Tail Dependence of a Pareto Process

Marta Ferreira

Abstract

Heavy-tailed autoregressive processes defined with minimum or maximum operator are good alternatives to classic linear ARMA with heavy tail noises, in what concerns extreme values modeling. In this paper we present a full characterization of the tail dependence of the autoregressive minima process, Yeh–Arnold–Robertson Pareto(III).

1 Introduction

Extreme value theory (EVT) provides tools that enable to estimate the probability of events that are more extreme than any that have already been observed. The classical result in EVT states that if the maximum of an independent and identically distributed (i.i.d.) sequence of random variables (r.v.'s) converges to some nondegenerate function G_γ, then it must be the generalized extreme value (*GEV*) function,

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

with the usual continuity correction $G_0(x) = exp(-e^{-x})$. The shape parameter γ, known as the tail index, determines the tail behavior: 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).

M. Ferreira (✉)
University of Minho, DMA/CMAT, Braga, Portugal
e-mail: msferreira@math.uminho.pt

The first results in EVT were developed under independence but, more recently, models for extreme values have been constructed under more realistic assumption of temporal dependence.

MARMA processes (*maximum autoregressive moving average*) with Fréchet marginals, in particular ARMAX [or MARMA(1,0)], given by,

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

with $0 < c < 1$ and $\{W_i\}_{i \geq 1}$ i.i.d., have been successfully applied to time series modeling in alternative to classical linear heavy-tailed ARMA (see [2] and references therein). Generalizations of MARMA processes and respective applications to financial time series can be seen in, e.g., [13] and [4]. Here we shall focus on autoregressive Pareto processes, i.e., an autoregressive process whose marginal distributions are of the Pareto or generalized Pareto form. As Pareto observed [10], 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, whose tail function decreases at a negative power of x as $x \to \infty$, i.e., $1 - F(x) \sim cx^{-\alpha}$, as $x \to \infty$. Generalizations of Pareto's distribution have been proposed for modeling economic variables (a survey can be seen in [1]).

We consider autoregressive Pareto(III) processes, more precisely, the Yeh–Arnold–Robertson Pareto(III) [12], in short YARP(III)(1), given by

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

where innovations $\{\varepsilon_n\}_{n \geq 1}$ are i.i.d. r.v.'s with distribution Pareto(III)$(0,\sigma,\alpha)$, i.e., a generalized type III Pareto, such that

$$1 - F_\varepsilon(x) = \left[1 + \left(\frac{x}{\sigma}\right)^\alpha\right]^{-1}, \quad x > 0.$$

with $\sigma, \alpha > 0$. The sequence $\{U_n\}_{n \geq 1}$ has i.i.d. r.v.'s with a Bernoulli(p) distribution (independent of the innovations). We interpret $1/0$ as $+\infty$. By conditioning on U_n, it is readily verified that the YARP(III)(1) process has a Pareto(III)$(0,\sigma,\alpha)$ stationary distribution and will be completely stationary if the distribution of the starting r.v. X_0 is also Pareto(III)$(0,\sigma,\alpha)$.

In this paper we analyze the dependence behavior of the YARP(III)(1) process in the right tail (the most used for applications). This process is almost unknown in literature but has large potential as it presents a quite similar tail behavior to ARMAX and more robust parameters estimation [3]. We characterize the lag-m tail dependence ($m = 1, 2, \ldots$) by computing several coefficients considered in [5, 6], defined under a temporal approach. The lag-m tail dependence allows a characterization of the process in time, analogous to the role of the ACF of a linear time series. In addition, these measures are also important in applications, such as risk assessment in financial time series or in engineering, to investigate how the best performer in a system is attracted by the worst one.

2   Measures of Tail Dependence

The *tail-dependence coefficient* (TDC), usually denoted λ, was the first tail dependence concept appearing in literature in a paper by Sibuya, who has shown that, no matter how high we choose the correlation of normal random pairs, if we go far enough into the tail, extreme events tend to occur independently in each margin [11]. It measures the probability of occurring extreme values for one r.v. given that another assumes an extreme value too. More precisely,

$$\lambda = \lim_{t \downarrow 0} P(F_1(X_1) > 1 - t | F_2(X_2) > 1 - t),$$

where F_1 and F_2 are the distribution functions (d.f.'s) of r.v.'s X_1 and X_2, respectively. It characterizes the dependence in the tail of a random pair (X_1, X_2), in the sense that, $\lambda > 0$ corresponds to tail dependence whose degree is measured by the value of λ, whereas $\lambda = 0$ means tail independence. Modern risk management is highly interested in assessing the amount of tail dependence. As an example, the Value-at-Risk at probability level $1 - t$ (VaR$_{1-t}$) of a random asset Z is given by the quantile function evaluated at $1 - t$, $F_Z^{-1}(1 - t) = \inf\{x : F_Z(x) \geq 1 - t\}$, and estimation is highly sensitive towards the tail behavior and the tail dependence of the portfolio's asset-return distribution. Observe that the TDC can be formulated as

$$\lambda = \lim_{t \downarrow 0} P(X_1 > VaR_{1-t}(X_1) | X_2 > VaR_{1-t}(X_2)).$$

Generalizations of the TDC have been considered with several practical applications. In [6], for integers s and k such that $1 \leq s < d - k + 1 \leq d$, it was considered the *upper s, k-extremal coefficient* of random vector $\mathbf{X} = (X_1, \ldots, X_d)$, defined by

$$\lambda_U(X_{s:d} | X_{d-k+1:d}) \equiv \lambda_U(U_{s:d} | U_{d-k+1:d})$$
$$= \lim_{t \downarrow 0} P(U_{s:d} > 1 - t | U_{d-k+1:d} > 1 - t),$$

where $U_{1:d} \leq \ldots \leq U_{d:d}$ are the order statistics of $(F_1(X_1), \ldots, F_d(X_d))$ and $X_{i:d}$ the inverse probability integral transform of $U_{i:d}$. In engineering, the coefficient $\lambda_U(X_{s:d} | X_{d-k+1:d})$ can be interpreted as the limiting probability that the sth worst performer in a system is attracted by the kth best one, provided the latter has an extremely good performance. In mathematical finance, $\lambda_U(X_{s:d} | X_{d-k+1:d})$ can be viewed as the limiting conditional probability that $X_{s:d}$ violates its value-at-risk at level $1 - t$, given that $X_{d-k+1:d}$ has done so. If $s = k = 1$, we obtain the *upper extremal dependence coefficient*, ϵ^U, considered in [7].

The study of systemic stability is also an important issue within the context of extreme risk dependence. The fragility of a system has been addressed through the *Fragility Index* (FI) introduced in [8]. More precisely, consider a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with d.f. F and $N_x := \sum_{i=1}^{d} \mathbf{1}_{\{X_i > x\}}$ the number of exceedances among $X_1, \ldots, X_d$ above a threshold x. The FI corresponding to $\mathbf{X}$

is the asymptotic conditional expected number of exceedances, given that there is at least one exceedance, i.e., $FI = \lim_{x \to \infty} E(N_x | N_x > 0)$. The stochastic system $\{X_1, \ldots, X_d\}$ is called fragile whenever $FI > 1$. In [5] it can be seen as a generalization of the FI that measures the stability of a stochastic system divided into blocks. More precisely, the block-FI of a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ relative to a partition $\mathscr{D} = \{I_1, \ldots, I_s\}$ of $D = \{1, \ldots, d\}$ is

$$FI(\mathbf{X}, \mathscr{D}) = \lim_{x \to \infty} E(N_{\mathbf{x}} | N_{\mathbf{x}} > 0),$$

where $N_{\mathbf{x}}$ is the number of blocks where it occurs at least one exceedance of $\mathbf{x}$, i.e.,

$$N_{\mathbf{x}} = \sum_{j=1}^{s} \mathbf{1}_{\{\mathbf{X}_{I_j} \not\leq \mathbf{x}_{I_j}\}},$$

and where $\mathbf{X}_{I_j}$ is a sub-vector of $\mathbf{X}$ whose components have indexes in I_j, with $j = 1, \ldots, s$ (i.e., $\mathbf{X}_{I_j}$ is the jth block of random vector $\mathbf{X}$) and $\mathbf{x}_{I_j}$ is a vector of length $|I_j|$ with components equal to $x \in \mathbf{R}$. Observe that if we consider a partition $\mathscr{D}^* = \{I_j = \{j\} : j = 1, \ldots d\}$, then the coefficient $FI(\mathbf{X}, \mathscr{D}^*)$ is the FI introduced in [8]. All operations and inequalities on vectors are meant componentwise.

Here we shall consider the abovementioned tail dependence coefficients defined in a time series perspective. More precisely, consider a stationary process $\{X_i\}_{i \geq 1}$ with marginal d.f. F_X. The lag-m TDC ($m = 1, 2, \ldots$) is given by

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

measuring the probability of occurring one extreme value observation given that another assumes an extreme value too, whenever separated in time by a lag-m. Analogously, we define the *lag-m upper s, k-extremal coefficient*,

$$\lambda_U(X_{s:m} | X_{m-k+1:m}) \equiv \lambda_U(U_{s:m} | U_{m-k+1:m})$$
$$= \lim_{t \downarrow 0} P(U_{s:m} > 1 - t | U_{m-k+1:m} > 1 - t),$$

a measure of the probability that, for a horizon of m successive time instants, the sth worst performer is attracted by the kth best one, provided the latter has an extremely good performance. If $s = k = 1$, we obtain the *lag-m upper extremal dependence coefficient*, ϵ_m^U. Finally, the lag-m block-FI relative to a partition $\mathscr{D}_m$ of $D_m = \{1, \ldots, m\}$ is

$$FI(\mathbf{X}, \mathscr{D}_m) = \lim_{x \to \infty} E(N_{\mathbf{x}} | N_{\mathbf{x}} > 0),$$

where, for a horizon of m successive time instants, $N_{\mathbf{x}}$ is the number of blocks where it occurs at least one exceedance of $\mathbf{x}$. Hence it measures the stability within

m successive time instants of a stochastic process divided into blocks. Analogously we define the $FI(\mathbf{X}, \mathscr{D}_m^*)$ for a partition $\mathscr{D}^* = \{I_j = \{j\} : j = 1, \ldots m\}$ as the lag-m FI version of [8].

3 Tail Dependence of YARP(III)(1)

In this section we shall present a characterization of the dependence structure and tail behavior of the YARP(III)(1) process. We start with the reference to some existing results and then we compute the above mentioned measures.

In order to determine the distribution of the maximum, $M_n = \max_{0 \le i \le n} X_i$, it is convenient to consider a family of level crossing processes $\{Z_n(x)\}$ indexed by $x > 0$, defined by

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

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

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

Hence, we have

$$F_{M_n}(x) = P(M_n \le x) = P(Z_0(x) = 0, Z_1(x) = 0, \ldots, Z_n(x) = 0)$$
$$= P(X_0 \le x)P(Z_i(x) = 0|Z_{i-1}(x) = 0)^n = \frac{\left(\tfrac{x}{\sigma}\right)^\alpha}{1 + \left(\tfrac{x}{\sigma}\right)^\alpha}\left(\frac{p + \left(\tfrac{x}{\sigma}\right)^\alpha}{1 + \left(\tfrac{x}{\sigma}\right)^\alpha}\right)^n$$

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

In [3] it was proved that the YARP(III)(1) process presents a β-mixing dependence structure. Hence, it satisfies the local dependence condition $D(u_n)$ of Leadbetter [9] for any real sequence $\{u_n\}_{n \ge 1}$ and so, for each $\tau > 0$ such that $n(1 - F_X(u_n)) \to \tau$, as $n \to \infty$, we have $P(M_n \le u_n) \to e^{-\theta \tau}$ as $n \to \infty$, with $\theta = 1 - p$ (Proposition 2.2 of [3]). The parameter θ, known in literature as *extremal index*, is associated with the tendency of clustering of high levels: in case $\theta < 1$ large values tend to occur in clusters, i.e., near each other and tail dependence takes place. Indeed, the YARP(III)(1) process presents tail dependence with lag-m TDC, $\lambda_m = p^m$ (see Proposition 2.8 of [3]).

The one-step transition probability function (tpf) of the YARP(III)(1) process is given by:

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

Similarly, we derive the m-step tpf:

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

In the sequel we shall denote a_t the quantile function at $1 - t$, i.e.,

$$a_t \equiv F_X^{-1}(1-t) = \sigma(t^{-1} - 1)^{1/\alpha} \tag{2}$$

and, for a set A, $\alpha(A)$ and $\zeta(A)$ denote the maximum and the minimum of A, respectively.

Proposition 1. *The YARP(III)(1) process has lag-m upper s, k-extremal coefficient,*

$$\lambda_U(X_{s:m} | X_{m-k+1:m})$$
$$= \frac{\displaystyle\sum_{i=0}^{s-1} \sum_{I \in \mathscr{F}_i} \sum_{J \subset I} (-1)^{|J|} p^{\alpha(\overline{I} \cup J) - \zeta(\overline{I} \cup J)}}{\displaystyle -\sum_{\emptyset \neq J \subset D_m} (-1)^{|J|} p^{\alpha(J) - \zeta(J)} - \sum_{i=1}^{k-1} \sum_{I \in \mathscr{F}_i} \sum_{J \subset \overline{I}} (-1)^{|J|} p^{\alpha(I \cup J) - \zeta(I \cup J)}},$$

where $\mathscr{F}_i$ denotes the family of all subsets of $D_m = \{1, \ldots, m\}$ with cardinal equal to i and $\overline{I}$ the complement set of $I \in \mathscr{F}_i$ in D_m.

Proof. Consider notation $P_A(t) = P\left(\bigcap_{a \in A} \{F_X(X_a) > 1 - t\}\right)$, for any set A. From Propositions 2.1 and 2.9 in [6], we have

$$\lambda_U(X_{s:m} | X_{m-k+1:m}) = \lim_{t \downarrow 0} \frac{\displaystyle\sum_{i=0}^{s-1} \sum_{I \in \mathscr{F}_i} \sum_{J \subset I} (-1)^{|J|} P_{\overline{I} \cup J}(t)/t}{\displaystyle -\sum_{\emptyset \neq J \subset \{1, \ldots, m\}} (-1)^{|J|} P_J(t)/t - \sum_{i=1}^{k-1} \sum_{I \in \mathscr{F}_i} \sum_{J \subset \overline{I}} (-1)^{|J|} P_{I \cup J}(t)/t}.$$

Now just observe that, for $i_1 < i_2 < i_3$, we have successively

$$
\begin{aligned}
P_{\{i_1,i_2,i_3\}}(t) &= \int_{a_t}^{\infty} P(X_{i_3} > a_t, X_{i_2} > a_t | X_{i_1} = u_1) dF_X(u_1) \\
&= \int_{a_t}^{\infty} \int_{a_t}^{\infty} P(X_{i_3} > a_t | X_{i_2} = u_2) Q(u_1, du_2) dF_X(u_1) \\
&= \int_{a_t}^{\infty} \int_{a_t}^{\infty} [1 - Q^{i_3-i_2}(u_2,]0, a_t])] Q^{i_2-i_1}(u_1, du_2) dF_X(u_1),
\end{aligned}
$$

where a_t is given in (2). Applying (1), we obtain

$$
\begin{aligned}
P_{\{i_1,i_2,i_3\}}(t) &= t[t + p^{i_3-i_2}(1-t)] \int_{a_t}^{\infty} \int_{a_t}^{\infty} Q^{i_2-i_1}(u_1, du_2) dF_X(u_1) \\
&= t[t + p^{i_3-i_2}(1-t)] \int_{a_t}^{\infty} [1 - Q^{i_2-i_1}(u_1,]0, a_t])] dF_X(u_1) \\
&= [t + p^{i_3-i_2}(1-t)][t + p^{i_2-i_1}(1-t)]t.
\end{aligned}
$$

A similar reasoning leads us to, for $i_1 < i_2 < \ldots < i_k$,

$$
\begin{aligned}
&P_{\{i_1,\ldots,i_k\}}(t) \\
&= \int_{a_t}^{\infty} \ldots \int_{a_t}^{\infty} \left(1 - Q^{i_k-i_{k-1}}\left(u_{i_{k-1}},]0, a_t]\right)\right) \prod_{j=2}^{k-1} Q^{i_{k-j}-i_{k-j+1}}(u_{i_{k-j}}, du_{i_{k-j+1}}) dF_X(u_{i_1}) \\
&= \prod_{j=2}^{k}(t + p^{i_j-i_{j-1}}(1-t))t,
\end{aligned}
$$

and hence

$$
\lim_{t \downarrow 0} P_{\{i_1,\ldots,i_k\}}(t)/t = \lim_{t \downarrow 0} \prod_{j=2}^{k}(t + p^{i_j-i_{j-1}}(1-t)) = p^{i_k-i_1}. \tag{3}
$$

$\square$

Corollary 1. *The YARP(III)(1) process has lag-m upper extremal dependence coefficient,*

$$
\epsilon_m^U = \frac{p^{m-1}}{m - (m-1)p}.
$$

A positive ϵ_m^U means the existence of extremal dependence on a time horizon of m time instants.

Proposition 2. *The YARP(III)(1) process has lag-m block-FI, relative to a partition $\mathscr{D}_m$ of $D_m = \{1, \ldots, m\}$, given by*

$$FI(\mathbf{X}, \mathscr{D}_m) = \frac{\sum_{j=1}^{s} \sum_{k \in I_j} (-1)^{k-1} \sum_{J \subset I_j; |J|=k} p^{\alpha(J)-\zeta(J)}}{m - (m-1)p}.$$

Proof. Based on Propositions 3.1 and 5.2 in [5], we have

$$FI(\mathbf{X}, \mathscr{D}_m) = \lim_{t \downarrow 0} \frac{\sum_{j=1}^{s} P(\bigcup_{i \in I_j} \{F_X(X_i) > 1 - t\})}{1 - P(\bigcap_{i \in \{1, \dots m\}} \{F_X(X_i) < 1 - t\})}$$
$$= \lim_{t \downarrow 0} \frac{\sum_{j=1}^{s} \sum_{k \in I_j} (-1)^{k-1} \sum_{J \subset I_j; |J|=k} P(\bigcap_{i \in J} \{F_X(X_i) > 1 - t\})}{1 - F_{M_{m-1}}(a_t)}.$$

Now observe that, from (3), we have

$$\lim_{t \downarrow 0} P(\cap_{i \in J} \{F_X(X_i) > 1 - t\})/t = p^{\alpha(J)-\zeta(J)}$$

and from (1) and (2), we have

$$\lim_{t \downarrow 0} (1 - F_{M_{m-1}}(a_t))/t = \lim_{t \downarrow 0} \frac{1}{t} \left(1 - \frac{t^{-1}-1}{t^{-1}} \left(\frac{p + t^{-1} - 1}{t^{-1}} \right)^{m-1} \right) = m - (m-1)p.$$

$\square$

Corollary 2. *The YARP(III)(1) process has lag-m FI,*

$$FI(\mathbf{X}, \mathscr{D}_m^*) = \frac{m}{m - (m-1)p}.$$

Therefore, on a time horizon of m ($m > 1$) time instants the process is strongly fragile since $FI > 1$.

We remark that the tail measures given above only depend on the parameter p of the YARP(III)(1) process and thus can be estimated through this latter. For a survey on the estimation of p, see [3].

References

1. Arnold, B.C.: Pareto Distributions. International Cooperative Publishing House, Fairland (1983)
2. Davis, R., Resnick, S.: Basic properties and prediction of max-ARMA processes. Adv. Appl. Probab. **21**, 781–803 (1989)
3. Ferreira, M.: On the extremal behavior of a pareto process: an alternative for armax modeling. Kybernetika **48**(1), 31–49 (2012)
4. Ferreira, M., Canto e Castro, L.: Modeling rare events through a pRARMAX process. J. Stat. Plann. Inference **140**(11), 3552–3566 (2010)
5. Ferreira, M., Ferreira, H.: Fragility index block tailed vectors. J. Stat. Plann. Inference **142**(7), 1837–1848 (2012)

6. Ferreira, M., Ferreira, H.: Tail dependence between order statistics. J. Multivariate Anal. **105**(1), 176–192 (2012)
7. Frahm G.: On the extremal dependence coefficient of multivariate distributions. Stat. Probab. Lett. **76**, 1470–1481 (2006)
8. Geluk, J.L., De Haan, L., De Vries, C.G.: Weak and strong financial fragility. Tinbergen Institute Discussion Paper, TI 2007-023/2 (2007)
9. Leadbetter, M.R.: On extreme values in stationary sequences. Z. Wahrsch. verw. Gebiete **28**, 289–303 (1974)
10. Pareto, V.: Cours d'economie Politique. Rouge, Lausanne (1897)
11. Sibuya, M.: Bivariate extreme statistics. Ann. Inst. Stat. Math. **11** 195–210 (1960)
12. Yeh, H.C., Arnold, B.C., Robertson, C.A.: Pareto processes. J. Appl. Probab. **25**, 291–301 (1988)
13. Zhang, Z., Smith, R.L.: On the estimation and application of max-stable processes. J. Stat. Plann. Inference **140**(5), 1135–1153 (2010)