

The stopped clock model

Helena Ferreira

Universidade da Beira Interior, Centro de Matemática e Aplicações
(CMA-UBI), Avenida Marquês d'Avila e Bolama, 6200-001 Covilhã,

Portugal

`helena.ferreira@ubi.pt`

Marta Ferreira

Center of Mathematics of Minho University

Center for Computational and Stochastic Mathematics of University of

Lisbon

Center of Statistics and Applications of University of Lisbon, Portugal

`msferreira@math.uminho.pt`

Abstract

The extreme values theory presents specific tools for modeling and predicting extreme phenomena. In particular, risk assessment is often analyzed through measures

for tail dependence and high values clustering. Despite technological advances allowing an increasingly larger and more efficient data collection, there are sometimes failures in the records, which causes difficulties in statistical inference, especially in the tail where data are scarcer. In this article we present a model with a simple and intuitive failures scheme, where each record failure is replaced by the last record available. We will study its extremal behavior with regard to local dependence and high values clustering, as well as the temporal dependence on the tail.

keywords: extreme values; stationary sequences; failures model; extremal index; tail dependence coefficient.

AMS 2000 Subject Classification: 60G70

1 Introduction

Let $\{X_n\}_{n \in \mathbb{Z}}$ and $\{U_n\}_{n \in \mathbb{Z}}$ be stationary sequences of real random variables on the probability space (Ω, \mathcal{A}, P) and $P(U_n \in \{0, 1\}) = 1$. We define, for $n \geq 1$,

$$Y_n = \begin{cases} X_n & , U_n = 1 \\ Y_{n-1} & , U_n = 0. \end{cases} \quad (1)$$

Sequence $\{Y_n\}_{n \geq 1}$ corresponds to a model of failures on records of $\{X_n\}_{n \in \mathbb{Z}}$ replaced by the last available record, which occurs in some random past instant, if we interpret n as time. Thus, if for example it occurs $\{U_1 = 1, U_2 = 0, U_3 = 1, U_4 = 0, U_5 = 0, U_6 = 0, U_7 = 1\}$, we will have $\{Y_1 = X_1, Y_2 = X_1, Y_3 = X_3, Y_4 = X_3, Y_5 = X_3, Y_6 = X_3, Y_7 = X_7\}$. This constancy of some variables of $\{X_n\}_{n \in \mathbb{Z}}$ for random

periods of time motivates the designation of “stopped clock model” for sequence $\{Y_n\}_{n \geq 1}$.

Failure models studied in the literature from the point of view of extremal behavior do not consider the stopped clock model (Hall and Hüsler, [4] 2006; Ferreira *et al.*, [3] 2019 and references therein).

The model we will study can also be represented by $\{X_{N_n}\}_{n \geq 1}$ where $\{N_n\}_{n \geq 1}$ is a sequence of positive integer variables representable by

$$N_n = nU_n + \sum_{n-1 \geq i \geq 1} \left(\prod_{j=0}^{i-1} (1 - U_{n-j}) \right) U_{n-i}(n-i), \quad n \geq 1.$$

We can also state a recursive formulation for $\{Y_n\}_{n \geq 1}$ through

$$Y_n = X_n U_n + \sum_{n-1 \geq i \geq 1} \left(\prod_{j=0}^{i-1} (1 - U_{n-j}) \right) U_{n-i} X_{n-i} + \prod_{i \geq 0} (1 - U_{n-i}) Y_{n-\kappa}, \quad n \geq 1, \quad \kappa \geq 1.$$

Under any of the three possible representations (failures model, random index sequence or recursive sequence), we are not aware of an extremal behavior study of $\{Y_n\}_{n \geq 1}$ in the literature.

Our departure hypotheses about the base sequence $\{X_n\}_{n \in \mathbb{Z}}$ and about sequence $\{U_n\}_{n \in \mathbb{Z}}$ are:

- (1) $\{X_n\}_{n \in \mathbb{Z}}$ is a stationary sequence of random variables almost surely distinct and, without loss of generality, such that $F_{X_n}(x) := F(x) = \exp(-1/x)$, $x > 0$, i.e., standard Fréchet distributed.
- (2) $\{X_n\}_{n \in \mathbb{Z}}$ and $\{U_n\}_{n \in \mathbb{Z}}$ are independent.
- (3) $\{U_n\}_{n \in \mathbb{Z}}$ is stationary and $p_{n_1, \dots, n_s}(i_1, \dots, i_s) := P(U_{n_1} = i_1, \dots, U_{n_s} = i_s)$, $i_j \in \{0, 1\}$, $j = 1, \dots, s$, is such that $p_{n, n+1, \dots, n+\kappa-1}(0, \dots, 0) = 0$, for some

$\kappa \geq 1$.

The trivial case $\kappa = 1$ corresponds to $Y_n = X_n$, $n \geq 1$. Hypothesis (3) means that we are assuming that it is almost impossible to lose κ or more consecutive values of $\{X_n\}_{n \in \mathbb{Z}}$. We remark that, along the paper, the summations, products and intersections is considered to be non-existent whenever the end of the counter is less than the beginning. We will also use notation $a \vee b = \max(a, b)$.

Example 1.1. Consider an independent and identically distributed sequence $\{W_n\}_{n \in \mathbb{Z}}$ of real random variables on (Ω, \mathcal{A}, P) and a Borelian set A . Let $p = P(A_n)$ where $A_n = \{W_n \in A\}$, $n \in \mathbb{Z}$. The sequence of Bernoulli random variables

$$U_n = \mathbf{1}_{\{\bigcap_{i=1}^{\kappa-1} \bar{A}_{n-i}\}} + (1 - \mathbf{1}_{\{\bigcap_{i=1}^{\kappa-1} \bar{A}_{n-i}\}}) \mathbf{1}_{\{A_n\}}, \quad n \in \mathbb{Z}, \quad (2)$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function, defined for some fixed $\kappa \geq 2$, is such that $p_{n, n+1, \dots, n+\kappa-1}(0, \dots, 0) = 0$, i.e., it is almost sure that after $\kappa - 1$ consecutive variables equal to zero, the next variable takes value one. We also have

$$\begin{aligned} p_n(0) &= P\left(\mathbf{1}_{\{\bigcap_{i=1}^{\kappa-1} \bar{A}_{n-i}\}} = 0, \mathbf{1}_{\{A_n\}} = 0\right) = P\left(\bigcup_{i=1}^{\kappa-1} A_{n-i} \cap \bar{A}_n\right) \\ &= P(\bar{A}_n) - P\left(\bigcap_{i=0}^{\kappa-1} \bar{A}_{n-i}\right) = 1 - p - (1 - p)^\kappa, \end{aligned}$$

since the independence of random variables W_n implies the independence of events A_n , and, for $\kappa > 2$,

$$\begin{aligned} p_{n-1, n}(0, 0) &= P\left(\left(\bigcup_{i=1}^{\kappa-1} A_{n-i} \cap \bar{A}_n\right) \cap \left(\bigcup_{i=1}^{\kappa-1} A_{n-1-i} \cap \bar{A}_{n-1}\right)\right) \\ &= P(\bar{A}_n \cap \bar{A}_{n-1}) - P\left(\bar{A}_n \cap \bar{A}_{n-1} \cap \left(\bigcap_{i=1}^{\kappa-1} \bar{A}_{n-i} \cup \bigcap_{i=1}^{\kappa-1} \bar{A}_{n-1-i}\right)\right) \\ &= P(\bar{A}_n)P(\bar{A}_{n-1}) - P\left(\bigcap_{i=0}^{\kappa-1} \bar{A}_{n-i}\right) = (1 - p)^2 - (1 - p)^\kappa, \end{aligned}$$

$$p_{n-1, n}(1, 0) = p_n(0) - p_{n-1, n}(0, 0) = p(1 - p).$$

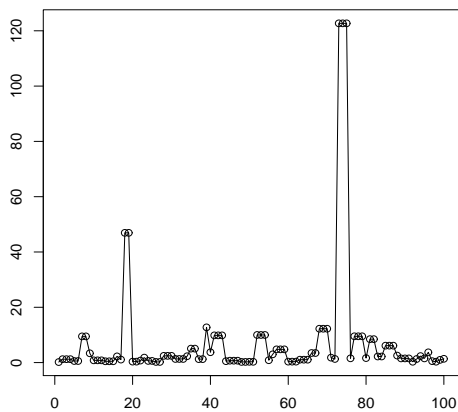


Figure 1: Sample path of 100 observations simulated from $\{Y_n\}$ defined in (1) based on independent standard Fréchet $\{X_n\}$ and on $\{U_n\}$ given in (2) where we take random variables $\{W_n\}$ standard exponential distributed, $A =]0, 1/2]$ and thus $p = 0.3935$ and considering $\kappa = 3$.

In Figure 1 we illustrate with a particular example based on independent standard Fréchet $\{X_n\}_{n \in \mathbb{Z}}$, $\{W_n\}_{n \in \mathbb{Z}}$ with standard exponential marginals, $A =]0, 1/2]$ and thus $p = 0.3935$ and considering $\kappa = 3$. Therefore, $p_{n,n+1,n+2}(0,0,0) = 0$, $p_{n,n+1,n+2}(1,0,0) = p_{n,n+1}(0,0) = p(1-p)^2$.

In the next section we propose an estimator for probabilities $p_{n,\dots,n+s}(1,0,\dots,0)$, $0 \leq s < \kappa - 1$. In Section 3 we analyse the existence of the extremal index for $\{Y_n\}_{n \geq 1}$, an important measure to evaluate the tendency to occur clusters of its high values (see, e.g., Kulik and Solier [6], 2020 and references therein). A characterization of the tail dependence will be presented in Section 4. The results are illustrated with an ARMAX sequence.

For the sake of simplicity, we will omit the variation of n in sequence notation whenever there is no doubt, taking into account that we will keep the designation $\{Y_n\}$ for the stopped clock model and $\{X_n\}$ and $\{U_n\}$ for the sequences that generate

it.

2 Inference on $\{U_n\}$

Assuming that $\{U_n\}$ is not observable, as well as the values of $\{X_n\}$ that are lost, it is of interest to retrieve information about these sequences from the available sequence $\{Y_n\}$.

Since, for $n \geq 1$ and $s \geq 1$, we have

$$p_n(1) = E(\mathbf{1}_{\{Y_n \neq Y_{n-1}\}}), \quad p_n(0) = E(\mathbf{1}_{\{Y_n = Y_{n-1}\}})$$

$$\text{and} \quad p_{n-s, n-s+1, \dots, n}(1, 0, \dots, 0) = E(\mathbf{1}_{\{Y_{n-s-1} \neq Y_{n-s} = Y_{n-s+1} = \dots = Y_n\}}),$$

we propose to estimate these probabilities from the respective empirical counterparts of a random sample $(\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_m)$ from $\{Y_n\}$, i.e.,

$$\hat{p}_n(1) = \frac{1}{m} \sum_{i=2}^m \mathbf{1}_{\{\hat{Y}_i \neq \hat{Y}_{i-1}\}}, \quad \hat{p}_n(0) = \frac{1}{m} \sum_{i=2}^m \mathbf{1}_{\{\hat{Y}_i = \hat{Y}_{i-1}\}}$$

$$\text{and} \quad \hat{p}_{n-s, n-s+1, \dots, n}(1, 0, \dots, 0) = \frac{1}{m} \sum_{i=s+2}^m \mathbf{1}_{\{\hat{Y}_{i-s-1} \neq \hat{Y}_{i-s} = \hat{Y}_{i-s+1} = \dots = \hat{Y}_i\}},$$

which are consistent by the *weak law of large numbers*. The value of κ can be inferred from

$$\hat{\kappa} = \bigvee_{i=s+2}^m \bigvee_{s \geq 1} s \mathbf{1}_{\{\hat{Y}_{i-s-1} \neq \hat{Y}_{i-s} = \hat{Y}_{i-s+1} = \dots = \hat{Y}_i\}}.$$

In order to evaluate the finite sample behavior of the estimators above, we have simulated 1000 independent replicas with size $m = 100, 1000, 5000$ of the model in Example 1.1. The absolute bias (abias) and root mean squared error (rmse) are presented in Table 1. The results reveal a good performance of the estimators, even

Table 1: The absolute bias (abias) and root mean squared error (rmse) obtained from 1000 simulated samples with size $m = 100, 1000, 5000$ of the model in Example 1.1.

		abias	rmse
$\widehat{p}_n(0)$	$m = 100$	0.0272	0.0335
	$m = 1000$	0.0087	0.0108
	$m = 5000$	0.0039	0.0048
$\widehat{p}_{n-1,n}(1, 0)$	$m = 100$	0.0199	0.0253
	$m = 1000$	0.0065	0.0080
	$m = 5000$	0.0030	0.0037
$\widehat{p}_{n-2,n-1,n}(1, 0, 0)$	$m = 100$	0.0160	0.0200
	$m = 1000$	0.0051	0.0064
	$m = 5000$	0.0022	0.0028

in the case of smaller sample sizes. Parameter κ was always estimated with no error.

3 The extremal index of $\{Y_n\}$

The sequence $\{Y_n\}$ is stationary because the sequences $\{X_n\}$ and $\{U_n\}$ are stationary and independent from each other. In addition, the common distribution for Y_n , $n \geq 1$, is also standard Fréchet, as is the common distribution for X_n , since

$$\begin{aligned} F_{Y_n}(x) &= \sum_{i=1}^{\kappa-1} P(X_{n-i} \leq x, U_{n-i} = 1, U_{n-i+1} = 0 = \dots = U_n) + P(X_n \leq x)P(U_n = 1) \\ &= F(x) \left(p_n(1) + \sum_{i=1}^{\kappa-1} p_{n-i, \dots, n}(1, 0, \dots, 0) \right) = F(x). \end{aligned}$$

For any $\tau > 0$, if we define $u_n \equiv u_n(\tau) = n/\tau$, $n \geq 1$, it turns out that $E\left(\sum_{i=1}^n \mathbf{1}_{\{Y_i > u_n\}}\right) = nP(Y_1 > u_n) \xrightarrow{n \rightarrow \infty} \tau$ and $nP(X_1 > u_n) \xrightarrow{n \rightarrow \infty} \tau$, so we refer to these levels u_n by *normalized levels* for $\{Y_n\}$ and $\{X_n\}$.

In this section, in addition to the general assumptions about the model presented in Section 1, we start by assuming that $\{X_n\}$ and $\{U_n\}$ present dependency structures such that variables sufficiently apart can be considered approximately in-

dependent. Concretely, we assume that $\{U_n\}$ satisfies the strong-mixing condition (Rosenblat [9] 1956) and $\{X_n\}$ satisfies condition $D(u_n)$ (Leadbetter [7] 1974) for normalized levels u_n .

Proposition 3.1. *If $\{U_n\}$ is strong-mixing and $\{X_n\}$ satisfies condition $D(u_n)$ then $\{Y_n\}$ also satisfies condition $D(u_n)$.*

Proof. For any choice of $p + q$ integers, $1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq n$ such that $j_1 \geq i_p + l$, we have that

$$\left| P \left(\bigcap_{s=1}^p X_{i_s} \leq u_n, \bigcap_{s=1}^q X_{j_s} \leq u_n \right) - P \left(\bigcap_{s=1}^p X_{i_s} \leq u_n \right) P \left(\bigcap_{s=1}^q X_{j_s} \leq u_n \right) \right| \leq \alpha_{n,l},$$

with $\alpha_{n,l_n} \rightarrow 0$, as $n \rightarrow \infty$, for some sequence $l_n = o(n)$, and

$$|P(A \cap B) - P(A)P(B)| \leq g(l),$$

with $g(l) \rightarrow 0$, as $l \rightarrow \infty$, where A belongs to the σ -algebra generated by $\{U_i, i = 1, \dots, i_p\}$ and B belongs to the σ -algebra generated by $\{U_i, i = j_1, j_1 + 1, \dots\}$. Thus, for any choice of $p + q$ integers, $1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq n$ such that $j_1 \geq i_p + l + \kappa$, we will have

$$\begin{aligned} & \left| P \left(\bigcap_{s=1}^p Y_{i_s} \leq u_n, \bigcap_{s=1}^q Y_{j_s} \leq u_n \right) - P \left(\bigcap_{s=1}^p Y_{i_s} \leq u_n \right) P \left(\bigcap_{s=1}^q Y_{j_s} \leq u_n \right) \right| \\ & \leq \sum_{\substack{i_s - \kappa < i_s^* \leq i_s \\ j_s - \kappa < j_s^* \leq j_s}} \left| P \left(\bigcap_{s=1}^p X_{i_s^*} \leq u_n, \bigcap_{s=1}^q X_{j_s^*} \leq u_n \right) P(A^* \cap B^*) \right. \\ & \quad \left. - P \left(\bigcap_{s=1}^p X_{i_s^*} \leq u_n \right) P \left(\bigcap_{s=1}^q X_{j_s^*} \leq u_n \right) P(A^*) P(B^*) \right|, \end{aligned}$$

where $A^* = \bigcap_{s=1}^p \{U_{i_s} = 0 = \dots = U_{i_s^*+1}, U_{i_s^*} = 1\}$ and $B^* = \bigcap_{s=1}^q \{U_{j_s} = 0 = \dots =$

$U_{j_s^*+1}, U_{j_s^*} = 1\}$ and $j_1^* > j_1 - \kappa \geq i_p^* + l$. Therefore, the last summation above is upper limited by

$$\begin{aligned} & \sum_{\substack{i_s - \kappa < i_s^* \leq i_s \\ j_s - \kappa < j_s^* \leq j_s}} \left(\left| P \left(\bigcap_{s=1}^p X_{i_s^*} \leq u_n, \bigcap_{s=1}^q X_{j_s^*} \leq u_n \right) - P \left(\bigcap_{s=1}^p X_{i_s^*} \leq u_n \right) P \left(\bigcap_{s=1}^q X_{j_s^*} \leq u_n \right) \right| \right. \\ & \quad \left. + |P(A^* \cap B^*) - P(A^*)P(B^*)| \right) \\ & \leq \sum_{\substack{i_s - \kappa < i_s^* \leq i_s \\ j_s - \kappa < j_s^* \leq j_s}} (\alpha_{n,l} + g(l)), \end{aligned}$$

which allows to conclude that $D(u_n)$ holds for $\{Y_n\}$ with $l_n^{(Y)} = l_n + \kappa$. \square

The tendency for clustering of values of $\{Y_n\}$ above u_n depends on the same tendency within $\{X_n\}$ and the propensity of $\{U_n\}$ for consecutive null values. The clustering tendency can be assessed through the extremal index (Leadbetter, [7] 1974). More precisely, $\{X_n\}$ is said to have extremal index $\theta_X \in (0, 1]$ if

$$\lim_{n \rightarrow \infty} P \left(\bigvee_{i=1}^n X_i \leq n/\tau \right) = e^{-\theta_X \tau}. \quad (3)$$

If $D(u_n)$ holds for $\{X_n\}$, we have

$$\lim_{n \rightarrow \infty} P \left(\bigvee_{i=1}^n X_i \leq u_n \right) = \lim_{n \rightarrow \infty} P^{k_n} \left(\bigvee_{i=1}^{\lfloor n/k_n \rfloor} X_i \leq u_n \right),$$

for any integers sequence $\{k_n\}$, such that,

$$k_n \rightarrow \infty, \quad k_n l_n / n \rightarrow 0 \quad \text{and} \quad k_n \alpha_{n, l_n} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (4)$$

We can therefore say that

$$\theta_X \tau = \lim_{n \rightarrow \infty} k_n P \left(\bigvee_{i=1}^{\lfloor n/k_n \rfloor} X_i > u_n \right).$$

Now we compare the local behavior of sequences $\{X_n\}$ and $\{Y_n\}$, i.e., of X_i and Y_i for $i \in \left\{ (j-1) \left\lfloor \frac{n}{k_n} \right\rfloor + 1, \dots, j \left\lfloor \frac{n}{k_n} \right\rfloor \right\}$, $j = 1, \dots, k_n$, with regard to the oscillations of their values in relation to u_n . To this end, we will use local dependency conditions $D^{(s)}(u_n)$. We say that $\{X_n\}$ satisfies $D^{(s)}(u_n)$, $s \geq 2$, whenever

$$\lim_{n \rightarrow \infty} n \sum_{j=s}^{\lfloor n/k_n \rfloor} P(X_1 > u_n, X_j \leq u_n < X_{j+1}) = 0,$$

for some integers sequence $\{k_n\}$ satisfying (4). Condition $D^{(1)}(u_n)$ translates into

$$\lim_{n \rightarrow \infty} n \sum_{j=2}^{\lfloor n/k_n \rfloor} P(X_1 > u_n, X_j > u_n) = 0.$$

Observe that if $D^{(s)}(u_n)$ holds for some $s \geq 1$, then $D^{(m)}(u_n)$ also holds for $m > s$. Condition $D^{(1)}(u_n)$ is known as $D'(u_n)$ (Leadbetter *et al.*, [8] 1983) and relates to a unit extremal index, i.e., absence of extreme values clustering. In particular, this is the case of independent variables. Although $\{X_n\}$ satisfies $D'(u_n)$, this condition is not generally valid for $\{Y_n\}$. Observe that

$$\begin{aligned} & n \sum_{j=2}^{\lfloor n/k_n \rfloor} P(Y_1 > u_n, Y_j > u_n) \\ &= \sum_{i=2-\kappa}^1 n \sum_{j=2}^{\lfloor n/k_n \rfloor} \sum_{j^*=i \vee (j-\kappa+1)}^j P(X_i > u_n, X_{j^*} > u_n) \cdot \\ & \quad \cdot p_{i, \dots, 1, j^*, j^*+1, \dots, j}(0, \dots, 0, 1, 0, \dots, 0). \end{aligned}$$

For $i = 1$ and $j = \kappa$, we have $j^* = 1$ and the corresponding term becomes $nP(X_1 > u_n) \rightarrow \tau > 0$, as $n \rightarrow \infty$, reason why, in general $\{Y_n\}$ does not satisfy $D'(u_n)$ even if $\{X_n\}$ satisfies it.

Proposition 3.2. *The following statements hold:*

(i) *If $\{Y_n\}$ satisfies $D^{(s)}(u_n)$, $s \geq 2$, then $\{X_n\}$ satisfies $D^{(s)}(u_n)$.*

(ii) *If $\{X_n\}$ satisfies $D^{(s)}(u_n)$, $s \geq 2$, then $\{Y_n\}$ satisfies $D^{(s+\kappa-1)}(u_n)$.*

(iii) *If $\{X_n\}$ satisfies $D'(u_n)$, then $\{Y_n\}$ satisfies $D^{(2)}(u_n)$.*

Proof. Consider $r_n = [n/k_n]$. We have that

$$\begin{aligned}
& n \sum_{j=s}^{r_n} P(Y_1 > u_n, Y_j \leq u_n < Y_{j+1}) \\
&= \sum_{i=2-\kappa}^1 n \sum_{j=s}^{r_n} P(X_i > u_n, Y_j \leq u_n < X_{j+1}, U_i = 1, U_{i+1} = 0 = \dots = U_1, U_{j+1} = 1) \\
&= \sum_{i=2-\kappa}^1 n \sum_{j=s}^{r_n} \sum_{j^*=(i+1) \vee (j-\kappa+1)}^j P(X_i > u_n, X_{j^*} \leq u_n < X_{j+1}) \cdot \\
&\quad \cdot p_{i,i+1,\dots,1,j^*,j^*+1,\dots,j,j+1}(1, 0, \dots, 0, 1, 0, \dots, 0, 1).
\end{aligned} \tag{5}$$

Since $\{Y_n\}$ satisfies $D^{(s)}(u_n)$, with $s \geq 2$, and thus the first summation in (5) converges to zero, as $n \rightarrow \infty$, then all the terms in the last summations also converge to zero. In particular, when $i = 1$ and $j^* = j$, we have $n \sum_{j=s}^{r_n} P(X_1 > u_n, X_j \leq u_n < X_{j+1}) \rightarrow 0$, as $n \rightarrow \infty$, which proves (i).

On the other hand, writing the first summation in (5) with j starting at $s + \kappa - 1$,

we have

$$\begin{aligned}
& n \sum_{j=s+\kappa-1}^{r_n} P(Y_1 > u_n, Y_j \leq u_n < Y_{j+1}) \\
&= \sum_{i=2-\kappa}^1 n \sum_{j=s+\kappa-1}^{r_n} \sum_{j^*=j-\kappa+1}^j P(X_i > u_n, X_{j^*} \leq u_n < X_{j+1}) \cdot \\
&\quad \cdot p_{i,i+1,\dots,1,j^*,j^*+1,\dots,j,j+1}(1, 0, \dots, 0, 1, 0, \dots, 0, 1) \\
&= \sum_{i=2-\kappa}^1 n \sum_{j=s+\kappa-1}^{r_n} \sum_{j^*=j-\kappa+1}^j \sum_{i^*=j^*}^j P(X_i > u_n, X_{j^*} \leq u_n, \dots, X_{i^*} \leq u_n, X_{j+1} > u_n) \cdot \\
&\quad \cdot p_{i,i+1,\dots,1,j^*,j^*+1,\dots,j,j+1}(1, 0, \dots, 0, 1, 0, \dots, 0, 1), \quad (6)
\end{aligned}$$

where the least of distances between i and i^* corresponds to the case $i = 1$ and $i^* = j^* = s$. Therefore, if $\{X_n\}$ satisfies $D^{(s)}(u_n)$ for some $s \geq 2$ then each term of (6) converges to zero, as $n \rightarrow \infty$, and thus $\{Y_n\}$ satisfies $D^{(s+\kappa-1)}(u_n)$, proving (ii).

As for (iii), observe that

$$\begin{aligned}
& n \sum_{j=2}^{r_n} P(Y_1 > u_n, Y_j \leq u_n < Y_{j+1}) \\
&= \sum_{i=2-\kappa}^1 n \sum_{j=2}^{r_n} P(X_i > u_n, Y_j \leq u_n < X_{j+1}, U_i = 1, U_{i+1} = 0 = \dots = U_1, U_{j+1} = 1) \\
&\leq \sum_{i=2-\kappa}^1 n \sum_{j=2}^{r_n} P(X_i > u_n, X_{j+1} > u_n) = \sum_{i=2-\kappa}^1 n \sum_{j=2}^{r_n} P(X_1 > u_n, X_{j-i+2} > u_n) \\
&\leq \kappa n \sum_{j=2}^{r_n} P(X_1 > u_n, X_j > u_n). \quad (7)
\end{aligned}$$

If $\{X_n\}$ satisfies $D'(u_n)$, then (7) converges to zero, as $n \rightarrow \infty$, and $D^{(2)}(u_n)$ holds for $\{Y_n\}$. \square

Under conditions $D(u_n)$ and $D^{(s)}(u_n)$ with $s \geq 2$, we can also compute the

extremal index θ_X defined in (3) by (Chernick *et al.*, [1] 1991; Corollary 1.3)

$$\theta_X = \lim_{n \rightarrow \infty} P(X_2 \leq u_n, \dots, X_s \leq u_n | X_1 > u_n). \quad (8)$$

If $\{X_n\}$ and $\{Y_n\}$ have extremal indexes θ_X and θ_Y , respectively, then $\theta_Y \leq \theta_X$, since $P(\bigvee_{i=1}^n X_i \leq n/\tau) \leq P(\bigvee_{i=1}^n Y_i \leq n/\tau)$. This corresponds to the intuitively expected, if we remember that the possible repetition of variables X_n leads to larger clusters of values above u_n . In the following result, we establish a relationship between θ_X and θ_Y .

Proposition 3.3. *Suppose that $\{U_n\}$ is strong-mixing and $\{X_n\}$ satisfies conditions $D(u_n)$ and $D^{(s)}(u_n)$, $s \geq 2$, for normalized levels $u_n \equiv u_n(\tau)$. If $\{X_n\}$ has extremal index θ_X then $\{Y_n\}$ has extremal index θ_Y given by*

$$\theta_Y = \theta_X \sum_{j=0}^{\kappa-1} p_{1,2,\dots,j+1,j+2}(1, 0, \dots, 0, 1) \beta_j,$$

where

$$\beta_j = \lim_{n \rightarrow \infty} P(X_{s+j} > u_n | X_1 \leq u_n, \dots, X_{s-1} \leq u_n < X_s).$$

Proof. By Proposition 3.1, $\{Y_n\}$ also satisfies condition $D(u_n)$. Thus we have

$$\lim_{n \rightarrow \infty} P\left(\bigvee_{i=1}^n Y_i \leq u_n\right) = \exp\left\{-\lim_{n \rightarrow \infty} k_n P\left(\bigvee_{i=1}^{\lfloor n/k_n \rfloor} Y_i > u_n\right)\right\}$$

and

$$\begin{aligned}
& \lim_{n \rightarrow \infty} k_n P \left(\bigvee_{i=1}^{[n/k_n]} Y_i > u_n \right) = \lim_{n \rightarrow \infty} k_n P \left(Y_1 \leq u_n, \bigvee_{i=1}^{[n/k_n]} \{Y_i > u_n\} \right) \\
&= \lim_{n \rightarrow \infty} k_n P \left(\bigcup_{i=1}^{[n/k_n]} \{Y_i \leq u_n < Y_{i+1}\} \right) = \lim_{n \rightarrow \infty} k_n P \left(\bigcup_{i=1}^{[n/k_n]} \{Y_i \leq u_n < X_{i+1}, U_{i+1} = 1\} \right) \\
&= \lim_{n \rightarrow \infty} k_n P \left(\bigcup_{i=1}^{[n/k_n]} \bigcup_{j=0}^{\kappa-1} \{X_{i-j} \leq u_n < X_{i+1}, U_{i-j} = 1, U_{i-j+1} = 0 = \dots = U_i, U_{i+1} = 1\} \right) \\
&= \lim_{n \rightarrow \infty} k_n P \left(\bigcup_{i=1}^{[n/k_n]} \bigcup_{j=0}^{\kappa-1} \{X_i \leq u_n < X_{i+j+1}, U_i = 1, U_{i+1} = 0 = \dots = U_{i+j}, U_{i+j+1} = 1\} \right) \\
&= \lim_{n \rightarrow \infty} k_n \sum_{i=1}^{[n/k_n]} \sum_{j=0}^{\kappa-1} P(X_1 \leq u_n, \dots, X_i \leq u_n < X_{i+1}, X_{i+j+1} > u_n) \cdot \\
&\quad \cdot p_{i, i+1, \dots, i+j, i+j+1}(1, 0, \dots, 0, 1) \tag{9} \\
&= \lim_{n \rightarrow \infty} k_n \sum_{i=1}^{[n/k_n]} \sum_{j=0}^{\kappa-1} P(X_{i-s+2} \leq u_n, \dots, X_i \leq u_n < X_{i+1}, X_{i+j+1} > u_n) \cdot \\
&\quad \cdot p_{1, 2, \dots, j+1, j+2}(1, 0, \dots, 0, 1)
\end{aligned}$$

since $\{X_n\}$ satisfies condition $D^{(s)}(u_n)$ for some $s \geq 2$. The stationarity of $\{X_n\}$

leads to

$$\begin{aligned}
& \lim_{n \rightarrow \infty} k_n \sum_{i=1}^{\lfloor n/k_n \rfloor} \sum_{j=0}^{\kappa-1} P(X_{i-s+2} \leq u_n, \dots, X_i \leq u_n < X_{i+1}, X_{i+j+1} > u_n) \cdot \\
& \quad \cdot p_{1,2,\dots,j+1,j+2}(1, 0, \dots, 0, 1) \\
&= \lim_{n \rightarrow \infty} k_n \sum_{i=1}^{\lfloor n/k_n \rfloor} \sum_{j=0}^{\kappa-1} P(X_1 \leq u_n, \dots, X_{s-1} \leq u_n < X_s, X_{s+j} > u_n) \cdot \\
& \quad \cdot p_{1,2,\dots,j+1,j+2}(1, 0, \dots, 0, 1) \\
&= \lim_{n \rightarrow \infty} \sum_{j=0}^{\kappa-1} n P(X_1 \leq u_n, \dots, X_{s-1} \leq u_n < X_s, X_{s+j} > u_n) \cdot \\
& \quad \cdot p_{1,2,\dots,j+1,j+2}(1, 0, \dots, 0, 1) \\
&= \lim_{n \rightarrow \infty} \sum_{j=0}^{\kappa-1} n P(X_1 \leq u_n, \dots, X_{s-1} \leq u_n < X_s) P(X_{s+j} > u_n | X_1 \leq u_n, \dots, X_{s-1} \leq u_n < X_s) \cdot \\
& \quad \cdot p_{1,2,\dots,j+1,j+2}(1, 0, \dots, 0, 1) \\
&= \tau \theta_X \sum_{j=0}^{\kappa-1} p_{1,2,\dots,j+1,j+2}(1, 0, \dots, 0, 1) \beta_j,
\end{aligned}$$

where the last step follows from (8). \square

Observe that $\sum_{j=0}^{\kappa-1} p_{1,2,\dots,j+1,j+2}(1, 0, \dots, 0, 1) = p_n(1) = P(U_n = 1)$ and thus $\theta_Y \leq \theta_X p_n(1) \leq \theta_X$, as expected.

Proposition 3.4. *Suppose that $\{U_n\}$ is strong-mixing and $\{X_n\}$ satisfies conditions $D(u_n)$ and $D'(u_n)$, for normalized levels $u_n \equiv u_n(\tau)$. Then $\{Y_n\}$ has extremal index θ_Y given by $\theta_Y = p_{1,2}(1, 1)$.*

Proof. By condition $D'(u_n)$, the only term to consider in (9) corresponds to $j = 0$,

and we obtain

$$\begin{aligned}
& \lim_{n \rightarrow \infty} k_n P \left(\bigvee_{i=1}^{\lfloor n/k_n \rfloor} Y_n \leq u_n \right) \\
&= \lim_{n \rightarrow \infty} k_n \sum_{i=1}^{\lfloor n/k_n \rfloor} P(X_1 \leq u_n, \dots, X_{s-1} \leq u_n < X_s) p_{1,2}(1, 1) \\
&= \lim_{n \rightarrow \infty} n P(X_s > u_n) p_{1,2}(1, 1) = \tau p_{1,2}(1, 1).
\end{aligned}$$

□

Observe that we can obtain the above result by applying Proposition 3.2 (iii) and calculating directly $\tau \theta_Y = \lim_{n \rightarrow \infty} n P(Y_1 \leq u_n < Y_2)$. More precisely, we have that $\{Y_n\}$ satisfies $D^{(2)}(u_n)$ and by applying (8), we obtain

$$\begin{aligned}
\tau \theta_Y &= \lim_{n \rightarrow \infty} n P(Y_1 \leq u_n < Y_2) \\
&= \lim_{n \rightarrow \infty} n P(Y_1 \leq u_n < X_2, U_2 = 1) \\
&= \lim_{n \rightarrow \infty} n P \left(\bigcup_{j=0}^{\kappa-1} X_{1-j} \leq u_n < X_2, U_{1-j} = 1, U_{1-j+1} = 0 = \dots = U_1, U_2 = 1, \right) \\
&= \lim_{n \rightarrow \infty} n P \left(\bigcup_{j=0}^{\kappa-1} X_{2-\kappa} \leq u_n, \dots, X_{1-j} \leq u_n < X_{2-j}, X_2 > u_n \right) \\
&\quad \cdot p_{1-j, 1-j+1, \dots, 1, 2}(1, 0, \dots, 0, 1) \\
&= \lim_{n \rightarrow \infty} n P(X_1 \leq u_n < X_2) p_{1,2}(1, 1) \\
&= \lim_{n \rightarrow \infty} n P(X_2 > u_n) p_{1,2}(1, 1) = \tau p_{1,2}(1, 1).
\end{aligned}$$

The same result can also be seen as a particular case of Proposition 3.3 where, if we take $s = 1$, we have $\beta_j = 0$, for $j \neq 0$, and we obtain $\theta_Y = \theta_X \beta_0 p_{1,2}(1, 1) = p_{1,2}(1, 1)$, since $\beta_0 = 1$ and under $D'(u_n)$ it comes $\theta_X = 1$.

Example 3.1. Consider $\{Y_n\}$ such that $\{X_n\}$ is an ARMAX sequence, i.e., $X_n = \phi X_{n-1} \vee (1 - \phi) Z_n$, $n \geq 1$, where $\{Z_n\}$ is an independent sequence of random

variables with standard Fréchet marginal distribution and $\{X_n\}$ and $\{Z_n\}$ are independent. We have that $\{X_n\}$ has also standard Fréchet marginal distribution, satisfies condition $D^{(2)}(u_n)$ and has extremal index $\theta_X = 1 - \phi$ (see e.g. Ferreira and Ferreira [2] 2012 and references therein).

Observe that, for normalized levels $u_n \equiv n/\tau$, $\tau > 0$, we have

$$\begin{aligned}
\beta_1 &= \lim_{n \rightarrow \infty} P(X_3 > u_n | X_1 \leq u_n < X_2) \\
&= \lim_{n \rightarrow \infty} \frac{P(X_1 \leq u_n) - P(X_1 \leq u_n, X_2 \leq u_n) - P(X_1 \leq u_n, X_3 \leq u_n) + P(X_1 \leq u_n, X_2 \leq u_n, X_3 \leq u_n)}{P(X_1 \leq u_n) - P(X_1 \leq u_n, X_2 \leq u_n)} \\
&= \lim_{n \rightarrow \infty} \frac{1 - \frac{\tau}{n} - (1 - \frac{\tau}{n}(2 - \phi)) - (1 - \frac{\tau}{n}(2 - \phi^2)) + 1 - \frac{\tau}{n}(3 - 2\phi)}{1 - \frac{\tau}{n} - (1 - \frac{\tau}{n}(2 - \phi))} \\
&= \phi.
\end{aligned}$$

Analogous calculations lead to $\beta_2 = \phi^2$. Considering $\kappa = 3$, we have $\theta_Y = (1 - \phi)(p_{1,2}(1, 1) + \phi p_{1,2,3}(1, 0, 1) + \phi^2 p_{1,2,3,4}(1, 0, 0, 1))$.

The observed sequence is $\{Y_n\}$, therefore results that allow retrieving information about the extreme behavior of the initial sequence $\{X_n\}$, subject to the failures determined by $\{U_n\}$, may be of interest.

If we assume that $\{Y_n\}$ satisfies $D^{(s)}(u_n)$ then $\{X_n\}$ also satisfies $D^{(s)}(u_n)$ by Proposition 3.2 (i), thus coming

$$\begin{aligned}
\tau \theta_X &= \lim_{n \rightarrow \infty} n P(X_1 \leq u_n, \dots, X_{s-1} \leq u_n < X_s) \\
&= \lim_{n \rightarrow \infty} n P(Y_1 \leq u_n, \dots, Y_{s-1} \leq u_n < Y_s | U_1 = \dots = U_s = 1) \\
&= \lim_{n \rightarrow \infty} n P(Y_1 \leq u_n, \dots, Y_{s-1} \leq u_n < Y_s | Y_0 \neq Y_1 \neq \dots \neq Y_s).
\end{aligned}$$

Thereby, we can write

$$\theta_X = \lim_{n \rightarrow \infty} \frac{P(Y_1 \leq u_n, \dots, Y_{s-1} \leq u_n < Y_s | Y_0 \neq Y_1 \neq \dots \neq Y_s)}{P(Y_1 > u_n)}.$$

4 Tail dependence

Now we will analyse the effect of this failure mechanism on the dependency between two variables, Y_n and Y_{n+m} , $m \geq 1$. More precisely, we are going to evaluate the lag- m tail dependence coefficient

$$\lambda(Y_{n+m}|Y_n) = \lim_{x \rightarrow \infty} P(Y_{n+m} > x | Y_n > x),$$

which incorporates the tail dependence between X_n and X_{n+j} , with j regulated by the maximum number of failures $\kappa - 1$ and by the relation between m and κ . In particular, independent variables present null tail dependence coefficients. If $m = 1$ we obtain the tail dependence coefficient in Joe ([5] 1997). For results related to lag- m tail dependence in the literature, see, e.g., Zhang ([11] 2005) and Zhang ([10] 2021).

Proposition 4.1. *Sequence $\{Y_n\}$ has lag- m tail dependence coefficient, with $m \geq 1$,*

$$\begin{aligned} \lambda(Y_{n+m}|Y_n) = & p_{1,\dots,m}(0, \dots, 0) \mathbf{1}_{\{m \leq \kappa-1\}} + \sum_{i=1 \vee (m-\kappa+1)}^m \sum_{i^*=0}^{\kappa-1} \lambda(X_{n+i+i^*}|X_n) \cdot \\ & \cdot p_{1,2,\dots,i^*+1,i^*+1+i,i^*+2+i,\dots,i^*+1+m}(1, 0, \dots, 0, 1, 0, \dots, 0), \end{aligned} \quad (10)$$

provided all coefficients $\lambda(X_{n+i+i^}|X_n)$ exist.*

Proof. Observe that

$$\begin{aligned}
& P(Y_n > x, Y_{n+m} > x) \\
&= P(Y_n > x, U_{n+1} = 0 = \dots = U_{n+m}) \mathbf{1}_{\{m \leq \kappa-1\}} \\
&\quad + \sum_{i=1 \vee (m-\kappa+1)}^m P(Y_n > x, X_{n+i} > x, U_{n+i} = 1, U_{n+i+1} = 0 = \dots = U_{n+m}) \\
&= \sum_{i=0}^{\kappa-1-m} P(X_{n-i} > x) p_{n-i, n-i+1, \dots, n+m}(1, 0, \dots, 0) \mathbf{1}_{\{m \leq \kappa-1\}} \\
&\quad + \sum_{i=1 \vee (m-\kappa+1)}^m \sum_{i^*=0}^{\kappa-1} P(X_{n-i^*} > x, X_{n+i} > x) p_{n-i^*, n-i^*+1, \dots, n, n+i, n+i+1, \dots, n+m}(1, 0, \dots, 0, 1, 0, \dots, 0)
\end{aligned}$$

$$\text{and } \sum_{i=0}^{\kappa-1-m} p_{1,2,\dots,m+i+1}(1, 0, \dots, 0) = p_{1,\dots,m}(0, \dots, 0). \quad \square$$

Taking $m = 1$ in (10), we obtain the tail dependence coefficient

$$\lambda(Y_{n+1}|Y_n) = p_n(0) + \sum_{i=0}^{\kappa-1} \lambda(X_{n+1+i}|X_n) p_{1,2,\dots,i+1,i+2}(1, 0, \dots, 0, 1),$$

provided all coefficients $\lambda(X_{n+1+i}|X_n)$ exist.

If $\{X_n\}$ is lag- m^* tail independent for all integer $m^* \geq 1 \vee (m - \kappa + 1)$, we have $\lambda(X_{n+i+i^*}|X_n) = 0$ in the second ter of (10) and thus $\lambda(Y_{n+m}|Y_n) = p_{1,\dots,m}(0, \dots, 0) \mathbf{1}_{\{m \leq \kappa-1\}}$ and $\{Y_n\}$ is lag- m tail independent for all integer $m \geq \kappa$.

Example 4.1. Consider again $\{Y_n\}$ based on ARMAX sequence $\{X_n\}$ as in Example 3.1. We have that $\{X_n\}$ has lag- m tail dependence coefficient $\lambda(X_{n+m}|X_n) = \phi^m$ (Ferreira and Ferreira [2] 2012) and thus

$$\begin{aligned}
\lambda(Y_{n+m}|Y_n) &= p_{1,\dots,m}(0, \dots, 0) \mathbf{1}_{\{m \leq \kappa-1\}} \\
&\quad + \sum_{i=1 \vee (m-\kappa+1)}^m \sum_{i^*=0}^{\kappa-1} \phi^{i+i^*} p_{1,2,\dots,i^*+1,i^*+1+i,i^*+2+i,\dots,i^*+1+m}(1, 0, \dots, 0, 1, 0, \dots, 0).
\end{aligned}$$

Acknowledgements

The first author was partially supported by the research unit Centre of Mathematics and Applications of University of Beira Interior UIDB/00212/2020 - FCT (Fundação para a Ciência e a Tecnologia). The second author was financed by Portuguese Funds through FCT - Fundação para a Ciência e a Tecnologia within the Projects UIDB/00013/2020 and UIDP/00013/2020 of Centre of Mathematics of the University of Minho, UIDB/00006/2020 of Centre of Statistics and its Applications of University of Lisbon, UIDB/04621/2020 and UIDP/04621/2020 of Center for Computational and Stochastic Mathematics and PTDC/MAT-STA/28243/2017.

Conflict of interest statement: Authors state no conflict of interest.

References

- [1] Chernick M.R., Hsing T. and W.P. McCormick (1991). Calculating the extremal index for a class of stationary sequences. *Adv. Appl. Probab.* 23, 835–850.
- [2] Ferreira, M. and H. Ferreira (2012). On extremal dependence: some contributions. *TEST* 21(3), 566–583.
- [3] Ferreira, H., Martins, A.P. and M. da Graça Temido (2021). Extremal behaviour of a periodically controlled sequence with imputed values. *Statist. Papers* 62, 2991–3013.
- [4] Hall, A. and J. Hüsler (2006). Extremes of stationary sequences with failures. *Stoch. Models* 22, 537–557.

- [5] Joe, H. (1997) *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability 73, Chapman and Hall, London.
- [6] Kulik, R. and P. Soulier (2020). *Heavy-Tailed Time Series*. Berlin: Springer.
- [7] Leadbetter M.R. (1974). On extreme values in stationary sequences. *Z. Wahrsch. Verw. Gebiete* 28(4), 289–303.
- [8] Leadbetter, M.R., Lindgren, G. and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag.
- [9] Rosenblatt M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* 42(1), 43–47.
- [10] Zhang, T. (2021). Asymptotics of sample tail autocorrelations for tail dependent time series: phase transition and visualization. *Biometrika*, asab038.
- [11] Zhang, Z. (2005). A new class of tail-dependent time series models and its applications in financial time series. *Adv. Econ.* 20(B), 323–358.