



ELSEVIER

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi



1

On the max-semistable limit of maxima of stationary sequences with missing values[☆]

3

Andreia Hall^a, Maria da Graça Temido^{b,*}

5

^aUIMA, Departamento de Matemática, Universidade de Aveiro, Campo Universitário de Santiago, 3810-193 Aveiro, Portugal

^bCMUC, Departamento de Matemática, Largo D. Dinis, Universidade de Coimbra, Campo Universitario de Santiago 3001-454 Coimbra, Portugal

ARTICLE INFO

Article history:

Received 22 December 2007

Received in revised form

30 April 2008

Accepted 20 May 2008

Keywords:

Extreme value theory

Integer-valued models

Max-semistable laws

ABSTRACT

Let $\{X_n\}$ be a stationary sequence with marginal distribution in the domain of attraction of a max-semistable distribution. This includes all distributions in the domain of attraction of any max-stable distribution and also other distributions like some integer-valued distributions with exponential type tails such as the Negative Binomial case. We consider the effect of missing values on the distribution of the maximum term. The pattern of occurrence of the missing values must be either iid or strongly mixing. We obtain the expression of the extremal index for the resulting sequence.

The results generalize and extend the ones obtained for the max-stable domain of attraction.

© 2008 Published by Elsevier B.V.

7

1. Introduction

9

Integer-valued time series have received increasing attention in the probabilistic and statistical literature over the past two decades because of its applicability in many different areas such as the natural sciences, the social sciences, international tourism demand and economy. We refer to Hall and Scotto (2006) for several references on the subject and to McKenzie (2003) for an overview of the early work in this area.

11

13

Within the integer-valued models proposed in the literature, little is known about its extremal properties. In part, this is due to the fact that many integer-valued distributions do not belong to the domain of attraction on any extreme-value distribution. Anderson (1970) gave an important contribution to this limitation by obtaining upper and lower bounds for the limiting distribution of the maximum term of independent and identically distributed (iid) sequences with distributions exhibiting an exponentially decaying tail. He proved that an integer-valued distribution function (df) F , with infinite right endpoint, satisfies

15

17

$$\lim_{n \rightarrow +\infty} \frac{1 - F(n-1)}{1 - F(n)} = r, \quad r \in]1, +\infty[, \tag{1}$$

19

if and only if

$$\begin{cases} \limsup_{n \rightarrow \infty} F^n(x + b_n) \leq \exp(-r^{-x}), \\ \liminf_{n \rightarrow \infty} F^n(x + b_n) \geq \exp(-r^{-(x-1)}) \end{cases}$$

21

for any real x and b_n appropriately chosen. We shall say that a distribution belongs to Anderson's class if it satisfies (1). An example of a well known member of this class is the Negative Binomial distribution.

[☆] Contract/grant sponsors: FCT plurianual funding.

* Corresponding author. Tel.: +351 239791172.

E-mail addresses: andreia.hall@ua.pt (A. Hall), mgtm@mat.uc.pt (M.G. Temido).

1 Based on Anderson's work several stationary models have been studied with respect to the extremes and similar limiting
 2 bounds were obtained for the limiting distribution of the maximum term in presence of different dependence structures. For
 3 instance McCormick and Park (1992) considered a first-order autoregressive sequence with Negative Binomial marginal dis-
 4 tribution, Hall (1996) considered a class of integer-valued max-autoregressive models, Hall (2003) considered a general class
 5 of infinite moving average models, and Hall and Moreira (2006) considered a particular type of moving average models with
 6 geometric marginal distribution introduced by McKenzie (1986). Q1

7 In an attempt to overcome the presence of limiting bounds instead of a well defined limiting distribution, Temido (2000)
 8 proved that (1) is necessary and sufficient for the existence of a nondecreasing positive integer sequence $\{k_n\}$ satisfying

$$9 \quad \lim_{n \rightarrow +\infty} \frac{k_{n+1}}{k_n} = r, \quad r \in]1, +\infty[, \quad (2)$$

10 and of a real sequence $\{u_n\}$ such that $k_n(1 - F(u_n)) \rightarrow \tau > 0$, as $n \rightarrow \infty$, for some $\tau > 0$. So, if instead of looking at the maximum
 11 term of the first n observations we look at the maximum term of the first k_n observations, where $\{k_n\}$ satisfies (2), we can obtain
 12 a well defined limiting distribution for the maximum term. The limiting distribution is not a max-stable (MS) distribution but
 13 belongs to a larger class of distributions known as max-semistable (MSS) distributions. This class was introduced by Pancheva
 14 (1992) and will be described below.

15 Using the MSS class Hall and Temido (2007) have studied the limiting distribution of the maximum term of several stationary
 16 models with margins in Anderson's class.

17 Over the last years attention has been given to the effect of sub-sampling on the extremes of stationary sequences. This is
 18 important for the analysis of environmental and financial processes. We refer to Hall and Scotto (2006) for an overview of the
 19 work in this area. Many of the results consider deterministic sub-sampling of the sequences. However, there are several real
 20 data applications where sub-sampling is the result of a (random) occurrence of missing values. One reason for the interest in
 21 extremes observed at random sampling rates comes from the need to compare schemes for monitoring systems with breakdowns
 22 or systems with automatic replacement of devices in case of failures. Examples are encountered, for instance, in ocean engineering
 23 and environmental studies. In these areas, missing observations appear when the measuring equipment is not working properly
 24 or is out of service. The effect of missing values on the extremes of stationary sequences has been considered by Hall and Hüsler
 25 (2006) and by Hall and Scotto (2008). The former consider models with marginal distribution in the domain of attraction of any
 26 MS distribution and also distributions in Anderson's class. The later considers models with marginal distribution in the domain
 27 of attraction of a Fréchet distribution and with a moving average structure. In this work we consider the setup given in Hall and
 28 Hüsler and study the limiting distribution of the maximum term using MSS distributions.

29 What happens when a missing value occurs? In the majority of situations, one of two things may happen: either the observation
 30 is replaced by a fixed value (for instance a code), or the observation is completely lost and the data sample will be sub-sampled
 31 resulting in a smaller (and random) sample size. Occasionally, it may be of interest to avoid the occurrence of missing values and
 32 an automatic replacement of a device or machine may be available. In this case the resulting sample will be a mixture of two
 33 original samples.

In this paper we consider three different models which were motivated by the situations described above.

34 Let $\{X_n\}$ be a strictly stationary sequence of random variables (rvs) with marginal df F and $M_n = \max\{X_1, \dots, X_n\}$. Without loss
 35 of generality let the upper endpoint x_F of F be positive. Assume that $\{U_n\}$ is another stationary sequence, independent of $\{X_n\}$,
 36 having Bernoulli marginal distribution with parameter β , $0 \leq \beta \leq 1$. Based on the sequences $\{X_n\}$ and $\{U_n\}$ we define the following
 37 models:

38 M1—Model with missing values: $Y_n = U_n X_n$. In this case the marginal distribution of $\{Y_n\}$ is $1 - \beta + \beta F$ for nonnegative argument
 39 values. Missing values are replaced by zeros.

40 M2—Model with sub-sampling: $\{Z_n\}_{n \geq 1}$ with $Z_n = X_{i_n}$, where i_n represents the indices of the sequence $\{X_n\}$ for which $U_n = 1$.
 41 Let $N_n = \sup\{j : i_j \leq n\} = \sum_{j=1}^n U_j$ and let us only consider the first N_n values of $\{Z_n\}$. The marginal distribution of $\{Z_n\}$ is F . In
 42 this model missing values are lost.

43 M3—Model with replaced missing values: $W_n = U_n X_n + (1 - U_n) X_n^{(1)}$ where $\{X_n^{(1)}\}$ is an independent replica of $\{X_n\}$. In this case
 44 the marginal distribution of $\{W_n\}$ is also F . Missing values are replaced by a substituting sequence.

In this paper we only consider situations where $\{X_n\}$ and $\{U_n\}$ are independent.

45 We will be interested in the limiting distribution of the maximum term of these models. We define $M_n(Y) = \max\{Y_1, \dots, Y_n\}$
 46 and $M_n(W) = \max\{W_1, \dots, W_n\}$. For convenience of notation we define also $M_n(Z) = \max\{Z_1, \dots, Z_{N_n}\}$ because for the sequence Z_n
 47 we are only interested in the first N_n variables.

48 The sequences $\{U_n\}$ considered in this paper will be considered stationary and strongly mixing. Furthermore, we assume that
 49 some kind of dependence conditions hold for $\{X_n\}$. We consider an asymptotic independence condition similar to Leadbetter
 50 et al.'s (1983) $D(u_n)$ condition. We also consider two types of local dependence conditions similar to conditions $D^{(2)}(u_n)$ and
 51 $D^{(3)}(u_n)$ of Chernick et al. (1991). Each of these conditions will ensure that the clusters of exceedances formed in the stationary
 52 sequence $\{X_n\}$ have a particular type of pattern. Throughout this work $\{u_n\}$ represents a sequence of linearly normalized levels
 53 $u_n = a_n x + b_n$, $a_n > 0$, $b_n \in \mathbb{R}$ such that $u_n \rightarrow x_F$.

1 We obtain the expression of the extremal index and hence the limiting distribution of the maximum term of the transformed
 2 sequences (models M1, M2 and M3). The results generalize and extend the ones obtained for the MS domain of attraction given
 3 by Hall and Hüsler (2006).

2. The MSS class and stationary sequences

5 If instead of considering n rvs we consider k_n rvs, where $\{k_n\}$ satisfies

$$\lim_{n \rightarrow +\infty} \frac{k_{n+1}}{k_n} = r \quad \text{with } r \text{ in } [1, +\infty[, \tag{3}$$

7 then we obtain a larger class of possible limiting distributions for the maxima, known as the MSS class. This class, introduced by
 8 Pancheva (1992), includes the MS distributions and also nondegenerate limiting dfs for the maxima of iid rvs with either discrete
 9 or multimodal continuous dfs which are not MS. Following Pancheva (1992) we will say that a real df G is MSS if there are reals
 10 $r > 1, a = a(r) > 0$ and $b = b(r)$ such that $G(x) = G^r(ax + b), x \in \mathbb{R}$, or equivalently, if there exist a sequence of iid rvs with df F and two
 11 real sequences $\{a_n > 0\}$ and $\{b_n\}$ for which $\lim_{n \rightarrow +\infty} F^{k_n}(a_n x + b_n) = G(x)$, for each continuity point of G , with $\{k_n\}$ satisfying (3).
 12 In this case we will say that F belongs to the domain of attraction of G .

13 Note the subtle difference between the definitions of MS and MSS distributions. While in the definition of the MS distributions,
 14 G is MS if for all reals $r > 1$, there exist $a = a(r) > 0$ and $b = b(r)$ such that $G(x) = G^r(ax + b), x \in \mathbb{R}$, in the definition of the MSS
 15 distribution the equality may hold only for some values of $r > 1$. More precisely, if G is MSS but not MS, the equality holds only for
 16 some r and all its integer powers.

17 Analytically, a df in the class MSS can be written as follows:

$$G_{\gamma, \nu}(x) = \begin{cases} \exp\{-(1 + \gamma x)^{-1/\gamma} \nu(\log(1 + \gamma x))\}, & x \in \mathbb{R}, 1 + \gamma x > 0 \text{ and } \gamma \neq 0, \\ \mathbb{1}_{]-\infty, 0[}(\gamma), & x \in \mathbb{R}, 1 + \gamma x \leq 0 \text{ and } \gamma \neq 0, \\ \exp\{-e^{-x} \nu(x)\}, & x \in \mathbb{R} \text{ and } \gamma = 0, \end{cases}$$

19 where ν is a positive, bounded and periodic function with period $p = |\log a| = |\gamma| \log r$, when $\gamma \neq 0$, and $p = b = \log r$, when $\gamma = 0$. If
 20 the function ν is a suitable constant, we get the MS class.

21 We recall that for iid sequences with common df $F, \lim_{n \rightarrow \infty} k_n(1 - F(u_n)) = \tau$ is equivalent to $\lim_{n \rightarrow \infty} F^{k_n}(u_n) = e^{-\tau}$. Consequently,
 22 in the sequel, we shall deal with levels $u_n := u_n(\tau, k_n)$ satisfying these limits. In this context we will consider the set

$$23 \quad \Gamma(F, k_n) = \left\{ \tau > 0 : \exists \{u_n\} : \lim_{n \rightarrow +\infty} k_n(1 - F(u_n)) = \tau \right\},$$

24 introduced in Temido (2000), and note that if F is discrete and (3) holds, then $\Gamma(F, k_n)$ is not necessarily the interval $]0, +\infty[$. In
 25 fact, if for some $\tau > 0$ there exists $u_n(\tau, k_n)$, for another $\tau' > 0$ there exists $u_n(\tau', k_n)$ if and only if $\tau = r^m \tau'$, for some integer m . This
 26 enables us to conclude that if any discrete df belongs to the domain of attraction of G , with $\{k_n\}$ satisfying (2), then G must be
 27 discrete.

28 Temido (2002) proved that if F is an integer-valued df, with upper endpoint $x_F = +\infty$, and there exist sequences $\{k_n\}$ satisfying
 29 (2), $\{a_n > 0\}$ and $\{b_n\}$ such that $F^{k_n}(a_n x + b_n) \rightarrow G(x), n \rightarrow +\infty$, then $G(x) = \exp(-\eta r^{-[x]}), x \in \mathbb{R}$, for some $\eta > 0$, if and only if (1)
 30 holds (for any real $x, [x]$ denotes the greatest integer not exceeding x). In this case, with $k'_n = [k_n/\eta]$ we get

$$31 \quad \lim_{n \rightarrow +\infty} F^{k'_n}(a_n x + b_n) = \exp(-r^{-[x]}), \quad x \in \mathbb{R} \setminus \mathbb{Z}.$$

32 Temido and Canto e Castro (2003) consider stationary sequences $\{X_n\}$, satisfying a dependence restriction, $D_{k_n}(u_n)$, which
 33 extends Leadbetter et al.'s (1983) $D(u_n)$ condition.

34 **Definition 2.1** (Temido and Canto e Castro, 2003). Let $\{k_n\}$ be a nondecreasing sequence of positive integers. The sequence of rvs
 35 $\{X_n\}$ satisfies condition $D_{k_n}(u_n)$ if for any integers $1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq k_n$, for which $j_1 - i_p > \ell_n$, we have

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

37 where $\lim_{n \rightarrow +\infty} \alpha_{n, \ell_n} = 0$ for some sequence $\ell_n = o_n(k_n)$.

38 Considering stationary sequences $\{X_n\}$ satisfying this long range condition, $D_{k_n}(u_n)$, Temido and Canto e Castro (2003) prove
 39 that the limiting distribution of M_{k_n} is MSS, whenever it exists. Namely, if there is $\{k_n\}$ as above and $\{a_n > 0\}$ and $\{b_n\}$ such that
 40 condition $D(a_n x + b_n)$ holds for the stationary sequence $\{X_n\}$, the sequence $\{k_n(1 - F(a_n x + b_n))\}$ is bounded, and $P(a_n(M_{k_n} - b_n) \leq x)$
 41 converges to $G(x)$, for each continuity point of the nondegenerate df G , then G is a MSS df. Furthermore, following the same authors
 we present the definition of extremal index.

1 **Definition 2.2** (Temido and Canto e Castro, 2003). We shall say that $\{X_n\}$ has an extremal index θ , with θ in $[0, 1]$, if there exists a
 3 nondecreasing positive integer sequence $\{k_n\}$ satisfying (3) such that, for all $\tau \in \Gamma(F, k_n)$ and all corresponding $u_n(\tau, k_n)$, we have
 $\lim_{n \rightarrow +\infty} P(M_{k_n} \leq u_n(\tau, k_n)) = e^{-\theta\tau}$.

5 In Hall and Hüsler (2006) an important lemma is introduced. Indeed, the authors prove that if $\{U_n\}$ is a Bernoulli strongly
 mixing stationary sequence and the long range condition $D(u_n)$ holds for $\{X_n\}$, then it also holds for $\{Y_n\}$ in Model M1. For the
 sequel and *mutatis mutandis* we have the following result.

7 **Lemma 2.1.** Let $\{k_n\}$ be a positive integer and nondecreasing sequence and suppose that $\{U_n\}$ is a Bernoulli strongly mixing stationary
 sequence. If condition $D_{k_n}(u_n)$ holds for $\{X_n\}$ then it also holds for $\{Y_n\}$.

9 **3. Independent missing values**

11 **Proposition 3.1.** Let $\{U_n\}$ be an iid Bernoulli (β) sequence, and $\{X_n\}$ a stationary sequence with extremal index θ in the sense of
 definition (2.2) and cluster size distribution π . Define

$$\theta^* = \theta(1 - \Pi(1 - \beta))/\beta \quad \text{and} \quad \tau^* = \tau\beta,$$

$$\theta^{**} = \theta(1 - \Pi(1 - \beta) + 1 - \Pi(\beta)),$$

13 where $\Pi(h) = \sum_{i=1}^{\infty} \pi(i)h^i$.
 15 If $P\{M_{k_n} \leq u_n(\tau)\} \xrightarrow{n \rightarrow \infty} e^{-\theta\tau}$, $\tau \in \Gamma(F, k_n)$, then,

$$P\{M_{k_n}(V) \leq u_n(\tau)\} \xrightarrow{n \rightarrow \infty} e^{-\theta_V\tau_V},$$

17 where $\theta_V = \theta^*$, $\tau_V = \tau^*$ for the sequences $\{Y_n\}$ and $\{Z_n\}$, and $\theta_V = \theta^{**}$, $\tau_V = \tau$, for the sequence $\{W_n\}$.

Proof. Let $\{s_n\}$ be a sequence of positive integers satisfying

$$\lim_{n \rightarrow \infty} s_n^{-1} = \lim_{n \rightarrow \infty} \frac{s_n \ell_n}{k_n} = \lim_{n \rightarrow \infty} s_n \alpha_{n, \ell_n} = 0. \tag{4}$$

19 The idea for this proof is to divide the first k_n elements of the sequence $\{Y_n\}$ into blocks of size $r_n := [k_n/s_n]$, calculate the
 21 probability of no exceedances in those blocks and then use Lemma 4 from Temido and Canto e Castro (2003).

Let us consider the first r_n elements of the sequences $\{X_n\}$ and $\{Y_n\}$. Assuming $u_n > 0$, then,

$$\begin{aligned} P\{M_{r_n}(Y) \leq u_n\} &= P\{\{Y_1, \dots, Y_{r_n}\} \text{ contains no exceedances of } u_n\} \\ &= P\{M_{r_n} \leq u_n\} + \sum_{j=1}^{r_n} P\{\{X_1, \dots, X_{r_n}\} \text{ contains } j \text{ exceedances all withdrawn}\}. \end{aligned}$$

23 From the definition of extremal index we get $P\{M_{k_n} \leq u_n\} - e^{-k_n P\{X_1 > u_n\}\theta} \xrightarrow{n \rightarrow \infty} 0$. On the other hand, from Lemma 4 from Temido
 25 and Canto e Castro (2003) we have

$$P\{M_{k_n} \leq u_n\} - P^{s_n}\{M_{r_n} \leq u_n\} \xrightarrow{n \rightarrow \infty} 0.$$

27 Combining both results we obtain

$$P\{M_{r_n} \leq u_n\} = 1 - \frac{k_n}{s_n} \theta P\{X_1 > u_n\} + o\left(\frac{1}{s_n}\right),$$

29 as long as $\lim_{n \rightarrow \infty} k_n P\{X_1 > u_n\} \neq 0$.

Since $k_n P\{X_1 > u_n\} \xrightarrow{n \rightarrow \infty} \tau > 0$ we have

$$P\{M_{r_n}(Y) \leq u_n\} = 1 - \frac{\theta\tau}{s_n} \left(1 - \sum_{j=1}^{r_n} \frac{s_n}{\theta\tau} P\{\{X_1, \dots, X_{r_n}\} \text{ contains } j \text{ exceedances all withdrawn}\} \right) + o\left(\frac{1}{s_n}\right).$$

31 But, with $\lambda_{n,i}(\cdot) := \mathbb{1}_{\{X_i > u_n\}}(\cdot)$, we also have

$$\begin{aligned}
 &P(\{X_1, \dots, X_{r_n}\} \text{ contains } j \text{ exceedances all withdrawn}) \\
 &= \sum_{i_1 < i_2 < \dots < i_j} P \left\{ \bigcap_{m=1}^j \{X_{i_m} > u_n\}, \text{ all other } X_i \leq u_n, \bigcap_{m=1}^j \{U_{i_m} = 0\} \right\} \\
 &= \sum_{i_1 < i_2 < \dots < i_j} P \left\{ \bigcap_{m=1}^j \{X_{i_m} > u_n\}, \text{ all other } X_i \leq u_n \right\} P(U_1 = 0)^j \\
 &= P \left(\sum_{i=1}^{r_n} \chi_{n,i} = j \right) (1 - \beta)^j.
 \end{aligned}$$

Now, using the above arguments again,

$$\begin{aligned}
 P \left(\sum_{i=1}^{r_n} \chi_{n,i} = j \right) &= \pi(j) P \left(\sum_{i=1}^{r_n} \chi_{n,i} > 0 \right) (1 + o(1)) \\
 &= \pi(j) \left(1 - P \left(\sum_{i=1}^{r_n} \chi_{n,i} = 0 \right) \right) (1 + o(1)) \\
 &= \pi(j) (1 + o(1)) \frac{\theta\tau}{s_n}.
 \end{aligned}$$

Therefore

$$P\{M_{r_n}(Y) \leq u_n\} = 1 - \frac{\theta\tau}{s_n} \left(1 - \sum_{j=1}^{r_n} \pi(j) (1 - \beta)^j (1 + o(1)) \right) + o\left(\frac{1}{s_n}\right).$$

Considering Lemmas 2.1 and 4 from [Temido and Canto e Castro \(2003\)](#), we use dominated convergence to obtain

$$P\{M_{k_n}(Y) \leq u_n\} \xrightarrow{n \rightarrow \infty} e^{-\theta\tau(1-\Pi(1-\beta))},$$

where $\Pi(s) = \sum_{j=1}^{\infty} \pi(j)s^j$ represents the probability generating function of the cluster size distribution.

As for the sequences $\{Z_n\}$ and $\{W_n\}$ it suffices to notice that $P\{M_{k_n}(Z) \leq u_n\} = P\{M_{k_n}(Y) \leq u_n\}$ and that $P\{M_{k_n}(W) \leq u_n\} = P\{M_{k_n}(Y) \leq u_n\} P\{M_{k_n}(\tilde{Y}) \leq u_n\}$ where $\{\tilde{Y}_n\}$ is a sequence defined by $\tilde{Y}_n = (1 - U_n)X_n$. \square

4. The effect of missing values under condition $D_{k_n}^{(2)}(u_n)$

We now consider a stationary sequence $\{X_n\}$ satisfying a local dependence condition, $D_{k_n}^{(2)}(u_n)$, and with marginal df F in the domain of attraction of some MSS df.

The natural extension of the local dependence condition $D^{(2)}(u_n)$ of [\(Chernick et al., 1991\)](#), which is similar to the condition $D''(u_n)$ (defined in [Leadbetter and Nandagopalan, 1989](#)), in this new context is given in the following definition.

Definition 4.1 ([Temido, 2000](#)). Let $\{k_n\}$ be a nondecreasing positive integer sequence such that $\lim_{n \rightarrow \infty} k_n = +\infty$. The stationary sequence $\{X_n\}$ satisfies the condition $D_{k_n}^{(2)}(u_n)$ if $D_{k_n}(u_n)$ holds and, for some positive integer-valued sequence $\{s_n\}$ satisfying (4), we have

$$\lim_{n \rightarrow \infty} k_n P\{X_1 > u_n \geq X_2, M_{3,r_n} > u_n\} = 0, \tag{5}$$

where $r_n = \lfloor k_n/s_n \rfloor$ and $M_{ij} = \max\{X_k, k = i, \dots, j\}$.

Clearly, (5) is implied by the condition

$$\lim_{n \rightarrow +\infty} k_n \sum_{j=3}^{r_n} P\{X_1 > u_n, X_{j-1} \leq u_n < X_j\} = 0,$$

which in some situations is easier to check.

Recall that under condition $D^{(2)}(u_n)$ the clusters of exceedances form groups of consecutive observations. Once the sequence falls below the high threshold u_n the probability of a new upcrossing in the near future is negligible.

1 Under $D_{k_n}^{(2)}(u_n)$ we can compute the extremal index applying the following result.

3 **Proposition 4.1** (Temido, 2000). Let $\{k_n\}$ be a nondecreasing positive integer sequence satisfying (1) and $\{X_n\}$ a stationary sequence under condition $D_{k_n}^{(2)}(u_n(\tau, k_n))$, for all τ in $\Gamma(F, k_n)$ and all corresponding $u_n(\tau, k_n)$. Then $\{X_n\}$ has extremal index θ if and only if

$$\lim_{n \rightarrow +\infty} P(X_2 \leq u_n(\tau, k_n) | X_1 > u_n(\tau, k_n)) = \theta.$$

5 The following proposition, our main result of this section, is a generalization of Theorem 4 in Hall and Hüsler (2006) concerning normalized levels $u_n(\tau, k_n)$. We use $\{V_n\}$ to denote any of the sequences $\{Y_n\}$, $\{Z_n\}$ and $\{W_n\}$ and recall that in the model M2 we consider the maximum of the first N_{k_n} rvs.

9 **Proposition 4.2.** Let $\{X_n\}$ be a stationary sequence with df F and $\{U_n\}$ be a Bernoulli $B(\beta)$ strongly mixing stationary sequence independent of $\{X_n\}$. Let $\{k_n\}$ be a nondecreasing positive integer sequence satisfying (3). Suppose that $\{X_n\}$ satisfies $D_{k_n}^{(2)}(u_n(\tau, k_n))$, for all $\tau \in \Gamma(F, k_n)$ and all corresponding $u_n(\tau, k_n)$. In addition suppose that $\{X_n\}$ has extremal index $\theta > 0$ associated with a limiting cluster size distribution π . For $i \geq 1$, take

$$\nabla(i) = P(U_1 = 0, U_2 = 0, \dots, U_i = 0), \quad \bar{\nabla}(i) = P(U_1 = 1, U_2 = 1, \dots, U_i = 1),$$

13
$$\theta^* = \theta \left(1 - \sum_{j=1}^{\infty} \pi(j) \nabla(j) \right) / \beta \quad \text{and} \quad \theta^{**} = \beta \theta^* + \theta \left(1 - \sum_{j=1}^{\infty} \pi(j) \bar{\nabla}(j) \right).$$

Then

15
$$\lim_{n \rightarrow \infty} P(M_{k_n}(V) \leq u_n(\tau, k_n)) = e^{-\theta_V \tau_V},$$

where $\theta_V = \theta^*$, $\tau_V = \tau\beta$ for the sequences $\{Y_n\}$ and $\{Z_n\}$, and $\theta_V = \theta^{**}$, $\tau_V = \tau$ for the sequence $\{W_n\}$, for all $\tau \in \Gamma(F, k_n)$.

17 The proof of this proposition follows the arguments used in Theorem 4 of Hall and Hüsler (2006) where, for F integer-valued, $e^{-\theta\tau(x-1)} \leq \liminf_{n \rightarrow \infty} P\{M_n \leq u_n\} \leq \limsup_{n \rightarrow \infty} P\{M_n \leq u_n\} \leq e^{-\theta\tau(x)}$ is replaced by $\lim_{n \rightarrow \infty} P\{M_{k_n} \leq u_n(\tau, k_n)\} = e^{-\theta\tau(x)}$. Observe that, as a consequence of what was said before, if F is an integer-valued df satisfying (1) and $u_n(\tau, k_n) = a_n x + b_n$, for some $a_n > 0$ and b_n , then, there exist $\{k_n\}$ satisfying (2) such that $\lim_{n \rightarrow +\infty} P(M_{k_n} \leq u_n(\tau, k_n)) = \exp(-\theta r^{-|x|})$, $x \in \mathbb{R} \setminus \mathbb{Z}$.

21 **5. The effect of missing values under condition $D_{k_n}^{(3)}(u_n)$**

23 In this section we consider that the stationary sequence $\{X_n\}$ satisfies an extension of the local dependence condition $D^{(3)}(u_n)$ of Chernick et al. (1991) defined by Temido (2000).

25 **Definition 5.1.** Let $\{k_n\}$ be a nondecreasing positive integer sequence such that $\lim_{n \rightarrow \infty} k_n = +\infty$. The stationary sequence $\{X_n\}$ satisfies the condition $D_{k_n}^{(3)}(u_n)$ if $D_{k_n}(u_n)$ holds and, for some positive integer-valued sequence $\{s_n\}$ satisfying (4), we have

$$\lim_{n \rightarrow +\infty} k_n P(X_1 > u_n \geq M_{2,3}, M_{4,r_n} > u_n) = 0, \tag{6}$$

27 where $r_n = \lfloor k_n/s_n \rfloor$ and $M_{ij} = \max\{X_k, k = i, \dots, j\}$.

As in the previous paragraph, (6) is implied by the condition

29
$$\lim_{n \rightarrow +\infty} k_n \sum_{i=4}^{r_n} P(X_1 > u_n, M_{i-2, i-1} \leq u_n < X_i) = 0,$$

which in some situations is easier to deal with.

31 We also suppose that

$$\lim_{n \rightarrow \infty} k_n P(X_1 > u_n, X_2 > u_n) = 0, \tag{7}$$

33 which together with the previous condition implies that the clusters have an alternating pattern above and below the threshold u_n .

35 As mentioned before, under condition $D_{k_n}^{(2)}(u_n)$ the clusters of exceedances are formed by runs of consecutive observations over u_n . However, under $D_{k_n}^{(3)}(u_n)$ the clusters may exhibit a wide variety of patterns which have in common the following

property: within a cluster, the sequence does not stay below u_n for more than one time instance. One particular pattern which may occur under $D_{k_n}^{(3)}(u_n)$ is the purely oscillating pattern which is characteristic of several time series models such as first-order autoregressive models with negative coefficient. We shall only consider such type of patterns and for that we must impose the additional condition (7).

Following Chernick et al. (1991) and O'Brien (1987), Temido (2000) proved that under condition $D_{k_n}^{(3)}(u_n)$, $\lim_{n \rightarrow \infty} P(M_{k_n} \leq u_n) - \exp(-k_n P(X_1 > u_n, X_2 \leq u_n, X_3 \leq u_n)) = 0$ and that the extremal index is obtained as

$$\theta = \lim_{n \rightarrow \infty} P(X_2 \leq u_n(\tau, k_n), X_3 \leq u_n(\tau, k_n) | X_1 > u_n(\tau, k_n)).$$

However, considering in addition condition (7) we can compute the extremal index in a different way. Indeed, like in the case when $D_{k_n}^{(2)}(u_n)$ holds, a bivariate tail distribution suffices to determine the extremal index.

Lemma 5.1. Let $\{k_n\}$ be a nondecreasing positive integer sequence satisfying (3) and $\{X_n\}$ a stationary sequence satisfying $D_{k_n}(u_n)$, for some real sequence $\{u_n\}$ such that $\limsup k_n(1 - F(u_n)) < +\infty$:

1. If $\{X_n\}$ satisfies $D_{k_n}^{(3)}(u_n)$ and (7) then

$$\lim_{n \rightarrow \infty} P(M_{k_n} \leq u_n) - \exp(-k_n P(X_1 > u_n, X_3 \leq u_n)) = 0. \tag{8}$$

2. If $\{X_n\}$ satisfies $D_{k_n}^{(3)}(u_n(\tau, k_n))$ and $\lim_{n \rightarrow \infty} k_n P(X_1 > u_n(\tau, k_n), X_2 > u_n(\tau, k_n)) = 0$, for all $\tau \in \Gamma(F, k_n)$ and all corresponding $u_n(\tau, k_n)$, then $\{X_n\}$ has extremal index θ if and only if

$$\lim_{n \rightarrow \infty} P(X_3 \leq u_n(\tau, k_n) | X_1 > u_n(\tau, k_n)) = \theta. \tag{9}$$

Proof. 1. Using the arguments of O'Brien (1987) we prove that, under $D_{k_n}(u_n)$,

$$P(M_{k_n} \leq u_n) - \exp(-k_n P(X_1 > u_n, M_{2,r_n} \leq u_n)) \rightarrow 0, \quad n \rightarrow \infty.$$

But, under the assumptions of the lemma, we have

$$\begin{aligned} P(X_1 > u_n, M_{2,r_n} \leq u_n) &= P(X_1 > u_n \geq M_{2,3}, M_{4,r_n} \leq u_n) \\ &= P(X_1 > u_n \geq M_{2,3}) - P(X_1 > u_n \geq M_{2,3}, M_{4,r_n} > u_n) \\ &= P(X_1 > u_n) - P(X_1 > u_n, X_2 > u_n) - P(X_1 > u_n, X_3 > u_n) \\ &\quad + P(X_1 > u_n, X_2 > u_n, X_3 > u_n) + o_n\left(\frac{1}{k_n}\right) \\ &= P(X_1 > u_n) - P(X_1 > u_n, X_3 > u_n) + o_n\left(\frac{1}{k_n}\right) \\ &= P(X_1 > u_n, X_3 \leq u_n) + o_n\left(\frac{1}{k_n}\right) \end{aligned}$$

and thus (8) holds.

2. If $\{X_n\}$ has extremal index θ we have

$$\lim_{n \rightarrow \infty} P(M_{k_n} \leq u_n(\tau, k_n)) = \exp(-\theta\tau), \tag{10}$$

for all $\tau \in \Gamma(F, k_n)$. Hence, by the first part of the lemma we get

$$\lim_{n \rightarrow \infty} k_n P(X_1 > u_n(\tau, k_n), X_3 \leq u_n(\tau, k_n)) = \theta\tau,$$

or, equivalently, the limit (9) occurs.

The proof of the converse is similar. \square

After this result we can establish the limit in distribution of $M_{k_n}(V)$ for all the models considered in this work.

In what follows we write $A_{t,h}^{(n)} := \{X_t > u_n, X_{t+1} \leq u_n < X_{t+2}, \dots, X_{h-1} \leq u_n < X_h\}$ and $B_{t,h}^{(n)} := \{X_t > u_n, X_{t+2} > u_n, \dots, X_{h-2} > u_n, X_h > u_n\}$ where $h = t + 2m$ for some nonnegative integer m . We also consider $M_{l,k} = -\infty$ if $k < l$.

Proposition 5.1. Let $\{X_n\}$ be a stationary sequence with df F and $\{U_n\}$ be a Bernoulli $B(\beta)$ strongly mixing stationary sequence independent of $\{X_n\}$. Let $\{k_n\}$ be a nondecreasing positive integer sequence satisfying (3). Suppose that $\{X_n\}$ satisfies $D_{k_n}^{(3)}(u_n(\tau, k_n))$ and

1 $\lim_{n \rightarrow \infty} k_n P(X_1 > u_n(\tau, k_n), X_2 > u_n(\tau, k_n)) = 0$, for all $\tau \in \Gamma(F, k_n)$ and all corresponding $u_n(\tau, k_n)$. In addition suppose that $\{X_n\}$ has extremal index $\theta > 0$ associated with a limiting cluster size distribution π . For $i \geq 1$, take

3
$$\nabla_{\text{even}}(i) = P(U_2 = 0, U_4 = 0, \dots, U_{2i} = 0), \quad \bar{\nabla}_{\text{even}}(i) = P(U_2 = 1, U_4 = 1, \dots, U_{2i} = 1),$$

$$\theta^* = \theta \left(1 - \sum_{j=1}^{\infty} \pi(j) \nabla_{\text{even}}(j) \right) / \beta \quad \text{and} \quad \theta^{**} = \beta \theta^* + \theta \left(1 - \sum_{j=1}^{\infty} \pi(j) \bar{\nabla}_{\text{even}}(j) \right).$$

5 Then

$$\lim_{n \rightarrow \infty} P(M_{k_n}(V) \leq u_n(\tau, k_n)) = e^{-\theta_V \tau_V},$$

7 where $\theta_V = \theta^*$, $\tau_V = \tau \beta$ for the sequences $\{Y_n\}$ and $\{Z_n\}$, and $\theta_V = \theta^{**}$, $\tau_V = \tau$ for the sequence $\{W_n\}$, for all $\tau \in \Gamma(F, k_n)$.

Proof. Consider the first r_n elements of the sequences $\{X_n\}$ and $\{Y_n\}$. Take $u_n := u_n(\tau, k_n) > 0$.

9 As in the proof of Proposition 3.1 we obtain

$$P(M_{r_n}(Y) \leq u_n) = 1 - \frac{\theta \tau}{s_n} \left(1 - \sum_{j=1}^{r_n} \frac{s_n}{\theta \tau} P(\{X_1, \dots, X_{r_n}\} \text{ contains } j \text{ exceedances all withdrawn}) \right) + o\left(\frac{1}{s_n}\right).$$

11 But,

$$\begin{aligned} & P(\{X_1, \dots, X_{r_n}\} \text{ contains } j \text{ exceedances all withdrawn}) \\ &= P(A_{1,2j-1}^{(n)}, M_{2j,r_n} \leq u_n) P\left\{ \bigcap_{i=1}^j \{U_{2i} = 0\} \right\} \\ &+ P(X_1 > u_n, A_{2,2j-2}^{(n)}, M_{2j-1,r_n} \leq u_n) P\left\{ U_1 = 0, \bigcap_{i=1}^{j-1} \{U_{2i} = 0\} \right\} \\ &+ P(M_{1,r_n-2j+1} \leq u_n, A_{r_n-2j+2,r_n}^{(n)}) P\left\{ \bigcap_{i=1}^j \{U_{2i} = 0\} \right\} \\ &+ P(M_{1,r_n-2j+2} \leq u_n, A_{r_n-2j+3,r_n-1}^{(n)}, X_{r_n} > u_n) P\left\{ \bigcap_{i=1}^{j-1} \{U_{2i-1} = 0\}, U_{2j-2} = 0 \right\} \\ &+ \sum_{i=2}^{r_n-2j-1} P(M_{1,i} \leq u_n, A_{i+1,i+2j-1}^{(n)}, M_{i+2j,r_n} \leq u_n) P\left\{ \bigcap_{i=1}^j \{U_{2i} = 0\} \right\} \\ &+ \sum \text{remaining terms} \\ &\times P\{U_1, \dots, U_{r_n} \text{ equals zero where the exceedances occur}\}. \end{aligned}$$

13 The first and third terms of the right-hand side of the equality are clearly $o(1/s_n)$ and the second and fourth are $o(1/k_n)$ by (7). The last term is bounded by

$$\begin{aligned} & \sum \text{remaining terms} \\ & \leq r_n P(X_1 > u_n, X_2 > u_n) + r_n P(X_1 > u_n \geq M_{2,3}, M_{4,r_n} > u_n) \\ & = o\left(\frac{1}{s_n}\right), \end{aligned}$$

15

by condition $D_{k_n}^{(3)}(u_n)$ and (7). On the other hand, by the same condition,

17
$$\sum_{i=2}^{r_n-2j-1} P(M_{1,i} \leq u_n, A_{i+1,i+2j-1}^{(n)}, M_{i+2j,r_n} \leq u_n) = \pi(j)(1 + o(1)) \frac{\theta \tau}{s_n}.$$

Therefore

19
$$\begin{aligned} & \frac{s_n}{\theta \tau} P(\{X_1, \dots, X_{r_n}\} \text{ contains } j \text{ exceedances all withdrawn}) \\ & = \nabla_{\text{even}}(j) \pi(j) (1 + o(1)) \end{aligned}$$

1 and

$$P\{M_{r_n}(Y) \leq u_n\} = 1 - \frac{\theta\tau}{s_n} \left(1 - \sum_{j=1}^{r_n} \pi(j) \nabla_{\text{even}(j)} (1 + o(1)) \right) + o\left(\frac{1}{s_n}\right).$$

3 Since the sequence $\{Y_n\}$ satisfies condition $D_{k_n}(u_n)$, once again by Lemma 4 of Temido and Canto e Castro (2003) and dominated convergence we obtain

$$5 \quad P\{M_{k_n}(Y) \leq u_n\} \xrightarrow{n \rightarrow \infty} e^{-\theta\tau(1 - \sum_{j=1}^{\infty} \pi(j) \nabla_{\text{even}(j)})}.$$

7 As for the sequences $\{Z_n\}$ and $\{W_n\}$ it suffices to notice that $P\{M_{k_n}(Z) \leq u_n\} = P\{M_{k_n}(Y) \leq u_n\}$ and that $P\{M_{k_n}(W) \leq u_n\} = P\{M_{k_n}(Y) \leq u_n\}P\{M_{k_n}(\bar{Y}) \leq u_n\}$ where $\{\bar{Y}_n\}$ is a sequence defined by $\bar{Y}_n = (1 - U_n)X_n$. \square

9 We now establish a result which enables us to compute the limiting cluster size distribution $\pi(\cdot)$ for stationary sequences satisfying (7) and $D_{k_n}^{(3)}(u_n)$.

11 **Lemma 5.2.** Let $\{X_n\}$ be a stationary sequence that satisfies condition $D_{k_n}^{(3)}(u_n(\tau, k_n))$ and $\lim_{n \rightarrow \infty} k_n P(X_1 > u_n(\tau, k_n), X_2 > u_n(\tau, k_n)) = 0$, for all $\tau \in \Gamma(F, k_n)$ and all corresponding $u_n(\tau, k_n)$. In addition suppose that $\{X_n\}$ has extremal index $\theta > 0$ associated with a limiting cluster size distribution π . Then, for $j \geq 1$,

$$13 \quad \pi(j) = \frac{\lim_{n \rightarrow \infty} k_n P(X_1 \leq u_n, X_3 > u_n, X_5 > u_n, \dots, X_{2j+1} > u_n, X_{2j+3} \leq u_n)}{\theta\tau}.$$

Proof. Observe first that

$$15 \quad \pi_n(j) = P\left(\sum_{i=1}^{r_n} \chi_{n,i} = j \mid \sum_{i=1}^{r_n} \chi_{n,i} > 0\right) = \frac{P(\sum_{i=1}^{r_n} \chi_{n,i} = j)}{P(M_{r_n} > u_n)}.$$

Thus, attending to the proof of Proposition 5.1 we have

$$17 \quad \begin{aligned} \pi_n(j) &= \frac{P(\sum_{i=1}^{r_n} \chi_{n,i} = j)}{\theta\tau/s_n + o(1/s_n)} \\ &= \frac{\sum_{i=2}^{r_n-2j-1} P(M_{1,i} \leq u_n, A_{i+1,i+2j-1}^{(n)}, M_{i+2j,r_n} \leq u_n) + o(1/s_n)}{\theta\tau/s_n + o(1/s_n)}. \end{aligned}$$

If $i \geq 3$ we have

$$19 \quad \begin{aligned} &P(X_{i-1} \leq u_n, X_i \leq u_n, A_{i+1,i+2j-1}^{(n)}, M_{i+2j,r_n} \leq u_n) \\ &\quad - P(M_{1,i-2} \leq u_n, X_{i-1} \leq u_n, X_i \leq u_n, A_{i+1,i+2j-1}^{(n)}, M_{i+2j,r_n} \leq u_n) \\ &\leq \sum_{l=1}^{i-2} P(X_l > u_n, X_{i-1} \leq u_n, X_i \leq u_n, A_{i+1,i+2j-1}^{(n)}) = o(1/k_n), \end{aligned}$$

by condition $D_{k_n}^{(3)}(u_n)$ and stationarity. Moreover, for $i \geq 2$, we get

$$21 \quad \begin{aligned} &P(X_{i-1} \leq u_n, A_{i+1,i+2j-1}^{(n)}, M_{i+2j,r_n} \leq u_n) \\ &\quad - P(X_{i-1} \leq u_n, X_i \leq u_n, A_{i+1,i+2j-1}^{(n)}, M_{i+2j,r_n} \leq u_n) \\ &\leq P(X_{i-1} \leq u_n, X_i > u_n, X_{i+1} > u_n) = o(1/k_n), \end{aligned}$$

by (7). Then, for $i \geq 2$,

$$23 \quad \begin{aligned} &P(M_{1,i-2} \leq u_n, X_{i-1} \leq u_n, X_i \leq u_n, A_{i+1,i+2j-1}^{(n)}, M_{i+2j,r_n} \leq u_n) \\ &= P(X_{i-1} \leq u_n, A_{i+1,i+2j-1}^{(n)}, M_{i+2j,r_n} \leq u_n) + o(1/k_n). \end{aligned}$$

1 On the other hand, for $i \leq r_n - 2j - 2$, we have

$$\begin{aligned} & P(X_{i-1} \leq u_n, A_{i+1, i+2j-1}^{(n)}, X_{i+2j} \leq u_n, X_{i+2j+1} \leq u_n) \\ & - P(X_{i-1} \leq u_n, A_{i+1, i+2j-1}^{(n)}, X_{i+2j} \leq u_n, X_{i+2j+1} \leq u_n, M_{i+2j+2, r_n} \leq u_n) \\ & \leq \sum_{l=i+2j+2}^{r_n} P(X_{i-1} \leq u_n, A_{i+1, i+2j-1}^{(n)}, X_{i+2j} \leq u_n, X_{i+2j+1} \leq u_n, X_l > u_n) \\ & = o(1/k_n), \end{aligned}$$

3 once again by $D_{k_n}^{(3)}(u_n)$ and stationarity. In the same way we deduce

$$\begin{aligned} & P(X_{i-1} \leq u_n, A_{i+1, i+2j-1}^{(n)}, X_{i+2j} \leq u_n, X_{i+2j+1} \leq u_n) \\ & = P(X_{i-1} \leq u_n, A_{i+1, i+2j-1}^{(n)}, X_{i+2j+1} \leq u_n) + o(1/k_n) \end{aligned}$$

5 and

$$\begin{aligned} & P(X_{i-1} \leq u_n, B_{i+1, i+2j-1}^{(n)}, X_{i+2j+1} \leq u_n) \\ & = P(X_{i-1} \leq u_n, A_{i+1, i+2j-1}^{(n)}, X_{i+2j+1} \leq u_n) + o(1/k_n). \end{aligned}$$

7 Thus, the stationarity of the process enables us to conclude that

$$\begin{aligned} \pi_n(j) &= \frac{\sum_{i=2}^{r_n-2j-1} (P(X_{i-1} \leq u_n, B_{i+1, i+2j-1}^{(n)}, X_{i+2j+1} \leq u_n) + o(1/k_n)) + o(1/s_n)}{\theta\tau/s_n + o(1/s_n)} \\ &= \frac{(r_n - 2j - 2)(P(X_1 \leq u_n, B_{3, 2j+1}^{(n)}, X_{2j+3} \leq u_n) + o(1/k_n)) + o(1/s_n)}{\theta\tau/s_n + o(1/s_n)} \\ &\rightarrow \frac{\lim_{n \rightarrow \infty} k_n P(X_1 \leq u_n, B_{3, 2j+1}^{(n)}, X_{2j+3} \leq u_n)}{\theta\tau} := \pi(j). \quad \square \end{aligned}$$

9 **6. Examples**

11 In this section we give several examples of application of the results of the previous sections. We shall consider five types of stationary sequences with extremal index $\theta < 1$. For the first two examples condition $D_{k_n}^{(2)}(u_n)$ holds, while for the last three condition $D_{k_n}^{(2)}(u_n)$ fails but condition $D_{k_n}^{(3)}(u_n)$ holds.

We consider two particular types of sequences $\{U_n\}$:

13

1. The first is the simplest case where $\{U_n\}$ is iid with $P(U_n = 1) = \beta, \beta \in]0, 1[$.
2. The second case is a homogeneous Markov chain with one-step transition probabilities

$$\begin{cases} P\{U_n = 1 | U_{n-1} = 1\} = \eta, \\ P\{U_n = 1 | U_{n-1} = 0\} = \mu. \end{cases}$$

17

In this model $\{U_n\}$ defines a sequence where the probability of failure ($U_n = 0$) depends only on whether a failure has just occurred. $\{U_n\}$ forms geometric blocks of consecutive zeros followed by geometric blocks of consecutive ones. In order to have a stationary chain the initial distribution is such that $P\{U_0 = 1\} = (1 - \eta)/(1 - \eta + \mu)$.

19

Hence, given any values of $\eta, \mu \in [0, 1]$

21

$$\beta = \frac{\mu}{1 - \eta + \mu}. \tag{11}$$

Hall and Hüsler (2006) proved that $\{U_n\}$ is strongly mixing.

23

For the $\{U_n\}$ Markov chain Hall and Hüsler (2006) have determined the expressions for $\nabla(j)$ and $\bar{\nabla}(j)$:

$$\nabla(j) = \frac{(1 - \mu)^{j-1}(1 - \eta)}{1 - \eta + \mu}, \quad j \geq 1,$$

25

$$\bar{\nabla}(j) = P\{U_1 = 1, \dots, U_j = 1\} = \frac{\mu\eta^{j-1}}{1 - \eta + \mu}, \quad j \geq 1.$$

1 Using the properties of homogeneous Markov chains we obtain

$$\nabla_{\text{even}}(j) = (1 - \mu(1 - \mu + \eta))^{j-1} \frac{1 - \eta}{1 - \eta + \mu}, \quad j \geq 1,$$

3
$$\bar{\nabla}_{\text{even}}(j) = (\eta^2 + \mu(1 - \eta))^{j-1} \frac{\mu}{1 - \eta + \mu}, \quad j \geq 1.$$

5 We now consider different types of stationary sequences and compute the limiting distribution of the maximum term of models M1, M2 and M3, whenever these sequences are transformed through any of the sequences $\{U_n\}$ described above.

6.1. First-order max-autoregressive model (multiplicative)

7 We first consider a multiplicative max-autoregressive model

$$X_n = k \max\{X_{n-1}, \varepsilon_n\}, \quad n \geq 1,$$

9 where $\{\varepsilon_n\}$ is a sequence of iid rvs with df F , $k \in]0, 1[$ and X_0 is independent of $\{\varepsilon_n\}$. Alpuim (1988) has proved that $\{X_n\}$ is strong mixing and has stationary df H if and only if $w_F > 0$ and, for some $x_0 > 0$, $F(x_0/k) > 0$ and $\sum_{j=1}^{+\infty} -\ln F(x_0/k^j) < +\infty$. In this case

11 $H(x) = \prod_{j=1}^{+\infty} F(x/k^j)$. Considering $\{\varepsilon_n\}$ with an MSS df

$$F(x) = \exp\{-x^{-\alpha} v(\log x)\}, \quad \alpha > 0, \quad x > 0$$

13 and $r = k^{-\alpha/m}$, where m is a positive integer, Temido and Canto e Castro (2003) proved that $\{X_n\}$ has extremal index $\theta = 1 - r^{-m}$ and $H(x) = F^\gamma(x)$, with $\gamma = r^{-m}/(1 - r^{-m})$, which is again MSS. Due to $H^{[r^n]}(r^{n/\alpha}x) \rightarrow H(x), n \rightarrow +\infty$, we get

15
$$P(M_{[r^n]} \leq r^{n/\alpha}x) \rightarrow H^{1-r^{-m}}(x), \quad n \rightarrow +\infty.$$

Moreover, since

17
$$P(X_1 > u_n, X_{j-1} \leq u_n < X_j) \leq P(X_1 > u_n, X_{j-1} \leq u_n, \varepsilon_j > u_n/k) \leq (1 - H(u_n))(1 - F(u_n/k))$$

condition $D_{k_n}^{(2)}(r^{n/\alpha}x)$ holds. Following Alpuim (1988) the cluster size distribution is geometric given by $\pi(j) = (1 - r^{-m})(r^{-m})^{j-1}, j \geq 1$.

19 Considering the two different types of sequences $\{U_n\}$ described in the beginning of the section we obtain the following results:

21 1. IID missing values: Applying Proposition 3.1 we obtain

$$\theta^* = \frac{1 - r^{-m}}{1 - (1 - \beta)r^{-m}} \quad \text{and} \quad \theta^{**} = (1 - r^{-m}) \frac{1 - r^{-m}(1 - 2\beta(1 - \beta))}{(1 - r^{-m}(1 - \beta))(1 - r^{-m}\beta)},$$

23 2. Missing values through a Markov chain: Since $\{X_n\}$ satisfies condition $D_{k_n}^{(2)}(u_n)$ we may apply Proposition 4.2 and obtain

25
$$\theta^* = \frac{(1 - r^{-m})(1 - r^{-m}(\eta - \mu))}{1 - r^{-m}(1 - \mu)},$$

$$\theta^{**} = \frac{(1 - r^{-m})(1 - r^{-m}(\eta - \mu))}{1 - \eta + \mu} \left(\frac{\mu}{1 - r^{-m}(1 - \mu)} + \frac{1 - \eta}{1 - r^{-m}\eta} \right).$$

In either case

27
$$\lim_{n \rightarrow \infty} P(M_{[r^n]}(V) \leq r^{n/\alpha}x) = e^{-\theta_V \tau_V},$$

where $\theta_V = \theta^*$, $\tau_V = \tau\beta$ for the sequences $\{Y_n\}$ and $\{Z_n\}$, and $\theta_V = \theta^{**}$, $\tau_V = \tau$ for the sequence $\{W_n\}$, $\tau = x^{-\alpha} v(\log x)$.

29 6.2. First-order max-autoregressive model (additive)

31 Now suppose that $\{X_n\}$ is an integer-valued stationary additive max-autoregressive sequence with marginal df H satisfying (1) (Anderson's class). More precisely,

$$X_n = \max\{X_{n-1}, \varepsilon_n\} - c, \quad n \geq 1,$$

33 where $c \in \mathbb{N}$, $\{\varepsilon_n\}$ is an iid sequence with df F and X_0 is independent of $\{\varepsilon_n\}$. Hall (1996) proved that $\{X_n\}$ is strong mixing and has stationary df $H(x) = \prod_{j=1}^{+\infty} F(x + jc)$. She also proved that condition $D^{(2)}(u_n)$ holds for $\{X_n\}$. Thus $D_{k_n}^{(2)}(u_n, k_n)$ also holds. Moreover,

1 with the arguments of Hall (1996), we prove that $\{X_n\}$ has extremal index $\theta = 1 - r^{-c}$. It is also known, due to Alpuim (1988), that
 2 the cluster size distribution is geometric, $\pi(j) = (1 - r^{-c})r^{-(j-1)c}$, $j \geq 1$.

3 For instance, if H is the df of the Negative Binomial $NB(m, p)$ distribution, we have

$$P(M_{[p^{-n}]} \leq x + b_n) \rightarrow \exp(-(1 - p^c)p^{\lfloor x \rfloor}), \quad n \rightarrow +\infty, \quad x \in \mathbb{R} \setminus \mathbb{Z},$$

5 with $b_n = n - 1 - 1/\log p \{(m - 1) \log n + \log((1 - p)^{m-1}/(m - 1)!)\}$ and $\pi(j) = (1 - p^c)p^{(j-1)c}$, $j \geq 1$.

Again, considering the two different types of sequences $\{U_n\}$ described in the beginning of the section we obtain the following
 6 results:

7 1. *IID missing values*: Applying Proposition 3.1 we obtain

$$9 \quad \theta^* = \frac{(1 - r^{-c})}{1 - (1 - \beta)r^{-c}} \quad \text{and} \quad \theta^{**} = (1 - r^{-c}) \frac{1 - r^{-c}(1 - 2\beta(1 - \beta))}{(1 - r^{-c}(1 - \beta))(1 - r^{-c}\beta)}.$$

2. *Missing values through a Markov chain*: Since $\{X_n\}$ satisfies condition $D_{k_n}^{(2)}(u_n)$ we may apply Proposition 4.2 and obtain

$$11 \quad \theta^* = \frac{(1 - r^{-c})(1 - r^{-c}(\eta - \mu))}{(1 - r^{-c}(1 - \mu))}$$

$$12 \quad \theta^{**} = \frac{(1 - r^{-c})(1 - r^{-c}(\eta - \mu))}{1 - \eta + \mu} \left(\frac{\mu}{1 - r^{-c}(1 - \mu)} + \frac{1 - \eta}{1 - r^{-c}\eta} \right).$$

13 In either case

$$\lim_{n \rightarrow \infty} P(M_{[r^n]}(V) \leq x + b_n) = e^{-\theta_V \tau_V}, \quad x \in \mathbb{R} \setminus \mathbb{Z},$$

15 where $\theta_V = \theta^*$, $\tau_V = \tau\beta$ for the sequences $\{Y_n\}$ and $\{Z_n\}$, and $\theta_V = \theta^{**}$, $\tau_V = \tau$ for the sequence $\{W_n\}$, $\tau = r^{-\lfloor x \rfloor}$ and $\{b_n\}$ is an appropriate
 16 sequence of constants satisfying $\lim_{n \rightarrow \infty} b_n = +\infty$.

17 6.3. *Second-order max-autoregressive model (additive)*

Now suppose that $\{X_n\}$ is an integer-valued stationary additive second-order max-autoregressive sequence with marginal df
 18 H satisfying (1) (Anderson's class). More precisely,

$$X_n = \max\{X_{n-2}, \varepsilon_n\} - c, \quad n \geq 2,$$

21 where $c \in \mathbb{N}$ and $\{\varepsilon_n\}$ is an iid sequence with df F satisfying $(1 - F(n - 1))/(1 - F(n)) = r$, with r in $]1, +\infty[$. Assuming that X_0, X_1 and
 22 $\{\varepsilon_n\}$ are independent Hall (1998) proved that $\{X_n\}$ is strong mixing and has a stationary df given by $H(x) = \prod_{j=1}^{+\infty} F(x/k^j)$. Moreover,

23 Hall and Temido (2007) proved that condition $D_{k_n}^{(3)}(u_n)$ holds and $\{X_n\}$ has extremal index $\theta = 1 - r^{-c}$. Consequently

$$P(M_{[r^n]} \leq x + b_n) \rightarrow \exp\{-(1 - r^{-c})r^{-\lfloor x \rfloor}\}, \quad n \rightarrow +\infty, \quad x \in \mathbb{R} \setminus \mathbb{Z}.$$

25 Furthermore, since X_0 and X_1 are independent, condition (7) holds for $\{X_n\}$. Then, applying Lemma 5.2, by induction, we prove
 26 that the cluster size distribution is geometric given by $\pi(j) = (1 - r^{-c})(r^{-c})^{j-1}$, $j \geq 1$.

27 Considering the two different types of sequences $\{U_n\}$ described in the beginning of the section we obtain the following results:

29 1. *IID missing values*: Applying Proposition 3.1 we obtain

$$\theta^* = \frac{1 - r^{-c}}{1 - (1 - \beta)r^{-c}} \quad \text{and} \quad \theta^{**} = (1 - r^{-c}) \frac{1 - r^{-c}(1 - 2\beta(1 - \beta))}{(1 - r^{-c}(1 - \beta))(1 - r^{-c}\beta)}.$$

31 2. *Missing values through a Markov chain*: Since $\{X_n\}$ satisfies condition $D_{k_n}^{(3)}(u_n)$ and (7) we may apply Proposition 5.1 and obtain

$$32 \quad \theta^* = \frac{(1 - r^{-c})}{\beta} \left(1 - \frac{(1 - r^{-c})(1 - \beta)}{1 - r^{-c}(1 - \mu(1 - \mu + \eta))} \right),$$

$$33 \quad \theta^{**} = (1 - r^{-c}) \left(1 - \frac{(1 - r^{-c})(1 - \beta)}{1 - r^{-c}(1 - \mu(1 - \mu + \eta))} + 1 - \frac{(1 - r^{-c})\beta}{1 - r^{-c}(\eta^2 + \mu(1 - \eta))} \right),$$

where $\beta = \mu/1 - \eta + \mu$.

1 In either case

$$\lim_{n \rightarrow \infty} P(M_{[rn]}(V) \leq x + b_n) = e^{-\theta_V \tau_V}, \quad x \in \mathbb{R} \setminus \mathbb{Z},$$

3 where $\theta_V = \theta^*$, $\tau_V = \tau$ for the sequences $\{Y_n\}$ and $\{Z_n\}$, and $\theta_V = \theta^{**}$, $\tau_V = \tau$ for the sequence $\{W_n\}$, $\tau = r^{-[x]}$ and $\{b_n\}$ is an appropriate real sequence satisfying $\lim_{n \rightarrow \infty} b_n = +\infty$.

5 The last two examples consist of sequences with marginal distribution in the domain of attraction of a MS distribution. Hence, it suffices to consider $k_n = n$ in these cases. Although max-semistability is not required to study the effect of missing values on their extremes, we include them in the present work because the condition $D_{k_n}^{(2)}(u_n)$ does not hold and the results of Hall and Hüsler (2006) are not enough to obtain the limiting distribution of the maximum. Due to their alternating nature, conditions $D_{k_n}^{(3)}(u_n)$ and (7) hold, and hence the present results allow us to obtain the desired MS limiting distributions.

6.4. Negative AR(1) model with uniform margins

11 This example concerns the negatively correlated uniform AR(1) defined by

$$X_n = -\frac{1}{\varrho} X_{n-1} + \varepsilon_n, \quad n \geq 1,$$

13 where $\varrho > 1$, $X_0 \sim \mathcal{U}(0, 1)$, $\{\varepsilon_n\}$ is a sequence of iid rvs with $P(\varepsilon_1 = j/\varrho) = 1/\varrho$, for $j \in \{1, 2, \dots, \varrho\}$ and X_0 is independent of $\{\varepsilon_n\}$. Asymptotic results for the extremes from this model were originally obtained in Chernick and Davis (1982). Chernick et al. (1991) proved that $\{X_n\}$ satisfies condition $D^{(3)}(1 - x/n)$ and that the extremal index is given by $\theta = 1 - 1/\varrho^2$.

15 Take $u_n = 1 - x/n$ and observe that $\tau := \tau(x) = x$. In view of

$$\begin{aligned} nP(X_1 > u_n, X_2 > u_n) &= nP\left(X_1 > u_n, \varepsilon_2 > u_n \left(1 + \frac{1}{\varrho}\right)\right) \\ &= nP(X_1 > u_n)P\left(\varepsilon_2 > u_n \left(1 + \frac{1}{\varrho}\right)\right) \rightarrow 0, \quad n \rightarrow +\infty, \end{aligned}$$

17

and hence condition (7) holds. Using this result we can compute the extremal index in a simpler way. Indeed due to (9) and attending that

19

$$\lim_{n \rightarrow \infty} P\left(-\frac{1}{\varrho} \varepsilon_j + \varepsilon_{j+1} > u_n(1 - \varrho^2)\right) = P(\varepsilon_{j+1} = 1, \varepsilon_j = 1/\varrho) = 1/\varrho^2, \quad j \geq 1,$$

21 we deduce

$$\begin{aligned} \theta &= \lim_{n \rightarrow \infty} \frac{P\left(X_1 > u_n, \frac{1}{\varrho^2} X_1 - \frac{1}{\varrho} \varepsilon_2 + \varepsilon_3 \leq u_n\right)}{P(X_1 > u_n)} \\ &= \lim_{n \rightarrow \infty} P\left(-\frac{1}{\varrho} \varepsilon_2 + \varepsilon_3 \leq u_n(1 - 1/\varrho^2)\right) \\ &= 1 - 1/\varrho^2. \end{aligned}$$

23

Using the previous results we obtain

$$P(M_n \leq 1 - x/n) \rightarrow \exp(-(1 - \varrho^{-2})x), \quad n \rightarrow +\infty, \quad x > 0.$$

25

Now, in order to compute the limit cluster size $\pi(j)$ we first prove, by induction, that

$$P(B_{1,2j+1}^{(n)}) = \left(\frac{1}{\varrho^2}\right)^j \frac{x}{n} (1 + o_n(1)), \quad j \geq 1.$$

27 In fact

$$\begin{aligned} P(X_1 > u_n, X_3 > u_n) &= P(X_1 > u_n) - P(X_1 > u_n, X_3 \leq u_n) \\ &= \frac{x}{n} (1 - \theta) (1 + o_n(1)) = \frac{x}{n} \frac{1}{\varrho^2} (1 + o_n(1)) \end{aligned}$$

29 and

$$\begin{aligned} P(B_{1,2j-1}^{(n)}, X_{2j+1} > u_n) &= P\left(B_{1,2j-1}^{(n)}, \frac{1}{\varrho^2} X_{2j-1} - \frac{1}{\varrho} \varepsilon_{2j} + \varepsilon_{2j+1} > u_n\right) \\ &= P(B_{1,2j-1}^{(n)}) P\left(-\frac{1}{\varrho} \varepsilon_{2j} + \varepsilon_{2j+1} > u_n \left(1 - \frac{1}{\varrho^2}\right)\right) = \left(\frac{1}{\varrho^2}\right)^{j-1} \frac{x}{n} \frac{1}{\varrho^2} (1 + o_n(1)). \end{aligned}$$

1 Thus, by stationarity we obtain

$$\begin{aligned} P(X_1 \leq u_n, B_{3,2j+1}^{(n)}, X_{2j+3} > u_n) &= P(B_{3,2j+1}^{(n)}) - 2P(B_{1,2j+1}^{(n)}) + P(B_{1,2j+3}^{(n)}) \\ &= \left(\left(\frac{1}{\varrho^2}\right)^{j-1} \frac{x}{n} - 2\left(\frac{1}{\varrho^2}\right)^j \frac{x}{n} + \left(\frac{1}{\varrho^2}\right)^{j+2} \frac{x}{n} \right) (1 + o_n(1)) \\ &= \frac{x}{n} \left(\frac{1}{\varrho^2}\right)^{j-1} \left(1 - \frac{1}{\varrho^2}\right)^2 (1 + o_n(1)) \end{aligned}$$

3 and so

$$\pi(j) = \frac{\lim_{n \rightarrow \infty} \frac{x}{n} \left(\frac{1}{\varrho^2}\right)^{j-1} \left(1 - \frac{1}{\varrho^2}\right)^2 (1 + o_n(1))}{x \left(1 - \frac{1}{\varrho^2}\right)} = \left(\frac{1}{\varrho^2}\right)^{j-1} \left(1 - \frac{1}{\varrho^2}\right) \quad \text{for } j \geq 1,$$

5 which corresponds to a geometric distribution.

We now obtain the limiting distribution of the maximum term of models M1, M2 and M3, whenever the negative uniform AR(1) model is transformed through either the sequences $\{U_n\}$ described in the beginning of the section.

7 1. *IID missing values:* Applying Proposition 3.1 we obtain

$$\theta^* = \frac{(1 - \varrho^{-2})}{1 - (1 - \beta)\varrho^{-2}} \quad \text{and} \quad \theta^{**} = (1 - \varrho^{-2}) \frac{1 - \varrho^{-2}(1 - 2\beta(1 - \beta))}{(1 - \varrho^{-2}(1 - \beta))(1 - \varrho^{-2}\beta)}.$$

2. *Missing values through a Markov chain:* Since $\{X_n\}$ satisfies condition $D_{kn}^{(3)}(u_n)$ and (7) we may apply Proposition 5.1 and obtain

$$\begin{aligned} \theta^* &= \frac{(1 - \varrho^{-2})}{\beta} \left(1 - \frac{(1 - \varrho^{-2})(1 - \beta)}{1 - \varrho^{-2}(1 - \mu(1 - \mu + \eta))} \right), \\ \theta^{**} &= (1 - \varrho^{-2}) \left(1 - \frac{(1 - \varrho^{-2})(1 - \beta)}{1 - \varrho^{-2}(1 - \mu(1 - \mu + \eta))} + 1 - \frac{(1 - \varrho^{-2})\beta}{1 - \varrho^{-2}(\eta^2 + \mu(1 - \eta))} \right), \end{aligned}$$

13 where $\beta = \mu/(1 - \eta + \mu)$.

In either case

$$\lim_{n \rightarrow \infty} P(M_n(V) \leq 1 - x/n) = e^{-\theta_V \tau_V},$$

where $\theta_V = \theta^*$, $\tau_V = \tau\beta$ for the sequences $\{Y_n\}$ and $\{Z_n\}$, and $\theta_V = \theta^{**}$, $\tau_V = \tau$ for the sequence $\{W_n\}$, with $\tau = x > 0$.

17 6.5. *Negative AR(1) model with regularly varying tails*

Our last example concerns again an AR(1) stationary process

$$19 \quad X_n = -\psi X_{n-1} + Z_n, \quad n \geq 1,$$

21 where $\psi \in]0, 1[$ and $\{Z_n\}$ is a sequence of iid rvs independent of X_0 . Following Scotto et al. (2003) we consider that the margins of $\{Z_n\}$ possesses regularly varying balanced tails

$$\lim_{t \rightarrow \infty} \frac{P(|Z_1| > tx)}{P(|Z_1| > t)} = x^{-\alpha},$$

23 for $\alpha > 0$ and $x > 0$, and satisfies the tail balancing conditions

$$\lim_{x \rightarrow \infty} \frac{P(Z_1 > x)}{P(|Z_1| > x)} = p \in [0, 1]$$

25 and

$$\lim_{x \rightarrow \infty} \frac{P(Z_1 < -x)}{P(|Z_1| > x)} = 1 - p.$$

27 We now prove that this sequence satisfies (7) and $D^{(3)}(u_n(\tau, n))$. In fact

$$\begin{aligned} nP(X_1 > u_n, X_2 > u_n) &= nP(X_1 > u_n, Z_2 > u_n + \psi X_1) \\ &= nP(Z_2 > u_n(1 + \psi))P(X_1 > u_n) \rightarrow 0, \quad n \rightarrow +\infty, \end{aligned}$$

1 because $\psi > 0$. On the other hand

$$\begin{aligned} &P(X_1 > u_n, X_{j-2} \leq u_n, X_{j-1} \leq u_n, X_j > u_n) \\ &= P(X_1 > u_n, X_{j-2} \leq u_n, X_{j-1} \leq u_n, \psi^2 X_{j-2} - \psi Z_{j-1} + Z_j > u_n) \\ &= P(X_1 > u_n, X_{j-2} \leq u_n, X_{j-1} \leq u_n, -\psi Z_{j-1} + Z_j > u_n(1 - \psi^2)) \\ &\leq P(X_1 > u_n, X_{j-2} \leq u_n, X_{j-1} \leq u_n, |-\psi Z_{j-1} + Z_j| > u_n(1 - \psi^2)) \\ &\leq P(X_1 > u_n, X_{j-2} \leq u_n, X_{j-1} \leq u_n, \psi |Z_{j-1}| + |Z_j| > u_n(1 - \psi^2)) \\ &\leq P(X_1 > u_n, X_{j-2} \leq u_n, X_{j-1} \leq u_n, 2 \max\{\psi |Z_{j-1}|, |Z_j|\} > u_n(1 - \psi^2)) \\ &\leq P\left(X_1 > u_n, X_{j-2} \leq u_n, X_{j-1} \leq u_n, \psi |Z_{j-1}| > \frac{u_n(1 - \psi^2)}{2}\right) \\ &\quad + P\left(X_1 > u_n, X_{j-2} \leq u_n, X_{j-1} \leq u_n, |Z_j| > \frac{u_n(1 - \psi^2)}{2}\right). \end{aligned}$$

3 Thus

$$\begin{aligned} &n \sum_{j=2}^{r_n} P(X_1 > u_n, X_{j-2} \leq u_n, X_{j-1} \leq u_n, X_j > u_n) \\ &\leq nr_n P(X_1 > u_n) P\left(|Z_2| > u_n \frac{1 - \psi^2}{2\psi}\right) + nr_n P(X_1 > u_n) P\left(|Z_3| > u_n \frac{(1 - \psi^2)}{2}\right) \\ &= r_n P(X_1 > u_n) n P(|Z_2| > u_n) \frac{P\left(|Z_2| > u_n \frac{1 - \psi^2}{2\psi}\right)}{P(|Z_2| > u_n)} + r_n P(X_1 > u_n) n P(|Z_3| > u_n) \frac{P\left(|Z_3| > u_n \frac{1 - \psi^2}{2}\right)}{P(|Z_3| > u_n)} \\ &\rightarrow 0 \times \tau \times \left(\frac{1 - \psi^2}{2\psi}\right)^{-\alpha} + 0 \times \tau \times \left(\frac{1 - \psi^2}{2}\right)^{-\alpha} = 0, \quad n \rightarrow +\infty. \end{aligned}$$

5 The extremal index of $\{X_n\}$ is given by $\theta = 1 - \psi^{2\alpha}$ and $\tau := \tau(x) = (p + q\psi^\alpha)/(1 - \psi^{2\alpha})x^{-\alpha}$ (Scotto et al., 2003) and

$$P(M_n \leq a_n x) \rightarrow \exp\{-(p + q\psi^\alpha)x^{-\alpha}\}, \quad n \rightarrow +\infty, \quad x > 0,$$

7 with $\{a_n\}$ satisfying $nP(|Z_1| > a_n) \rightarrow 1, n \rightarrow +\infty$.

Furthermore, with the arguments used in the last example we deduce that

$$9 \quad \pi(j) = \psi^{2\alpha(j-1)}(1 - \psi^{2\alpha}), \quad \text{for } j \geq 1.$$

We observe that this coincides with the result of Scotto et al. (2003) for $\pi_1(j)$, taking into account the differences of the definitions of $\pi(j)$ used in this work (given by Leadbetter and Nandagopalan, 1989) and of $\pi_1(j)$ used by those authors (which is the same as in Embrechts et al., 1997, p. 273).

13 We now obtain the limiting distribution of the maximum term of models M1, M2 and M3, whenever the negative heavy tailed AR(1) model is transformed through either the sequences $\{U_n\}$ described in the beginning of the section.

15 1. IID missing values: Applying Proposition 3.1 we obtain

$$\theta^* = \frac{1 - \psi^{2\alpha}}{1 - (1 - \beta)\psi^{2\alpha}} \quad \text{and} \quad \theta^{**} = (1 - \psi^{2\alpha}) \frac{1 - \psi^{2\alpha}(1 - 2\beta(1 - \beta))}{(1 - \psi^{2\alpha}(1 - \beta))(1 - \psi^{2\alpha}\beta)}.$$

17 2. Missing values through a Markov chain: Since $\{X_n\}$ satisfies condition $D_{k_n}^{(3)}(u_n)$ and (7) we may apply Proposition 5.1 and obtain

$$\begin{aligned} \theta^* &= \frac{(1 - \psi^{2\alpha})}{\beta} \left(1 - \frac{(1 - \psi^{2\alpha})(1 - \beta)}{1 - \psi^{2\alpha}(1 - \mu(1 - \mu + \eta))}\right), \\ \theta^{**} &= (1 - \psi^{2\alpha}) \left(1 - \frac{(1 - \psi^{2\alpha})(1 - \beta)}{1 - \psi^{2\alpha}(1 - \mu(1 - \mu + \eta))} + 1 - \frac{(1 - \psi^{2\alpha})\beta}{1 - \psi^{2\alpha}(\eta^2 + \mu(1 - \eta))}\right), \end{aligned}$$

where $\beta = \mu/1 - \eta + \mu$.

In either case

$$\lim_{n \rightarrow \infty} P(M_n(V) \leq a_n x) = e^{-\theta_V \tau_V}, \quad x > 0,$$

23 where $\theta_V = \theta^*$, $\tau_V = \tau\beta$ for the sequences $\{Y_n\}$ and $\{Z_n\}$, and $\theta_V = \theta^{**}$, $\tau_V = \tau$ for the sequence $\{W_n\}$, with τ as above.

1 **7. Uncited references**

Canto e Castro et al. (2002).

Q2

3 **References**

- Alpuim, M.T., 1988. Contribuições à teoria de extremos em sucessões dependentes. Ph.D. Thesis, University of Lisbon.
- Anderson, C.W., 1970. Extreme value theory for a class of discrete distribution with applications to some stochastic processes. *J. Appl. Probab.* 7, 99–113.
- Canto e Castro, L., de Haan, L., Temido, M.G., 2002. Rarely observed sample maxima. *Theory Probab. Appl.* 45 (4), 779–782.
- Chernick, M.R., Davis, R., 1982. Extremes in autoregressive processes with uniform marginal distributions. *Statist. Probab. Lett.* 1, 85–88.
- Chernick, M.R., Hsing, T., McCormick, W., 1991. Calculating the extremal index for a class of stationary sequences. *Adv. Appl. Probab.* 6, 711–731.
- Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. *Modelling Extremal Events*. Springer, Berlin.
- Hall, A., 1996. Maximum term of a particular sequence with discrete margins. *Comm. Statist. Theory Methods* 28, 721–736.
- Hall, A., 2003. Extremes of integer-valued moving averages models with exponential type-tails. *Extremes* 6, 361–379.
- Hall, A., Hüslér, J., 2006. Extremes of stationary sequences with failures. *Stochastic Models* 22, 537–557.
- Hall, A., Scotto, M., 2006. Extremes of periodic integer-valued sequences with exponential type tails. *Revstat Statist. J.* 4, 249–273.
- Hall, A., Scotto, M., 2008. On the extremes of randomly sub-sampled time series. *Revstat Statist. J.*, to appear.
- Hall, A., Temido, M.G., 2007. On the maximum term of MA and Max-AR models with margins in Anderson's class. *Theory Probab. Appl.* 51, 291–304.
- Leadbetter, M.R., Nandagopalan, S., 1989. On exceedance point processes for stationary sequences under mild oscillation restriction. In: Hüslér, J., Reiss, R.D. (Eds.), *Extreme Values*. Springer, New York.
- Leadbetter, M.R., Lindgren, G., Rootzén, H., 1983. *Extremes and Related Properties of Random Sequences and Processes*. Springer, Berlin.
- McCormick, W., Park, Y., 1992. Asymptotic analysis of extremes from autoregressive negative binomial processes. *J. Appl. Probab.* 29, 904–920.
- McKenzie, E., 1986. Auto regressive-moving-average processes with negative binomial and geometric marginal distribution. *Adv. Appl. Probab.* 18, 679–705.
- McKenzie, E., 2003. Discrete variate time series. In: Shanbhag, D.N., Rao, C.R. (Eds.), *Handbook of Statistics*, vol. 21. Elsevier Science Pub. Co., pp. 573–606.
- Pancheva, E., 1992. Multivariate max-semistable distributions. *Theory Probab. Appl.* 18, 679–705.
- Temido, M.G., 2000. *Classes de leis limite em teoria de valores extremos-estabilidade e semiestabilidade*. Ph.D. Thesis, University of Coimbra.
- Temido, M.G., 2002. Domínios de atracção de funções de distribuição discretas, *Novos Rumos em Estatística*. In: *Proceedings of "IX Congresso da Sociedade Portuguesa de Estatística"*, Edition of the Portuguese Society of Statistics.
- Temido, M.G., Canto e Castro, L., 2003. Max-semistable laws in extremes of stationary random sequences. *Theory Probab. Appl.* 47, 365–374.

Q3