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

Marta Ferreira

U. Minho/DMA-CMAT, Braga - Portugal

msferreira@math.uminho.pt

Abstract: 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, in what concerns extreme values modeling (Davis and Resnick 1989 [8], Ferreira and Canto e Castro 2010 [13]). In this paper we present a complete characterization of the tail behavior of the autoregressive Pareto process, *Yeh-Arnold-Robertson Pareto(III)* (Yeh *et al.* 1988 [31]). 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.

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,

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

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

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

($G_0(x)=exp(-e^{-x})$) 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 on 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.,

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

where L 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 1989 [8]), and also the particular case MAR(1) or ARMAX, given by,

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

with $0 < c < 1$ and $\{Z_i\}_{i \geq 1}$ i.i.d. (Alpuim 1989 [1]; Canto e Castro 1992 [5]; Lebedev 2008 [22]) 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 are more convenient for analysis as their finite-dimensional distributions can easily be written explicitly (Davis and Resnick 1989 [8]). Actually, MARMA processes and their generalizations have been applied to various phenomena, e.g., a solar thermal energy storage system (Daley and Haslett 1982 [9]), the water density in a sill fjord (Helland and Nielsen 1976 [15]) or financial series (Zhang and Smith 2001[32]). Heavy tailed power max-autoregressive processes have also been developed with successful application to financial time series modeling (Ferreira and Canto e Castro 2010 [13]).

Here we shall focus on 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 (1897) [25] 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, 1983 [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.* 1988 [31]). We shall characterize the right tail behavior and conclude that it is similar to the process ARMAX (Section 2.1). 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. 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 $X_0 \frown \mathscr{P}(III)(0,\sigma,\alpha)$. 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

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

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.* (1988) [31] 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 1975 [16]) has good properties such as consistency and

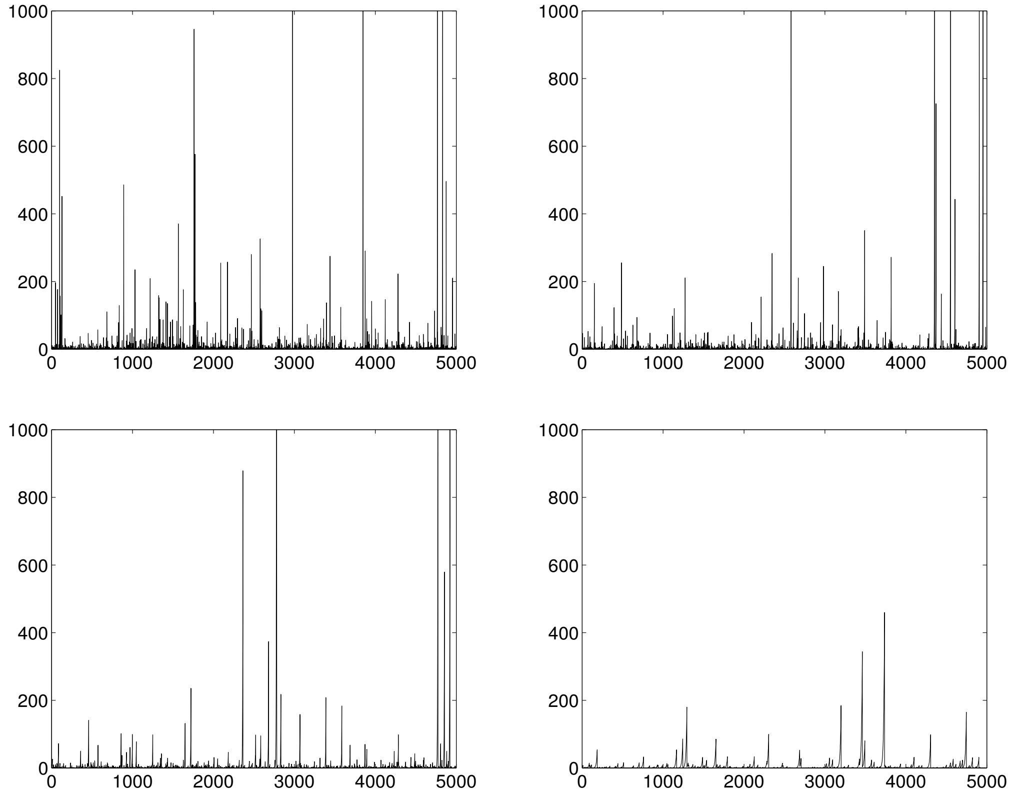

Figure 1: Simulated sample paths of YARP(III)(1) processes with marginals $\mathscr{P}(III)(0,1,1)$ for $p = 0.3$ (top-left), $p = 0.5$ (top-right), $p = 0.7$ (bottom-left), $p = 0.9$ (bottom-right)

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 $T_n \stackrel{d}{=} \min_{i \le N} \varepsilon_i$, where ε_i, $i \ge 1$, are i.i.d. Pareto(III)$(0, \sigma, \alpha)$ and N, independent of ε_i is such that $N - 1 \frown Binomial(n, 1-p)$. It follows by conditioning on N that

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

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 state 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

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

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, 2001 [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:

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

Similarly, we derive the m-step tpf:

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

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:

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

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 stated 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

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

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

$$Q^m(x,B) \geq \epsilon\lambda(B),\ x \in \mathbb{R} \tag{15}$$

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

$$Q^{m+1}(x,B) \geq \epsilon_1\lambda(B) \ \text{ and } \ Q^m(x,B) \geq \epsilon_2\lambda(B),\ \forall x \in \mathbb{R}, \tag{16}$$

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 (2010) [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

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

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

$$Q(x,B) \geq \int_{B\cap S} dQ(x,z) = \int_{B\cap S} (1-p)dF_\varepsilon(z) = \epsilon\lambda(B), \tag{17}$$

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 (17) still holds. Hence condition (15) holds. Observe now that

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

and by (12),

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

with $\epsilon_1 = \epsilon(1-p)P(\varepsilon \in S)$. Hence, condition (16) 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 (1974) [19] 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 (1981) [6], if for each $\tau > 0$ there is a real sequence $\{u_n\}_{n\geq 1}$ satisfying

$$n(1 - F_X(u_n)) \to \tau,\ n \to \infty, \tag{18}$$

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 $\theta = 1 - p$.*

Proof. First observe that the quantile function is given by

$$F_X^{-1}(t) = \sigma((1-t)^{-1} - 1)^{1/\alpha} \tag{19}$$

and that real levels $\{u_n\}_{n\geq 1}$ satisfying (18) 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.* (1983) [20] 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 2.1. *Condition $D'(u_n)$ 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*

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

This also means that a (right) tail dependence takes place, i.e., the coefficient of tail dependence

$$\lambda = \lim_{x \to x_F} P(Y > x | X > x), \tag{20}$$

is positive, provided the limit exists (x_F is the right endpoint). If $\lambda = 0$ the variables are said to be tail independent. Loosely stated, λ is the probability of one variable being extreme given that the other is extreme. According to Ferreira and Ferreira (2010) [14] (Proposition 3.2), we can relate θ with λ, as we shall see below.

Several local dependence conditions provide formulas for the computation of θ. For instance, the family of conditions $D^{(k)}(u_n)$, for $k \geq 1$ (Chernick *et al.* 1991 [7]) are sufficient to

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

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

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

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 1989 [21], defined below.

Definition 2.2. *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*

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

We shall see that the YARP(III)(1) process satisfies $D''(u_n)$ and hence $D^{(2)}(u_n)$.

Under condition $D^{(2)}(u_n)$ we are able to conclude that the tail coefficient λ stated in (20) is given by p (we have $\lambda = 1-\theta$, as can be seen in Ferreira and Ferreira (2010) [14], Proposition 3.2).

Proposition 2.3. *Condition $D''(u_n)$ holds for process YARP(III)(1), for levels u_n satisfying (18).*

Proof. Observe that

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

where we have considered $\min(p^{-\frac{j-1}{\alpha}} X_1, \frac{p^{-\frac{j-2}{\alpha}}\varepsilon_2}{1-U_2}, ..., \frac{\varepsilon_j}{1-U_j}) \ne p^{-\frac{j-1}{\alpha}} X_1$ and $U_k \ne 1$, 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 (18) 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 $F_\varepsilon(p^{\frac{j-k}{\alpha}} u_n) \le F_\varepsilon(u_n)$, we have

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

□

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 (2006) [12]). 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 (2006) [12] states the upcrossings index, $\vartheta\in[0,1]$, a measure for clustering of upcrossings of high levels. The family $\widetilde{D}^{(k)}(u_n)$, for $k\ge 1$, provide a way to compute ϑ too.

In what concerns tail dependence, a similar reasoning applied to coefficient λ, i.e., replacing exceedances by upcrossings, lead us to the upcrossings tail dependence coefficient

$$
\mu=\lim_{x\to x_F} P(Y_1\le x<Y_2|X_1\le x<X_2). \tag{24}
$$

provided the limit exist.

Proposition 2.4. *In YARP(III)(1) processes the upcrossings tail dependence coefficient μ is null.*

Proof. Under condition $D''(u_n)$ for levels u_n satisfying $nP(X_1\le u_n<X_2)\to\varsigma$ as $n\to\infty$, the upcrossings tail dependence coefficient μ is

null (Ferreira and Ferreira 2010, [14], Proposition 3.1). This happens to be our case since $D''(u_n)$ holds for levels satisfying (18) which imply $nP(X_1 \leq u_n < X_2) \to \varsigma$, with $\varsigma = \tau(1-p)$:

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

□

Corollary 2.5. *In YARP(III)(1) processes the upcrossings index ϑ is unit.*

Proof. Straightforward by Proposition 3.3 in Ferreira and Ferreira 2010, [14] since $\mu = 0$ as stated in Proposition 2.4. □

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

In order to graduate the “strength" of dependence in the tail, in the case of asymptotic independence (i.e., $\lambda = 0$), Ledford and Tawn [23, 24] (1996, 1997) have considered the following formulation that states the rate of convergence towards zero in the tail:

$$P(X > x, Y > x) \underset{x \to x_F}{\sim} P(X > x)^{1/\eta_{X,Y}} L_{\eta_{X,Y}}(1/P(X > x)), \tag{26}$$

where $L_{\eta_{X,Y}}$ is a slowly varying function at ∞. Coefficient $\eta_{X,Y} \in (0, 1]$ is usually denoted “*Ledford and Tawn coefficient*". In case of tail dependence (i.e., $\lambda > 0$) we have $\eta = 1$ and we have asymptotic independence when $\eta_{X,Y} < 1$. If $\eta_{X,Y} > 1/2$ we have positive extremal dependence, $\eta_{X,Y} < 1/2$ corresponds to negative dependence and $\eta_{X,Y} = 1/2$ to (almost) independence.

Similarly, Ferreira and Ferreira 2010, [14] consider a formulation stating the convergence rate of $P(Y_1 \leq x < Y_2, X_1 \leq x < X_2)$ to 0, as $x \to x_F$, in order to graduate the “strength" of dependence in the tail within asymptotic upcrossings independence. More precisely,

$$P(X_1 \leq x < X_2, Y_1 \leq x < Y_2) \underset{x \to x_F}{\sim} P(X_1 \leq x < X_2)^{1/\nu_{Y|X}} L_{\nu_{Y|X}}(1/P(X_1 \leq x < X_2)), \tag{27}$$

where $L_{\nu_{Y|X}}$ is a slowly varying function at ∞, and similar conclusions as for exceedances concerning the Ledford and Tawn coefficient $\eta_{X,Y}$, are also derived.

The above mentioned tail dependence measures can be stated 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 2010, [14]). This formulation is interesting for model diagnosis purposes, similar to the role of autocorrelation function in linear models.
Considering marginal uniform normalization, we have the lag-m *tail dependence coefficient*,

$$\lambda_m = \lim_{t \downarrow 0} P(X_{1+m} > F_X^{-1}(1-t) | X_1 > F_X^{-1}(1-t)), \tag{28}$$

the lag-m *Ledford and Tawn coefficient*, η_m,

$$P(X_1 > F_X^{-1}(1-t), X_{1+m} > F_X^{-1}(1-t)) \sim t^{1/\eta_m} L_m(t), \text{ as } t \downarrow 0 \tag{29}$$

where $L_m(t)$ is a slowly varying function at 0, the lag-m *upcrossings tail dependence coefficient*

$$\mu_m = \lim_{t \downarrow 0} P(X_{2+m} \leq F_X^{-1}(1-t) < X_{3+m} | X_1 \leq F_X^{-1}(1-t) < X_2), \tag{30}$$

and also

$$P(X_1 \leq F_X^{-1}(1-t) < X_2, X_{2+m} \leq F_X^{-1}(1-t) < X_{3+m}) \sim P(X_1 \leq F_X^{-1}(1-t) < X_2)^{1/\nu_m} L_{\nu_m}(t), \tag{31}$$

as $t \downarrow 0$, with function $L_{\nu_m}(t)$ slowly varying at 0.

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

Proposition 2.6. *The YARP(III)(1) process has lag-m tail dependence coefficient* $\lambda_m = p^m$.

Proof. We have

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

where in last step we have applied (13). Considering the quantile function given in (19), we have

$$\begin{aligned} &1 - Q^m(u,]0, a_t]) = \prod_{j=0}^{m-1} [\overline{F}_\varepsilon(p^{j/\alpha} a_t)(1-p) + p] \\ &= \prod_{j=0}^{m-1} \left[\tfrac{t(1-p)}{t + p^j(1-t)} (1-p) + p \right] = t + p^m(1-t). \end{aligned} \tag{32}$$

Thus being we obtain

$$\begin{aligned} &P(X_1 > a_t, X_{1+m} > a_t) = t + p^m(1-t) \int_{a_t}^{\infty} dF_X(u) \\ &= t[t + p^m(1-t)] = t^2(1-p^m) + tp^m \sim tp^m. \end{aligned} \tag{33}$$

The result follows from (28). □

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

Proposition 2.7. *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 2.6 and (29). □

This result is expected since our process is tail dependent, i.e., $\lambda_m = p^m > 0$ by Proposition 2.6.

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

Proposition 2.8. *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 (33) we obtain, respectively,

$$P(X_1 \le a_t < X_2) = t - t[t + p(1-t)] \tag{34}$$

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

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

Applying (32), we obtain

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

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

$$\begin{aligned}
&P(X_1 \le a_t < X_2, X_{2+m} \le a_t < X_{3+m}) \\
&= t[t + p^{m+1}(1-t)] - [t + p^{m+1}(1-t)][t + p(1-t)]t \\
&-[t + p^m(1-t)][t + p(1-t)]t + [t + p^m(1-t)][t + p(1-t)]^2 t \\
&= (1-p)^2(1-p^m)(1-t)^2 t^2
\end{aligned} \tag{35}$$

By (34) and (35) 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 (30) is null. □

Proposition 2.9. *The YARP(III)(1) process has lag-m coefficient* $\nu_m = 1/2$*, for all positive integer* m*.*

Proof. By (35), we have

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

and the result follows from (31), corresponding to tail upcrossings (almost) independence. □

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 $\{Z_i\}_{i \ge 1}$ 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 1989[1] and Canto e Castro, L. (1992) [5]). They have also $\lambda_m = c^{m\alpha}$, $\eta_m = 1$, $\mu_m = 0$ and $\nu_m = 1/2$, for all $m \in \mathbb{N}$ (Ferreira and Ferreira (2010) [14]). Moreover, based on the m-step transition probability function given by, $Q^m(x,]0, y]) = \prod_{i=0}^{m-1} F_Z(y/c^i)\mathbf{1}_{\{x<y/c^m\}}$, where $\mathbf{1}_{\{\cdot\}}$ is the indicator function, we have

$$P(X_{n-1} < X_n) = \int_0^\infty P(X_n > X_{n-1} | X_{n-1} = x) dF_X(x) =$$

$$\int_0^\infty (1 - Q(x,]0, x]) dF_X(x) = \int_0^\infty (1 - F_Z(x)) dF_X(x) = \frac{1 - c^{1/\alpha}}{2 - c^{1/\alpha}}.$$

Every mentioned coefficients involving parameter c depends on α as well. 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 2008 [22]). 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:

$$\widehat{f}_m = \frac{1}{n-m}\sum_{j=m+1}^{n} \mathbf{1}_{\{X_{j-m}<X_j\}},\ m \geq 1. \tag{36}$$

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,*

$$n^{1/2}(\widehat{p^m} - p^m) \xrightarrow{D} N(0, 4\sigma_m^2) \tag{37}$$

where

$$\sigma_m^2 = f_m(1-f_m)(1-2f_m+\chi_m)/(1-\chi_m), \tag{38}$$

with f_m given in (14) and

$$\chi_m = \tfrac{P(X_{j-m}<X_j, X_{j-m-1}<X_{j-1})}{f_m}. \tag{39}$$

Proof. Observe that $\widehat{f}_m$ is the mean of Bernoulli trials with Markov dependence. From Klotz (1973) [18], 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 (38), 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 (37) is straightforward by the Delta Method. $\square$

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, the tail index of the marginal distribution of the YARP(III)(1) process. There are several estimators in literature such as, Hill estimator (1975) [16], Pickands' estimator (1975) [26], maximum likelihood estimator (Smith 1987 [30]), moments estimator (Dekkers *et al.* 1989 [10]), generalized weighted moments (Hosking and Wallis 1987 [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.* 1990 [29], Resnick and Stărică (1995, 1996) [27, 28], Drees (2003) [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.* 1990 [29]). 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 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 (38), f_m by $\widehat{f_m}$ stated in (36) and χ_m by the empirical counterpart

$$\widehat{\chi_m} = \frac{1}{n-m-1} \sum_{j=m+2}^{n} \mathbf{1}_{\{X_{j-m}<X_j, X_{j-m-1}<X_{j-1}\}} \Big/ \widehat{f_m}. \tag{40}$$

	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$
p^m	0.3	0.09	0.027	0.0081	0.00243
$\widehat{p^m}$	0.297459	0.10124	0.026216	0.005204	—
IC($\widehat{\lambda}$)	(0.288591, 0.306328)	(0.078965, 0.123516)	(-0.00334, 0.055773)	(-0.0273,0.037711)	—
IC($\widetilde{\lambda}$)	(0.288586, 0.306333)	(0.07897, 0.123511)	(-0.00334,0.055772)	(0.037717,0.005204)	—
$\widehat{p}$ ($p=0.3$)	0.297459	0.318183	0.297067	0.268589	—
p^m	0.5	0.25	0.125	0.0625	0.03125
$\widehat{p^m}$	0.5011	0.264506	0.125475	0.06245	0.015015
IC($\widehat{\lambda}$)	(0.492552, 0.509649)	(0.238415, 0.290597)	(0.087946, 0.163005)	(0.01409, 0.11081)	(-0.03683,0.06686)
IC($\widetilde{\lambda}$)	(0.492532, 0.509668)	(0.238369, 0.290642)	(0.087874, 0.163077)	(0.013983, 0.110917)	(-0.0369, 0.066926)
$\widehat{p}$ ($p=0.5$)	0.5011	0.514301	0.500633	0.4999	0.431822
p^m	0.7	0.49	0.343	0.2401	0.16807
$\widehat{p^m}$	0.690338	0.47459	0.321193	0.226581	0.157157
IC($\widehat{\lambda}$)	(0.683922, 0.696754)	(0.449006, 0.500174)	(0.276366, 0.366019)	(0.166672,0.286491)	(0.085266, 0.229049)
IC($\widetilde{\lambda}$)	(0.683919, 0.696757)	(0.448991, 0.500189)	(0.276332, 0.366054)	(0.166613, 0.28655)	(0.085179, 0.229135)
$\widehat{p}$ ($p=0.7$)	0.690338	0.688905	0.684839	0.689932	0.690664
p^m	0.9	0.81	0.729	0.6561	0.59049
$\widehat{p^m}$	0.908382	0.82513	0.751851	0.684147	0.624424
IC($\widehat{\lambda}$)	(0.905993, 0.91077)	(0.812521, 0.837739)	(0.724173, 0.77953)	(0.637454, 0.73084)	(0.558071, 0.690777)
IC($\widetilde{\lambda}$)	(0.905971, 0.910792)	(0.812461,0.837799)	(0.724053, 0.77965)	(0.63725,0.731044)	(0.557768, 0.691081)
$\widehat{p}$ ($p=0.9$)	0.908382	0.908367	0.909307	0.909468	0.910114

Table 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 (40) and $\widetilde{\lambda}$ given in (41); non filled cells mean that a $\widehat{f_m}$ less than 0.5 was obtained.

or alternatively, use the estimator proposed by Klotz (1973) [18],

$$\widetilde{\chi_m} = \frac{r-\widehat{q_m}(2s-t)+(n-1)\widehat{f_m}+((r-\widehat{q_m}(2s-t)+(n-1)\widehat{f_m})^2+4r(1-2\widehat{f_m})(n-1)\widehat{f_m})^{1/2}}{2(n-1)(1-\widehat{f_m})} \quad (41)$$

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 (1973) [18], 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.

References

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

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

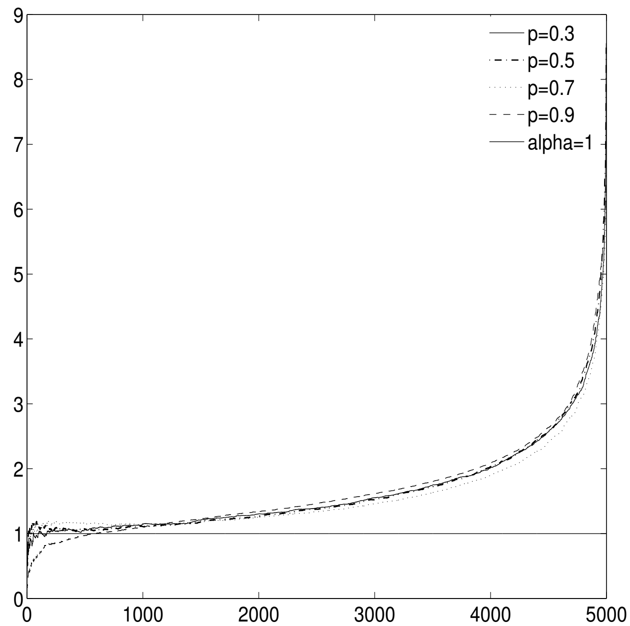

Figure 2: Hill sample paths of YARP(III)(1) process, with marginal Pareto(III)(0,1,1), for $p = 0.3$, $p = 0.5$, $p = 0.7$ and $p = 0.9$.

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

[4] Asmussen, S. (1987). *Applied Probability and Queues.* John Wiley & Sons.

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

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

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

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

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

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

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

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

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

[14] Ferreira, M., Ferreira, H. (2010). On extremal dependence: some contributions. Notes and Communications CEAUL 01/10.

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

[16] Hill, B.M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *Ann. Stat.* 3, 1163-1174.

[17] Hosking, J.R.M., Wallis, J.R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics* 29, 339-349.

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

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

[20] Leadbetter, M.R., Lindgren, G., Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, New-York, Heidelberg-Berlin.

[21] Leadbetter, M.R., Nandagopalan, S. (1989) On Exceedance Point Processes for Stationary Sequences under Mild Oscillation Re-

strictions. *Extreme Value Theory.* (J. Hüsler e R.-D. Reiss, eds.), Springer-Verlag, 69-80.

[22] Lebedev, A.V. (2008). Statistical Analysis of First-Order MARMA Processes. *Mat. Zametki* 83(4), 552-558.

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

[24] Ledford, A. and Tawn, J.A. (1997). Modelling Dependence within joint tail regions. *J. Royal Statist. Soc.Ser. B* 59, 475-499.

[25] Pareto, V. (1897). *Cours d'economie Politique*, Vol.II. F. Rouge, Lausanne.

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

[27] Resnick, S., Stărică, C. (1995). Consistency of Hill's estimator for dependent data. J Appl. Probab. 32, 139-167.

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

[29] Rootzén, H., Leadbetter, M.R., De Haan, L. (1990). Tail and Quantile Estimation for Strongly Mixing Stationary Sequences. Technical rept, UNC Center for Stochastic Processes.

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

[31] Yeh, H.C., Arnold, B.C., Robertson, C.A. (1988). Pareto Processes. *J. Appl. Probab.* 25, 291-301.

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