View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Repositório do Instituto Politécnico de Castelo Branco

Methods for estimating the upcrossings index: improvements and comparison

A.P. Martins *
Departamento de Matemática
Universidade da Beira Interior
Portugal

J.R. Sebastião †
Escola Superior de Gestão
Instituto Politécnico de Castelo Branco
Portugal

arXiv:1508.00053v1 [math.PR] 31 Jul 2015

Abstract: The upcrossings index $0 \leq \eta \leq 1$, a measure of the degree of local dependence in the upcrossings of a high level by a stationary process, plays, together with the extremal index θ, an important role in extreme events modelling. For stationary processes, verifying a long range dependence condition, upcrossings of high thresholds in different blocks can be assumed asymptotically independent and therefore blocks estimators for the upcrossings index can be easily constructed using disjoint blocks. In this paper we focus on the estimation of the upcrossings index via the blocks method and properties such as consistency and asymptotic normality are studied. We also enlarge the family of runs estimators of η and provide an empirical way of checking local dependence conditions that control the clustering of upcrossings to improve the estimates obtained with the runs method.

We compare the performance of a range of different estimators for η and illustrate the methods using simulated data and financial data.

Keywords: Stationary sequences, upcrossings index, blocks estimators, consistency and asymptotic normality.

Mathematics Subject Classification (2000) 60G70

1 Introduction

Extreme Value Theory aims to predict occurrence of rare events but with disastrous impact, making an adequate estimation of the parameters related with such events of primordial importance in Statistics of Extremes. The extremal index $\theta \in (0, 1]$ and the upcrossings index $\eta \in (0, 1]$ play an important role when modelling extreme events. The knowledge of these parameters, θ and η, entails in particular the understanding of the way in which exceedances and upcrossings of high levels, respectively, cluster in time. They provide different but complementary information concerning the grouping characteristics of rare events.

Suppose that we have n observations from a strictly stationary process $\mathbf{X} = \{X_n\}_{n\geq 1}$, with marginal distribution function F and finite or infinite right endpoint $x_F = \sup\{x : \ F(x) < 1\}$. $\mathbf{X}$ is said to have

*E-mail:amartins@ubi.pt
†E-mail:jrenato@ipcb.pt

extremal index $\theta \in [0,1]$ if for each $\tau > 0$ there exists a sequence of thresholds $\{u_n^{(\tau)}\}_{n\geq 1}$ such that

$$n(1 - F(u_n^{(\tau)})) \xrightarrow[n\to+\infty]{} \tau \qquad \text{and} \qquad F_n(u_n^{(\tau)}) = P(M_n \leq u_n^{(\tau)}) \xrightarrow[n\to+\infty]{} e^{-\theta\tau}, \tag{1.1}$$

where $M_n = \max\{X_1, \ldots, X_n\}$ (Leadbetter (1983) [16]). This parameter is a measure of clustering tendency of extremes, more precisely θ^{-1} is the limiting mean cluster size in the point process of exceedance times over a high threshold, under an appropriate long range dependence condition (Hsing *et al.* (1988) [14]). For independent and identically distributed (i.i.d.) processes, extremes are isolated and hence $\theta = 1$, whereas for stationary processes, the stronger the dependence the larger the clusters of exceedances and therefore the smaller θ.

Many results in extreme value theory may be naturally discussed in terms of point processes. When looking at the point process of upcrossing times of a threshold u_n, that is

$$\widetilde{N}_n(u_n)(B) = \sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq u_n < X_{i+1}\}} \delta_{\frac{i}{n}}(B), \quad B \subset (0,1], \tag{1.2}$$

where $\mathbb{I}_A$ denotes the indicator of event A and δ_a the unit mass at a, Ferreira (2006) [6] showed that, under a long range dependence condition, if the sequence of point processes of upcrossings $\widetilde{N}_n(u_n)$ converges in distribution (as a point process on (0,1]), then the limit is necessarily a compound Poisson process and the Poisson rate of the limiting point process is $\eta\nu$, when the limiting mean number of upcrossings of u_n is $\nu > 0$. That is, there is a clustering of upcrossings of high levels, where the underlying Poisson values represent cluster positions and the multiplicities are the cluster sizes. In such a situation there appears the upcrossings index η which has an important role in obtaining the sizes of the clusters of upcrossings.

The stationary process $\mathbf{X}$ is said to have upcrossings index $\eta \in [0,1]$ if for each $\nu > 0$ there exists a sequence of thresholds $\{u_n^{(\nu)}\}_{n\geq 1}$ such that

$$nP(X_1 \leq u_n^{(\nu)} < X_2) \xrightarrow[n\to+\infty]{} \nu \qquad \text{and} \qquad P(\widetilde{N}_n(u_n^{(\nu)})((0,1]) = 0) \xrightarrow[n\to+\infty]{} e^{-\eta\nu}. \tag{1.3}$$

The upcrossings index η exists if and only if there exists the extremal index θ (Ferreira (2006) [6]) and in this case

$$\eta = \frac{\tau}{\nu}\,\theta. \tag{1.4}$$

Note that for a high level u_n such that $nP(X_1 > u_n) = \tau'$ and $nP(X_1 \leq u_n < X_2) = \nu'$ we have $\tau' \geq \nu' \geq 0$ and therefore $\eta \geq \theta$.

Let us now consider the moving maxima sequence of Ferreira (2006) [6] in the following example:

Example 1.1 Let $\{Y_n\}_{n\geq -2}$ be a sequence of independent uniformly distributed on [0,1] variables with common distribution function F. Define the 3-dependent moving maxima sequence as $X_n = \max\{Y_n, Y_{n-2}, Y_{n-3}\}$, $n \geq 1$, and consider $\{Z_n\}_{n\geq 1}$ a sequence of i.i.d. random variables with common distribution F^3. The underlying distribution of X_n is also F^3 and it has an extremal index $\theta = 1/3$ and upcrossings index $\eta = 1/2$. Moreover, for $u_n = 1 - \frac{\tau}{n}$, $\tau > 0$, it holds $nP(X_1 > u_n) \xrightarrow[n\to+\infty]{} 3\tau$ and $nP(X_1 \leq u_n < X_2) \xrightarrow[n\to+\infty]{} 2\tau$, agreeing with (1.4). In Figure 1 we can see the sizes equal to 3 and 2, respectively for the clusters of exceedances and upcrossings of a high level by the max-autoregressive sequence X_n. Note that X_n verifies condition $D^{(3)}(u_n)$ of Chernick *et al.* (1991) [3] (see Ferreira (2006)

[6]), therefore two runs of exceedances separated by one single non-exceedance are considered in the same cluster.

A shrinkage of the largest and smallest observations for the 3-dependent sequence can in general also be observed, despite the fact that we have the same model underlying both sequences.

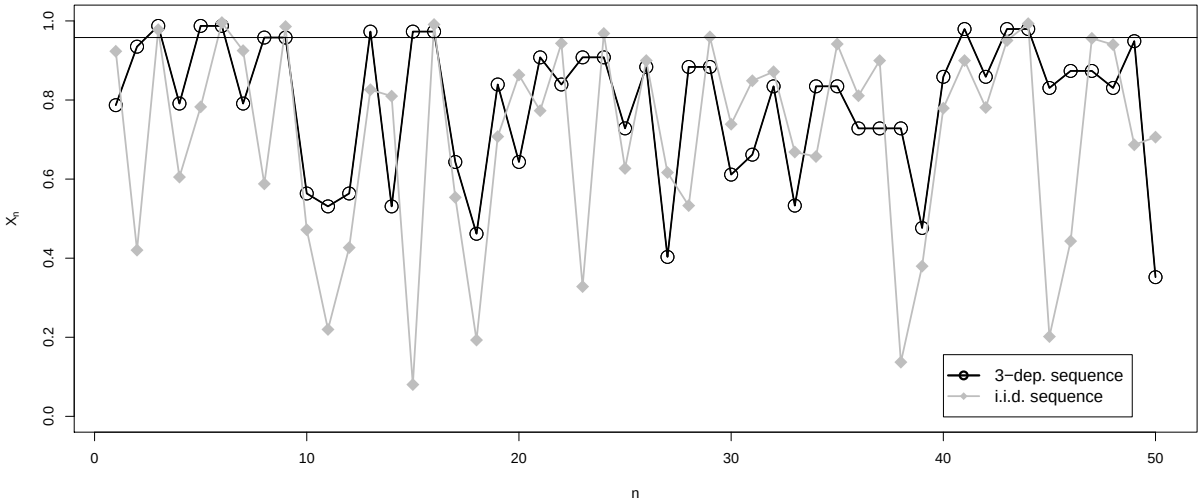

Figure 1: Sample paths from an i.i.d. process and a 3-dependent process. The horizontal line corresponds the the 80th percentile of $n = 50$ realizations of the 3-dependent sequence $X_n,\ n \geq 1$.

In the i.i.d. setting we obviously have an upcrossings index $\eta = 1$. This also holds for processes satisfying the local dependence condition $D''(u_n)$ of Leadbetter and Nandagopalan (1989) [17], which enables clustering of upcrossings. Other properties of the upcrossings index can be found in Ferreira (2006, 2007) [6], [7] and Sebastião *et al.* (2010) [23], relations with upcrossings-tail dependence coefficients are presented in Ferreira and Ferreira (2012) [8].

It is clear that clustering of extremes can have catastrophic consequences and upcrossings of a high threshold can also lead to adverse situations. These motivate an increasing interest in the grouping characteristics of rare events. Applications include, for example, the temporal distribution of large financial crashes or the evaluation of oscillation periods in financial markets. A search for reliable tools to describe these features has become constant in the past years, a key aspect being the estimation of the extremal index θ and the upcrossings index η which govern, respectively, the clustering of exceedances and upcrossings of a high level by an univariate observational series.

Several estimators for the extremal index θ can be found in the literature (see Ancona-Navarrete and Tawn (2000) [1], Robert *et al.* (2009) [21] and references therein). Most of the proposed estimators are constructed by the blocks method or by the runs method. These methods identify clusters and construct estimates for θ based on these clusters. The way the clusters are identified distinguishes both methods. The blocks estimators are usually constructed by using disjoint blocks because exceedances over high thresholds of different blocks can be assumed asymptotically independent. More recently, an inter-exceedance times method has been proposed that obviates the need for a cluster identification scheme parameter (Ferro and Segers (2003) [9]). In what concerns the estimation of the upcrossings index, little

has yet been done. In Sebastião *et al.* (2013) [24] we find a runs estimator and a blocks estimator derived from a disjoint blocks estimator for the extremal index, given in [21], for this parameter. It was shown that the runs estimator has a smaller asymptotic variance and a better performance, nevertheless it assumes the validation of a local dependence condition which does not hold for many well known processes and can be cumbersome to verify.

This paper focuses on the estimation of η. The main novelty is our proposal of new blocks estimators and the study of their properties, namely consistency and asymptotic normality, as well as an improvement of the existing runs estimator. We examine the behaviour of the new and existing estimators of η and assess their performance for a range of different processes. Assuming the validation of local dependence conditions $\widetilde{D}^{(k)}(u_n)$, $k \geq 3$, given in Ferreira (2006) [6] we define new runs estimators for η. With these new estimators we obtain better estimates of η, comparatively to the ones obtained in Sebastião *et al.* (2013) [24] where $k = 3$ was considered. We analyze a way of checking conditions $\widetilde{D}^{(k)}(u_n)$ preliminary to estimation, which helps considerably when using the runs estimators to estimate η.

We consider the problem of estimating the upcrossings index in Section 2. There we suggest an estimator motivated by the relation

$$E[\widetilde{N}_{r_n}(u_n^{(\nu)}) \mid \widetilde{N}_{r_n}(u_n^{(\nu)}) > 0] = \sum_{j=1}^{+\infty} j\widetilde{\pi}_n(j; u_n^{(\nu)}) \xrightarrow[n\to+\infty]{} \eta^{-1}, \tag{1.5}$$

where

$$\widetilde{\pi}_n(j; u_n) := P(\widetilde{N}_{r_n}(u_n) = j \mid \widetilde{N}_{r_n}(u_n) > 0)$$

is the conditional cluster size distribution and $\widetilde{N}_{r_n}(u_n) \equiv \widetilde{N}_n(u_n)((0, r_n/n])$, under suitable choices of $r_n \xrightarrow[n\to+\infty]{} +\infty$. This relation is a consequence of Lemma 2.1 of Ferreira (2006) [6], that states that under condition $\Delta(u_n)$ of Hisng *et al.* (1988) [14] upcrossings over disjoint blocks are asymptotically independent, and definition (1.3) of the upcrossings index η. It is extremely important in the interpretation of the upcrossings index, since it tells us that the reciprocal of the upcrossings index can be viewed as the limiting mean cluster size of upcrossings.

Other blocks estimators are also presented and their properties studied in Section 2. In subsection 2.2 we revisit the runs declustering method considered in Sebastião *et al.* (2013) [24] and define new runs estimators for η under the validation of any dependence condition $\widetilde{D}^{(k)}(u_n)$, $k \geq 3$. A way of checking conditions $\widetilde{D}^{(k)}(u_n)$, $k \geq 3$, preliminary to estimation is also given.

We compare, in Section 3, the performance of the estimators through a simulation study for a range of different processes with $\eta < 1$ to $\eta = 1$ and, in Section 4, by the application of the methods to the DAX daily log returns. Conclusions are drawn in Section 5 and proofs are given in the appendix.

2 Estimation of the upcrossings index

We shall now consider the problem of estimating the upcrossings index in the following setting. As previously mentioned $\mathbf{X} = \{X_n\}_{n\geq 1}$ is a stationary sequence of random variables with a continuous marginal distribution function F and a nonzero upcrossings index η. Assume also that condition $\Delta(u_n^{(\nu)})$ of Hsing *et al.* (1988) [14] holds for $\mathbf{X}$.

2.1 Blocks declustering

The blocks declustering scheme consists in choosing a block length r_n and partitioning the n observations $(X_1, \ldots, X_n)$ of $\mathbf{X}$ into $k_n = [n/r_n]$ disjoint blocks, where $[y]$ denotes the integer part of y. Each block that contains at least one upcrossing is treated as one cluster. If in Figure 1 we consider blocks of size 10 ($r_n = 10$) we can identify 3 clusters.

2.1.1 A disjoint blocks estimator

Under relation (1.5), η^{-1} is the (asymptotic) mean cluster size of upcrossings, and so a natural way to estimate η is as the reciprocal of the sample average cluster size, that is, count the total number of blocks with at least one upcrossing and divide it by the total number of upcrossings of a high threshold u. We then propose the following blocks estimator for η :

$$\widehat{\eta}_n^B(u_n) := \frac{\sum_{i=1}^{k_n} \mathbb{1}_{\{\widetilde{N}_{r_n}^{(i)}(u_n)>0\}}}{\sum_{i=1}^{k_n} \widetilde{N}_{r_n}^{(i)}(u_n)} = \left(\sum_{j=1}^{+\infty} \widehat{\widetilde{\pi}}_n(j; u_n) \right)^{-1}, \tag{2.6}$$

where

$$\widehat{\widetilde{\pi}}_n(j; u_n) := \sum_{i=1}^{k_n} \mathbb{1}_{\{\widetilde{N}_{r_n}^{(i)}(u_n)=j\}} \Big/ \sum_{i=1}^{k_n} \mathbb{1}_{\{\widetilde{N}_{r_n}^{(i)}(u_n)>0\}} \tag{2.7}$$

is the empirical conditional distribution of $\widetilde{N}_{r_n}^{(i)}(u_n)$ and $\widetilde{N}_{r_n}^{(i)}(u_n) \equiv \widetilde{N}_n(u_n)([((i-1)r_n+1)/n,\ ir_n/n]) = \sum_{j=(i-1)r_n+1}^{ir_n} \mathbb{1}_{\{X_j \le u_n < X_{j+1}\}}$, $1 \le i \le k_n$, represents the number of upcrossings of the level u_n in the $ith-$block of length r_n.

As we can see from (2.6) and (2.7) the average number of upcrossings in the blocks estimates η and the proportion of blocks with j upcrossings estimates $\widetilde{\pi}(j; u_n) = \lim_{n\to+\infty} \widetilde{\pi}_n(j; u_n)$.

The level u_n in (2.6) is a tuning constant that determines the quality of the estimate. If u_n is too high there will be just a few upcrossings of the threshold so estimates will be highly variable, but if u_n is too low the approximation to the limiting characterization may be poor so the estimates may be biased. The sample size is important here as it influences what is a too high threshold.

Note that when considering n finite the sequences $\{r_n\}_{n\ge1}$ and $\{k_n\}_{n\ge1}$ determine the cluster identification. The values of r_n are determined by the dependence structure of the process. If a process has weak dependence, the high level upcrossings will occur in well defined isolated groups, so small values for r_n will suffice, similarly large values of r_n are required with strong dependence.

As we shall see in the next section, the quality of the estimate depends heavily on the choice of both parameters, u_n and r_n.

2.1.1.1 Consistency and asymptotic normality

Consistency of the estimator $\widehat{\eta}_n^B$ can only be achieved with lower thresholds than the ones considered in (1.3), since for these levels there are insufficient upcrossings to give statistical "consistency" for the estimator. That is, as n increases the value of $\widehat{\eta}_n^B$ does not necessarily converge appropriately to the value η. We shall therefore consider thresholds $v_n = u_{[n/c_n]}^{(\nu)}$, for some fixed $\nu > 0$, that satisfy

$$nP(X_1 \le v_n < X_2) \sim c_n\nu \quad \text{as } n \to +\infty, \tag{2.8}$$

where $\{c_n\}_{n\geq 1}$ is a sequence of real numbers such that $c_n \xrightarrow[n\to+\infty]{} +\infty$, $n/c_n \xrightarrow[n\to+\infty]{} +\infty$ and $c_n/k_n \xrightarrow[n\to+\infty]{} 0$, for a sequence of integers $\{k_n\}_{n\geq 1}$ satisfying $k_n \xrightarrow[n\to+\infty]{} +\infty$.

The following result shows that for thresholds v_n, satisfying (2.8), convergence (1.5) still holds.

Theorem 2.1 *Let $\{r_n\}_{n\geq 1}$, $\{k_n\}_{n\geq 1}$ and $\{c_n\}_{n\geq 1}$ be sequences of real numbers such that*

$$c_n \to +\infty, \quad k_n \to +\infty, \quad r_n \to +\infty, \quad c_n/k_n \to 0, \quad k_n r_n \leq n \quad k_n r_n \sim n, \quad \text{as } n \to +\infty$$

and $\{v_n\}_{n\geq 1}$ a sequence of thresholds defined by (2.8). Suppose there exists $l_n = o(r_n)$ such that $\frac{k_n}{c_n}\alpha_{n,l_n} \xrightarrow[n\to+\infty]{} 0$, where α_{n,l_n}, $n \geq 1$, are the mixing coefficients of the $\Delta(v_n)$ condition.

Then

$$\lim_{n\to+\infty} \frac{k_n}{c_n} P(\widetilde{N}_{r_n}(v_n) > 0) = \eta\nu, \tag{2.9}$$

and (1.5) holds for v_n, that is

$$E[\widetilde{N}_{r_n}(v_n) \mid \widetilde{N}_{r_n}(v_n) > 0] \xrightarrow[n\to+\infty]{} \eta^{-1}.$$

Note that from Theorem 2.1 if (2.8) holds then the order of magnitude of the expected number of blocks having upcrossings of v_n is $c_n \xrightarrow[n\to+\infty]{} +\infty$. This justifies the need to consider lower levels satisfying (2.8) in order to guarantee the consistency of the estimator $\widehat{\eta}_n^B$. Such levels were also considered in Sebastião *et al.* (2013) [24] to obtain the consistency of the runs estimator of η.

Following the style of proofs used by Hsing (1991) [15], we show in the next results that the blocks estimator of the upcrossings index $\widehat{\eta}_n^B$ is a consistent estimator for η and that it is asymptotically normal. These results require assumptions on the limiting behaviour of the first and second moments of $\widetilde{N}_{r_n}(v_n)$.

Theorem 2.2 *Suppose that the conditions of Theorem 2.1 hold and for some $\nu > 0$,*

(1) $\frac{k_n}{c_n} E[\widetilde{N}_{r_n}(v_n) \mathbb{1}_{\{\widetilde{N}_{r_n}(v_n) > c_n\}}] \xrightarrow[n\to+\infty]{} 0$,

(2) $\frac{k_n}{c_n^2} E[\widetilde{N}_{r_n}^2(v_n) \mathbb{1}_{\{\widetilde{N}_{r_n}(v_n) \leq c_n\}}] \xrightarrow[n\to+\infty]{} 0$.

Then

$$\widehat{\eta}_n^B(v_n) \xrightarrow[n\to+\infty]{P} \eta.$$

The underlying idea to obtain the asymptotic distribution of the estimator $\widehat{\eta}_n^B$ is to split each block into a small block of length l_n and a big block of length $r_n - l_n$. Since $l_n = o(r_n)$ and $\frac{k_n}{c_n}\alpha_{n,l_n} \xrightarrow[n\to+\infty]{} 0$ it is ensured that l_n is sufficiently large such that blocks that are not adjacent are asymptotically independent, but does not grow to fast such that the contributions of the small blocks are negligible.

Theorem 2.3 *If conditions of Theorem 2.1 hold and for some $\nu > 0$,*

(1) $\frac{k_n}{c_n} E[\widetilde{N}_{r_n}^2(v_n) \mathbb{1}_{\{\widetilde{N}_{r_n}^2(v_n) > \epsilon c_n\}}] \xrightarrow[n\to+\infty]{} 0$, *for all $\epsilon > 0$,*

(2) $\frac{k_n}{c_n} E[\widetilde{N}_{r_n}^2(v_n)] \xrightarrow[n\to+\infty]{} \eta\nu\sigma^2$, *for some $\sigma^2 < +\infty$.*

Then

$$c_n^{-1/2}\left[\begin{array}{c}\sum_{i=1}^{k_n}\left(\widetilde{N}_{r_n}^{(i)}(v_n)-E[\widetilde{N}_{r_n}^{(i)}(v_n)]\right)\\ \sum_{i=1}^{k_n}\left(1\!\!1_{\{\widetilde{N}_{r_n}^{(i)}(v_n)>0\}}-E[1\!\!1_{\{\widetilde{N}_{r_n}^{(i)}(v_n)>0\}}]\right)\end{array}\right]\xrightarrow[n\to+\infty]{d}\mathcal{N}\left(\left[\begin{array}{c}0\\0\end{array}\right],\left[\begin{array}{cc}\nu\eta\sigma^2 & \nu\\ \nu & \nu\eta\end{array}\right]\right). \tag{2.10}$$

Remark 1 *From assumption* (2) *of the previous theorem we have that* $E[\widetilde{N}_{r_n}^2(v_n)\mid\widetilde{N}_{r_n}^2(v_n)>0]\xrightarrow[n\to+\infty]{}\sigma_B^2$, *for some* $\sigma_B^2>0$, *which means that* $Var[\widetilde{N}_{r_n}^2(v_n)\mid\widetilde{N}_{r_n}^2(v_n)>0]\xrightarrow[n\to+\infty]{}\sigma_B^2-\eta^{-2}$, *the asymptotic variance of a cluster size.*

The asymptotic normality of $\widehat{\eta}_n^B$ is now an immediate consequence of the previous result.

Corollary 2.1 *If the conditions of Theorem 2.3 hold then*

$$\sqrt{c_n}(\widehat{\eta}_n^B-\eta_n)\xrightarrow[n\to+\infty]{d}\mathcal{N}\left(0,\frac{\eta}{\nu}(\eta^2\sigma^2-1)\right),$$

where $\eta_n=E[1\!\!1_{\{\widetilde{N}_{r_n}(v_n)>0\}}]/E[\widetilde{N}_{r_n}(v_n)]$.

From the previous result it is clear that the asymptotic distribution of $\widehat{\eta}_n^B$ is not well defined when $\sigma^2\leq 1/\eta^2$, since in this case a smaller or equal to zero variance is obtained.

Remark 2 *The asymptotic distribution found in Corollary 2.1 differs from the one obtained with the runs estimator in Sebastião et al. (2013) [24], only on the value of* σ^2. *Thus, the value of* σ^2 *determines which of the two is the most efficient estimator of* η.

In order to control the bias of the upcrossings index estimator $\widehat{\eta}_n^B$, lets assume that the block sizes are sufficiently large so that as $n\to+\infty$

$$\sqrt{c_n}(\eta_n-\eta)\xrightarrow[n\to+\infty]{}0. \tag{2.11}$$

The asymptotic variance of $\widehat{\eta}_n^B$ will depend on η, ν and σ^2 which can easily be estimated with consistent estimators as shown in the next result. This result enables the construction of approximate confidence intervals or a hypothesis test regarding η. From a practical viewpoint it is more useful than Corollary 2.1 and follows readily from the previous results.

Corollary 2.2 *If the conditions of Corollary 2.1 hold, (2.11) and*

$$\frac{k_n}{c_n^2}E[\widetilde{N}_{r_n}^4(v_n)1\!\!1_{\{\widetilde{N}_{r_n}^2(v_n)\leq c_n\}}]\xrightarrow[n\to+\infty]{}0.$$

Then

$$\sqrt{\frac{\sum_{i=1}^{k_n}\widetilde{N}_{r_n}^{(i)}(v_n)}{\widehat{\eta}_n^B((\widehat{\eta}_n^B\widehat{\sigma}_n)^2-1)}}\,(\widehat{\eta}_n^B-\eta)\xrightarrow[n\to+\infty]{}\mathcal{N}(0,1), \tag{2.12}$$

where $\widetilde{N}_{r_n}^{(i)}(v_n)=\sum_{j=(i-1)r_n+1}^{ir_n}1\!\!1_{\{X_j\leq v_n<X_{j+1}\}}$ *and* $\widehat{\sigma}_n^2=\frac{\sum_{i=1}^{k_n}(\widetilde{N}_{r_n}^{(i)}(v_n))^2}{\sum_{i=1}^{k_n}1\!\!1_{\{\widetilde{N}_{r_n}^{(i)}(v_n)>0\}}}$.

Given a random sample $(X_1, \ldots, X_n)$ of $\mathbf{X}$ and a threshold u we obtain, from Corollary 2.2, the asymptotic $100(1-\alpha)$ percent two-sided confidence intervals for η

$$\left(\widehat{\eta}_n^B - z_{1-\alpha/2}\sqrt{\frac{\widehat{\eta}_n^B((\widehat{\eta}_n^B\widehat{\sigma}_n)^2-1)}{\sum_{i=1}^{k_n}\widetilde{N}_{r_n}^{(i)}(u)}},\ \widehat{\eta}_n^B + z_{1-\alpha/2}\sqrt{\frac{\widehat{\eta}_n^B((\widehat{\eta}_n^B\widehat{\sigma}_n)^2-1)}{\sum_{i=1}^{k_n}\widetilde{N}_{r_n}^{(i)}(u)}}\right), \tag{2.13}$$

where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ standard normal quantile, $\widehat{\eta}_n^B = \frac{\sum_{i=1}^{k_n}\mathbb{1}_{\{\widetilde{N}_{r_n}^{(i)}(u)>0\}}}{\sum_{i=1}^{k_n}\widetilde{N}_{r_n}^{(i)}(u)}$ and $\widehat{\sigma}_n^2 = \frac{\sum_{i=1}^{k_n}(\widetilde{N}_{r_n}^{(i)}(u))^2}{\sum_{i=1}^{k_n}\mathbb{1}_{\{\widetilde{N}_{r_n}^{(i)}(u)>0\}}}$.

The intervals (2.13) are approximations of the true confidence intervals for finite samples.

The finite sample properties of $\widehat{\eta}_n^B$ are now analyzed in simulated samples from three different processes with $\eta < 1$ to $\eta = 1$. The results for all examples were obtained from 5000 simulated sequences of size $n = 5000$, of each process. In each example, the tables contain Monte-Carlo approximations to biases and root mean squared errors (RMSE), in brackets, of the estimator $\widehat{\eta}_n^B(u)$ and the 95% confidence intervals based on the asymptotic Normal approximation for the blocks estimates (2.13), for different values of r_n and different thresholds corresponding to the range between the 0.7 and 0.99 empirical quantiles, $q_\bullet$. Note that the confidence intervals (2.13) can only be computed when $\widehat{\sigma}_n^2 > 1/(\widehat{\eta}_n^B)^2$.

Example 2.1 (MM process) Let $X_n = \max\{Y_n,\ Y_{n-2},\ Y_{n-3}\},\ n \geq 1$, denote the moving maxima process, where $\{Y_n\}_{n\geq -2}$ is a sequence of independent and uniformly distributed on $[0,1]$ variables, considered in Example 1.1.

r_n	$q_{0.7}$	$q_{0.8}$	$q_{0.9}$	$q_{0.95}$	$q_{0.99}$
5	0.1983 (0.1986)	0.1956 (0.1960)	0.1964 (0.1972)	0.1983 (0.2000)	0.2001 (0.2093)
	(0.6775,0.7191)	(0.6716,0.7196)	(0.6641,0.7287)	(0.6536,0.7430)	(0.6029,0.7974)
7	0.0991 (0.0997)	0.1081 (0.1088)	0.1228 (0.1240)	0.1322 (0.1345)	0.1402 (0.1514)
	(0.5788,0.6195)	(0.5850,0.6312)	(0.5927,0.6530)	(0.5916,0.6729)	(0.5546,0.7258)
10	-0.0034 (0.0102)$^+$	0.0215 (0.0244)	0.0561 (0.0583)	0.0770 (0.0801)	0.0948 (0.1069)
	(0.4765,0.5167)	(0.4984,0.5446)	(0.5266,0.5856)	(0.5389,0.6151)	(0.5208,0.6687)
15	-0.1154 (0.1157)	-0.0726 (0.0733)	-0.0126 (0.0196)$^+$	0.0250 (0.0326)	0.0583 (0.0727)
	(0.3658,0.4035)	(0.4042,0.4506)	(0.4570,0.5178)	(0.4873,0.5627)	(0.4960,0.6207)
25	-0.2418 (0.2419)	-0.1882 (0.1883)	-0.0995 (0.1004)	-0.0369 (0.0420)	0.0243 (0.0451)
	(0.2430,0.2733)	(0.2906,0.3331)	(0.3688,0.4322)	(0.4229,0.5032)	(0.4699,0.5787)
50	-0.3652 (0.3652)	-0.3228 (0.3228)	-0.2246 (0.2249)	-0.1300 (0.1314)	-0.0136 (0.0385)
	(0.1261,0.1436)	(0.1625,0.1919)	(0.2461,0.3047)	(0.3261,0.4138)	(0.4296,0.5432)
100	-0.4324 (0.4324)	-0.4095 (0.4095)	-0.3403 (0.3403)	-0.2400 (0.2404)	-0.0568 (0.0693)
	(0.0632,0.0719)	(0.0826,0.0984)	(0.1392,0.1801)	(0.2193,0.3007)	(0.3713,0.5152)
1000	-0.4932 (0.4932)	-0.4909 (0.4909)	-0.4836 (0.4836)	-0.4685 (0.4685)	-0.3476 (0.3478)
	(0.0064,0.0071)	(0.0084,0.0097)	(0.0145,0.0183)	(0.0260,0.0369)	(0.0946,0.2102)

Table 1: Bias and root mean squared error (RMSE), in brackets, of the estimator $\widehat{\eta}_n^B(u)$ (first line) and the 95% confidence intervals based on the asymptotic Normal approximation for the blocks estimates (2.13) (second line), for an MM process with $\theta = 1/3$ and $\eta = 1/2$. Best values are ticked with a plus.

Example 2.2 (AR(1) process) Let $X_n = -\frac{1}{\beta}X_{n-1} + \epsilon_n,\ n \geq 1$, denote a negatively correlated autoregressive process of order one, where $\{\epsilon_n\}_{n\geq 1}$ is a sequence of i.i.d. random variables, such that, for a fixed integer $\beta \geq 2$, $\epsilon_n \sim U\{\frac{1}{\beta}, \ldots, \frac{\beta-1}{\beta}, 1\}$ and $X_0 \sim U(0,1)$ independent of ϵ_n.

Condition $\widetilde{D}^{(3)}(u_n)$ holds for this stationary sequence, it has extremal index $\theta = 1 - 1/\beta^2$ and upcrossings index $\eta^{(1,\mathbf{X})} = 1 - 1/\beta^2$, $\beta \geq 2$ (see Sebastião *et al.* (2010) [23]).

r_n	$q_{0.7}$	$q_{0.8}$	$q_{0.9}$	$q_{0.95}$	$q_{0.99}$
5	-0.1290 (0.1291)	0.0143 (0.0170)	0.0881 (0.0894)	0.1008 (0.1030)	0.1031 (0.1128)
	(0.6064,0.6355)	(0.7438,0.7848)	(0.8092,0.8670)	(0.8105,0.8910)	(0.7659,0.9402)
7	-0.2834 (0.2834)	-0.1245 (0.1248)	0.0098 (0.0190)$^+$	0.0566 (0.0611)	0.0747 (0.0902)
	(0.4547,0.4784)	(0.6052,0.6457)	(0.7279,0.7916)	(0.7617,0.8514)	(0.7281,0.9213)
10	-0.4176 (0.4176)	-0.2732 (0.2732)	-0.0874 (0.0888)	0.0009 (0.0235)$^+$	0.0497 (0.0730)
	(0.3237,0.3411)	(0.4591,0.4944)	(0.6290,0.6962)	(0.7022,0.7996)	(0.6964,0.9030)
15	-0.5278 (0.5278)	-0.4197 (0.4197)	-0.2139 (0.2143)	-0.0776 (0.0810)	0.0220 (0.0605)
	(0.2165,0.2279)	(0.3172,0.3435)	(0.5037,0.5685)	(0.6211,0.7237)	(0.6635,0.8805)
25	-0.6166 (0.6166)	-0.5500 (0.5500)	-0.3762 (0.3763)	-0.1994 (0.2004)	-0.0161 (0.0598)
	(0.1300,0.1367)	(0.1919,0.2081)	(0.3471,0.4004)	(0.4996,0.6015)	(0.6197,0.8482)
50	-0.6833 (0.6833)	-0.6500 (0.6500)	-0.5508 (0.5508)	-0.3889 (0.3891)	-0.0881 (0.1041)
	(0.0650,0.0683)	(0.0960,0.1040)	(0.1834,0.2149)	(0.3197,0.4025)	(0.5414,0.7823)
100	-0.7167 (0.7167)	-0.7000 (0.7000)	-0.6500 (0.6500)	-0.5519 (0.5519)	-0.1983 (0.2041)
	(0.0325,0.0341)	(0.0480,0.0520)	(0.0921,0.1080)	(0.1727,0.2234)	(0.4311,0.6723)
1000	-0.7467 (0.7467)	-0.7450 (0.7450)	-0.7400 (0.7400)	-0.7300 (0.7300)	-0.6500 (0.6500)
	(0.0033,0.0034)	(0.0048,0.0052)	(0.0093,0.0107)	(0.0178,0.0223)	(0.0712,0.1288)

Table 2: Bias and root mean squared error (RMSE), in brackets, of $\widehat{\eta}_n^B(u)$ (first line) and the 95% confidence intervals based on the asymptotic Normal approximation for the blocks estimates (2.13) (second line), for a negatively correlated AR(1) process with $\beta = 2$ and $\theta = \eta = 0.75$. Best values are ticked with a plus.

Example 2.3 (MAR(1) process) Let $X_n = \alpha \max\{X_{n-1},\ \epsilon_n\},\ n \geq 1,\ 0 < \alpha < 1$, denote the max-autoregressive process of order one where X_0 is a random variable with d.f. H_0, independent of the sequence $\{\epsilon_n\}_{n\geq 1}$ of i.i.d. random variables with unit Fréchet distribution. This process, which is a special case of the general MARMA(p,q) processes introduced by Davis and Resnick (1989) [4], verifies condition $D''(u_n)$, has extremal index $\theta = 1 - \alpha,\ 0 < \alpha < 1$ (see Alpuim (1989) [2] and Ferreira (1994) [5]) and upcrossings index $\eta = 1$.

r_n	$q_{0.7}$	$q_{0.8}$	$q_{0.9}$	$q_{0.95}$	$q_{0.99}$
5	-0.0061 (0.0092)	-0.0039 (0.0075)	-0.0019 (0.0065)	-0.0008 (0.0056)$^+$	-0.0003 (0.0075)$^+$
	(0.9843,1.0035)	(0.9893,1.0029)	(0.9947,1.0016)	(0.9978,1.0007)	(0.9992,1.0002)
7	-0.0143 (0.0177)	-0.0096 (0.0140)	-0.0048 (0.0110)	-0.0024 (0.0100)	-0.0009 (0.0122)
	(0.9678,1.0036)	(0.9759,1.0048)	(0.9871,1.0033)	(0.9934,1.0017)	(0.9982,1.0001)
10	-0.0305 (0.0341)	-0.0202 (0.0250)	-0.0100 (0.0174)	-0.0050 (0.0148)	-0.0013 (0.0149)
	(0.9413,0.9976)	(0.9546,1.0049)	(0.9736,1.0064)	(0.9861,1.0039)	(0.9968,1.0006)
15	-0.0659 (0.0695)	-0.0446 (0.0494)	-0.0232 (0.0317)	-0.0112 (0.0236)	-0.0029 (0.0214)
	(0.8935,0.9747)	(0.9161,0.9948)	(0.9446,1.0090)	(0.9702,1.0074)	(0.9931,1.0012)
25	-0.1495 (0.1525)	-0.1044 (0.1090)	-0.0549 (0.0636)	-0.0280 (0.0427)	-0.0055 (0.0301)
	(0.7946,0.9063)	(0.8378,0.9534)	(0.8887,1.0015)	(0.9303,1.0138)	(0.9866,1.0025)
50	-0.3521 (0.3539)	-0.2624 (0.2658)	-0.1466 (0.1542)	-0.0766 (0.0924)	-0.0153 (0.0515)
	(0.5837,0.7122)	(0.6621,0.8130)	(0.7653,0.9416)	(0.8385,1.0082)	(0.9629,1.0065)
100	-0.6059 (0.6066)	-0.4966 (0.4982)	-0.3123 (0.3177)	-0.1751 (0.1893)	-0.0365 (0.0838)
	(0.3447,0.4435)	(0.4328,0.5740)	(0.5818,0.7936)	(0.6988,0.9510)	(0.9148,1.0122)
1000	-0.9591 (0.9592)	-0.9431 (0.9431)	-0.8932 (0.8933)	-0.7921 (0.7932)	-0.3382 (0.3843)
	(0.0363,0.0454)	(0.0487,0.0651)	(0.0835,0.1301)	(0.1420,0.2738)	(0.4526,0.8711)

Table 3: Bias and root mean squared error (RMSE), in brackets, of $\widehat{\eta}_n^B(u)$ (first line) and the 95% confidence intervals based on the asymptotic Normal approximation for the blocks estimates (2.13) (second line), for a MAR(1) process with $\alpha = 0.9$, $\theta = 0.1$ and $\eta = 1$. Best values are ticked with a plus.

Absolute biases tend to be smallest at small block sizes, *i.e.*, $r_n = 5$, 7, 10 and 15. Furthermore, the

results suggest that generally speaking the absolute bias of $\widehat{\eta}_n^B(u)$ increases with r_n but decreases with u and the variance of $\widehat{\eta}_n^B(u)$ increases with u but decreases with r_n. The quality of the estimate depends strongly on the choice of the two parameters u and r_n, as previously stated. Nevertheless, the blocks estimator has a better performance at the 90% and the 95% thresholds. The poor performance at the 99% threshold is justified by the few observed threshold upcrossings.

Remark 3 *We point out from the previous examples that the stronger (weaker) the dependence between upcrossings, i.e. the bigger (smaller) clustering of upcrossings, which implies the smaller (bigger) η, the bigger (smaller) values of r_n are required.*

Remark 4 *In the estimation of the extremal index the block size r_n has commonly been chosen as the square root of the sample size, i.e. $r_n \approx \sqrt{n}$ (Gomes (1993) [11], Ancona-Navarrete and Tawn (2000) [1], among others). The previous examples show that such a large block size is not a reasonable choice in the estimation of the upcrossings index. Choices ranging from $r_n = [n/(\log n)^{3.2}]$ to $r_n = [n/(\log n)^{2.7}]$ seem more reasonable.*

2.1.1.2 Some considerations about the choice of the levels

In practical applications the deterministic levels v_n previously considered will have to be estimated, *i.e.*, they will have to be replaced by random levels suggested by the relation $\frac{n}{c_n}P(X_1 \leq v_n < X_2) \sim \nu$, as $n \to +\infty$. Nevertheless, these random levels cannot be represented by an appropriate order statistics, contrarily to the random levels used in the estimation of the extremal index.

As in Sebastião *et al.* (2013) [24] we shall consider the random levels $\widehat{v}_n = X_{n-[c_n\tau]:n}$, used in the estimation of the extremal index, more precisely, we shall consider the blocks estimator

$$\widetilde{\eta}_n^B(\widehat{v}_n) := \frac{\sum_{i=1}^{k_n} \mathbb{1}_{\{\widetilde{N}_{r_n}^{(i)}(\widehat{v}_n)>0\}}}{\sum_{i=1}^{k_n} \widetilde{N}_{r_n}^{(i)}(\widehat{v}_n)} = \left(\sum_{j=1}^{+\infty} j\widehat{\widetilde{\pi}}_n(j;\widehat{v}_n)\right)^{-1}. \tag{2.14}$$

For these random levels $\widehat{v}_n$, we show in the following result that the blocks estimator $\widetilde{\eta}_n^B(\widehat{v}_n)$ is also a consistent estimator of η.

Theorem 2.4 *Suppose that for each $\nu > 0$ there exists $v_n = u_{n/c_n}^{(\nu)} = u_{n/c_n}^{(\tau)}$ ($\frac{n}{c_n}P(X_1 \leq v_n < X_2) \sim \nu$ and $\frac{n}{c_n}P(X_1 > v_n) \sim \tau$ as $n \to +\infty$) for some $\tau > 0$, the conditions of Theorem 2.2 hold for all $\nu > 0$, then*

$$\widehat{\eta}_n^B(\widehat{v}_n) \xrightarrow[n\to+\infty]{} \eta.$$

Remark 5 *If $\mathbf{X}$ has extremal index $\theta > 0$ and upcrossings index $\eta > 0$, then there exists $v_n = u_{n/c_n}^{(\nu)} = u_{n/c_n}^{(\tau)}$ and $\eta\nu = \theta\tau$ (Ferreira (2006) [6]).*

The high level v_n, considered in Theorem 2.4, must be such that $nP(X_1 > v_n) = c_n\tau = \tau_n$, $\tau_n \xrightarrow[n\to+\infty]{} +\infty$ and $\tau_n/n \xrightarrow[n\to+\infty]{} 0$. In the simulation study, presented further ahead, we shall consider the deterministic levels $u_n = X_{n-s:n}$, corresponding to the $(s+1)$th top order statistics associated

to a random sample $(X_1, \ldots, X_n)$ of $\mathbf{X}$. The upcrossings index estimator will become in this way a function o s. In fact, s is replacing $\tau_n = c_n\tau$, $c_n \xrightarrow[n\to+\infty]{} +\infty$. Consistency is attained only if s is intermediate, *i.e.*, $s \to +\infty$ and $s = o(n)$ as $n \to +\infty$.

The choice of s, the number of top order statistics to be used in the estimation procedure is a sensitive and complex problem in extreme value applications. In Neves *et al.* (2014) ([19]) an heuristic algorithm is used for the choice of the optimal sample fraction in the estimation of the extremal index. The optimal sample fraction is chosen by looking at the sample path of the estimator of the extremal index as a function of s and finding the largest range of s values associated with stable estimates. Since the estimates of the extremal index and the upcrossings index sometimes do not have the largest stability region around the target value, this algorithm may give misleading results.

2.1.2 Short note on other blocks estimators

If the stationary sequence $\mathbf{X}$ also has extremal index $\theta > 0$ and there exists a sequence of thresholds $\{u_{r_n}\}_{r_n\geq 1}$ such that $\lim\limits_{r_n\to+\infty} r_nP(X_1 > u_{r_n}) = \tau$ and $\lim\limits_{r_n\to+\infty} r_nP(X_1 \leq u_{r_n} < X_2) = \nu$ then, since (1.4) holds, it follows that

$$\eta_{r_n}(u_{r_n}) = -\frac{\log F_{r_n}(u_{r_n})}{r_nP(X_1 \leq u_{r_n} < X_2)} \xrightarrow[r_n\to+\infty]{} \eta, \tag{2.15}$$

where $M_{r_n} = \max\{X_1, \ldots, X_{r_n}\}$ and $F_{r_n}(u_{r_n}) = P(M_{r_n} \leq u_{r_n})$.

The distribution function of the block maximum F_{r_n} in (2.15) can be estimated using maxima of $k_n = [n/r_n]$ disjoint blocks or using maxima of $n - r_n + 1$ sliding blocks, as suggested by Robert *et al.* (2009) [21], in the following ways

$$\widehat{F}_n^{dj}(u_{r_n}) = \frac{1}{k_n}\sum_{i=1}^{k_n} \mathbb{1}_{\{M_{(i-1)r_n+1,ir_n}\leq u_{r_n}\}}, \quad \widehat{F}_n^{sl}(u_{r_n}) = \frac{1}{n-r_n+1}\sum_{i=1}^{n-r_n+1} \mathbb{1}_{\{M_{i,r_n+i-1}\leq u_{r_n}\}}$$

where $M_{i,j} = \max\{X_i, \ldots, X_j\}$, for $0 < i < j$. Robert *et al.* (2009) [21] showed that $\widehat{F}_n^{dj}(u_{r_n}) - \widehat{F}_n^{sl}(u_{r_n})$ has a non-negligible asymptotic variance and is asymptotically uncorrelated with $\widehat{F}_n^{sl}(u_{r_n})$. Thus, the sliding blocks estimator is the most efficient convex combination of the disjoint and sliding blocks estimators for $F_{r_n}(u_{r_n})$.

We can now define a disjoint and a sliding blocks estimator for the upcrossings index η, respectively, as

$$\widehat{\eta}_n^{dj}(u_{r_n}) := -\frac{\log \widehat{F}_n^{dj}(u_{r_n})}{\frac{1}{k_n}\sum_{i=1}^{k_n} \widetilde{N}_{r_n}^{(i)}(u_{r_n})}, \tag{2.16}$$

and

$$\widehat{\eta}_n^{sl}(u_{r_n}) := -\frac{\log \widehat{F}_n^{sl}(u_{r_n})}{\frac{1}{k_n}\sum_{i=1}^{k_n} \widetilde{N}_{r_n}^{(i)}(u_{r_n})}. \tag{2.17}$$

The estimators $\widehat{\eta}_n^{dj}$ and $\widehat{\eta}_n^{sl}$ in (2.16) and (2.17) are not defined when all blocks have at least one upcrossing and $\widehat{\eta}_n^{dj}$ has already appeared in Sebastião *et al.* (2013) [24].

Remark 6 *If for each $\nu > 0$ there exists $u_n^{(\nu)}$ and $u_n^{(\nu)} = u_n^{(\tau)}$ for some $\tau > 0$, then $P(\max\{X_1, \ldots, X_n\} \leq u_n^{(\tau)}) - P(\widetilde{N}_n(u_n^{(\nu)}) = 0) \xrightarrow[n\to+\infty]{} 0$ (Ferreira (2006) [6]). Therefore, for large n we can replace the*

probability in the numerator of (2.15) with $P(\widetilde{N}_{r_n}(u_{r_n})=0)$ and estimate it by its empirical counterpart. Nevertheless, estimator (2.16) creates synergies in the estimation of the pair $(\theta,\ \eta)$. We recall that the extremal index θ and the upcrossings index η provide different but complementary information concerning the occurrence of rare events.

Imposing on $\mathbf{X}$ a stronger condition than condition $\Delta(u_n)$, involving its maximal correlation coefficients, and considering the threshold sequence $\{u_{r_n}\}_{r_n\geq 1}$ deterministic, we have, *mutatis mutandis*, from the arguments used in Robert *et al.* (2009) [21], that $\widehat{\eta}_n^{dj}$ and $\widehat{\eta}_n^{sl}$ are consistent estimators of η. If in addition, there exists a constant $p>1$ such that $E\left[\widetilde{N}_{r_n}^{2p}(u_{r_n})\right]=O(1)$, as $n\to+\infty$, and $\sqrt{k_n}(\eta_{r_n}(u_{r_n})-\eta)\xrightarrow[r_n\to+\infty]{}0$, then from the central limit theorem for triangular arrays it can be shown that

$$\sqrt{k_n}\begin{pmatrix}\widehat{\eta}_n^{dj}(u_{r_n})-\eta\\ \widehat{\eta}_n^{sl}(u_{r_n})-\eta\end{pmatrix}\xrightarrow[r_n\to+\infty]{d}N(0,\mathbf{V})$$

where $\mathbf{V}=[v_{i,j}]_{2\times 2}$ is a symmetric matrix with

$$\begin{aligned} v_{11} &= \frac{\eta}{\nu}\left(\frac{e^{\eta\nu}}{\eta\nu}-\frac{1}{\eta\nu}+c^2-1\right),\\ v_{22}=v_{12} &= \frac{\eta}{\nu}\left(2\times\frac{e^{\eta\nu}-1}{\eta^2\nu^2}-\frac{1}{\eta\nu}+c^2-1\right)\end{aligned}$$

and $c^2=\frac{\sum_{j\geq 1}j^2\widetilde{\pi}(j;u_{r_n})-\eta^{-2}}{\eta^{-2}}$.

Remark 7 *Since $v_{22}\leq v_{11}$ the sliding blocks estimator $\widehat{\eta}_n^{sl}$ is more efficient than its disjoint version. On the other hand, considering that $c_n/k_n\to 0$, as $n\to+\infty$, the blocks estimator $\widehat{\eta}_n^{B}$ is a more efficient estimator than $\widehat{\eta}_n^{dj}$.*

2.2 Runs declsutering

Runs declustering assumes that upcrossings belong to the same cluster if they are separated by fewer than a certain number of non-upcrossings of the threshold. More precisely, if the process $\mathbf{X}$ verifies condition $\widetilde{D}^{(k)}(u_n)$, for some $k\geq 2$, of Ferreira (2006) [6], that locally restricts the dependence of the sequence but still allows clustering of upcrossings, runs of upcrossings in the same cluster must be separated at most by $k-2$ non-upcrossings. In Figure 1 we can identify 3 clusters of size 2, since the process verifies condition $\widetilde{D}^{(3)}(u_n)$.

Condition $\widetilde{D}^{(k)}(u_n)$, $k\geq 2$, is said to be satisfied if

$$\lim_{n\to+\infty} nP(X_1\leq u_n<X_2,\ \widetilde{N}_{3,k}=0,\ \widetilde{N}_{k+1,r_n}>0)=0, \tag{2.18}$$

for some sequence $r_n=[n/k_n]$ with $\{k_n\}_{n\geq 1}$ satisfying $k_n\xrightarrow[n\to+\infty]{}+\infty$, $\frac{k_nl_n}{n}\xrightarrow[n\to+\infty]{}0$, $k_n\alpha_{n,l_n}\xrightarrow[n\to+\infty]{}0$, where α_{n,l_n} are the mixing coefficients of the $\Delta(u_n)$ condition, and $\widetilde{N}_{i,j}\equiv\widetilde{N}_n(u_n)([i/n,j/n])$ with $\widetilde{N}_{i,j}=0$ if $j<i$.

This family of local dependence conditions is slightly stronger than $D^{(k)}(u_n)$ of Chernick *et al.* (1991) [3] and (2.18) is implied by

$$n \sum_{i=k+1}^{r_n-1} P(X_1 \leq u < X_2,\ X_i \leq u < X_{i+1}) \xrightarrow[n\to+\infty]{} 0.$$

When $k = 2$ we find the slightly weakened condition $D''(u_n)$ of Leadbetter and Nandagopalan (1989) [17].

Under this hierarchy of increasingly weaker mixing conditions $\widetilde{D}^{(k)}(u_n)$, $k \geq 2$, the upcrossings index can be computed as follows:

Proposition 2.1 (Ferreira (2006) [6]) *If* $\mathbf{X}$ *satisfies condition* $\Delta(u_n)$ *and for some* $k \geq 2$, *condition* $\widetilde{D}^{(k)}(u_n^{(\nu)})$ *holds for some* $\nu > 0$, *then the upcrossings index of* $\mathbf{X}$ *exists and is equal to* η *if and only if*

$$P(\widetilde{N}_{3,k}(u_n^{(\nu)}) = 0 \mid X_1 \leq u_n^{(\nu)} < X_2) \xrightarrow[n\to+\infty]{} \eta$$

for each $\nu > 0$.

2.2.1 The runs estimators

The estimators considered up to now treat each block as one cluster, but under a local dependence condition $\widetilde{D}^{(k)}(u_n)$, $k \geq 3$, clusters can be identified in a different way, for example, runs of upcrossings or runs of upcrossings separated by at most one non-upcrossings of a certain threshold may define a cluster. More precisely, if the process $\mathbf{X}$ verifies condition $\widetilde{D}^{(3)}(u_n)$, upcrossing clusters may be simply identified asymptotically as runs of consecutive upcrossings and the cluster sizes as run lengths (Ferreira (2007) [7]). Therefore, from Propositon 2.1 if for some $k \geq 3$ condition $\widetilde{D}^{(k)}(u_n)$ holds, the upcrossings index can be estimated by the ratio between the total number of $k-2$ non-upcrossings followed by an upcrossing and the total number of upcrossings, *i.e.*, the runs estimator given by

$$\widehat{\eta}_n^{R|k} = \widehat{\eta}_n^{R|k}(u_n) := \frac{\sum_{i=1}^{n-k} \mathbb{1}_{\{\widetilde{N}_{i,i+k-3}=0,\ X_{i+k-1}\leq u_n < X_{i+k}\}}}{\sum_{i=1}^{n-1} \mathbb{1}_{\{X_i \leq u_n < X_{i+1}\}}}, \tag{2.19}$$

with $\widetilde{N}_{i,j} \equiv \widetilde{N}_n(u_n)([i/n, j/n])$ and $\widetilde{N}_{i,j} = 0$ if $j < i$.

When $k = 3$ the runs estimator corresponds to the one fairly studied in Sebastião *et al.* (2013) [24]. There it was shown that the runs estimator has smaller bias and mean squared error when compared to the disjoint blocks estimator $\widehat{\eta}_n^{dj}$.

Properties such as consistency and asymptotic normality can be proved for the estimators $\widehat{\eta}_n^{R|k}$, $k \geq 4$, with similar arguments to the ones used in Sebastião *et al.* (2013) [24]. Nevertheless, their validation is restricted to the validation of a local dependence condition $\widetilde{D}^{(k)}(u_n)$, $k \geq 4$.

In what follows we propose a way of checking conditions $\widetilde{D}^{(k)}(u_n)$, $k \geq 3$, preliminary to estimation, which turns out to be the only possible solution when dealing with real data. Even when the underlying model is known, the verification of conditions $\widetilde{D}^{(k)}(u_n)$, $k \geq 3$, can be cumbersome, so the following procedure can in these situations be an important auxiliary tool. A similar approach has been followed by Süveges (2007) [25] to check condition $D^{(2)}(u_n)$ of Chernick *et al.* (1991) [3] which is slightly stronger than condition $D''(u_n)$ of Leadbetter and Nandagopalan (1989) [17].

Condition $\widetilde{D}^{(k)}(u_n)$, $k \geq 3$, essential for the validity of the runs estimator $\widehat{\eta}_n^{R|k}$, may be checked by calculating

$$p^{(k)}(u,r) = \frac{\sum_{i=1}^{n-r+1} \mathbb{1}_{\{X_i \leq u < X_{i+1},\ \widetilde{N}_{i+2,i+k-1}=0,\ \widetilde{N}_{i+k,i+r-1}>0\}}}{\sum_{i=1}^{n-1} \mathbb{1}_{\{X_i \leq u < X_{i+1}\}}} \tag{2.20}$$

for a sample $(X_1, \ldots, X_n)$ of $\mathbf{X}$, u a high threshold and r the size of the blocks into which the sample is partitioned.

The limit condition $\widetilde{D}^{(k)}(u_n)$, for some $k \geq 3$, will be satisfied if there exists a path (u_i, r_j) with $u_i \to +\infty$ and $r_j \to +\infty$ for which the $p^{(k)}(u_i, r_j) \to 0$. With (2.20) we are looking for the so called anti-$\widetilde{D}^{(k)}(u_n)$, $k \geq 3$, events $\{\widetilde{N}_{i+2,i+2} = 0,\ \widetilde{N}_{i+3,i+r-1} > 0 \mid X_i \leq u < X_{i+1}\}$ among the upcrossings for a range of thresholds and block sizes. This proportion of anti-$\widetilde{D}^{(k)}(u_n)$, $k \geq 3$, events gives us information on how many clusters are misidentified as two or more clusters and therefore information of the bias. If such clusters are few, it is plausible to accept the relatively small upward bias despite the possible failure of $\widetilde{D}^{(k)}(u_n)$, $k \geq 3$, and use $\widehat{\eta}_n^{R|k}$ to estimate η. This gives us a way to choose, in practical applications, among the estimators $\widehat{\eta}_n^{R|k}$, $k \geq 3$, the one to estimate η.

In Figures 2 and 3 we plot the proportions $p(u,r) \equiv p^{(3)}(u,r)$, considering u the empirical quantiles with probabilities ranging from 0.95 to 0.995 and r ranging from 5 to 20, for sequences of size $n = 10000$, of the processes considered in the examples of Section 2, of an AR(2) process and a GARCH(1,1) process with Student-t distributed innovations. For the sake of clarity we consider $\widehat{F}(u)$ instead of u in the plots.

Similar results were obtained for other sample sizes, namely $n = 5000$ and 30000. For the AR(2) process the proportions $p^{(4)}(u,r)$ and $p^{(5)}(u,r)$ did not differ much from $p^{(3)}(u,r)$ which suggest that this process also does not verify conditions $\widetilde{D}^{(4)}(u_n)$ and $\widetilde{D}^{(5)}(u_n)$. The anti-$\widetilde{D}^{(k)}(u_n)$, $k = 3,4,5,6$, proportions for the GARCH(1,1) process, plotted in Figure 3, show a more prominent decrease, towards zero, for $k = 5$ and $k = 6$. This suggests that conditions $\widetilde{D}^{(3)}(u_n)$ and $\widetilde{D}^{(4)}(u_n)$ are unlikely to hold for this process and that $k = 5$ seems a reasonable choice.

We recall that the conditions $\widetilde{D}^{(k)}(u_n)$, $k \geq 3$, form a hierarchy of increasingly weaker mixing conditions and therefore, if a process verifies condition $\widetilde{D}^{(k')}(u_n)$ for a fixed k', then it verifies all conditions $\widetilde{D}^{(k)}(u_n)$ with $k \geq k'$. Accordingly we can use any of the estimators $\widehat{\eta}_n^{R|k}$, $k = k', k'+1, \ldots$, to estimate η. We simulated for the estimators $\widehat{\eta}_n^{R|k}$, $k = 3,4,5$, the mean (E) and the root mean squared error (RMSE) from 5000 samples of sizes $n = 5000$ of the MM process given in Example 2.1. In Figure 4, we find the estimated values as functions of $s \geq 1$, corresponding to the estimates obtained with $\widehat{\eta}_n^{R|k}(X_{n-s:n})$, $k = 3,4,5$, where $X_{n-s:n}$ denotes the $(s+1)$th top order statistics.

As we can see, from Figure 4, for the MM process the estimator $\widehat{\eta}_n^{R|4}$ outperforms the estimators $\widehat{\eta}_n^{R|3}$ and $\widehat{\eta}_n^{R|5}$, improving the results found in Sebastião *et al.* (2013) [24]. The estimates have a big stability region near the true value $\eta = 0.5$ and the RMSE presents a desirable wide "bathtub" pattern. For the processes of Examples 2.2 and 2.3 the best results were obtained with $\widehat{\eta}_n^{R|3}$. The results show that even if we validate condition $\widetilde{D}^{(k)}(u_n)$ for some value $k \geq 3$ we might obtain better estimates for η with a runs estimator with a higher value of k. It is therefore convenient to plot sample paths of the runs estimator $\widehat{\eta}_n^{R|k}(X_{n-s:n})$ for different values of k, as a function of s, and analyze the stability regions. Note however that the choice of a too large k may reduce the precision of the new estimators, as can be seen by the

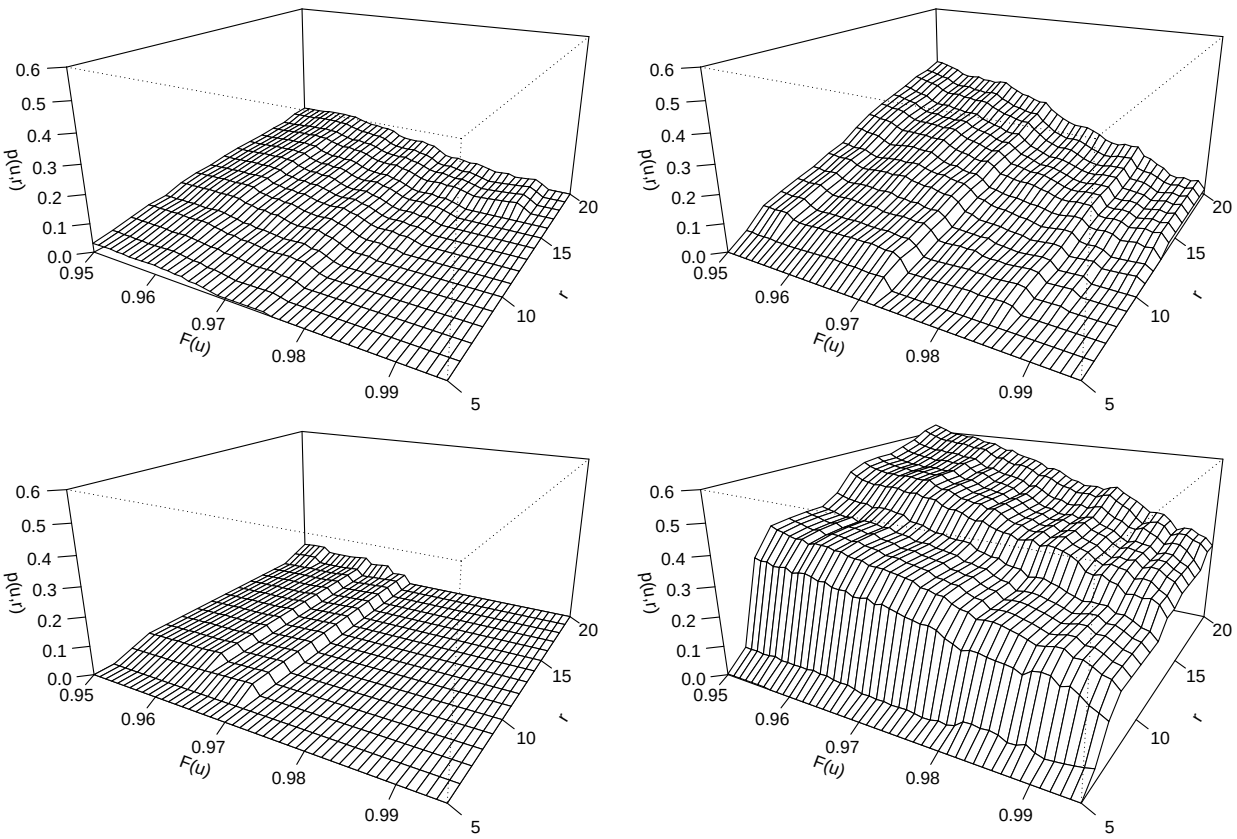

Figure 2: Observed proportions $p(u,r) \equiv p^{(3)}(u,r)$ for the MM process (top left panel), negatively correlated AR(1) process with $s = 2$ (top right panel), MAR(1) process (bottom left panel) and AR(2) process with $\phi_1 = 0.93$ and $\phi_2 = -0.86$ (bottom right panel).

values plotted in Figure 4.

3 Simulation study

We compare the performance of the several estimators previously presented for a range of different processes, having various dependence structures which give both $\eta < 1$ and $\eta = 1$

In this comparison study we include the estimator that naturally arises from relation (1.4), by replacing θ, τ and ν with corresponding consistent estimators. Such an estimator has already been considered in Sebastião *et al.* (2013) [24], where θ was estimated with Ferro and Seger's ([9]) intervals estimator, $\widehat{\theta}^{FS}$, and τ and ν by their empirical counterparts. This estimator is formally defined as follows

$$\widehat{\eta}_n^{EI} = \widehat{\eta}_n^{EI}(u_n) := \widehat{\theta}^{FS} \frac{\sum_{i=1}^{n} \mathbb{I}_{\{X_i \geq u_n\}}}{\sum_{i=1}^{n-1} \mathbb{I}_{\{X_i \leq u_n < X_{i+1}\}}}, \tag{3.21}$$

where

$$\widehat{\theta}^{FS} = \begin{cases} 1 \wedge \hat{\theta}_n^1(u_n) & , \quad \max\{T_i : \ 1 \leq i \leq N-1\} \leq 2 \\ 1 \wedge \hat{\theta}_n^2(u_n) & , \quad \max\{T_i : \ 1 \leq i \leq N-1\} > 2 \end{cases}$$

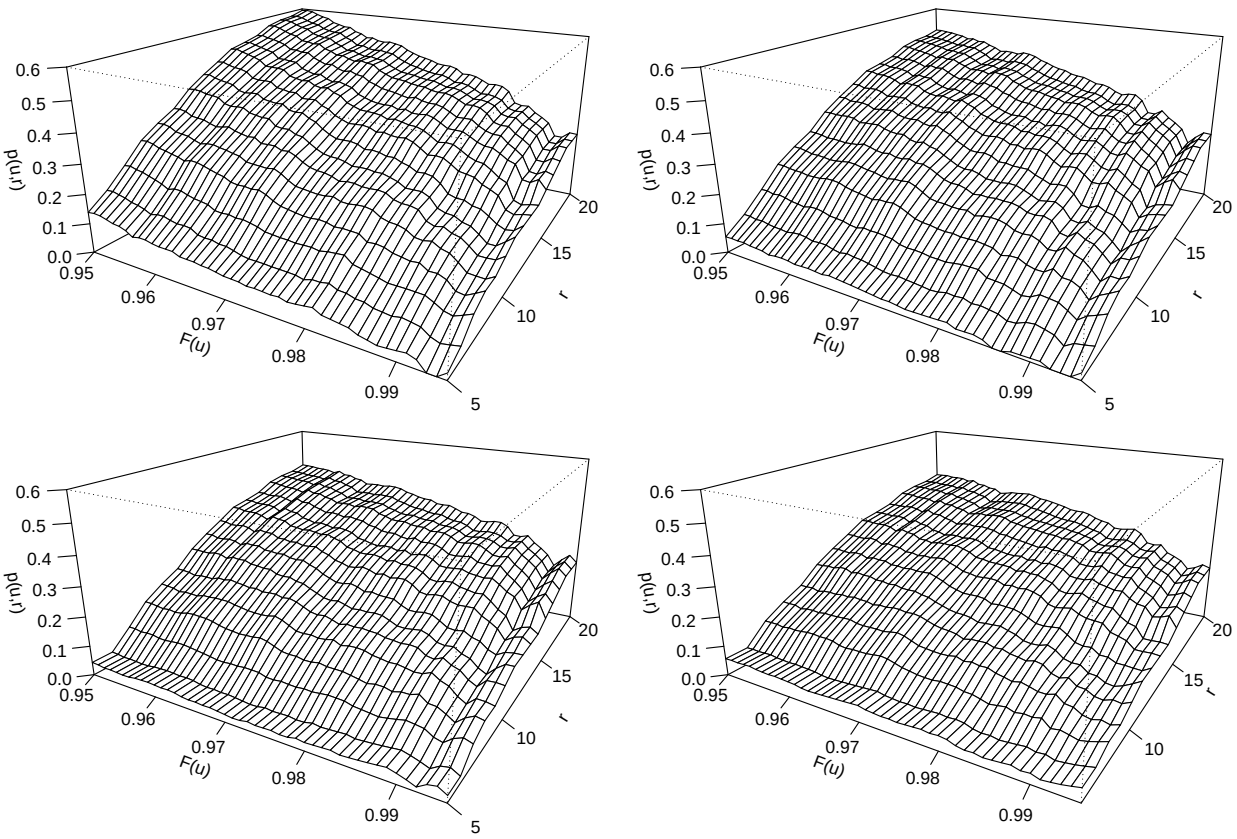

Figure 3: From left to right and top to bottom, observed proportions $p(u,r) \equiv p^{(k)}(u,r)$, $k = 3,4,5$ and 6 for a GARCH(1,1) process with autoregressive parameter $\alpha = 0.08$, variance parameter $\beta = 0.87$ and innovations t_7

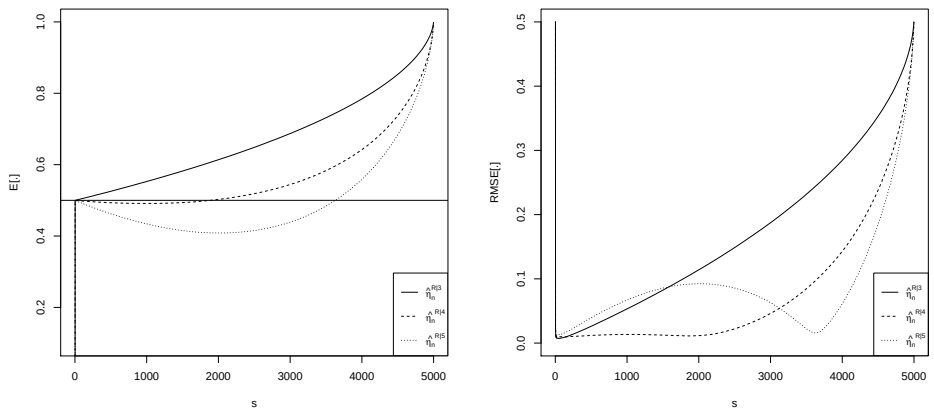

Figure 4: Mean values (left) and root mean squared errors (right) of the estimators $\widehat{\eta}_n^{R|k}$, $k = 3,4,5$, for the MM process in Example 2.1 ($\eta = 0.5$).

with

$$\hat{\theta}_n^1(u_n) = \frac{2\left(\sum_{i=1}^{N-1} T_i\right)^2}{(N-1)\sum_{i=1}^{N-1} T_i^2} \quad \text{and} \quad \hat{\theta}_n^2(u_n) = \frac{2\left\{\sum_{i=1}^{N-1}(T_i-1)\right\}^2}{(N-1)\sum_{i=1}^{N-1}(T_i-1)(T_i-2)},$$

denoting T_i, $i = 1, \ldots, N-1$, the observed interexceedance times, *i.e.*, $T_i = S_{i+1} - S_i$, where $1 \leq S_1 < \ldots < S_N \leq n$ are the exceedance times.

The different estimation methods are compared via the estimated bias and variance. Furthermore, to analyze the bias-variance trade-off we compare the corresponding root mean squared errors, RMSE. To measure the degree of possible misinterpretation caused by considering the variance/standard error as a criteria for the quality of the estimator we consider the mean error to standard errors MESE = Bias/$\sqrt{\text{Var}}$ (see Frahm *et al.* (2005) [10] for further details).

The values for the statistical quantities considered were obtained with 5000 independent replications of the estimation procedures for samples of size $n = 5000$. Since all estimators of η depend on a threshold, we considered all simulated values for each estimator computed at thresholds corresponding to the 90% and 95% empirical quantiles. To reduce sampling variation, for each process all the estimators are evaluated using the same simulated data sets.

For the blocks estimators we compare different block sizes, found to be the best choices from independent simulations to those reported here and from the results obtained in Tables 1-3.

The performance of the estimators for the processes given in Examples 2.1, 2.2 and 2.3 are presented, respectively, in Tables 4, 5 and 6.

For a better comparison of the blocks estimators $\widehat{\eta}_n^B$ with $r_n = 5, 10, 15$ and the runs estimators $\widehat{\eta}_n^{R|k}$, $k = 3, 4$, which had the best performances in the previous simulations, we plot in Figure 5 the estimated mean values and RMSE for levels given by the $(s+1)$th, $s = 1, ..., n-1$, top order statistics.

Some overall conclusions:

- In general, all estimation methods have small sample bias. For the three processes the blocks estimators $\widehat{\eta}_n^B$, $r_n = 5$, 10,15 and the runs estimators $\widehat{\eta}_n^{R|k}$, $k = 3, 4, 5$, lead to the smallest absolute sample biases. The biggest sample biases are obtained with the sliding blocks estimators $\widehat{\eta}_n^{sl}$ and the disjoint blocks estimators $\widehat{\eta}_n^{dj}$. The latter has a bit worse performance, except for the AR(1) process at the 90% threshold where the sliding blocks estimator and the naive estimator $\widehat{\eta}_n^{FS}$ perform worst.

- The sample variances for all estimators and processes are smaller than 0.05. In the group of the smallest observed variance values we find once again the the blocks estimators $\widehat{\eta}_n^B$, $r_n = 5$, 10,15 and the runs estimators $\widehat{\eta}_n^{R|k}$, $k = 3, 4, 5$. The conclusions drawn in Remark 7 can be pictured in Tables 4, 5 and 6. When η is on the boundary of the parameter space there is no clear overall change in efficiency of the estimators with threshold level.

- The conclusions previously drawn for the bias and variance carry over to the RMSE, for obvious reasons. For the MAR(1) process, with $\eta = 1$, we obtain the smallest values of the RMSE for all estimators, in fact the blocks estimator $\widehat{\eta}_n^B$, $r_n = 5$, and the runs estimator $\widehat{\eta}_n^{R|3}$ have a RMSE of 0.006.

- A large sample bias relative to the sample variance translates into a large MESE, which is true for the blocks estimators $\widehat{\eta}_n^B$, $r_n = 5, 10$, with the MM and the AR(1) processes. The runs estimators present in the majority of the cases values smaller than 1, which indicates that the true upcrossings index η lies within 1σ confidence band, where σ denotes the standard error.

Table 4: Simulation results for the estimators of η with the MM process of Example 2.1.

Estimator	Bias	Var	RMSE	MESE	Bias	Var	RMSE	MESE
	$q_{0.90}$				$q_{0.95}$			
$\widehat{\eta}_n^B$ $(r_n = 5)$	0.1964^-	0.00033	0.1972^-	10.8028^-	0.1983^-	0.00069	0.2000^-	7.5571^-
$\widehat{\eta}_n^B$ $(r_n = 10)$	0.0561	0.00025	0.0583	3.5137	0.0770	0.00049	0.0801	3.4779^-
$\widehat{\eta}_n^B$ $(r_n = 15)$	-0.0126^+	0.00022	0.0196^+	0.8459^+	0.0250	0.00044	0.0326	1.1946
$\widehat{\eta}_n^{dj}$ $(r_n = 50)$	0.1195	0.00404	0.1353	1.8797	0.0717	0.00228	0.0862	1.5002
$\widehat{\eta}_n^{dj}$ $(r_n = 60)$	0.1179	0.00602	0.1411	1.5204	0.0667	0.00286^-	0.0855	1.2471
$\widehat{\eta}_n^{dj}$ $(r_n = 70)$	0.1172	0.00898^-	0.1507^-	1.2366	0.0660	0.00361	0.0893	1.0990
$\widehat{\eta}_n^{sl}$ $(r_n = 50)$	0.1156	0.00202	0.1241	2.5719	0.0705	0.00129	0.0791	1.9663
$\widehat{\eta}_n^{sl}$ $(r_n = 60)$	0.1124	0.00283	0.1243	2.1139	0.0660	0.00164	0.0774	1.6305
$\widehat{\eta}_n^{sl}$ $(r_n = 70)$	0.1117	0.00396	0.1282	1.7748	0.0633	0.00205	0.0778	1.3993
$\widehat{\eta}_n^{R\|3}$	0.0255	0.00004^+	0.0262	4.0038^-	0.0126^+	0.00004^+	0.0141^+	1.9755
$\widehat{\eta}_n^{R\|4}$	-0.0067^+	0.00009^+	0.0117^+	0.6892^+	-0.0039^+	0.00009^+	0.0104^+	0.4029^+
$\widehat{\eta}_n^{R\|5}$	-0.0376	0.00013	0.0394	3.2787	-0.0202	0.00014	0.0234	1.7051
$\widehat{\eta}_n^{FS}$	0.0506	0.00183	0.0663	1.1827	0.0300	0.00308^-	0.0631	0.5415^+

Best values are ticked with a plus and worst are ticked with a minus.

- All but the blocks estimator $\widehat{\eta}_n^B$ have a better performance the higher the threshold used, since in essence they are estimating the non-upcrossing or non-exceedance of the threshold.
- Estimators obtained from asymptotic characterizations of the upcrossings index have a better performance than estimators obtained from the relation with the extremal index.
- The good performance of the blocks estimator $\widehat{\eta}_n^B$ makes it the best alternative for the runs estimators $\widehat{\eta}_n^{R|k}$ which need the validation of local dependence conditions. This statement is reinforced by Figure 5, where we see that for an adequate block size the estimated mean of the blocks estimator presents a well defined stability region near the true value and the RMSE has a wide "bathtub" shape.

Table 5: Simulation results for the estimators of η with the AR(1) process with $\beta = 2$ of Example 2.2.

Estimator	Bias	Var	RMSE	MESE	Bias	Var	RMSE	MESE
	$q_{0.90}$				$q_{0.95}$			
$\widehat{\eta}_n^B$ $(r_n = 5)$	0.0881	0.00022	0.0894	5.9123	0.1008	0.00045^+	0.1030	4.7764^-
$\widehat{\eta}_n^B$ $(r_n = 10)$	-0.0874	0.00024	0.0888	5.6748	0.0009	0.00055^+	0.0235^+	0.0399
$\widehat{\eta}_n^B$ $(r_n = 15)$	-0.2139	0.00015^+	0.2143	17.6109^-	-0.0776	0.00055^+	0.0810	3.3176^-
$\widehat{\eta}_n^{dj}$ $(r_n = 50)$	0.1404	0.00393	0.1538	2.2390	0.1974	0.01331	0.2286	1.7104
$\widehat{\eta}_n^{dj}$ $(r_n = 60)$	-0.0221	0.00095	0.0380	0.7167	0.2093^-	0.02289^-	0.2582^-	1.3834
$\widehat{\eta}_n^{dj}$ $(r_n = 70)$	-0.1428	0.00020^+	0.1434	10.1522	0.2048	0.02340^-	0.2556^-	1.3385
$\widehat{\eta}_n^{sl}$ $(r_n = 50)$	0.3980^-	0.04007^-	0.4455^-	1.9883	0.1890	0.00698	0.2066	2.2622
$\widehat{\eta}_n^{sl}$ $(r_n = 60)$	0.3100^-	0.02936^-	0.3542^-	1.8094	0.1949	0.01014	0.2194	1.9349
$\widehat{\eta}_n^{sl}$ $(r_n = 70)$	0.1953	0.01983	0.2408	1.3870	0.2050^-	0.01545	0.2397	1.6490
$\widehat{\eta}_n^{R\|3}$	0.0006^+	0.00039	0.0198^+	0.0279^+	0.0007	0.00077	0.0278^+	0.0255
$\widehat{\eta}_n^{R\|4}$	0.0004^+	0.00039	0.0198^+	0.0193^+	0.0006^+	0.00077	0.0278^+	0.0208^+
$\widehat{\eta}_n^{R\|5}$	-0.0622	0.00044	0.0657	2.9574	0.0004^+	0.00077	0.0278^+	0.0159^+
$\widehat{\eta}_n^{FS}$	0.2424	0.00040	0.2432	12.0808^-	0.1566	0.00394	0.1687	2.4965

Best values are ticked with a plus and worst are ticked with a minus.

4 Application to financial data

We consider the performance of the estimators under study which have previously better performed, *i.e.* the blocks estimator $\widehat{\eta}_n^B$ and the runs estimators $\widehat{\eta}_n^{R|k}$, $k \geq 3$, when applied to the analysis of the log-returns associated with the daily closing prices of the German stock market index DAX, collected from 3 January 2000 to 29 June 2012. This series was analysed previously by Sebastião *et al.* (2013) [24] and is plotted in Figure 6: DAX daily closing prices over the mentioned period, x_t, and the log-returns, $100 \times (\ln x_t - \ln x_{t-1})$, the data to be analyzed, after removing null log-returns ($n = 3177$).

As stated in Sebastião *et al.* (2013) [24], Klar et al. (2012) [20] have analyzed the DAX German stock market index time series and concluded that the GARCH(1,1) process with Student-t distributed innovations is a good model to describe these data. The simulated values of the anti-$\widetilde{D}^{(k)}(u_n)$, $k = 3, 4, 5, 6$ proportions found in Figure 3 for the GARCH(1,1) process and the conclusions there taken led us to consider the runs estimators $\widehat{\eta}^{R|k}$ with $k = 4$, 5 and 6. For the blocks estimator, the choice of $r_n = 5$, 7 and 10 seemed adequate for the sample size considered.

Table 6: Simulation results for the estimators of η with the MAR(1) process with $\alpha = 0.9$ of Example 2.3.

Estimator	Bias	Var	RMSE	MESE	Bias	Var	RMSE	MESE
	$q_{0.90}$				$q_{0.95}$			
$\widehat{\eta}_n^B$ $(r_n = 5)$	-0.0019^+	0.00004^+	0.0065^+	0.3006^+	-0.0008^+	0.00003^+	0.0056^+	0.1420^+
$\widehat{\eta}_n^B$ $(r_n = 10)$	-0.0100	0.00020	0.0174	0.7001	-0.0050	0.00019	0.0148	0.3583
$\widehat{\eta}_n^B$ $(r_n = 15)$	-0.0232	0.00046	0.0317	1.0804	-0.0112	0.00043	0.0236	0.5397
$\widehat{\eta}_n^{dj}$ $(r_n = 50)$	0.3304^-	0.01156	0.3474^-	3.0723	0.2586	0.01341	0.2834	2.2335
$\widehat{\eta}_n^{dj}$ $(r_n = 60)$	0.2986	0.01363	0.3206	2.5580	0.2248	0.01387	0.2538^-	1.9086
$\widehat{\eta}_n^{dj}$ $(r_n = 70)$	0.2759	0.01526^-	0.3023	2.2332	0.2066	0.01429^-	0.2386	1.7283
$\widehat{\eta}_n^{sl}$ $(r_n = 50)$	0.3270^-	0.00591	0.3359^-	4.2524^-	0.2552^-	0.00594	0.2666^-	3.3119^-
$\widehat{\eta}_n^{sl}$ $(r_n = 60)$	0.2949	0.00687	0.3063	3.5595^-	0.2241	0.00626	0.2377	2.8332
$\widehat{\eta}_n^{sl}$ $(r_n = 70)$	0.2726	0.00806	0.2870	3.0372	0.2028	0.00695	0.2192	2.4321
$\widehat{\eta}_n^{R\|3}$	-0.0013^+	0.00003^+	0.0055^+	0.2518^+	-0.0009^+	0.00003^+	0.0060^+	0.1451^+
$\widehat{\eta}_n^{R\|4}$	-0.0035	0.00007	0.0091	0.4140	-0.0021	0.00009^+	0.0095	0.2250
$\widehat{\eta}_n^{R\|5}$	-0.0066	0.00014	0.0135	0.5666	-0.0036	0.00014	0.0125	0.3019
$\widehat{\eta}_n^{FS}$	0.0324	0.01645^-	0.1323	0.2523	0.0426	0.02779^-	0.1720	0.2558

Best values are ticked with a plus and worst are ticked with a minus.

Clusters of upcrossings of high returns can bring serious risk to financial investors so estimates of the upcrossings index for the log-returns are plotted against $s \geq 1$ in Figure 7, in a linear scale and in logarithmic scale to better observe estimates at high levels. Note that Figure 7 also enhances the effect that the block size has on the estimates.

The sample paths of all estimators have a stability region around the value $\eta = 0.8$, which can be identified in the plot of the estimates in a logarithmic scale of Figure 7. The longest stability region is obtained with the blocks estimator for $r_n = 7$. The higher value of η obtained with the runs estimator $\widehat{\eta}^{R|4}$ comparatively to the ones obtained with the runs estimators $\widehat{\eta}^{R|5}$ and $\widehat{\eta}^{R|6}$ can be justified by the fact that condition $\widetilde{D}^{(4)}$ does not hold and therefore runs of upcrossings are over evaluated.

A value of η significatively less than 1 shows certain short range dependence reflected in the clustering of high level upcrossings. This may be interpreted as a pattern in the occurrence of the extreme values, that is, once a large profit in the asset return has occurred it will be followed by a loss (negative return) and we can expect a period of large profits followed by losses (values crossing some threshold). The

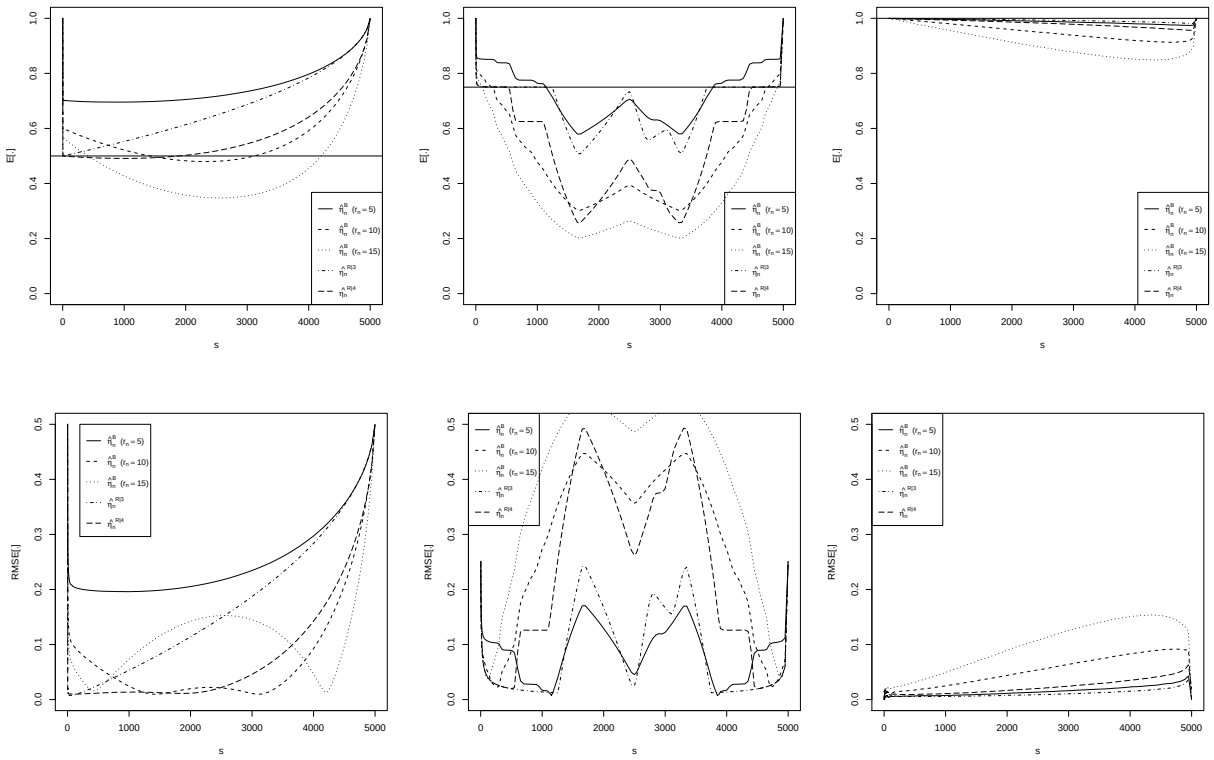

Figure 5: From left to right, mean values (top) and root mean squared errors (bottom) of the estimators $\widehat{\eta}_n^B$ with $r_n = 5, 10, 15$ and $\widehat{\eta}_n^{R|k}$, $k = 3, 4$, for the MM process in Example 2.1 ($\eta = 0.5$), the AR(1) process with $\beta = 2$ in Example 2.2 ($\eta = 0.75$) and the MAR(1) process with $\alpha = 0.9$ in Example 2.3 ($\eta = 1$).

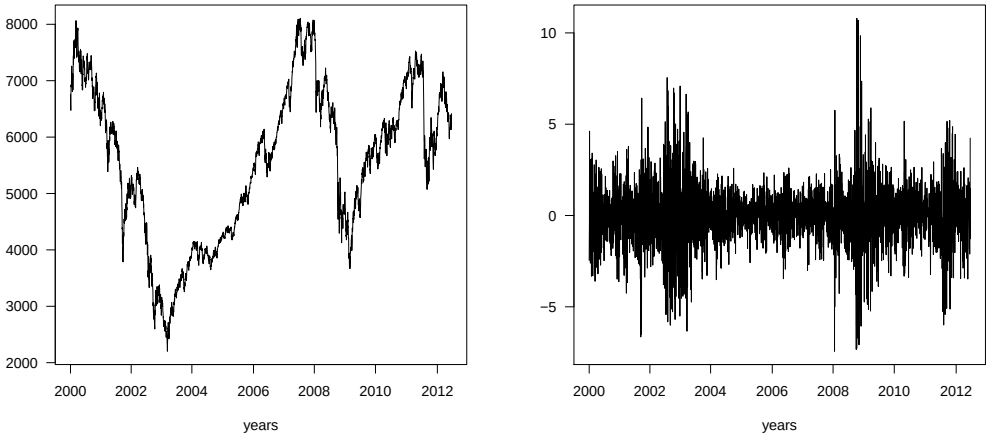

Figure 6: DAX daily closing prices (left) and daily log-returns (right), of DAX from 2000 to 2012, with 3177 observations (successive equal prices excluded).

average length of this period being the inverse of the upcrossings index.

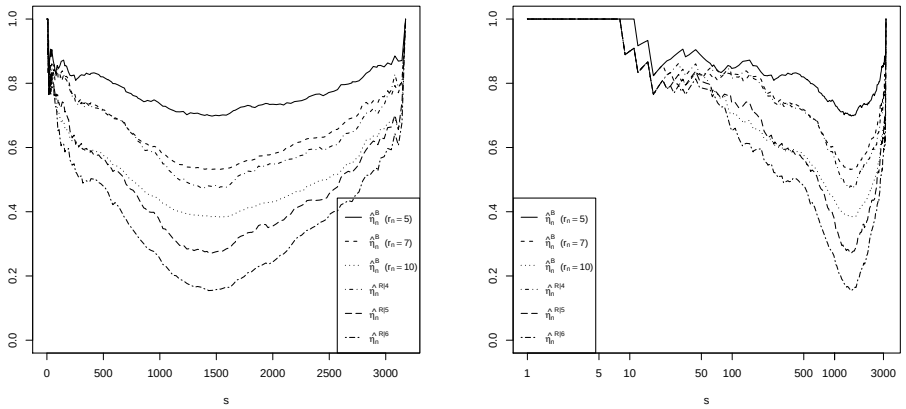

Figure 7: Estimates of the upcrossings index for the daily DAX log returns plotted against $s \geq 1$, in a linear scale (left) and in a logarithmic scale (right).

5 Conclusions

New blocks estimators for the upcrossings index η, a measure of the clustering of upcrossings of high thresholds, have been proposed in this paper. These estimators have the advantage of not requiring the validation of a local dependence condition, contrarily to the known runs estimator of the upcrossings index. The study of the properties of these blocks estimators, namely their consistency and asymptotic normality, has been carried out in Section 2. The family of runs estimators of η has been enlarged and an empirical way of checking local dependence conditions that control the clustering of upcrossings provided.

The estimators proposed in this paper improve substantially on existing methods of estimating the upcrossings index. Nevertheless, the simulations performed, in Sections 3 and 4, tell us that there is still space for improvements. In future, a discussion on the optimal choice of the key parameter r_n has to be considered as well as bias reduction techniques, namely the generalized Jackknife methodology.

Appendix A: Proofs for Section 2

A.1 Proof of Theorem 2.1

Let us consider the sample $X_1, \ldots, X_{[n/c_n]}$, divided into k_n/c_n disjoint blocks. We can then apply the arguments used in Lemma 2.1 of Ferreira (2006) [6] and conclude that

$$P(\widetilde{N}_{[n/c_n]}(v_n) = 0) \sim P^{k_n/c_n}(\widetilde{N}_{r_n}(v_n) = 0). \tag{A.1}$$

Now, from the definition of the upcrossings index η we have $P(\widetilde{N}_{[n/c_n]}(v_n) = 0) \xrightarrow[n\to+\infty]{} e^{-\eta\nu}$, hence applying logarithms on both sides of (A.1), (2.9) follows immediately.

For the conditional mean number of upcrossings in each block, we have from (2.9) and the definition

of the thresholds v_n

$$\begin{aligned} E[\widetilde{N}_{r_n}(v_n) \mid \widetilde{N}_{r_n}(v_n) > 0] &= \frac{\sum_{j\geq 1} j \times P(\widetilde{N}_{r_n}(v_n) = j,\ \widetilde{N}_{r_n}(v_n) > 0)}{P(\widetilde{N}_{r_n}(v_n) > 0)} \\ &= \frac{E[\widetilde{N}_{r_n}(v_n)]}{P(\widetilde{N}_{r_n}(v_n) > 0)} = \frac{r_n P(X_1 \leq v_n < X_2)}{P(\widetilde{N}_{r_n}(v_n) > 0)} \\ &\sim \frac{r_n c_n \nu/n}{c_n \eta \nu/k_n}, \end{aligned}$$

which completes the proof since $k_n r_n \sim n$. □

A.2 Proof of Theorem 2.2

It suffices to show that

$$\lim_{n\to+\infty} \sum_{j=1}^{+\infty} j\widetilde{\pi}_n(j; v_n) = \eta^{-1} \tag{A.2}$$

and

$$\sum_{j=1}^{+\infty} j(\widehat{\widetilde{\pi}}_n(j; v_n) - \widetilde{\pi}_n(j; v_n)) \xrightarrow[n\to+\infty]{P} 0. \tag{A.3}$$

(A.2) follows immediately from Theorem 2.1 since

$$\lim_{n\to+\infty} \sum_{j=1}^{+\infty} j\widetilde{\pi}_n(j; v_n) = \lim_{n\to+\infty} E[\widetilde{N}_{r_n}(v_n) \mid \widetilde{N}_{r_n}(v_n) > 0] = \eta^{-1}.$$

To show (A.3), lets start by noting that Theorem 2.2 of Hsing (1991) [15] holds for $\widehat{\widetilde{\pi}}_n(j; v_n)$ and $\widetilde{\pi}_n(j; v_n)$, if Y_{n1}, the number of exceedances of v_n in a block of size r_n, is replaced by $\widetilde{N}_{r_n}(v_n)$. The proof now follows from considering $\alpha = 1$ and $T(j) = j\mathbb{1}_{\{j\geq 1\}}$ in this theorem for $\widetilde{N}_{r_n}(v_n)$ and verifying that conditions (a), (c) and (d) of this theorem follow from the assumptions of Theorem 2.1, (2) and (3), (b) follows from Theorem 2.1 and (e) from the fact that

$$\lim_{n\to+\infty} \frac{k_n}{c_n} E[\widetilde{N}_{r_n}(v_n)] = \lim_{n\to+\infty} \frac{k_n}{c_n} r_n P(X_{\leq} v_n < X_2) = \nu.$$

□

A.3 Proof of Theorem 2.3

From Cramer-Wald's Theorem we know that a necessary and sufficient condition for (2.10) to hold is that for any $a, b \in \mathbb{R}$

$$\begin{aligned} c_n^{-1/2} \left(a \sum_{i=1}^{k_n} \left(\widetilde{N}_{r_n}^{(i)}(v_n) - E[\widetilde{N}_{r_n}^{(i)}(v_n)] \right) + b \sum_{i=1}^{k_n} \left(\mathbb{1}_{\{\widetilde{N}_{r_n}^{(i)}(v_n)>0\}} - E[\mathbb{1}_{\{\widetilde{N}_{r_n}^{(i)}(v_n)>0\}}] \right) \right) \\ \xrightarrow[n\to+\infty]{d} \mathcal{N}(0,\ \nu(a^2\eta\sigma^2 + 2ab + b^2\eta)). \end{aligned} \tag{A.4}$$

We shall therefore prove (A.4). For this, lets consider

$$U_{r_n-l_n}^{(i)}(v_n) = \sum_{j=(i-1)r_n+1}^{ir_n-l_n} \mathbb{1}_{\{X_j \leq v_n < X_{j+1}\}}, \qquad V_{l_n}^{(i)}(v_n) = \widetilde{N}_{r_n}^{(i)}(v_n) - U_{r_n-l_n}^{(i)}(v_n), \quad 1 \leq i \leq k_n$$

and note that

$$a\sum_{i=1}^{k_n}\left(\widetilde{N}_{r_n}^{(i)}(v_n)-E[\widetilde{N}_{r_n}^{(i)}(v_n)]\right)+b\sum_{i=1}^{k_n}\left(\mathbb{1}_{\{\widetilde{N}_{r_n}^{(i)}(v_n)>0\}}-E[\mathbb{1}_{\{\widetilde{N}_{r_n}^{(i)}(v_n)>0\}}]\right)$$

$$\begin{aligned}
&= a\sum_{i=1}^{k_n}\left(U_{r_n-l_n}^{(i)}(v_n)-E[U_{r_n-l_n}^{(i)}(v_n)]\right)+a\sum_{i=1}^{k_n}\left(V_{l_n}^{(i)}(v_n)-E[V_{l_n}^{(i)}(v_n)]\right)\\
&\quad +b\sum_{i=1}^{k_n}\left(\mathbb{1}_{\{U_{r_n-l_n}^{(i)}(v_n)>0\}}-E[\mathbb{1}_{\{U_{r_n-l_n}^{(i)}(v_n)>0\}}]\right)+\\
&\quad +b\sum_{i=1}^{k_n}\left(\mathbb{1}_{\{U_{r_n-l_n}^{(i)}(v_n)>0,\ U_{r_n-l_n}^{(i)}(v_n)=0\}}-E[\mathbb{1}_{\{U_{r_n-l_n}^{(i)}(v_n)>0,\ U_{r_n-l_n}^{(i)}(v_n)=0\}}]\right). \qquad \text{(A.5)}
\end{aligned}$$

Now, since (A.5) holds, to show (A.4) we have to prove the following

$$c_n^{-1/2}\left(a\sum_{i=1}^{k_n}\left(U_{r_n-l_n}^{(i)}(v_n)-E[U_{r_n-l_n}^{(i)}(v_n)]\right)+b\sum_{i=1}^{k_n}\left(\mathbb{1}_{\{U_{r_n-l_n}^{(i)}(v_n)>0\}}-E[\mathbb{1}_{\{U_{r_n-l_n}^{(i)}(v_n)>0\}}]\right)\right)$$
$$\xrightarrow[n\to+\infty]{d}\mathcal{N}(0,\ \nu(a^2\eta\sigma^2+2ab+b^2\eta)), \qquad \text{(A.6)}$$

$$c_n^{-1/2}\sum_{i=1}^{k_n}(V_{l_n}^{(i)}(v_n)-E[V_{l_n}^{(i)}(v_n)])\xrightarrow[n\to+\infty]{P}0, \qquad \text{(A.7)}$$

$$c_n^{-1/2}\sum_{i=1}^{k_n}\left(\mathbb{1}_{\{N_{r_n}^{(i)}>0,\ U_{r_n-l_n}^{(i)}(v_n)=0\}}-E[\mathbb{1}_{\{N_{r_n}^{(i)}>0,\ U_{r_n-l_n}^{(i)}(v_n)=0\}}]\right)\xrightarrow[n\to+\infty]{P}0. \qquad \text{(A.8)}$$

Lets first prove (A.6). The summands in $\sum_{i=1}^{k_n}(aU_{r_n-l_n}^{(i)}(v_n)+b\mathbb{1}_{\{U_{r_n-l_n}^{(i)}(v_n)>0\}})$ are functions of indicator variables that are at least l_n time units apart from each other, therefore for each $t\in\mathbb{R}$

$$\left|E\left[\exp\left(\mathrm{i}tc_n^{-1/2}\sum_{i=1}^{k_n}\left(aU_{r_n-l_n}^{(i)}(v_n)+b\mathbb{1}_{\{U_{r_n-l_n}^{(i)}(v_n)>0\}}\right)\right)\right]- \qquad \text{(A.9)}$$

$$-\prod_{i=1}^{k_n}E\left[\mathrm{i}tc_n^{-1/2}\left(aU_{r_n-l_n}^{(i)}(v_n)+b\mathbb{1}_{\{U_{r_n-l_n}^{(i)}(v_n)>0\}}\right)\right]\right|\le 16k_n\alpha_{n,l_n}, \qquad \text{(A.10)}$$

where, i is the imaginary unit, from repeatedly using a result in Volkonski and Rozanov (1959) [26] and the triangular inequality. Now, since condition $\Delta(v_n)$ holds for $\mathbf{X}$ (A.10) tends to zero and so we can assume that the summands are i.i.d.. Therefore, in order to apply Lindberg's Central Limit Theorem we need to verify that

$$\frac{k_n}{c_n}E\left[\left(aU_{r_n-l_n}(v_n)+b\mathbb{1}_{\{U_{r_n-l_n}(v_n)>0\}}\right)^2\right]\xrightarrow[n\to+\infty]{}\nu(a^2\eta\sigma^2+2ab+b^2\eta), \qquad \text{(A.11)}$$

since $\frac{k_n}{c_n}E^2[U_{r_n-l_n}(v_n)]\xrightarrow[n\to+\infty]{}0$, and Lindberg's Condition

$$\frac{k_n}{c_n}E\left[\left(aU_{r_n-l_n}(v_n)+b\mathbb{1}_{\{U_{r_n-l_n}(v_n)>0\}}\right)^2\mathbb{1}_{\{(aU_{r_n-l_n}(v_n)+b\mathbb{1}_{\{U_{r_n-l_n}(v_n)>0\}})^2>\epsilon c_n\}}\right]\xrightarrow[n\to+\infty]{}0,\quad \text{for all } \epsilon>0, \qquad \text{(A.12)}$$

with $U_{r_n-l_n}(v_n) = U^{(1)}_{r_n-l_n}(v_n)$.

From the definition of $\widetilde{N}^{(i)}_{r_n}(v_n)$ and $V^{(i)}_{l_n}(v_n)$ and assumption (2) we have that

$$\frac{k_n}{c_n}E[V^2_{r_n}(v_n)] \leq \left(\frac{r_n}{l_n}\right)^{-1}\frac{k_n}{c_n}E[\widetilde{N}^2_{r_n}(v_n)] \xrightarrow[n\to+\infty]{} 0, \tag{A.13}$$

with $V_{r_n}(v_n) = V^{(1)}_{r_n}(v_n)$. Now, by Cauchy-Schwarz's inequality

$$\frac{k_n}{c_n}E[\widetilde{N}_{r_n}(v_n)V_{r_n}(v_n)] \leq \sqrt{E[\widetilde{N}^2_{r_n}(v_n)]E[V^2_{r_n}(v_n)]} \xrightarrow[n\to+\infty]{} 0,$$

thus

$$\frac{k_n}{c_n}E[U^2_{r_n-l_n}(v_n)] = \frac{k_n}{c_n}E[(\widetilde{N}_{r_n}(v_n) - V_{r_n}(v_n))^2] \xrightarrow[n\to+\infty]{} \eta\nu\sigma^2. \tag{A.14}$$

On the other hand, since $k_n(r_n - l_n) \sim n$, Theorem 2.1 implies that

$$\frac{k_n}{c_n}E\left[1\!\!1_{\{U_{r_n-l_n}(v_n)>0\}}\right] = \frac{k_n}{c_n}P(U_{r_n-l_n}(v_n) > 0) \xrightarrow[n\to+\infty]{} \eta\nu. \tag{A.15}$$

Furthermore, by definition (2.8)

$$\frac{k_n}{c_n}E\left[U_{r_n-l_n}(v_n)1\!\!1_{\{U_{r_n-l_n}(v_n)>0\}}\right] = \frac{k_n}{c_n}E\left[U_{r_n-l_n}(v_n)\right] = \frac{k_n(r_n-l_n)}{c_n}P(X_1 \leq v_n < X_2) \xrightarrow[n\to+\infty]{} \nu. \tag{A.16}$$

(A.14)-(A.16) prove (A.11) and since Lindberg's Condition follows immediately from assumption (1), (A.6) is proven.

Finally, since Theorem 2.1 of Hsing (1991) [15] holds for $\widetilde{N}_{r_n}(v_n)$, it implies (A.7) and (A.8) because $\frac{k_n}{c_n}E[V^2_{l_n}(v_n)] \xrightarrow[n\to+\infty]{} 0$ by (A.13) and

$$\frac{k_n}{c_n}E[1\!\!1^2_{\{\widetilde{N}^{(i)}_{r_n}(v_n)>0,\ U^{(i)}_{r_n-l_n}(v_n)=0\}}] = \frac{k_n}{c_n}(P(\widetilde{N}^{(i)}_{r_n}(v_n) > 0) - P(U^{(i)}_{r_n-l_n}(v_n)) > 0) \xrightarrow[n\to+\infty]{} 0$$

by Theorem 2.1 and (A.15). This concludes the proof. □

A.4 Proof of Corollary 2.1

Since $\eta_n \xrightarrow[n\to+\infty]{} \eta$, by Theorem 2.1, $\frac{k_n}{c_n}E[\widetilde{N}_{r_n}(v_n)] \xrightarrow[n\to+\infty]{} \nu$, and $c_n^{-1}\sum_{i=1}^{k_n}(\widetilde{N}^{(i)}_{r_n}(v_n) - E[\widetilde{N}^{(i)}_{r_n}(v_n)]) \xrightarrow[n\to+\infty]{P} 0$, by Theorem 2.1 of Hsing (1991) [15] which holds for $\widetilde{N}r_n{}^{(i)}(v_n)$, the result now follows from the fact that

$$\sqrt{c_n}(\widehat{\eta}^B_n - \eta_n) = \frac{1}{c_n^{-1}\sum_{i=1}^{k_n}\widetilde{N}^{(i)}_{r_n}(v_n)}\left(c_n^{-1/2}\sum_{i=1}^{k_n}\left(1\!\!1_{\{\widetilde{N}^{(i)}_{r_n}(v_n)>0\}} - E[1\!\!1_{\{\widetilde{N}^{(i)}_{r_n}(v_n)>0\}}]\right) - \right.$$
$$\left. -\eta_n c_n^{-1/2}\sum_{i=1}^{k_n}\left(\widetilde{N}^{(i)}_{r_n}(v_n) - E[\widetilde{N}^{(i)}_{r_n}(v_n)]\right)\right)$$

and Theorem 2.3. □

A.5 Proof of Theorem 2.4

Since (A.2) holds, we only need to show that

$$\sum_{j=1}^{+\infty}(\widetilde{\pi}_n(j;\widehat{v}_n)-\widetilde{\pi}_n(j;v_n)) \xrightarrow[n\to+\infty]{P} 0. \tag{A.17}$$

Lets start by noting that for $v_n^{(\tau)}$ such that $P(X_1 > v_n^{(\tau)}) \sim c_n\tau/n$ as $n \to +\infty$ and $\epsilon > 0$ we have

$$\begin{aligned}
&\lim_{\epsilon\to 0}\lim_{n\to+\infty}\frac{k_n}{c_n}\left|E\left[\widetilde{N}_{r_n}(v_n^{(\tau+\epsilon)})\mathbb{1}_{\{\widetilde{N}_{r_n}(v_n^{(\tau+\epsilon)})>0\}}\right]-E\left[\widetilde{N}_{r_n}(v_n^{(\tau)})\mathbb{1}_{\{\widetilde{N}_{r_n}(v_n^{(\tau)})>0\}}\right]\right| \\
&= \lim_{\epsilon\to 0}\lim_{n\to+\infty}\frac{k_n r_n}{c_n}(P(X_1 \le v_n^{(\tau+\epsilon)} < X_2) - P(X_1 \le v_n^{(\tau)} < X_2)) \\
&= \lim_{\epsilon\to 0}\lim_{n\to+\infty}\frac{k_n r_n}{c_n}(P(X_1 > v_n^{(\tau+\epsilon)}) - P(X_1 > v_n^{(\tau)})) = 0.
\end{aligned} \tag{A.18}$$

(A.18) proves condition b) of Theorem 2.3 in Hsing (1991) [15] which holds for $\widetilde{N}_{r_n}(v_n)$, where $T(j) = j\mathbb{1}_{\{j\ge 1\}}$. The other conditions have been verified in the proof of Theorem 2.2 as well as the conditions of Corollary 2.4 in Hsing (1991) [15] for $\widetilde{N}_{r_n}(v_n)$. Therefore (A.17) holds, completing the proof. □

Acknowledgements We acknowledge the support from research unit "Centro de Matemática e Aplicações" of the University of Beira Interior, through the research project UID/MAT/00212/2013.

References

[1] Ancona-Navarrete, M.A. and Tawn, J.A. (2000). A comparison of methods for estimating the extremal index. *Extremes*, **3**(1), 5-38.

[2] Alpuim, M. (1989). An extremal Markovian sequence. *J. Appl. Prob.*, **26**, 219-232.

[3] Chernick, M., Hsing, T. and McCormick, W. (1991). Calculating the extremal index for a class of stationary sequences. *Adv. Appl. Prob.*, **23**, 835-850.

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

[5] Ferreira, H. (1994). Multivariate extreme values in T-periodic random sequences under mild oscillation restrictions. *Stochast. Process. Appl.*, **49**, 111-125.

[6] Ferreira. H. (2006). The upcrossing index and the extremal index. *J. Appl. Prob.*, **43**, 927-937.

[7] Ferreira, H. (2007). Runs of high values and the upcrossings index for a stationary sequence. In *Proceedings of the 56th Session of the ISI.*

[8] Ferreira, M. and Ferreira, H. (2012). On extremal dependence: some contributions. *Test.* **21**(3), 566-583.

[9] Ferro, C. and Segers, J. (2003). Inference for clusters of extreme values. *J. Royal Statist. Soc. B*, **65**, 545-556.

[10] Frahm, G., Junker, M. and Schmidt, R. (2005). *Insurance: Mathematics and Economics*, **37**, 80-100.

[11] Gomes, M. (1993). On the estimation of parameters of rare events in environmental time series. In V. Barnett and K. Turkman (Eds.), *Statistics for the environment 2: Water Related Issues*, 225-241. Wiley.

[12] Gomes, M.I., Hall, A. and Miranda, C. (2008). Subsampling techniques and the Jacknife methodology in the estimation of the extremal index. *J. Statist. Comput. Simulation*, **52**, 2022-2041.

[13] Gomes, M.I. and Oliveira, O. (2001). The boostrap methodology in Statistics of extremes - choice of the optimal sample fraction. *Extremes*, **4**(4), 331-358.

[14] Hsing, T., Hüsler, J. and Leadbetter, M.R. (1988). On the exceedance point process for a stationary sequence. *Prob. Th. Rel. Fields*, **78**, 97-112.

[15] Hsing, T. (1991). Estimating the parameters of rare events. *Stochast. Process. Appl.*, **37**(1), 117-139.

[16] Leadbetter, M.R. (1983). Extremes and local dependence in stationary processes. *Z. Wahrsch. verw. Gebiete*, **65**, 291-306.

[17] Leadbetter, M.R. and Nandagopalan, S. (1989). On exceedance point process for stationary sequences under mild oscillation restrictions. In J. Hüsler and D. Reiss (eds.). *Extreme Value Theory: Proceedings, Oberwolfach* 1987, Springer, New York, 69-80.

[18] Martins, A.P. and Ferreira, H. (2004). The extremal index of sub-sampled processes. *J. Statist. Plann. Inference*, **124**, 145-152.

[19] Neves, M., Gomes, M.I., Figueiredo, F. and Gomes, D. (2015). Modeling extreme events: sample fraction adaptive choice in parameter estimation. *J. Statist. Theory Pract.*, **9**(1), 184-199.

[20] Klar, B., Lindner, F. and Meintanis, S.G. (2011). Specification tests for the error distribution in Garch models. *Comput. Statist. Data Anal.* **56**, 3587-3598.

[21] Robert, C.Y., Segers, J. and Ferro, C. (2009). A sliding blocks estimator for the extremal index. *Elect. J. Statist.*, **3**, 993-1020.

[22] Robinson, M.E. and Tawn, J.A. (2000). Extremal analysis of processes sampled at different frequencies. *J. Roy. Statist. Soc. Ser. B*, **62**, 117-135.

[23] Sebastião, J., Martins, A.P., Pereira, L. and Ferreira. H., (2010). Clustering of upcrossings of high values. *J. Statist. Plann. Inference*, **140**, 1003-1012.

[24] Sebastião, J., Martins, A.P., Ferreira, H and Pereira, L., (2013). Estimating the upcrossings index. *Test.* **22**(4), 549-579.

[25] Süveges, M. (2007). Likelihood estimation of the extremal index. *Extremes*, **10**, 41-55.

[26] Volkonski, V.A. and Rozanov, Y.A. (1959). Some limit theorems for random function I. *Theory Probab. Appl.*, **4**, 178-197.