Extremal behaviour of a periodically controlled sequence with imputed values

Helena Ferreira · Ana Paula Martins · Maria da Graça Temido

Received: 27 February 2020 / Revised: 16 November 2020 / Accepted: 29 November 2020 / Published online: 6 February 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract
Extreme events are a major concern in statistical modeling. Random missing data can constitute a problem when modeling such rare events. Imputation is crucial in these situations and therefore models that describe different imputation functions enhance possible applications and enlarge the few known families of models that cover these situations. In this paper we consider a family of models \( \{Y_n\}, n \geq 1 \), that can be associated to automatic systems which have a periodic control, in the sense that at instants multiple of \( T, T \geq 2 \), no value is lost. Random missing values are here replaced by the biggest of the previous observations up to the one surely registered. We prove that when the underlying sequence is stationary, \( \{Y_n\} \) is \( T \)-periodic and, if it also verifies some local dependence conditions, then \( \{Y_n\} \) verifies one of the well known \( D_T^{(s)}(u_n) \), \( s \geq 1 \), dependence conditions for \( T \)-periodic sequences. We also obtain the extremal index of \( \{Y_n\} \) and relate it to the extremal index of the underlying sequence. A consistent estimator for the parameter that “controls” the missing values is here proposed and its finite sample properties are analysed. The obtained results are illustrated with Markovian sequences of recognized interest in applications.

Keywords Missing values · Periodic sequence · Local dependence conditions · Extremal index

Ana Paula Martins
amartins@ubi.pt
Helena Ferreira
helenaf@ubi.pt
Maria da Graça Temido
mgmt@mat.uc.pt

1 Centro de Matemática e Aplicações (CMA-UBI), Departamento de Matemática, Universidade da Beira Interior, Covilhã, Portugal
2 Departamento de Matemática, Universidade de Coimbra, CMUC, Coimbra, Portugal
1 Introduction and preliminary results

Data collection is prevalent in everyday life and is used in several domains, such as finance, climate observation, computer science. The main goal of any data collection effort is to compile quality data, but issues with missing data often arise in the process. The data unavailability may be caused by the failure of some system. The failure of a reading device, or simply the lack of retention due to the intrinsic properties of the data, e.g. financial or environmental data only reported at certain time instants (Hall and Hüsler 2006; Hall and Scotto 2008 and references therein).

Many analysis methods require the use of imputation, i.e. missing values to be replaced with reasonable values up-front, making it a current issue (e.g. Zha and Harel 2019; Crambes and Henchiri 2019). An overview on univariate time series imputation can be found in Moritz et al. (2015), whereas the R package `imputeTS` offers several imputation algorithms (see Moritz and Bartz-Beielstein 2017).

Falk et al. (2011) summarize the several strategies that are usually applied when missing data values occur in time series: (i) the missing value is replaced by a predefined value $x_0$ which can be sometimes 999 (if one is interested in small values and no such large values occur) or $-1$ (if one is interested in large values and no negative values occur); (ii) the data is completely lost and the time series is sub-sampled with a smaller (and random) sample size; (iii) an automatic measurement device is used to replace the missing data by a proxy value.

The extremal properties of sequences with random missing values replaced by 0 were studied by Falk et al. (2011). The sub-sample referred in strategy (ii) above may result from missing values that occur according to some deterministic pattern or occur randomly. The effect of deterministic missing values on the properties of strictly stationary (stationary) sequences has been studied by Ferreira and Martins (2003), Martins and Ferreira (2004), Scotto et al. (2003), among others. Random missing values have been considered by Weissman and Cohen (1995) for the case of constant failure probability and independent failures, and their results were generalized for situations where the failure pattern has a weak dependence structure by Hall and Hüsler (2006). This was pursued by Hall and Scotto (2008), when the underlying process is represented as a moving average driven by heavy-tailed innovations and the sub-sampling process is strongly mixing.

When the missing values are replaced by an automatic measurement device the resulting sample will be a mixture of two original samples. This case was considered by Hall and Hüsler (2006) and later by Hall and Temido (2009) in the context of max-semistability. The extremal properties of a model where missing values are replaced by independent replicas of the original values were discussed. Investigating the extremal properties of models that describe other imputation functions enhances possible applications in situations where it may be of interest to avoid the occurrence of missing values and where periodic controls guarantee some correctly observed values. Energy consumptions in households are examples of such situations, since meter readings are usually done by the company supplier on a periodic basis, once every two or three months, or communicated to the company by the householder. Daily consumptions will have missing data, corresponding to the absence of meter readings, but monthly consumptions can have more than one value when the meter readings...
are made by the supplier and by the householder, most probably on different days of
the same month. To minimize the risk, the energy supplier will benefit from taking
the maximum meter reading since the last company reading to calculate the monthly
value to charge the householder for the service supplied. A model that translates this
idea, where a missing value is replaced by the biggest of the previous observations up
to to the one at an instant multiple of \( T \), which is surely registered, is defined by

\[
Y_n = U_n X_n + (1 - U_n) \sqrt{U_i X_i}, \quad n \geq 1, \quad T \geq 2, \tag{1.1}
\]

where \([a]\) denotes the integer part of \( a \in \mathbb{R} \), \( \{U_n\}_{n \geq 0} \) is a sequence of independent
variables, such that, for all \( k \geq 0 \), \( U_{kT} = 1 \) almost surely, and \( U_n \) follows a Bernoulli
distribution with parameter \( p \in ]0, 1[ \), for all \( n \neq kT \). \( \{X_n\}_{n \geq 0} \) denotes a positive
stationary sequence, independent of \( U_n, n \geq 0 \), with marginal continuous distribution
function (d.f.) \( F \). Here and throughout \( \bigvee_{j=s}^t W_j \equiv \bigvee_{j=s}^t W_j \), for any \( s, t \in \mathbb{Z}, s < t \),
and random variables \( W_j \).

Model (1.1) can be associated to automatic systems which have a manual periodic
verification. As we can see, at each instant \( n \neq kT, k \geq 0 \), we can observe \( X_n \) or in the
case that it is not observed it is replaced by the maximum of the previous observations
up to the last one that was surely registered. The registration of the observations is
periodically controlled, with the guarantee that at instants \( kT, k \geq 0 \), no observations
are lost and therefore in the period \([\lfloor n-1 \rfloor T, \ldots, n-1]\) at least the observation with
index \([\lfloor n-1 \rfloor T, \ldots, n-1]\) is available. The case \( T = 1 \), if considered, would correspond to the
non-occurrence of missing data.

To better understand model (1.1) let us consider the following illustrative example.

\textbf{Example 1.1} Let \( \{Z_n\}_{n \geq -1} \) be a sequence of independent and identically distributed
(i.i.d.) random variables with unit Fréchet marginal d.f. \( F_Z \). With this sequence we
define the moving maxima \( X_n = \frac{1}{2} Z_n \lor \frac{1}{2} Z_{n-1}, n \geq 0 \), which is stationary and also
has unit Fréchet margins.

A model with the characteristics of (1.1) is given by

\[
Y_n = U_n X_n + (1 - U_n) U_{n-1} X_{n-1}, \quad n \geq 1, \tag{1.2}
\]

where \( \{U_n\}_{n \geq 0} \) is an independent Bernoulli sequence with parameter \( p \) for all \( n \neq 2k \)
and \( U_{2k} = 1, k \geq 0 \), almost surely.

Here \( \{Y_n\} \) is controlled at instants which are multiples of 2 \( (T = 2) \), so at these
instants we always have the guarantee that an observation of the moving maxima was
retained. At all the other instants we can have a moving maxima value, whenever it is
observed, or the maximum of the previous observations up to the last observation
“controlled”, which in this case corresponds only to the previous observation because
\( T = 2 \). In Fig. 1 this becomes clear with 100 observation of (1.2), since newly imputed
observations are marked differently than the rest of the series and the instants \( 2s, s \geq 1 \),
are highlighted (dashed vertical lines).
We shall return to this simple example to illustrate several of the results presented, namely in Section 2 where we deal with the estimation of the model parameter \( p \in [0, 1] \). We propose a consistent estimator for this parameter and analyse its finite sample behaviour.

In Section 3, we achieve our main goal which is the characterization of the extremal behaviour of \( \{Y_n\}_{n \geq 1} \) given in (1.1). We start by proving that \( \{Y_n\}_{n \geq 1} \) is a \( T \)-periodic sequence, i.e. there exists a positive integer \( T \) such that, for any choice of integers \( 1 \leq i_1 < i_2 < \ldots < i_p \), the vectors \( (Y_{i_1}, \ldots, Y_{i_p}) \) and \( (Y_{i_1+T}, \ldots, Y_{i_p+T}) \) are equally distributed. Extreme value theory known for periodic sequences can then be applied to this periodically controlled sequence with imputed values \( \{Y_n\}_{n \geq 1} \). Alpuim (1988) showed that under Leadbetter’s global mixing condition \( D(u_n) \), the only possible limit laws for the normalized maxima of a \( T \)-periodic sequence are the three extreme value distributions and generalized, as well, the definition of extremal index for such sequences. We recall that a \( T \)-periodic sequence \( \{Y_n\}_{n \geq 1} \) has extremal index \( \theta_Y \), \( 0 \leq \theta_Y \leq 1 \), defined by

\[
\tau = \lim_{n \to +\infty} n \frac{1}{T} \sum_{j=1}^{T} F_j(u_n) = \lim_{n \to +\infty} n \left( 1 - \frac{1}{T} \sum_{j=1}^{T} F_j(u_n) \right) , \tag{1.3}
\]

with \( F_j(x) = 1 - F_j(x) \) the tail functions of \( Y_j \) for \( j = 1, \ldots, T \).

Under local mixing conditions \( D_T^{(s)}(u_n) \), \( s = 1, 2 \), Ferreira (1994) studied the extremal behaviour of periodic sequences and under the weaker local mixing conditions \( D_T^{(s)}(u_n) \), \( s \geq 3 \), Ferreira and Martins (2003) obtained the expression for the extremal index of a \( T \)-periodic sequence from the joint distribution of \( s \) consecutive variables of the sequence.
To complete Section 3, we obtain necessary conditions, that rely on the underlying sequence \( \{X_n\}_{n \geq 0} \), for sequence \( \{Y_n\}_{n \geq 1} \) to satisfy Leadbetter’s \( D(u_n) \) condition, as well as some local dependence condition \( D_T^{(s)}(u_n), s \geq 1 \), for \( T \)-periodic sequences. The validation of these conditions will permit the determination of its extremal index expression. The results here obtained are illustrated with examples of recognized interest in applications, such as Markovian sequences.

### 2 Model parameter estimation

The proposed model (1.1) depends on an unknown parameter \( p \in ]0, 1[ \), that “controls” the number of missing values, and on an underlying stationary sequence with unknown marginal d.f. \( F \). The estimation of \( p \) and \( F \) is therefore essential for practical applications of this model.

The next result, that characterizes the probabilities \( P \left( \bigcap_{j=1}^{T-1} \{Y_{sT+j} = Y_{sT}\} \right), s \geq 1, T \geq 2 \), gives a simple procedure to estimate the parameter \( p \) involved in the definition of model (1.1).

**Theorem 2.1** For the sequence \( \{Y_n\}_{n \geq 1} \) defined in (1.1) it holds

\[
P \left( \bigcap_{j=1}^{T-1} \{Y_{sT+j} = Y_{sT}\} \right) = (1 - p)^{T-1}, \ s \geq 1, \ T \geq 2.
\]

**Proof** For all \( s \geq 1 \), \( Y_{sT} = X_{sT} \) almost surely and, if \( U_{sT+j} = 1 \) for some \( j \in \{1, \ldots, T-1\} \), we have \( P \left( \bigcap_{j=1}^{T-1} \{Y_{sT+j} = Y_{sT}\} \right) = 0 \), because the underlying d.f.’s are continuous. Therefore we can write

\[
P \left( \bigcap_{j=1}^{T-1} \{Y_{sT+j} = Y_{sT}\} \right) = P \left( \bigcap_{j=1}^{T-1} \left\{ U_{sT+j} X_{sT+j} + (1 - U_{sT+j}) \bigvee_{i=sT+j-1}^{sT+j-1} U_i X_i = X_{sT} \right\} \right)
\]

\[
= (1 - p)^{T-1} P \left( \bigcap_{j=1}^{T-1} \left\{ \bigvee_{i=sT+j-1}^{sT+j-1} U_i X_i = X_{sT} \right\} \right). \quad (2.4)
\]

Now since \( U_{sT+j} = 0 \), for all \( j = 1, \ldots, T-1 \), the several maxima in (2.4) are all equal to the variable \( X_{sT} \) and the result follows immediately. \( \square \)

The way to estimate parameter \( p \) of model (1.1) becomes clear from the previous result. So, if \( \{Y_1, \ldots, Y_n\} \) is a random sample of \( \{Y_n\}_{n \geq 1} \) with fixed \( T \geq 2 \), an estimator
for \( p \in ]0, 1[ \) is given by

\[
\hat{p}_n^{(T)} = 1 - \left( \frac{1}{\left\lfloor \frac{n+1}{T} \right\rfloor - 1} \sum_{s=1}^{\left\lfloor \frac{n+1}{T} \right\rfloor - 1} \mathbb{I}_{A_s} \right)^{\frac{1}{T-1}}, \tag{2.5}
\]

with \( A_s = \bigcap_{j=1}^{T-1} \{ Y_{sT+j} = Y_{sT} \} \).

Note that the estimator \( \hat{p}_n^{(T)} \) only depends on the sample size \( n \) since \( T \geq 2 \) is known from the model.

From the weak law of large numbers and the fact that \( E[\mathbb{I}_{\bigcap_{j=1}^{T-1} \{ Y_{sT+j} = Y_{sT} \}}] = P\left( \bigcap_{j=1}^{T-1} \{ Y_{sT+j} = Y_{sT} \} \right) \), we can state that estimator (2.5) is a consistent estimator for \( p \).

The d.f. \( F \) can be estimated from the observations of \( Y_{sT} \equiv X_{sT}, s \geq 1 \), with the empirical d.f. or a kernel estimator. A review of these estimators and an explanation on their functionality and applicability in \( \mathbb{R} \) can be found in Quintela-del-Río and Estévez-Pérez (2012).

### 2.1 Simulation results

We now analyse the finite sample properties of the estimator given in (2.5) with simulated data from model (1.2) given in Example 1.1 for which \( T = 2 \). Each simulated data set consists of 1000 independent realizations of a random sample \((Y_1, \ldots, Y_n)\) of (1.2) having one particular value of \( p \in ]0, 1[ \) out of five. Three different sample sizes are considered for each data set. The sample means \( \hat{\mu}(\hat{p}_n^{(2)}) \) and the sample standard deviations \( \hat{\sigma}(\hat{p}_n^{(2)}) \) of the estimates \( \hat{p}_n^{(2)} \), \( i = 1, \ldots, 1000 \), depending on the sample size \( n \), were computed. The bias and the root mean squared errors (RMSE(\( \hat{p}_n^{(2)} \))) were also determined. Table 1 summarizes the estimation results obtained. The estimator has a good behaviour even for small sample sizes.

### 3 Extremal behaviour

In order to characterize the extremal behaviour of \( \{ Y_n \}_{n \geq 1} \) in (1.1), let us start by noting that, for all \( k \geq 0 \) and for all \( x \in \mathbb{R} \), the marginal d.f.’s of \( \{ Y_n \}_{n \geq 1} \) satisfy

\[
F_{Y_{kT+j}}(x) = \begin{cases} 
F(x), & j \in \{0, 1\}, \ T \geq 2, \\
pF(x) + (1-p)G_j(x), & j \in \{2, \ldots, T-1\}, \ T \geq 3,
\end{cases}
\]

\( \mathcal{C} \) Springer
Table 1 Various statistical results for the estimation of $p \in [0, 1]$ in model (1.2)

| $p$  | $n$     | $\hat{\mu}(\hat{p}_n)$ | $BIAS(\hat{p}_n)$ | $\hat{\sigma}(\hat{p}_n)$ | $RMSE(\hat{p}_n)$ |
|------|---------|--------------------------|-------------------|---------------------------|------------------|
| 0.10 | $n = 250$ | 0.0989                   | -0.0011           | 0.0267                    | 0.0267           |
|      | $n = 1000$ | 0.1008                   | 0.0008            | 0.0137                    | 0.0137           |
|      | $n = 5000$ | 0.1001                   | 0.0001            | 0.0059                    | 0.0059           |
| 0.25 | $n = 250$ | 0.2492                   | -0.0008           | 0.0395                    | 0.0395           |
|      | $n = 1000$ | 0.2494                   | -0.0006           | 0.0198                    | 0.0198           |
|      | $n = 5000$ | 0.2495                   | -0.0005           | 0.0083                    | 0.0083           |
| 0.50 | $n = 250$ | 0.5025                   | 0.0025            | 0.0461                    | 0.0461           |
|      | $n = 1000$ | 0.4997                   | -0.0003           | 0.0222                    | 0.0223           |
|      | $n = 5000$ | 0.4999                   | 0.0001            | 0.0096                    | 0.0096           |
| 0.75 | $n = 250$ | 0.7510                   | 0.0010            | 0.0373                    | 0.0373           |
|      | $n = 1000$ | 0.7506                   | 0.0006            | 0.0188                    | 0.0188           |
|      | $n = 5000$ | 0.7498                   | -0.0002           | 0.0085                    | 0.0085           |
| 0.90 | $n = 250$ | 0.9000                   | 0.0000            | 0.0263                    | 0.0263           |
|      | $n = 1000$ | 0.9001                   | 0.0001            | 0.0129                    | 0.0129           |
|      | $n = 5000$ | 0.9001                   | 0.0001            | 0.0060                    | 0.0060           |

with $kT + j \neq 0$ and

$$G_j(x) = \sum_{\emptyset \subseteq S \subseteq [kT+1,\ldots,kT+j-1]} p^{|S|} (1-p)^{j-1-|S|} \times$$

$$\times F_{kT,kT+1,\ldots,kT+j-1}(x, x(\delta_{kT+1}(S))^{-1}, \ldots, x(\delta_{kT+j-1}(S))^{-1}),$$

where $\delta_{*}(B)$ denotes the Dirac measure on $B$, $F_{i_1,\ldots,i_p}$ denotes the d.f. of the vector $(X_{i_1}, \ldots, X_{i_p})$, $|S|$ stands for the cardinality of a set $S$ and we convention that $x \times \frac{1}{0} = +\infty$.

The sequence $\{Y_n\}_{n \geq 1}$ is a $T$-periodic sequence. Indeed, for any choice of integers $1 \leq i_1 < i_2 < \ldots < i_p$, we have

$$P \left( \bigcap_{s=1}^{p} \{ Y_{i_s+T} \leq x_s \} \right)$$

$$= P \left( \bigcap_{s=1}^{p} \left\{ \bigcup_{j=i_s+T-1}^{[t+T-1]} U_{i_s+T} X_{i_s+T} + (1 - U_{i_s+T}) U_j X_j \leq x_s \right\} \right)$$

$$= E \left( P \left( \bigcap_{s=1}^{p} \left\{ \bigcup_{j=i_s+T-1}^{[t+T-1]} U_{i_s+T} X_{i_s+T} + (1 - U_{i_s+T}) U_j X_j \leq x_s \right\} \right) \right)$$
since \( \{U_n\}_{n \geq 1} \) is a \( T \)-periodic sequence, \( \{X_n\}_{n \geq 0} \) is a stationary sequence and they are independent. From now on we use the notation \( F_j(x) := F_{Y_j}(x) \), for any \( j \geq 1 \), for which holds \( F_j(x) := F_{J+kT}(x) \), for all \( k \geq 1 \) and \( T \geq 2 \).

The study of the extremal behaviour of stationary or periodic sequences most often relies upon the verification of appropriate dependence conditions which assure that the limiting distribution of the maximum term is of the same type as the limiting distribution of the maximum of i.i.d. random variables. Usual conditions used in the literature for stationary and periodic sequences are Leadbetter’s \( D(u_n) \) condition (Leadbetter 1974), conditions \( D(s)(u_n), s \geq 1 \), of Chernick et al. (1991) and conditions \( D_T^{(s)}(u_n), s, T \geq 1 \), of Ferreira (1994) and Ferreira and Martins (2003).

For any sequence of real numbers \( \{u_n\}_{n \geq 1} \), the dependence condition \( D(u_n) \), for the sequence \( \{Y_n\}_{n \geq 1} \), of Leadbetter, states that \( \alpha_n, \ell_n \to 0 \), as \( n \to +\infty \), for some sequence \( \ell_n = o(n) \), where

\[
\alpha_{n, \ell_n} = \sup \{ |P(M_{Y_{i_1,i_1+p}}^{(Y)} \leq u_n, M_{Y_{j_1,j_1+q}}^{(Y)} \leq u_n) \bigcap P(M_{Y_{i_1,i_1+p}}^{(Y)} \leq u_n) \} : 1 \leq i_1 < i_1 + p + \ell_n \leq j_1 < j_1 + q \leq n \},
\]

and \( M_{s,s+t}^{(Y)} \) denotes \( \bigcup_{i=s}^{s+t} Y_i \).

In the next result we show that if condition \( D(u_n) \) holds for the stationary sequence \( \{X_n\}_{n \geq 0} \) then it also holds for the \( T \)-periodic sequence \( \{Y_n\}_{n \geq 1} \).

**Theorem 3.1** If for any positive sequence \( \{u_n\}_{n \geq 1} \), condition \( D(u_n) \) holds for the stationary sequence \( \{X_n\}_{n \geq 0} \) then it also holds for the sequence \( \{Y_n\}_{n \geq 1} \) defined in (1.1).

**Proof** Consider the sets of consecutive integers \( A_p = \{i_1, \ldots, i_1 + p\} \) and \( B_q = \{j_1, \ldots, j_1 + q\} \), with \( j_1 - (i_1 + p) \geq \ell_n \).
The definition of \( \{Y_n\}_{n \geq 1} \) induces over \( \{X_n\}_{n \geq 0} \) the corresponding sets of integers \( A_p' = \left\{ \left\lfloor \frac{i-1}{T} \right\rfloor T, \ldots, i_1 + p \right\} \) and \( B_q' = \left\{ \left\lfloor \frac{j-1}{T} \right\rfloor T, \ldots, j_1 + q \right\} \), such that \( \left\lfloor \frac{j-1}{T} \right\rfloor T - (i_1 + p) \geq \ell_n - T \). Hence, each realization \( u \) of \( U := \{U_i, i \in A_p' \cup B_q'\} \) gives rise to another pair of subsets of positive integers, say \( A_p(U) \) and \( B_q(U) \), and therefore

\[
P\left(M(X)(A_p) \leq u_n, \ M(X)(B_q) \leq u_n\right) = E\left(P\left(M(X)(A_p(U)) \leq u_n, \ M(X)(B_q(U)) \leq u_n \mid U\right)\right).
\]

Due to the fact that \( \{X_n\}_{n \geq 0} \) satisfies \( D(u_n) \) condition, the last average becomes

\[
E\left(P\left(M(X)(A_p(U)) \leq u_n \mid U\right)\right) P\left(M(X)(B_q(U)) \leq u_n \mid U\right) + O(\alpha_n, \ell_n - T)
\]

\[
= E\left(P\left(M(X)(A_p(U)) \leq u_n \mid U\right)\right) E\left(P\left(M(X)(B_q(U)) \leq u_n \mid U\right)\right) + O(\alpha_n, \ell_n - T)
\]

because \( \{U_n\}_{n \geq 0} \) is a sequence of independent variables.

Returning to the sequence \( \{Y_n\}_{n \geq 1} \) we deduce

\[
|P\left(M(Y)(A_p) \leq u_n, \ M(Y)(B_q) \leq u_n\right) - P\left(M(Y)(A_p) \leq u_n\right) P\left(M(Y)(B_q) \leq u_n\right) | \leq \beta_n, \ell_n,
\]

with \( \beta_n, \ell_n = \alpha_n, \ell_n - T \) and for \( \ell'_n = \ell_n + T = o(n) \) we have \( \beta_n, \ell'_n \xrightarrow{n \to +\infty} 0 \), as required.

If the \( T \)-periodic sequence \( \{Y_n\}_{n \geq 1} \) satisfies condition \( D(u_n) \), for all \( \{u_n\}_{n \geq 1} \), then we say that it also satisfies condition \( D_T^{(s)}(u_n), T, s \geq 1 \), when there exists a sequence of integers \( \{k_n\}_{n \geq 1} \) such that \( k_n \to +\infty, k_n \frac{\ell_n}{n} \to 0, k_n \ell_n, \ell_n \to 0 \), as \( n \to +\infty \), and

\[
\lim_{n \to +\infty} n \frac{1}{T} \sum_{i=1}^{T} P\left(Y_i > u_n \geq M_{i+1,i+s-1}^{(Y)}, \ M_{i+s+1}^{(Y)} \left\lfloor \frac{n}{k_n T} \right\rfloor T > u_n\right) = 0, \tag{3.7}
\]

where \( M_{i,j}^{(Y)} = -\infty \), for \( i < j \), and \( M_{i,i}^{(Y)} = Y_i \). These local dependence conditions were first defined in Ferreira (1994), for \( s = 1 \) and \( s = 2 \). This family was later enlarged by Ferreira and Martins (2003) with values of \( s \geq 3 \).

Observe that, when \( s \geq 2 \), condition (3.7) is implied by

\[
\lim_{n \to +\infty} n \frac{1}{T} \sum_{i=1}^{T} \sum_{j=i+s} P\left(Y_i > u_n, \ Y_{j-1} \leq u_n < Y_j\right) = 0.
\]
and it limits the distance between exceedances of level $u_n$, i.e., in each block there can only occur more than one exceedance of $u_n$ if separated by less than $s - 1$ non-exceedances of $u_n$. Consequently, the local dependence conditions $D_T^{(s)}(u_n)$, $s \geq 1$, become weaker as the value of $s$ increases and thereby enhance the number of processes to which our results apply. Condition (3.7), when $T = 1$, coincides with the one considered in $D_T^{(1)}(u_n)$ of Chernick et al. (1991) for stationary sequences.

Under some local dependence condition $D_T^{(s)}(u_n)$, $s \geq 1$, where $\{u_n\}_{n \geq 1}$ is a sequence of positive numbers satisfying (1.3), the extremal index of $\{Y_n\}_{n \geq 1}$, $\theta_Y$, can be computed from

$$
\theta_Y = \frac{1}{\tau} \lim_{n \to +\infty} n \left\{ \frac{1}{T} \sum_{i=1}^{T} P(Y_i > u_n \geq M_{i+1,i+s-1}^{(Y)}) \right\}.
$$

(3.8)

In order to apply the previous results we shall impose the following two conditions on the tail of the d.f.'s $F$ and $G_j$, $j = 2, \ldots, T - 1$, $T \geq 3$, where $G_j$ denotes the d.f. of $\bigvee_{i=j-1}^{[i/T]} U_i X_i$, defined by (3.6). Namely

$$
n \bar{F}(u_n) \xrightarrow{n \to +\infty} \tau X \quad \text{and} \quad n \bar{G}_j(u_n) \xrightarrow{n \to +\infty} \tau_j.
$$

(3.9)

The particular case $T = 2$ leads to $\tau = \tau X$. Under such conditions, for $T \geq 3$, we have

$$
n \frac{1}{T} \sum_{j=1}^{T} F_j(u_n) = n \frac{1}{T} \left\{ \sum_{j=2}^{T-1} \left( p \bar{F}(u_n) + (1 - p) P \left( \bigvee_{i=j-1}^{[i/T]} U_i X_i > u_n \right) \right) \right\} + 2 \bar{F}(u_n)
$$

$$
= \frac{1}{T} ((T - 2)p + 2) \bar{F}(u_n) + (1 - p) \frac{n}{T} \sum_{j=2}^{T-1} P \left( \bigvee_{i=j-1}^{[i/T]} U_i X_i > u_n \right)
$$

$$
= \frac{1}{T} ((T - 2)p + 2) \bar{F}(u_n) + (1 - p) \frac{1}{T} \sum_{j=2}^{T-1} n \bar{G}_j(u_n)
$$

$$
\xrightarrow{n \to +\infty} \frac{1}{T} \left\{ ((T - 2)p + 2) \tau X + (1 - p) \sum_{j=2}^{T-1} \tau_j \right\} = \tau.
$$

(3.10)

We derive now a relation between a slightly stronger condition than $D_T^{(T+1)}(u_n)$ of Chernick et al. (1991) for the underlying stationary sequence $\{X_n\}_{n \geq 0}$ and condition $D_T^{(T+1)}(u_n)$ for the $T$-periodic sequence $\{Y_n\}_{n \geq 1}$.
Theorem 3.2 If for any positive sequence \( \{u_n\}_{n \geq 1} \), the stationary sequence \( \{X_n\}_{n \geq 0} \) satisfies condition \( D(u_n) \) and
\[
 nP \left( M_{0,T}^{(X)} > u_n \geq X_{T+1}, \ M_{T+2, \left( \frac{n}{knT} \right) T + T}^{(X)} > u_n \right) \xrightarrow{n \to +\infty} 0 \tag{3.11}
\]
then condition \( D^{(T+1)}(u_n) \) holds for the sequence \( \{Y_n\}_{n \geq 1} \).

Proof Observe first that
\[
 n \frac{1}{T} \sum_{i=1}^{T} P \left( Y_i > u_n \geq M_{i+1,i+T}^{(Y)} \ M_{i+T+1, \left( \frac{n}{knT} \right) T}^{(Y)} > u_n \right)
\]
\[
 \leq n \frac{1}{T} \sum_{i=1}^{T} P \left( Y_i > u_n \geq Y_{i+1} \ M_{i+T+1, \left( \frac{n}{knT} \right) T}^{(Y)} > u_n \right)
\]
\[
 \leq n \frac{1}{T} \sum_{i=1}^{T} P \left( Y_i > u_n \geq Y_{i+1} \ M_{i+2, \left( \frac{n}{knT} \right) T}^{(Y)} > u_n \right).
\]
In what concerns the probability involved in this last sum, for \( i \in \{1, 2, \ldots, T-2\} \), we have
\[
P \left( Y_i > u_n \geq Y_{i+1}, \ M_{i+2, \left( \frac{n}{knT} \right) T}^{(X)} > u_n \left| U_i = 1, U_{i+1} = 1 \right. \right)
\]
\[
= P \left( X_i > u_n \geq X_{i+1}, \ M_{i+2, \left( \frac{n}{knT} \right) T}^{(X)} > u_n \right)
\]
\[
= P \left( X_T > u_n \geq X_{T+1}, \ M_{T+2, \left( \frac{n}{knT} \right) T + T - i}^{(X)} > u_n \right)
\]
\[
\leq P \left( \bigvee_{j=0}^{T} X_j > u_n \geq X_{T+1}, \ M_{T+2, \left( \frac{n}{knT} \right) T + T}^{(X)} > u_n \right),
\]
as well as
\[
P \left( Y_i > u_n \geq Y_{i+1}, \ M_{i+2, \left( \frac{n}{knT} \right) T}^{(X)} > u_n \left| U_i = 0, U_{i+1} = 1 \right. \right)
\]
\[
= P \left( \bigvee_{j=0}^{i-1} U_j X_j > u_n \geq X_{i+1}, \ M_{i+2, \left( \frac{n}{knT} \right) T}^{(X)} > u_n \right)
\]
\[
\leq P \left( \bigvee_{j=0}^{T} X_j > u_n \geq X_{T+1}, \ M_{T+2, \left( \frac{n}{knT} \right) T + T}^{(X)} > u_n \right). \tag{3.12}
\]
where we used the stationarity of \( \{X_n\} \). The remaining probabilities are equal to zero, since

\[
P \left( Y_i > u_n \geq Y_{i+1}, M^{(X)}_{i+2, \left[ \frac{n}{knT} \right] T} > u_n \mid U_i = 1, U_{i+1} = 0 \right) \]

\[
= P \left( X_i > u_n \geq \bigcup_{j=0}^{i} U_j X_j, M^{(X)}_{i+2, \left[ \frac{n}{knT} \right] T} > u_n \mid U_i = 1 \right) \]

\[
\leq P \left( X_i > u_n \geq X_i \right) = 0
\]

and

\[
P \left( Y_i > u_n \geq Y_{i+1}, M^{(X)}_{i+2, \left[ \frac{n}{knT} \right] T} > u_n \mid U_i = 0, U_{i+1} = 0 \right) \]

\[
= P \left( \bigcup_{j=0}^{i-1} U_j X_j > u_n \geq \bigcup_{j=0}^{i} U_j X_j, M^{(X)}_{i+2, \left[ \frac{n}{knT} \right] T} > u_n \mid U_i = 0 \right) \]

\[
\leq P \left( \bigcup_{j=0}^{i-1} U_j X_j > u_n \geq \bigcup_{j=0}^{i-1} U_j X_j \right) = 0.
\]

For \( i = T - 1 \) and \( i = T \) we respectively have

\[
P \left( Y_{T-1} > u_n \geq Y_T, M^{(X)}_{T+1, \left[ \frac{n}{knT} \right] T} > u_n \right) \]

\[
\leq P \left( X_{T-1} > u_n \geq X_T, M^{(X)}_{T+1, \left[ \frac{n}{knT} \right] T} > u_n \right) \]

\[
+ P \left( \bigcup_{j=0}^{T-2} X_j > u_n \geq X_T, M^{(X)}_{T+1, \left[ \frac{n}{knT} \right] T} > u_n \right) \]

\[
\leq 2P \left( \bigcup_{j=0}^{T} X_j > u_n \geq X_{T+1}, M^{(X)}_{T+2, \left[ \frac{n}{knT} \right] T+T} > u_n \right)
\]

and

\[
P \left( Y_T > u_n \geq Y_{T+1}, M^{(X)}_{T+2, \left[ \frac{n}{knT} \right] T} > u_n \right) \]

\[
\leq P \left( X_T > u_n \geq X_{T+1}, M^{(X)}_{T+2, \left[ \frac{n}{knT} \right] T} > u_n \right) \]

\[
+ P \left( X_T > u_n \geq X_T \right)
\]
\[ \leq P \left( \bigvee_{j=0}^{T} X_j > u_n \geq X_{T+1}, \ M_{T+2}^{(X)} \left[ \frac{n}{k T} \right]_{T+T} > u_n \right) . \]

Due to (3.12) with these last two upper bounds, we can now write
\[
\frac{1}{T} \sum_{i=1}^{T} P \left( Y_i > u_n \geq M_{i+1,i+T}^{(Y)}, \ M_{i+T+1}^{(Y)} \left[ \frac{n}{k T} \right]_T > u_n \right) \leq \frac{2T - 1}{T} n P \left( \bigvee_{j=0}^{T} X_j > u_n \geq X_{T+1}, \ M_{T+2}^{(X)} \left[ \frac{n}{k T} \right]_{T+T} > u_n \right) , \quad (3.13)
\]

Thus, condition \( D_T^{(T+1)}(u_n) \) holds for \( \{Y_n\}_{n \geq 1} \) since (3.13) goes to zero as \( n \to +\infty \), and we may conclude that \( \{Y_n\}_{n \geq 1} \), given in (1.2) of Example 1.1, satisfies condition \( D_2^{(3)}(u_n) \).

\[ \text{Remark 1} \quad \text{Condition (3.11) is implied by condition} \]
\[
n \sum_{j=0}^{T} \sum_{k=T+2}^{T+T} P \left( X_j > u_n, X_{k-1} \leq u_n < X_k \right) \to 0 .
\]

Under condition (3.11) for \( \{X_n\}_{n \geq 0} \) we shall see that it is possible to obtain a relation between the extremal index of \( \{Y_n\}_{n \geq 1} \), \( \theta_Y \), and the extremal index of \( \{X_n\}_{n \geq 0} \), \( \theta_X \). Before we present a required lemma.

**Lemma 3.1 (a)** For \( i \in \{1, \ldots , T - 1\} \) it holds
\[
P \left( Y_i > u_n, M_{i+1,i+T}^{(Y)} \leq u_n \right) \quad (3.14)
\]
\[ P \left( X_i > u_n, M(X)(S_i) \leq u_n, M(Y)(\overline{S}_i) \leq u_n \right) p P(A_i) \\
+ P \left( \bigvee_{j=0}^{i-1} U_j X_j > u_n, M(X)(S_i) \leq u_n, M(Y)(\overline{S}_i) \leq u_n \right) (1 - p) P(A_i) \]

with \( S_i = \{i + 1, \ldots, T\}, \overline{S}_i = \{T + 1, \ldots, T + i\} \) and \( A_i := \bigcap_{j \in S_i} \{U_j = 1\} \).

(b) \[ P \left( Y_T > u_n, M(Y)_{T+1,2T} \leq u_n \right) = P \left( X_T > u_n, M(X)_{T+1,2T} \leq u_n \right) p^{T-1}. \]

**Proof (a)** We trivially have, for all \( i \in \{1, \ldots, T - 1\} \),
\[
P \left( Y_i > u_n, M(Y)_{i+1,i+T} \leq u_n \right) \\
= P \left( X_i > u_n, M(Y)(S_i) \leq u_n, M(Y)(\overline{S}_i) \leq u_n \right) p \\
+ P \left( \bigvee_{j=0}^{i-1} U_j X_j > u_n, M(Y)(S_i) \leq u_n, M(Y)(\overline{S}_i) \leq u_n \right) (1 - p).
\]

Consider now that \( i \in \{1, \ldots, T - 2\} \). In order to deal with \( M(Y)(S_i) \), observe that if there is some \( m \in \{i + 1, \ldots, T - 1\} \) such that \( U_m = 0 \), then the probability (3.14) involves the following probabilities
\[
P \left( X_i > u_n, \ldots, \bigvee_{j=0}^{m-1} U_j X_j \leq u_n, \ldots \mid U_i = 1, \ldots, U_m = 0, \ldots \right) \\
\leq P(X_i \leq u_n < X_i) = 0
\]
and
\[
P \left( \bigvee_{j=0}^{i-1} U_j X_j > u_n, \ldots, \bigvee_{j=0}^{m-1} U_j X_j \leq u_n, \ldots \mid U_i = 0, \ldots, U_m = 0, \ldots \right) \\
\leq P \left( \bigvee_{j=0}^{i-1} U_j X_j \leq u_n < \bigvee_{j=0}^{i-1} U_j X_j \right) = 0
\]

since \( m \geq i + 1 \). The proof follows immediately because \( M(Y)(S_i) \) leads directly to \( M(X)(S_i) \) under the occurrence of \( A_i \). For \( i = T - 1 \) the result follows as well because \( S_i = \{T\} \) and \( M(Y)(S_i) = X_T \).

(b) Straightforward from the previously used arguments and the fact that \( U_{kT} = 1, \ k \geq 0, \ T \geq 1 \), almost surely. \( \square \)
Theorem 3.3 If \( \{X_n\}_{n \geq 0} \) satisfies condition D\((u_n)\) for all \( \{u_n\}_{n \geq 1} \) and condition (3.11) for some \( T \geq 2 \) and \( \{u_n\}_{n \geq 1} \) satisfying (3.9), then

\[
\theta_Y = \frac{\tau_X}{\tau_T} \theta_X \left( (T - 1)p^T + p^{T-1} \right) + \frac{1}{\tau_T} \sum_{i=1}^{T-1} P_{i,T}, \tag{3.15}
\]

with

\[
P_{i,T} = \lim_{n \to +\infty} n \sum_{i \subseteq \tilde{S}_i} P \left( X_i > u_n, M^{(X)}(S_i \cup I_i) \leq u_n \right) p^{\left| I_i \right| + T - i} (1 - p)^{i - \left| I_i \right|}
\]

\[
+ \lim_{n \to +\infty} n \sum_{J_i \subseteq \tilde{S}_i} \sum_{I_i \subseteq \tilde{S}_i} P \left( M^{(X)}(J_i \cup \{0\}) > u_n, M^{(X)}(S_i \cup I_i) \leq u_n \right)
\]

\[
\times \left( 1 - p \right)^{i - \left| I_i \right| - T - i} (1 - p)^{2i - \left| I_i \right| - \left| J_i \right|}
\]

where \( J_1 = \emptyset, S_i = \{i + 1, \ldots, T\}, \tilde{S}_i = \{T + 1, \ldots, T + i\} \) for \( i \in \{1, \ldots, T - 1\} \) and, for \( T \geq 3 \) and \( i \in \{2, \ldots, T - 1\}, \tilde{S}_i = \{1, \ldots, i - 1\} \).

Proof Since \( \{Y_n\}_{n \geq 1} \) satisfies D\(_T\)\((T + 1)\)(\(u_n\)), we deduce \( \theta_Y \) from (3.8). For \( i = T \) we have, from Lemma 3.1 (b),

\[
P \left( Y_i > u_n, M^{(X)}_{i+1,i+T} \leq u_n \right) = P \left( X_T > u_n, M^{(X)}_{T+1,2T} \leq u_n \right) p^{T-1}.
\]

For any \( i \in \{1, \ldots, T - 1\} \), taking into account Lemma 3.1 (a), where \( P(A_i) = p^{T-i-1} \), we can write

\[
P \left( Y_i > u_n, M^{(X)}_{i+1,i+T} \leq u_n \right)
\]

\[
= \sum_{i \subseteq \tilde{S}_i} P \left( X_i > u_n, M^{(X)}(S_i) \leq u_n, M^{(X)}(I_i) \leq u_n \right) p^{\left| I_i \right| + T - i} (1 - p)^{i - \left| I_i \right|}
\]

\[
+ \sum_{J_i \subseteq \tilde{S}_i} \sum_{I_i \subseteq \tilde{S}_i} P \left( \max\{X_0, M^{(X)}(J_i)\} > u_n, M^{(X)}(S_i) \leq u_n, M^{(X)}(I_i) \leq u_n \right)
\]

\[
\times \left( 1 - p \right)^{i - \left| I_i \right| + T - (i+1)} (1 - p)^{i - \left| I_i \right| + (i+1) - \left| J_i \right|}
\]

\[
= P \left( X_i > u_n, M^{(X)}_{i+1,i+T} \leq u_n \right) p^T
\]

\[
+ \sum_{i \subseteq \tilde{S}_i} P \left( X_i > u_n, M^{(X)}(S_i \cup I_i) \leq u_n \right) p^{\left| I_i \right| + T - i} (1 - p)^{i - \left| I_i \right|}
\]

\[
+ \sum_{J_i \subseteq \tilde{S}_i} \sum_{I_i \subseteq \tilde{S}_i} P \left( M^{(X)}(J_i \cup \{0\}) > u_n, M^{(X)}(S_i \cup I_i) \leq u_n \right)
\]

\[
\times \left( 1 - p \right)^{i - \left| I_i \right| - T - i} (1 - p)^{2i - \left| I_i \right| - \left| J_i \right|}
\]
Now, due to the fact that condition (3.11) implies condition $D^{(T+1)}(u_n)$ of Chernick et al. (1991) and $\{X_n\}_{n\geq 0}$ is a stationary sequence, the extremal index of $\{X_n\}_{n\geq 0}$ is given by

$$\lim_{n\to+\infty} P \left( M_{i+1, i+T}^{(X)} \leq u_n \mid X_i > u_n \right) = \theta_X. $$

Hence

$$\lim_{n\to+\infty} n P \left( X_i > u_n, M_{i+1, i+T}^{(X)} \leq u_n \right) = \tau \theta_X,$$

which concludes the proof. \[\Box\]

We shall now apply the previous result in the computation of the extremal index of $\{Y_n\}_{n\geq 1}$, given in (1.1), in two different scenarios. First, we shall consider a periodic control at instants multiple of $T = 2$, i.e. $\{Y_n\}_{n\geq 1}$ 2-periodic, and that the underlying stationary sequences $\{X_n\}_{n\geq 0}$ is the moving maxima of Example 1.1. Second, we consider a 3-periodic sequence $\{Y_n\}_{n\geq 1}$ where the underlying sequence $\{X_n\}_{n\geq 0}$ is a max-autoregressive sequence (ARMAX) of order one.

**Example 3.1** ($\{X_n\}_{n\geq 0}$ moving maxima)

As previously noted, the sequence $\{Y_n\}_{n\geq 1}$ of Example 1.1 satisfies condition $D^{(3)}_2(u_n)$ with $u_n = nx$, $x > 0$, therefore its extremal index, $\theta_Y$, can be computed from (3.8).

This yields $\theta_Y = \frac{1}{2}$ which is equal to $\theta_X$ as expected.

Since $\{X_n\}_{n\geq 0}$ satisfies condition (3.11) for $T = 2$ and $u_n = nx$, $x > 0$, the extremal index $\theta_Y$ can also be calculated from (3.15) of Theorem 3.3. Note that in this case $\tau_X = \tau$. Consequently

$$\theta_Y = \frac{1}{4} \left( p^2 + p \right) + \frac{1}{2 \tau} P_{1, 2} \quad (3.16)$$

with

$$P_{1, 2} = \lim_{n\to+\infty} n \left( P(X_1 > u_n, X_2 \leq u_n) p(1 - p) + \right.$$

$$\left. + P(X_0 > u_n, X_2 \leq u_n)(1 - p)^2 + P(X_0 > u_n, M_{2, 3}^{(X)} \leq u_n) p(1 - p) \right). \quad (3.17)$$

For the probabilities in $P_{1, 2}$ we have

$$nP(X_1 > u_n, X_2 \leq u_n) = n(1 - F_Z(2u_n)) F_Z^2(2u_n) \xrightarrow{n\to+\infty} \frac{\tau}{2},$$

$$nP(X_0 > u_n, X_2 \leq u_n) = n(1 - F_Z^2(2u_n)) F_Z^2(2u_n) \xrightarrow{n\to+\infty} \tau,$$

$$nP(X_0 > u_n, M_{2, 3}^{(X)} \leq u_n) = n(1 - F_Z^2(2u_n)) F_Z^3(2u_n) \xrightarrow{n\to+\infty} \tau.$$

Taking this all into account, we finally obtain from (3.16) that $\theta_Y = \frac{1}{2}$. 

\[\Box\] Springer
Example 3.2 (\([X_n]_{n \geq 0}\) ARMAX of order one)

If we have a solar thermal energy storage system where the temperature level in a tank is periodically controlled and eventually, for some reason, temperatures at certain time points are not retained, our model (1.1) can be used to describe the temperature in such a situation. According to Hasllett (1979), the model defined by

\[
X_j = \phi X_{j-1} \vee (\psi \phi X_{j-1} + W_j), \quad j \geq 1, \quad 0 \leq \psi < 1 < \phi < 1,
\]

can be used to describe the temperature level in a tank. The extremal behaviour of this first order ARMAX storage model, for the particular case \(\psi = 0\), was studied by Alpuim (1989) and in its multivariate version by Ferreira and Ferreira (2013).

We shall now consider in (1.1) \([Y_n]_{n \geq 1}\) to be a 3-periodic \((T = 3)\) positive sequence and the underlying sequence \([X_n]_{n \geq 0}\) the first order ARMAX process of Alpuim (1989),

\[
X_n = t \max\{X_{n-1}, W_n\}, \quad n \geq 1,
\]

where \(t \in ]0, 1[\) is a constant, \(X_0\) is a positive random variable with d.f. \(H_0\), independent of the sequence of i.i.d. positive random variables \([W_n]_{n \geq 1}\) with d.f. \(L\).

Let us assume that the Markovian sequence \([X_n]_{n \geq 1}\) is stationary, i.e., there exists \(x > 0\) such that \(L(x/t) > 0\) and

\[
0 < \sum_{s=1}^{+\infty} (1 - L(x/t^s)) < +\infty,
\]

as proved in Alpuim (1989). Therefore, the non-degenerate d.f. \(H\) of \(X_n, n \geq 0\), satisfies the following equation

\[
L(x) = \frac{H(tx)}{H(x)}, \quad x \geq \alpha(H)/t,
\]

where \(\alpha(H) = \inf\{x : H(x) > 0\} \geq 0\).

It can be easily verified that, for \(n \geq 1\),

\[
X_n = \max\left\{t^n X_0, \max_{1 \leq i \leq n} t^{n-i+1} W_i\right\}.
\]

Sequence \([X_n]_{n \geq 1}\) satisfies the condition \(D(u_n)\), for any sequence \([u_n]_{n \geq 1}\), because it is strong-mixing (see Alpuim 1988).

If \(H\) belongs to the max-domain of attraction of the Fréchet d.f with parameter \(\alpha > 0\), then the normalized levels \([u_n]_{n \geq 1}\) for \([X_n]_{n \geq 1}\), i.e., such that \(n(1 - H(u_n)) \xrightarrow{n \to +\infty} \tau x \geq 0\), satisfy

\[
n(1 - H (u_n/t)) \xrightarrow{n \to +\infty} \tau x t^{\alpha}.
\]
In this case \( \{X_n\}_{n \geq 1} \) has extremal index \( \theta_X = 1 - t^\alpha \) and

\[
H(u_n/t) = \frac{H(u_n)}{H(u_n/t)} \quad \text{and} \quad n \left( H(u_n) - H(u_n/t) \right) = n(1 - H(u_n/t)) \frac{1}{H(u_n/t)}
\]

\[
\tau_X(1 - t^\alpha) = \tau_X\theta_X,
\]

(see Ferreira and Ferreira (2013) for further details). Similarly, we establish that

\[
n \left( H\left(u_n/t^{j+1}\right) - H(u_n) \right) \xrightarrow{n \to +\infty} \tau_X\theta_X(1 + t^\alpha + \ldots + t^{j\alpha}), \quad j \geq 0,
\]

\[
n \left( 1 - L\left(u_n/t^j\right) \right) \xrightarrow{n \to +\infty} t^{(j-1)\alpha} \tau_X\theta_X, \quad j \geq 1,
\]

\[
n \left( 1 - L^j(u_n/t) L\left(u_n/t^2\right) \right) \xrightarrow{n \to +\infty} \tau_X\theta_X(j + t^\alpha), \quad j \geq 1.
\]

Condition (3.11) holds for \( \{X_n\}_{n \geq 0} \), for the normalized levels \( \{u_n\}_{n \geq 1} \) and any \( T \geq 2 \). Indeed, with \( r'_n = \left\lceil \frac{n}{k_n T} \right\rceil \) we have

\[
n P\left( \bigvee_{j=0}^T X_j > u_n \geq X_{T+1}, \ M_{T+2, r'_n T+T} > u_n \right)
\]

\[
\leq n \frac{1}{T} \sum_{j=0}^T \sum_{s=T+1}^{r'_n T+T} P(X_j > u_n, X_s \leq u_n < X_{s+1})
\]

\[
\leq n \frac{1}{T} \sum_{j=0}^T \sum_{s=T+1}^{r'_n T+T} P(X_j > u_n, X_s \leq u_n, W_{s+1} > u_n/t)
\]

\[
\leq n \frac{1}{T} \sum_{j=0}^T \sum_{s=T+1}^{r'_n T+T} P(X_j > u_n) P(W_{s+1} > u_n/t)
\]

\[
\leq n \frac{T + 1}{T} \left( \frac{n}{k_n} + T \right) (1 - H(u_n)) (1 - L(u_n/t)) \xrightarrow{n \to +\infty} 0,
\]

for any positive integer sequence \( \{k_n\} \) such that \( k_n \xrightarrow{n \to +\infty} +\infty \) and \( k_n/n \xrightarrow{n \to +\infty} 0 \).

The validation of condition (3.11) for sequence \( \{X_n\}_{n \geq 0} \) and for all \( T \geq 2 \) guarantees that it also holds for \( T = 3 \) and so condition \( D_3^{(4)}(u_n) \) holds for \( \{Y_n\}_{n \geq 1} \) (Theorem 3.2). Thus, its extremal index can be computed from the expression given in Theorem 3.3. Indeed, since all the factors with products containing \( L^j(u_n/t^m) \) tend
to 1, for all \( j, m \geq 1 \), we have

\[
\sum_{i=1}^{2} P_{i,3} = \lim_{n \to +\infty} n \left( H(u_n/t) - H(u_n) \right) (p + p^2 - 2p^3) + n \left( H(u_n/t^2) - H(u_n) \right) (1 - p) + n \left( 1 - L^2(u_n/t)L(u_n/t^2) \right) p(1 - p) = \tau_X \theta_X \left\{ \left( p + p^2 - 2p^3 \right) + (2 - \theta_X)p(1 - p) + (3 - \theta_X)p(1 - p) + (3 - 3\theta_X + \theta_X^2)(1 - p)^2 \right\}.
\]

Furthermore, \( G_1(x) = H(x) \) and \( G_2(x) = (1 - p)H(x) - pP(X_1 \leq x, X_0 \leq x) \), leading (3.9) to \( \tau_1 = \tau_X \) and \( \tau_2 = \theta_X + \rho \theta_X \tau_X \). Hence, by (3.10), it holds

\[
\tau = \tau_X \left( 1 + \frac{1}{3} \rho(1 - p)\theta_X \right).
\]

As a consequence, (3.15) becomes

\[
\theta_Y = \frac{\tau_X}{\tau_T} \theta_X \left( 2p^3 + p^2 \right) + \frac{\tau_X}{\tau_T} \theta_X \left( 3 - 2p^3 - p^2 + \theta_X(-3 + 4p - p^2) + \theta_X^2(1 - p)^2 \right) = \frac{3\theta_X + \theta_X^2(-3 + 4p - p^2) + \theta_X^3(1 - p)^2}{3 + p(1 - p)\theta_X}
\]

Figure 2 shows the effect that the extremal index of the underlying sequence \( \theta_X \in [0, 1] \) and the parameter \( p \in [0, 1] \) have on the extremal index of \( \{Y_n\}_{n \geq 1} \), \( \theta_Y \). As we can see, when \( p \) is very close to one we get \( \theta_Y \approx \theta_X \), since in this case there are almost no missing values and so \( Y_n = X_n, n \geq 1 \), almost surely. When \( p \) is very close to zero, almost all values of the underlying sequence are missing, except for the values at instants multiple of three, since \( T = 3 \). In this case, sequence \( \{Y_n\}_{n \geq 1} \) will have the following form \( X_0, X_0, X_3, X_3, X_3, X_6, X_6, X_6, \ldots, X_{3T}, X_{3T}, X_{3T}, \ldots \), and so if \( \theta_X \approx 1 \) (\( \theta_X = 1 \) occurs, for example, for i.i.d sequences), exceedances of high levels will form clusters of mean size approximately \( T = 3 \), yielding an extremal index approximately equal to \( 1/3 \) as observed in Fig. 2.

Moreover, considering that the periodic control took place at instants multiple of two, \( T = 2 \), then, with \( P_{1,2} \) given by (3.17) we would obtain, from Theorem 3.3,

\[
\theta_Y = \frac{\theta_X}{2} (p^2 + p) + \frac{\theta_X}{2} ((2 - \theta_X)(1 - p) + p(1 - p)) = \theta_X + \theta_X^2 \frac{p - 1}{2},
\]

since it was proved that condition (3.11) also holds for the underlying ARMAX sequence of order one when \( T = 2 \).
Fig. 2 \( \theta_Y \) as a function of \( \theta_X \) and \( p \) for an ARMAX underlying sequence

Imposing a stronger condition on the behaviour of the underlying stationary sequence \( \{X_n\}_{n \geq 0} \) than condition (3.11), we obtain, with similar arguments as used to prove Theorem 3.2, the validation of condition \( D_T(u_n) \) for \( \{Y_n\}_{n \geq 1} \) and consequently also of \( D_{T+1}(u_n) \), as stated in the next result.

**Theorem 3.4** If for any sequence \( \{u_n\}_{n \geq 1} \), the sequence \( \{X_n\}_{n \geq 0} \) satisfies condition \( D(u_n) \) and

\[
nP \left( M_{0,T}^{(X)} > u_n, M_{T+1}^{(X)} \left[ \frac{n}{\lambda_n^T} \right]_{T+1} > u_n \right) \xrightarrow{n \to +\infty} 0 \tag{3.18}
\]

then condition \( D_T(u_n) \) holds for the sequence \( \{Y_n\}_{n \geq 1} \).

Condition (3.18) is indeed more demanding than condition (3.11) as we can verify with the moving maxima sequence \( \{X_n\}_{n \geq 0} \) defined in Example 1.1. In this case condition (3.18) with \( T = 2 \) does not hold since

\[
nP \left( M_{0,2}^{(X)} > u_n, M_{2+1}^{(X)} \left[ \frac{n}{2\lambda_n^2} \right]_{2+2} > u_n \right) \\
\geq nP (X_2 > u_n, X_3 > u_n) \\
= n(1 - F_Z(2u_n)) \left( 1 + F_Z(2u_n) - F_Z^2(2u_n) \right) \xrightarrow{n \to +\infty} \frac{\tau}{2} > 0.
\]
Nevertheless, the associated sequence \( \{Y_n\}_{n \geq 1} \), satisfies condition \( D_2^{(2)}(u_n) \) which shows that the condition (3.18) is not a necessary condition.

**Remark 2** Condition (3.18) is implied by the following condition

\[
\frac{T}{T+1} \sum_{j=0}^{T} \sum_{k=T+1}^{n} P(X_j > u_n, X_k > u_n) \xrightarrow{n \to +\infty} 0.
\]

We now present an analogous result to Theorem 3.3, relating \( \theta_Y \) and \( \theta_X \) under the hypothesis that \( \{X_n\}_{n \geq 0} \) satisfies condition (3.18) and consequently \( \{Y_n\}_{n \geq 1} \) satisfies condition \( D_T^{(T)}(u_n) \). The main difference can be found in the first term of the expression for \( \theta_Y \) and in the definition of \( Si \).

**Theorem 3.5** If \( \{X_n\}_{n \geq 0} \) satisfies condition \( D(u_n) \), for all \( \{u_n\}_{n \geq 1} \), and (3.18), for some \( T \geq 2 \) and \( \{u_n\}_{n \geq 1} \) satisfying (3.9), then \( \{Y_n\}_{n \geq 1} \) satisfies condition \( D_T^{(T)}(u_n) \) and

\[
\theta_Y = \frac{\tau X}{\tau} \theta_X p^{T-1} + \frac{1}{\tau T} \sum_{i=1}^{T-1} P^*_i T
\]

with

\[
P^*_i T = \lim_{n \to +\infty} \frac{n}{T} \left( \sum_{I_i \subseteq S_i} P \left( X_i > u_n, M(X)(S_i \cup I_i) \leq u_n \right) \times \right.
\]

\[
\left. \times p^{|I_i|+T-i}(1-p)^{i-1-|I_i|} + \sum_{J_i \subseteq S_i} \sum_{L_i \subseteq S_i} P \left( M(X)(J_i \cup \{0\}) > u_n, M(X)(S_i \cup I_i) \leq u_n \right) \times \right.
\]

\[
\left. \times p^{|I_i|+|J_i|+T-i-1}(1-p)^{2i-1-|I_i|-|J_i|}, \right)
\]

where \( I_1 = J_1 = \emptyset \), \( S_i = \{i+1, \ldots, T\} \) for any \( i \in \{1, \ldots, T-1\} \) and, for \( T \geq 3 \) and any \( i \in \{2, \ldots, T-1\} \), \( \overline{S}_i = \{T+1, \ldots, T+i-1\} \) and \( \overline{S}_j = \{1, \ldots, i-1\} \).

We observe that any i.i.d. positive sequence \( \{X_n\}_{n \geq 0} \) trivially satisfies (3.18) for normalized levels \( u_n \), and therefore sequence \( \{Y_n\}_{n \geq 1} \) satisfies condition \( D_T^{(T)}(u_n) \). Then, considering for instance \( T = 3 \) and proceeding as before, we get \( \tau_2 = \tau_X(1+p), \) \( \tau = \frac{1}{3} \tau_X(3+p-p^2) \) and

\[
\sum_{i=1}^{2} P^*_i T = \lim_{n \to +\infty} n(1-F(u_n))(1-p^2) + \lim_{n \to +\infty} n(1-F^2(u_n))(p-p^2)
\]

\[
= \tau_X(1+2p-3p^2).
\]
Hence
\[
\theta_Y = \frac{\tau_X}{\tau} p^2 + \frac{1}{3\tau} \tau_X (1 + 2p - 3p^2) = \frac{1 + 2p}{3 + p(1 - p)}.
\]

4 Conclusions

Missing data is a cross-cutting problem in several fields. In this work we present a model \( \{Y_n\} \) where randomly occurring missing values are imputed with the maximum of the previous observations up to the last one surely registered due to a periodic control. This periodic control at instants multiple of \( T \geq 2 \) and the stationarity of the underlying sequence guarantee that \( \{Y_n\} \) is \( T \)-periodic. The known results that characterize the extremal behaviour of periodic sequences rely on the validation of long range and local dependence conditions. We obtain necessary conditions, which depend on the underlying sequence, that validate these dependence conditions for \( \{Y_n\} \) and therefore allow the computation of its extremal index.

A simple consistent estimator for the model parameter, that regulates the missing values, is proposed and its good performance shown through simulated examples.

Acknowledgements The first two authors were 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 third author was partially supported by the Centre for Mathematics of the University of Coimbra – UIDB/00324/2020, funded by the Portuguese Government through FCT/MCTES.

References

Alpuim MT (1988) Contribuições à teoria de valores extremos em sucessões dependentes. Ph.D. Thesis. DEIO, University of Lisbon
Alpuim MT (1989) An extremal Markovian sequence. J Appl Probab 26:219–232
Crambes C, Henchiri Y (2019) Regression imputation in the functional linear model with missing values in the response. J Stat Plann Inf 201:103–119
Chernick MR, Hsing T, McCormick WP (1991) Calculating the extremal index for a class of stationary sequences. Adv Appl Probab 23:835–850
Falk M, Hüsler J, Reiss R-D (2011) Laws of small numbers: extremes and rare events, 3rd edn. Birkhäuser, Basel
Ferreira H (1994) Multivariate extreme values in \( T \)-periodic random sequences under mild oscillation restrictions. Stoch Process Appl 49:111–125
Ferreira M, Ferreira H (2013) Extremes of multivariate ARMAX processes. Test 22(4):606–627
Ferreira H, Martins AP (2003) The extremal index of sub-sampled periodic sequences with strong local dependence. REVSTAT Stat J 1:15–24
Hall A, Hüsler J (2006) Extremes of stationary sequences with failures. Stoch Models 22:537–557
Hall A, Scotto M (2008) On the extremes of randomly sub-sampled time series. REVSTAT Stat J 6(2):151–164
Hall A, Temido MG (2009) On the max-semistable limit of maxima of stationary sequences with missing values. J Stat Plan Inference 139:875–890
Haslett J (1979) Problems in the stochastic storage of a solar thermal energy. In: Jacobs O (ed) Analysis and optimization of stochastic systems. Academic Press, London
Leadbetter MR (1974) On extreme values in stationary sequences. Z Wahrscheinlichkeitstheor Verw Geb 28(4):289–303
Martins AP, Ferreira H (2004) The extremal index of sub-sampled processes. J Stat Plan Inference 1:145–152

Springer
Moritz S, Sardá A, Bartz-Beielstein T, Zaefferer M, Stork J (2015) Comparison of different methods for univariate time series imputation in R. ArXiv e-prints
Moritz S, Bartz-Beielstein T (2017) imputeTS: time series missing value imputation in R. R J 9(1):207–218
Quintela-del-Río A, Estévez-Pérez R (2012) Nonparametric kernel distribution function estimation with kerdiest: an R package for bandwidth choice and applications. J Stat Softw 50:8
Scotto M, Turkman K, Anderson C (2003) Extremes OS some sub-sampled time series. J Time Ser Anal 24:579–590
Weissman I, Cohen U (1995) The extremal index and clustering of high values for derived stationary sequences. J Appl Probab 32:972–981
Zha R, Harel O (2019) Power calculation in multiply imputed data. Stat Pap. https://doi.org/10.1007/s00362-019-01098-8

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.