North Carolina, Chapel Hill, NC, 2001); http://www.stat.unc.edu/faculty/rs/ papers/RLS_Papers.html.

Marta Ferreira, Center of Mathematics of Minho University, Campus de Gualtar 4710 - 057 Braga, Portugal.

e-mail: msferreira@math.uminho.pt