МАТЕМАТИЧКИ ВЕСНИК
68, 2 (2016), 130–139
June 2016

оригинални научни рад
research paper

SEMI PARAMETRIC ESTIMATION OF EXTREMAL INDEX FOR ARMAX PROCESS WITH INFINITE VARIANCE

Hakim Ouadjed and Mami Tawfiq Fawzi

Abstract. We consider estimating the extremal index of a maximum autoregressive process of order one under the assumption that the distribution of the innovations has a regularly varying tail at infinity. We establish the asymptotic normality of the new estimator using the extreme quantile approach, and its performance is illustrated in a simulation study. Moreover, we compare, in terms of bias and mean squared error, our estimator with the estimator of Ferro and Segers [Inference for clusters of extreme values, J. Royal Stat. Soc., Ser. B, 65 (2003), 545–556] and Olmo [A new family of consistent and asymptotically-normal estimators for the extremal index, Econometrics, 3 (2015), 633–653].

1. Introduction

The extremal index parameter characterizes the degree of local dependence in the extremes of a stationary time series and has important applications in a number of areas, such as hydrology, telecommunications, finance and environmental studies. This parameter is the key for extending extreme value theory results from i.i.d. to stationary sequences.

Many applications as in insurance and finance, telecommunication and other areas of technical risk, usually exhibit a dependence structure. Leadbetter et al. [12] put a mixing condition $D(u_n)$ based on the probability of exceedances of a high threshold u_n, it limits the degree of long-term dependence of the sequence, providing asymptotic independence between far apart extreme observations.

Definition1.1. [$D(u_n)$ condition] A strictly stationary sequence $\{X_i\}$, whose marginal distribution F has upper support point $x_F = \sup\{x : F(x) < 1\}$, is said to satisfy $D(u_n)$ if, for any integers $i_1 < \cdots < i_p < j_1 < \cdots < j_q$ with $j_1 - i_p > l_n$,

$$\big|P\left\{X_{i_1} \le u_n, \dots, X_{i_p} \le u_n, X_{j_1} \le u_n, \dots, X_{j_q} \le u_n\right\} - P\left\{X_{i_1} \le u_n, \dots, X_{i_p} \le u_n\right\} P\left\{X_{j_1} \le u_n, \dots, X_{j_q} \le u_n\right\}\big| \le \delta(n, l_n),$$

where $\delta(n, l_n) \to 0$ for some sequences $l_n = o(n)$ and $u_n \to x_F$ as $n \to \infty$.

2010 Mathematics Subject Classification: 60G70, 62G32
Keywords and phrases: extreme value theory; max autoregressive processes; tail index estimation.

Let $X_1, \ldots, X_n$ be a strictly stationary sequence with marginal distribution F, and $\tilde{X}_1, \ldots, \tilde{X}_n$ an i.i.d. sequence of random variables with the same distribution F, define the following quantities $M_n = \max(X_1, \ldots, X_n)$ and $\tilde{M}_n = \max(\tilde{X}_1, \ldots, \tilde{X}_n)$. Under the $D(u_n)$ condition, with $u_n = a_n x + b_n$, if

$$\mathbf{P}[a_n^{-1}(\tilde{M}_n - b_n) \le x] \to G(x), \text{ as } n \to \infty, \tag{1.1}$$

for normalizing sequences $a_n > 0$ and $b_n \in \mathbf{R}$, Leadbetter et al. [12] showed that

$$\mathbf{P}[a_n^{-1}(M_n - b_n) \le x] \to [G(x)]^{\theta}, \text{ as } n \to \infty, \tag{1.2}$$

where G is one of the three extreme value types distributions:

Type I (Gumbel):

$$G(x) = \exp(-e^{-x}), \; x \in \mathbf{R},$$

Type II (Fréchet):

$$G(x) = \begin{cases} 0, & x \le 0 \\ \exp(-x^{-\alpha}), & x > 0, \alpha > 0, \end{cases}$$

Type III (Weibull):

$$G(x) = \begin{cases} \exp(-(-x)^{\alpha}), & x \le 0, \alpha > 0 \\ 1, & x > 0, \end{cases}$$

and $\theta \in (0, 1]$ is the extremal index, this parameter characterizes the short-range dependence of the maxima. In particular, θ^{-1} gives a measure of the degree of clustering of large values of the sequence. Theoretical properties of the extremal index have been studied fairly extensively (O'Brien [13]), Hsing et al. [11], and the references therein). The problem of estimating θ has also received some attention in the literature (see Smith and Weissman [16], Weissman and Novak [17]).

Ferro and Segers [8], using a moment estimator, obtained

$$\widehat{\theta}_{FS} = \begin{cases} 1 \wedge \widehat{\theta}_1, & \max\{T_i : 1 \le i \le N-1\} \le 2 \\ 1 \wedge \widehat{\theta}_2, & \max\{T_i : 1 \le i \le N-1\} > 2, \end{cases} \tag{1.3}$$

where T_i are the inter-exceedance times and N is the number of exceedances of a fixed high threshold u and

$$\widehat{\theta}_1 = \frac{2[\sum_{i=1}^{N-1} T_i]^2}{(N-1)\sum_{i=1}^{N-1} T_i^2}, \quad \widehat{\theta}_2 = \frac{2[\sum_{i=1}^{N-1}(T_i - 1)]^2}{(N-1)\sum_{i=1}^{N-1}(T_i - 1)(T_i - 2)}$$

The Ferro-Segers estimator is consistent for m-dependent strictly stationary sequences.

Recently, Olmo [14] introduces an estimator for this parameter as the ratio of the number of elements of two point processes defined by a partition of the sample in different blocks, and by the block maxima exceeding the corresponding thresholds v_n and u_n, with $v_n > u_n$. The estimator is consistent and converges to a normal distribution and is given by :

$$\widehat{\theta}_n(r_n) = \frac{B_{v_n}}{B_{u_n}} = \frac{\sum_{j=1}^{k_n} I(M_{(j-1)r_n+1,jr_n} > v_n)}{\sum_{j=1}^{k_n} I(M_{(j-1)r_n+1,jr_n} > u_n)},$$

with $I(X > u_n)$ the indicator function and dividing the data $\{X_i, i \geq 1\}$ of length n into k_n blocks of size r_n, with $k_n = o(n)$, $r_n = [n/k_n]$, where $[\cdot]$ is the integer part and

$$M_{(j-1)r_n+1,jr_n} = \max(X_{(j-1)r_n+1}, \ldots, X_{jr_n})$$

and v_n verified

$$E\Big[\sum_{j=1}^{r_n} I(X_j > v_n) | \sum_{j=1}^{r_n} I(X_j > u_n) \geq 1\Big] \to 1.$$

In practice he proposed to estimate v_n by $\widehat{v}_n = F^{-1}\Big(1 - \frac{B_{u_n}}{n}\Big)$ and the estimator $\widehat{\theta}_n(r_n)$ becomes

$$\widehat{\theta}_n^f(r_n) = \frac{B_{\hat{v}_n}}{B_{u_n}} \tag{1.4}$$

and we have

$$\sqrt{B_{u_n}}(\widehat{\theta}_n^f(r_n) - \theta) \xrightarrow{D} \mathcal{N}\left(0, \sigma_1^2\right), \tag{1.5}$$

where

$$\sigma_1^2 = \theta. \tag{1.6}$$

The rest of this paper is organized as follows. In Section 2, we discus of heavy-tailed ARMAX(1) properties and in Section 3 we construct an normal estimator of the extremal index θ for this process. In Section 4 we compare by simulation the performance of our estimator and their of Ferro and Segers [8] and Olmo [14]. Section 5 is devoted to the proofs.

2. ARMAX process

Infinite variance time series are popular modelling tools in the telecommunication industry. For a summary of results and applications, see Resnick [15] and the references therein. We consider the maximum autoregressive process of order one ARMAX(1). This process has been recommended as an alternative to AR(1) process with heavy-tailed innovations by Davis and Resnick [4] and they are more convenient for analysis extreme values because their dimensional finite-distributions can easily be written explicitly.

Let $X_1, X_2, \ldots, X_n$ be the process defined recursively as follows:

$$X_i = \max\left(\lambda\ X_{i-1}, Z_i\right), \tag{2.1}$$

where $0 < \lambda < 1$ and $Z_1, \ldots, Z_n$ are independent and identically distributed, with distribution function $F_Z(x) = \exp(-x^{-\alpha}), 0 < \alpha < 2$. We shall call it the maximal autoregressive process of order 1 (abbreviate to ARMAX(1)). Such processes have finite-dimensional distributions which are max-stable and hence are examples of max-stable processes.

The process $\{X_i\}$ defined in (2.1) has stationary distribution given by,

$$F_X(x) = \prod_{j=0}^{\infty} F_Z\left(x/\lambda^j\right) \tag{2.2}$$

and any d.f. that is solution of equation

$$F_X(x) = F_Z(x)F_X\left(x/\lambda\right) \tag{2.3}$$

is a stationary d.f. of $\{X_i\}$.

From Ferreira and Canto e Castro [7] we have for $x \to \infty$

$$\mathbf{P}(X > x) \sim (1-\lambda^{\alpha})^{-1}\mathbf{P}(Z > x) \sim (1-\lambda^{\alpha})^{-1}x^{-\alpha}. \tag{2.4}$$

The ARMAX process, have a weak dependence structure. In fact they verify the β-mixing condition (see Drees [6]), $D(u_n)$ condition. Hence the extremal index of $\{X_i\}$ is $\theta = 1-\lambda^{\alpha}$ (see Beirlant et al. [2]).

3. A semi parametric estimate of θ

We can rewrite the relation (2.4) as

$$\mathbf{P}(X > x) \sim \theta^{-1}x^{-\alpha}.$$

Hence, we can estimate θ^{-1} by $\frac{k}{n}X_{n-k,n}^{\widehat{\alpha}_X}$, where $k = k(n) \to \infty, k/n \to 0$ and

$$\widehat{\alpha}_X = \Big[\frac{1}{k}\sum_{i=1}^{k}\log X_{n-i+1,n} - \log X_{n-k,n}\Big]^{-1},$$

is the Hill estimator [10], with $X_{i,n}$ denotes the i-th ascending order statistics $1 \le i \le n$, associated to the random sample $(X_1, X_2, \dots, X_n)$. It easy to check that

$$\widehat{\theta}_n = \frac{n}{k}X_{n-k,n}^{-\widehat{\alpha}_X} \tag{3.1}$$

We note that from Theorem 2.2 of Drees [6], we have

$$\sqrt{k}(\widehat{\alpha}_X - \alpha) \xrightarrow{D} \mathcal{N}\left(0,\sigma^2\right), \tag{3.2}$$

where

$$\sigma^2 = \alpha^2\int_{(0,1]}\int_{(0,1]}(st)^{-(1+1/\alpha)}\widetilde{c}(s,t)\,\nu(ds)\,\nu(dt),$$

ν being the signed measure defined by $\nu(dt) = t^{\alpha^{-1}}dt - \delta_1(dt)$ and δ_1 the Dirac measure with mass 1 at 1 and where

$$\widetilde{c}(x,y) := \min(x,y) + \sum_{m=1}^{\infty}[c_m(x,y) + c_m(y,x)],$$

and

$$c_m(x,y)=\lim_{x\to\infty}\frac{n}{k}P\Big[X_1>F_X^{-1}\Big(1-\frac{k}{n}x\Big),X_{1+m}>F_X^{-1}\Big(1-\frac{k}{n}y\Big)\Big]$$

for all $m\in\mathbf{N},\ x>0,\ y\le 1+\varepsilon,\ \varepsilon>0$ and F^{-1} denoting the inverse function of F.

We note from Dress [5] that

$$\sigma^2=\alpha^2\widetilde{c}(1,1). \tag{3.3}$$

In the case of ARMAX(1) given by equation (2.1), Ferreira and Canto e Castro [7] showed that

$$\widetilde{c}(x,y):=\min(x,y)+\sum_{m=1}^{p-1}[c_m(x,y)+c_m(y,x)]+(x+y)\frac{\lambda^{p\alpha}}{1-\lambda^{\alpha}}$$

for

$$p\equiv p_{x,y}=[\max\{\alpha^{-1}\ln(x/y)/\ln\lambda,\alpha^{-1}\ln(y/x)/\ln\lambda\}]+1.$$

Hence the variance of Hill estimator in (3.3) becomes

$$\sigma^2=\alpha^2\Big(1+2\frac{\lambda^{\alpha}}{1-\lambda^{\alpha}}\Big). \tag{3.4}$$

The asymptotic normality of $\widehat{\theta}_n$ is established in the following theorem.

Theorem 3.1. *Suppose* (2.1) *and* $k=k_n$ *be such that* $k\to\infty,\ k/n\to 0$. *Then*

$$\frac{\sqrt{k}}{\log(n/k)}(\widehat{\theta}_n-\theta)\xrightarrow{D}\mathcal{N}\left(0,\sigma_2^2\right),$$

where

$$\sigma_2^2=\alpha^4\theta^3(2-\theta). \tag{3.5}$$

4. Proof

Let $U(t)=F_X^{-1}\left(1-\frac{1}{t}\right)$. Note that

$$\begin{aligned}&\frac{k}{n}X_{n-k,n}^{\widehat{\alpha}_X}-\theta^{-1}\\&\quad=\Big(\frac{k}{n}X_{n-k,n}^{\widehat{\alpha}_X}-\frac{k}{n}X_{n-k,n}^{\alpha}\Big)+\frac{k}{n}U^{\alpha}(n/k)\Big(\frac{X_{n-k,n}^{\alpha}}{U^{\alpha}(n/k)}-1\Big)+\frac{k}{n}U^{\alpha}(n/k)-\theta^{-1}.\end{aligned}$$

Using Mean-Value Theorem we find

$$\begin{aligned}\frac{k}{n}X_{n-k,n}^{\widehat{\alpha}_X}-\theta^{-1}&=\Big(\frac{k}{n}X_{n-k,n}^{\alpha}(\widehat{\alpha}_X-\alpha)\log X_{n-k,n}\Big)(1+o_P(1))\\&\quad+\frac{k}{n}U^{\alpha}(n/k)\Big(\frac{X_{n-k,n}^{\alpha}}{U^{\alpha}(n/k)}-1\Big)+\frac{k}{n}U^{\alpha}(n/k)-\theta^{-1}.\end{aligned}$$

From Theorem 2.1 of Drees [6] we have $\dfrac{X^{\alpha}_{n-k,n}}{U^{\alpha}(n/k)} = 1 + O_P(1/\sqrt{k})$ and using (3.2) we obtain

$$\frac{\sqrt{k}}{\log(n/k)}\Big(\frac{k}{n}X^{\widehat{\alpha}x}_{n-k,n} - \theta^{-1}\Big) \xrightarrow{D} \mathcal{N}(0, \alpha^2\sigma^2).$$

Applying the delta method, it follows that the estimator $\widehat{\theta}_n$ defined in (3.1) satisfies the following result

$$\frac{\sqrt{k}}{\log(n/k)}(\widehat{\theta}_n - \theta) \xrightarrow{D} \mathcal{N}\Big(0, \alpha^4\frac{(2-\theta)}{\theta}\Big[(f)'\Big(\frac{1}{\theta}\Big)\Big]^2\Big),$$

where $f(x) = 1/x$. This completes the proof of Theorem 3.1. ■

5. Simulation study

Tail index estimation depends for its accuracy on a precise choice of the sample fraction, i.e., the number of extreme order statistics on which the estimation is based. The most common methods of adaptive choice of the threshold k are based on the minimization of some kind of MSE's estimates

$$k_{opt} = \arg\min_{k} E(\widehat{\alpha} - \alpha)^2. \tag{5.1}$$

We mention the pioneering papers by Hall and Welsh [9], Danielsson et al. [3] and Beirlant et al. [1].

To obtain confidence intervals for our estimator $\widehat{\theta}_n$, we generate 100 replications of the time series $(X_1, \ldots, X_n)$ for different sample sizes $(1000, 3000)$, where X_t is an ARMAX(1) process satisfying

$$X_t = \max(\lambda X_{t-1}, Z_t), \quad 0 < \lambda < 1,\ t \geq 1, \tag{5.2}$$

where $\{Z_t\}_{t\geq 1}$ are i.i.d. with tail distribution $1 - F_Z(x) = 1 - \exp(-x^{-\alpha})$, we use (5.1) for compute k_{opt}. The simulation results are presented in Table 1 and Table 2, where lb and ub stand respectively for lower bound and upper bound of the confidence interval. We compare in Table 3, in terms of bias and root of the mean squared error (RMSE), the performances of our estimator $\widehat{\theta}_n$ and Ferro and segers estimator $\widehat{\theta}_{FS}$ in (1.3). We conclude that $\widehat{\theta_n}$ has smaller bias and RMSE and consequently it performs better than $\widehat{\theta}_{FS}$ in the case $0 < \alpha < 1$ which the distribution is very heavy tailed.

n	θ	$\widehat{\theta}_n$	lb	ub	length
1000	0.519	0.553	0.387	0.720	0.333
3000	0.519	0.479	0.317	0.642	0.325

Table 1. 95% confidence intervals for θ, with $\lambda = 0.4$ and tail index $\alpha = 0.8$.

n	θ	$\widehat{\theta}_n$	lb	ub	length
1000	0.696	0.707	0.013	1.400	1.387
3000	0.696	0.693	0.178	1.208	1.030

Table 2. 95% confidence intervals for θ, with $\lambda = 0.4$ and tail index $\alpha = 1.3$.

α	0.2		0.3		1.3		1.6	
θ	0.275		0.382		0.876		0.923	
n	1000	3000	1000	3000	1000	3000	1000	3000
$\widehat{\theta}_{FS}$	0.301	0.296	0.453	0.413	0.899	0.882	0.941	0.925
Bias	0.026	0.021	0.071	0.031	0.023	0.006	0.018	0.002
RMSE	0.210	0.167	0.228	0.153	0.093	0.080	0.077	0.064
$\widehat{\theta}_n$	0.259	0.260	0.359	0.361	0.909	0.888	0.949	0.905
Bias	-0.016	-0.010	-0.023	-0.021	0.033	0.012	0.026	-0.018
RMSE	0.123	0.084	0.170	0.120	0.181	0.150	0.252	0.174

Table 3. Comparison of $\widehat{\theta}_{FS}$ and $\widehat{\theta}_n$ for $\lambda = 0.2$.

Now, we compare the performance of our estimator $\widehat{\theta}_n$ and Olmo estimator $\widehat{\theta}^f_n(r_n)$ in (1.4). We choose $u_n = X_{kopt+1,n}$ and $\widehat{v}_n = X_{B_{u_n}+1,n}$ from the sequence $X_{1,n} \geq X_{2,n} \geq \cdots \geq X_{n,n}$.

First, our comparison will be based on the ratio of the asymptotic variances defined in (1.6) and (3.5), namely

$$R = \frac{\sigma_2^2}{\sigma_1^2} = \alpha^4\theta^2(2-\theta).$$

We investigate its behavior with the help of graphs.

Given a fixed index $0 < \alpha < 1$, we compare in the left of Fig. 1 the value of the above ratio with respect to 1, as the extremal index θ varies in the interval $]0,1[$, we remark that $\widehat{\theta}_n$ is more efficient than $\widehat{\theta}^f_n(r_n)$.

For $1 < \alpha < 2$ and after several trials, we noticed that there exists a real number θ_0 such that for $\theta_0 < \theta < 1$ the asymptotic variance of $\widehat{\theta}^f_n(r_n)$ is smaller than of $\widehat{\theta}_n$ (see the right of Fig. 1).

Second, we generate 100 replications of the time series $(X_1, \ldots, X_n)$ for different sample sizes $(200, 500, 1000, 2000)$, where X_t is an ARMAX(1) process satisfying (5.2). We plot in Figs. 2, 3 the absolute bias (abias) of $\widehat{\theta}^f_n(r_n)$, $r_n \in [1, 20]$, where the horizontal line is the absolute bias of $\widehat{\theta}_n$.

For $0 < \alpha < 1$, the (abias) of $\widehat{\theta}_n$ is always better than $\widehat{\theta}^f_n(r_n)$ (see Fig. 2).

For $1 < \alpha < 2$ and $n = 200$, the abias of $\widehat{\theta}^f_n(r_n)$ is better than $\widehat{\theta}_n$ in most cases unlike for $n = 500, 1000, 2000$ (see Fig. 3).

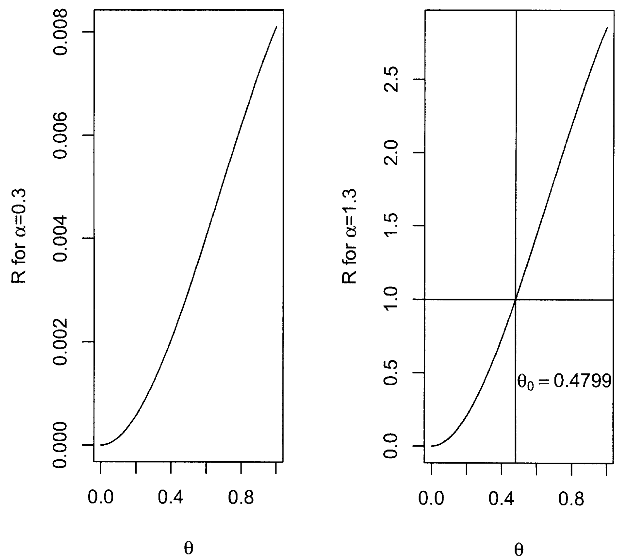

Fig. 1. Ratio of the asymptotic variance of the estimator $\widehat{\theta}_n$ over that of the $\widehat{\theta}_n^f(r_n)$ with $\alpha = 0.3$ (left) and $\alpha = 1.3$ (right)

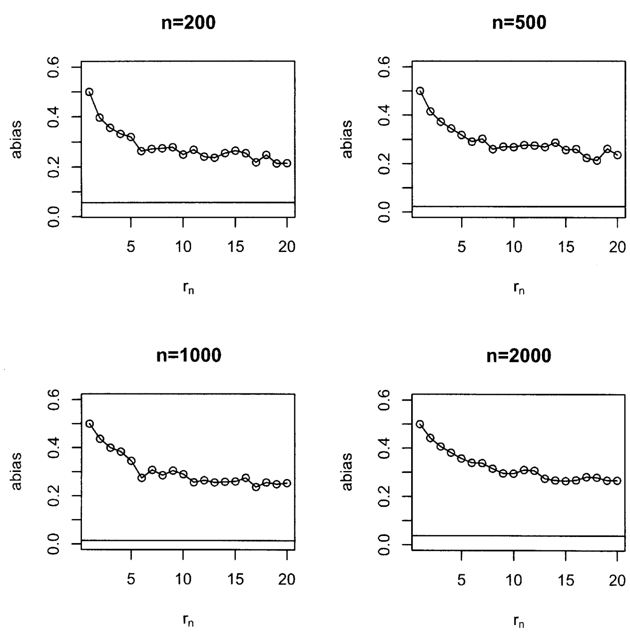

Fig. 2. The abias of $\widehat{\theta}_n^f(r_n)$ and $\widehat{\theta}_n$ for $\alpha = 0.5,\ \lambda = 0.25$

Acknowledgement. The authors would like to thank the referee for careful reading and for their comments which greatly improved the paper.

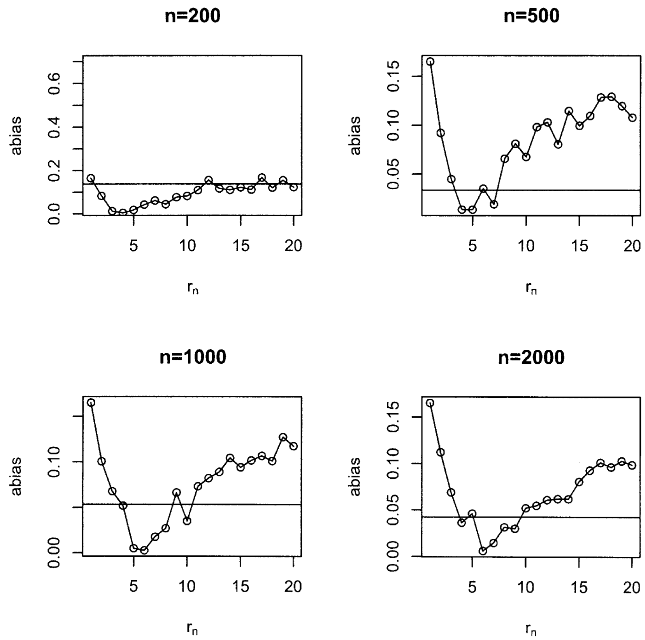

Fig. 3. The abias of $\widehat{\theta}_n^f(r_n)$ and $\widehat{\theta}_n$ for $\alpha = 1.3,\ \lambda = 0.25$

REFERENCES

[**1**] J. Beirlant, P. Vynckier and J.L. Teugels, *Tail index estimation, Pareto quantile plots, and regression diagnostics*, J. Amer. Stat. Assoc., **91** (1966), 1659–1667.

[**2**] J. Beirlant, Y. Goegebeur, J. Segers and J. Teugels, *Statistics of Extremes*, Wiley, 2004.

[**3**] J. Danielsson, L. de Haan, L. Peng and C.G. de Vries, *Using a bootstrap method to choose the sample fraction in tail index estimation*, J. Multivariate Anal., **76** (2001), 226–248.

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

[**5**] H. Drees, *Weighted approximations of tail processes for β-mixing random variables*, Ann. Appl. Probab., **10** (2000), 1274–1301.

[**6**] H. Drees, *Extreme quantile estimation for dependent data, with applications to finance*, Bernoulli, **4** (2003), 617–657.

[**7**] M. Ferreira, L. Canto e Castro, *Tail and dependence behavior of levels that persist for a fixed period of time*, Extremes, **11** (2008), 113–133.

[**8**] C.A.T. Ferro and J. Segers, *Inference for clusters of extreme values*, J. Royal Stat. Soc., Ser. B, **65** (2003), 545–556.

[**9**] P. Hall and A. Welsh, *Adaptive estimates of parameters of regular variation*, Ann. Statist., **13** (1985), 331–341.

[**10**] B.M. Hill, *A simple approach to inference about the tail of a distribution*, Ann. Statist. **3** (1975), 1136–1174.

[**11**] T. Hsing, J. Husler and M. Leadbetter, *On the exceedance point process for a stationary sequence*, Probability Theory & Related Fields, **78** (1988), 97–112.

[**12**] M.R. Leadbetter, G. Lindgren and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York, 1983.

[**13**] G. O'Brien, *Extreme values for stationary and Markov sequences*, Annals Probability, **15** (1987), 281–291.

[**14**] J. Olmo, *A new family of consistent and asymptotically-normal estimators for the extremal index*, Econometrics, **3** (2015), 633–653.

[**15**] S. Resnick, *Heavy tail modeling and teletraffic data*, Ann. Statist **25** (5) (1997), 1805–1869.

[**16**] R.L. Smith and I. Weissman, *Estimating the extremal index*, J. Royal Stat. Soc., Series B, **56** (1994), 515–528.

[**17**] I. Weissman and S.Y. Novak, *On blocks and runs estimators of the extremal index*, J. Statist. Plann. Inference, **66** (1998), 281–288.

(received 27.05.2015; in revised form 19.01.2016; available online 01.02.2016)

H.O.: Mascara University, Algeria

E-mail: `o_hakim77@yahoo.fr`

M.T.F.: Ain Temouchent University, Algeria

E-mail: `mami_math_sba@yahoo.fr`