Test (2011) 20:447–479
DOI 10.1007/s11749-011-0269-8

INVITED PAPER

Subsampling weakly dependent time series and application to extremes

Paul Doukhan · Silika Prohl · Christian Y. Robert

Published online: 1 November 2011

Abstract This paper provides extensions of the work on subsampling by Bertail et al. in J. Econ. 120:295–326 (2004) for strongly mixing case to weakly dependent case by application of the results of Doukhan and Louhichi in Stoch. Proc. Appl. 84:313–342 (1999). We investigate properties of smooth and rough subsampling estimators for sampling distributions of converging and extreme statistics when the underlying time series is η- or λ-weakly dependent.

Keywords Subsampling · Weakly dependent time series · Max-stable distributions

Mathematics Subject Classification (2000) 62G30 · 63G32 · 62G99

This invited paper is discussed in the comments available at doi:10.1007/s11749-011-0270-2, doi:10.1007/s11749-011-0271-1, doi:10.1007/s11749-011-0272-0, doi:10.1007/s11749-011-0273-z, doi:10.1007/s11749-011-0274-y.

P. Doukhan
French Universitary Institute, IUF University Cergy Pontoise, UFR Sciences-Techniques, Saint-Martin, 2, avenue Adolphe-Chauvin, B.P. 222 Pontoise, 95302 Cergy-Pontoise cedex, France
e-mail: doukhan@u-cergy.fr

S. Prohl
Department of Operations Research and Financial Engineering, Princeton University, Sherrerd Hall, Charlton Street, 08544 Princeton, NJ, USA
e-mail: sprohl@princeton.edu

C.Y. Robert (✉)
Institut de Science Financière et d'Assurances, Université de Lyon, Université Lyon 1, 50 Avenue Tony Garnier, 69007 Lyon, France
e-mail: christian.robert@univ-lyon1.fr

1 Introduction

Politis and Romano (1994) established the methodology of subsampling to approximate the sampling distributions of converging statistics when the underlying sequence is strongly mixing. Bertail et al. (2004) extended this technique to a subsampling estimator for distributions of diverging statistics. In particular, they constructed an approximation of the distribution of the sample maximum without any information on the tail of the stationary distribution. However, the assumption on the strong mixing properties of the time series is sometimes too strong as for the first-order autoregressive (AR(1)) process introduced and studied by Chernick (1981):

$$X_t = \frac{1}{r}(X_{t-1} + \varepsilon_t), \quad t \geq 1, \tag{1}$$

where $r \geq 2$ is an integer, $(\varepsilon_t)_{t\in\mathbb{N}}$ are i.i.d. and uniformly distributed on the set $\{0, 1, \ldots, r-1\}$, and X_0 is uniformly distributed on $[0, 1]$. Andrews (1984) and Ango Nze and Doukhan (2004) (see p. 1009 and Note 5 on p. 1028) provide tools to prove that such a class of processes is not strongly mixing. The results of Bertail et al. (2004) cannot be used, although the normalized sample maximum has a nondegenerate limiting distribution: let $M_n = \max(X_1, \ldots, X_n)$; then

$$\lim_{n\to\infty} \mathbb{P}\big(n(1-M_n) \leq x\big) = 1 - \exp\big(-r^{-1}(r-1)x\big) \quad \text{for all } x \geq 0$$

(see Theorem 4.1 in Chernick 1981).

This paper is aimed at weakening the strong mixing condition assumed in Bertail et al. (2004) and at studying a new smooth subsampling estimator adapted to the weak dependence conditions considered in Doukhan and Louhichi (1999). Indeed in this paper a wide dependence framework is introduced that turns out in particular to apply to the previous process and that widely improves the amount of potentially usable models (see Dedecker et al. 2007).

The content of the paper is organized as follows. Section 2 is devoted to a definition of the dependence structure and provides examples of weakly dependent sequences. In this paper we assume that the underlying process is either η- or λ-weakly dependent. The smooth and rough subsampling estimators for the sampling distributions of converging statistics are then introduced, and uniform almost sure convergence results are given. Section 3 gives sufficient conditions for convergence in distribution of suitably normalized sample maxima for weakly dependent time series. Then, we discuss how to estimate the normalizing factors and derive the asymptotic properties of the smooth subsampler. Section 4 presents and discusses a comparative simulation study of the finite sample behavior of the smooth and rough subsampling estimators on simulated data. Proofs are reported in the last section.

In the sequel, we use the following notation. For two sequences $a \equiv (a_n)$ and $b \equiv (b_n)$, $a \prec b$ says that there exists a positive constant c such that, for all n, $a_n \leq cb_n$. The maximum of the numbers a, b is denoted by $a \vee b$, the integer-valued part of the real number x by $\lfloor x \rfloor$, and the almost sure convergence by a.s. We denote by $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{R}$ the sets of nonnegative integers, integers, and the real line.

2 Subsampling the distribution of converging statistics for weakly dependent time series

2.1 Weak dependence

Doukhan and Louhichi (1999) proposed a new concept of weak dependence that makes explicit the asymptotic independence between past and future. Let us consider a strictly stationary time series $X = (X_t)_{t\in\mathbb{Z}}$ which (for simplicity) will be assumed to be real-valued. Let us denote by F its stationary distribution function. If X is a sequence of iid random variables, then for all $t_1 \neq t_2$, independence between X_{t_1} and X_{t_2} writes $\mathrm{Cov}(f(X_{t_1}), g(X_{t_2})) = 0$ for all f, g with $\|f\|_\infty, \|g\|_\infty \le 1$, where $\|f\|_\infty$ denotes the supremum norm of f. For a sequence of dependent random variables, we would like that $\mathrm{Cov}(f(\text{'}\mathit{past}\text{'}), g(\text{'}\mathit{future}\text{'}))$ is small when the distance between the past and the future is sufficiently large.

More precisely, for a real function $h : \mathbb{R}^u \to \mathbb{R}$ ($u \in \mathbb{N}^*$), define the Lipschitz modulus as

$$\mathrm{Lip}\, h = \sup_{(y_1,\ldots,y_u)\neq(x_1,\ldots,x_u)\in\mathbb{R}^u} \frac{|h(y_1,\ldots,y_u) - h(x_1,\ldots,x_u)|}{\|y_1 - x_1\| + \cdots + \|y_u - x_u\|}.$$

Definition 1 (Doukhan and Louhichi 1999) The process X is said to be (ε, Ψ)-weakly dependent if

$$\varepsilon(k) = \sup \frac{|\mathrm{Cov}(f(X_{s_1},\ldots,X_{s_u}), g(X_{t_1},\ldots,X_{t_v}))|}{\Psi(f,g)} \xrightarrow[k\to\infty]{} 0,$$

where the sup bound is relative to $u, v \ge 1$, $s_1 \le \cdots \le s_u \le t_1 \le \cdots \le t_v$ with $k = t_1 - s_u$, and $f : \mathbb{R}^u \to \mathbb{R}$, $g : \mathbb{R}^v \to \mathbb{R}$ satisfy $\mathrm{Lip}\, f, \mathrm{Lip}\, g < \infty$ and $\|f\|_\infty \le 1$, $\|g\|_\infty \le 1$.

The following distinct functions Ψ yield η- and λ-weak dependence coefficients:

$$\begin{aligned} \text{if } \Psi(f,g) &= u\mathrm{Lip}\, f + v\mathrm{Lip}\, g, \quad \text{then } \epsilon(k) = \eta(k),\\ &= u\mathrm{Lip}\, f + v\mathrm{Lip}\, g + uv\mathrm{Lip}\, f \cdot \mathrm{Lip}\, g, \quad \text{then } \epsilon(k) = \lambda(k). \end{aligned}$$

Note that λ-weak dependence includes η-weak dependence. The main feature of Definition 1 is that it incorporates a much wider range of classes of models than those that might be described through a mixing condition (i.e., α-mixing, β-mixing, ρ-mixing, ϕ-mixing, etc., see Doukhan 1994) or association condition (see Chaps. 1–3 in Dedecker et al. 2007). Limit theorems and very sharp results have been proved for this class of processes (see Chaps. 6–12 in Dedecker et al. 2007 for more information).

We now provide a nonexhaustive list of weakly dependent sequences with their weak-dependence properties. This will prove how wide the range of potential applications is.

Example 1

1. The Bernoulli shift with independent inputs $(\xi_t)_{t\in\mathbb{Z}}$ is defined as $X_t = H((\xi_{t-j})_{j\in\mathbb{Z}})$, $H:\mathbb{R}^{\mathbb{Z}}\to\mathbb{R}$, $(\xi_i)_{i\in\mathbb{Z}}$ i.i.d. The process $(X_t)_{t\in\mathbb{Z}}$ is η-weakly dependent with $\eta(k)=2\delta_{[k/2]}^{m\wedge 1}$ if $\mathbb{E}|X_t|^m<\infty$ and
$$\mathbb{E}\Big|H(\xi_j, j\in\mathbb{Z}) - H(\xi_j \mathbb{1}_{|j|<k}, j\in\mathbb{Z})\Big| \le \delta_k \downarrow 0 \quad (k\uparrow\infty).$$
Two following (causal) examples are given by:
   - The first-order autoregressive sequences with discrete innovations in $\{0, 1, \ldots, r-1\}$ given by (1). This process is not strongly mixing, but it is a η-weakly dependent process such that $\eta(k) = O(r^{-k})$.
   - The LARCH model with Rademacher i.i.d. inputs:
$$X_t = \xi_t(1 + aX_{t-1}), \qquad \mathbb{P}(\xi_t = \pm 1) = \frac{1}{2}. \tag{2}$$
If $a < 1$, there exists a unique stationary solution (see Dedecker et al. 2007). Doukhan et al. (2009) proved that if $a \in ((3-\sqrt{5})/2, 1/2]$, the stationary solution $X_t = \xi_t + \sum_{j\ge 1} a^j \xi_t \cdots \xi_{t-j}$ is not strongly mixing, but X is a η-weakly dependent process such that $\eta(k) = O(a^k)$.

2. If X is a GARCH(p,q) process or, more generally, a ARCH(∞) process such that $X_t = \rho_t\xi_t$ with $\rho_t^2 = b_0 + \sum_{j=1}^{\infty} b_j X_{t-j}^2$ for $t\in\mathbb{Z}$ and $\beta \equiv \sum_{j>0} b_j < 1$, then if
   - there exists $q>0$ such that $b_j = 0$ for $j>q$, then $(X_t)_{t\in\mathbb{Z}}$ is a η-weakly dependent process with $\eta(k) = O(e^{-ck})$ for some explicit $c\equiv c(q,\beta)>0$ (this is the case of ARCH(q) processes);
   - there exist $C>0$ and $\mu\in\,]0,1[$ such that $\forall j\in\mathbb{N}$, $0\le b_j\le C\mu^j$, then $(X_t)_{t\in\mathbb{Z}}$ is a η-weakly dependent process with $\eta(k) = O(e^{-c\sqrt{k}})$ for some explicit $c\equiv c(\mu,\beta)>0$;
   - there exist $C>0$ and $\nu>1$ such that $\forall j\in\mathbb{N}$, $0\le b_j\le Cj^{-\nu}$, then $(X_t)_{t\in\mathbb{Z}}$ is a η-weakly dependent process with $\eta(k) = O(k^{-c})$ for some explicit $c\equiv c(\nu,\beta)>0$.

   See Doukhan et al. (2006) for details.
3. If $(X_t)_{t\in\mathbb{Z}}$ is either a Gaussian or an associated process, then it is λ-weakly dependent, and
$$\lambda(k) = O\Big(\sup_{i\ge k}\big|\mathrm{Cov}(X_0, X_i)\big|\Big)$$
(see Doukhan and Louhichi 1999).

Other λ-weakly dependent processes are described in Doukhan and Wintenberger (2007) where it is proved that this dependence is essentially stable through Bernoulli shifts (consider, e.g., the LARCH(∞) models with bounded and dependent innovations). Other examples of weakly dependent processes may also be found in the monograph Dedecker et al. (2007) and in Doukhan et al. (2006).

2.2 Subsampling the distribution of converging statistics

To describe the present approach, let $(X_n)_{n\in\mathbb{Z}}$ be a strictly stationary real-valued sequence of random variables. Suppose that $S_n = s_n(X_1, \dots, X_n)$ is a statistic of interest that converges in distribution as n tends to infinity. Subsampling is used to approximate the sampling distribution of S_n, $\mathbb{K}_n(x) = \mathbb{P}(S_n \le x)$.

To obtain subsampling counterpart of S_n, we can use either the overlapping scheme of blocks of size b

$$Y_{b,i} = (X_{i+1}, \dots, X_{i+b}) \tag{3}$$

for $i = 1, \dots, N$ with $N = n - b$ (see Künsch 1989 and Politis and Romano 1994) or the nonoverlapping scheme

$$Y_{b,i} = (X_{(i-1)b+1}, \dots, X_{ib}) \tag{4}$$

for $i = 1, \dots, N$ with $N = \lfloor n/b \rfloor$ (see Carlstein 1986). In each collection of subsamples, computation of subsampling analogue of statistics is made for each subseries and used to construct subsampling estimators. We will see that the nonoverlapping scheme is more interesting when working with weakly dependent time series as it allows one to impose less restrictive dependence assumptions on the estimator.

In the remainder of the paper we consider a bandwidth $b \equiv (b_n)_{n\in\mathbb{N}}$ such that $b \to \infty$ and $\lim_{n\to\infty} n/b = \infty$.

To prove the asymptotic properties of subsampling estimators under the weak dependence condition, we assume that one of the two following conditions holds. The first condition assumes that $\mathbb{K}_n$ is a converging statistic in the sense that it has a non-degenerate limit denoted by $\mathbb{K}$ and that the density probability function of $\mathbb{K}$ exists.

C. 1 *Convergent statistics*: *suppose that there exists a positive sequence* $(r_b)_{b\in\mathbb{N}}$ *such that*

$$r_b = \sup_{x\in\mathbb{R}} \left|\mathbb{K}_b(x) - \mathbb{K}(x)\right| \xrightarrow[b\to\infty]{} 0, \qquad \|\mathbb{K}'\|_\infty < \infty, \tag{5}$$

where $\mathbb{K}'$ *denotes the density of the limit probability distribution function*.

The second one is a technical condition needed to derive uniform a.s. convergence results for the rough subsampling estimator. It controls the limiting variance and the higher order moments of the estimator.

C. 2 *Concentration condition*: *suppose that there exist suitable constants* c, $D(b) > 0$, $b = 1, 2, \dots,$ *such that*

$$\sup_{x\in\mathbb{R}} \mathbb{P}\big(S_b \in [x, x+z]\big) \le D(b) z^c \quad (\forall z > 0). \tag{6}$$

Unless the simple stationary Markov case for which the existence of a bounded transition probability density is enough to assert that $c = 1$, this condition is intricate to be established for more general setting (see Doukhan and Wintenberger 2007, 2008 for examples for which the concentration condition is satisfied).

Let us now define the subsampling estimators. We begin with a smooth subsampling estimator suitable for weakly dependent time series

$$\widetilde{\mathbb{K}}_{b,n}(x) = \frac{1}{N}\sum_{i=0}^{N-1}\varphi\left(\frac{S_{b,i}-x}{\epsilon_n}\right), \tag{7}$$

where $S_{b,i} = s_b(Y_{b,i})$, $\epsilon_n \downarrow 0$, and φ is the nonincreasing continuous function such that $\varphi = 1$ or 0 according to $x \le 0$ or $x \ge 1$ and which is affine between 0 and 1. The monotonicity of the function φ is essential to derive the uniform convergence of the estimator.

For completeness, we will also report asymptotic results for the rough subsampling estimator introduced and studied in Politis and Romano (1994)

$$\widehat{\mathbb{K}}_{b,n}(x) = \frac{1}{N}\sum_{i=0}^{N-1}\mathbb{1}_{\{S_{b,i}\le x\}}, \tag{8}$$

where $\mathbb{1}$ denotes the indicator function. It is worth noting that the rough subsampler is based on the indicator function, which is not a Lipschitz function.

Remark 1 Under the assumption of the convergent statistics (5), one can easily check that the bias of the smooth subsampling estimator is bounded in the following way:

$$\sup_{x\in\mathbb{R}}\left|\mathbb{E}\left[\widetilde{\mathbb{K}}_{b,n}(x)\right]-\mathbb{K}(x)\right| \le r_b + \epsilon_n\|\mathbb{K}'\|_\infty. \tag{9}$$

In order to prove either the uniform strong or weak laws of large numbers, we compute bounds for the absolute values of the p-centered moments of the subsampling estimators ($p \in \mathbb{N}$ and $p \ge 2$) defined by

$$\widetilde{\Delta}^{(p)}_{b,n}(x) = \left|\mathbb{E}\left[\widetilde{\mathbb{K}}_{b,n}(x) - \mathbb{E}\left[\widetilde{\mathbb{K}}_{b,n}(x)\right]\right]^p\right|,$$

$$\widehat{\Delta}^{(p)}_{b,n}(x) = \left|\mathbb{E}\left[\widehat{\mathbb{K}}_{b,n}(x) - \mathbb{E}\left[\widehat{\mathbb{K}}_{b,n}(x)\right]\right]^p\right|.$$

Indeed, for sufficiently large p such that $\sum_n \widetilde{\Delta}^{(p)}_{b,n}(x) < \infty$ and $\sum_n \widehat{\Delta}^{(p)}_{b,n}(x) < \infty$, Markov's inequality, together with the Borel–Cantelli lemma, allows us to conclude the a.s. convergence of the estimators. The uniformity with respect to x follows with the arguments of Dini's second theorem.

To characterize asymptotic properties of the subsampling estimators, we start by giving results for the smooth subsampler under convergence condition (5).

Theorem 1 (Smooth subsampler) *Assume that Condition* (C.1) *holds. Let* $\delta > 0$, $p \in \mathbb{N}^*$, *and* $L \equiv L(b) = \mathrm{Lip}\, s_b$. *Suppose moreover that respectively* the overlapping setting is used *and one of the following relations holds*:

$$\underline{\eta\text{-dependence}}:\quad \sum_{t=0}^{\infty}(t+1)^{p-2}\eta(t)<\infty,\qquad \frac{b}{n}\left[1\vee\frac{L}{\epsilon_n}\right]\prec n^{-\delta},\quad \textit{or}$$

$$\underline{\lambda\text{-dependence}}:\quad \sum_{t=0}^{\infty}(t+1)^{p-2}\lambda(t)<\infty,\qquad \frac{b}{n}\left[1\vee\frac{L}{\epsilon_n}\vee\frac{bL^2}{\epsilon_n^2}\right]\prec n^{-\delta},$$

or the nonoverlapping setting is used *and*

$$\underline{\eta\text{-dependence}}:\quad \sum_{t=0}^{n-1}(t+1)^{p-2}\eta(t)\prec b^{p-2},\qquad \frac{b}{n}\left[1\vee\frac{bL}{\epsilon_n}\right]\prec n^{-\delta},\quad \textit{or}$$

$$\underline{\lambda\text{-dependence}}:\quad \sum_{t=0}^{n-1}(t+1)^{p-2}\lambda(t)\prec b^{p-2},\qquad \frac{b}{n}\left[1\vee\frac{bL}{\epsilon_n}\vee\frac{bL^2}{\epsilon_n^2}\right]\prec n^{-\delta}.$$

Then

$$\widetilde{\Delta}_{b,n}^{(p)}(x)\prec n^{-\lfloor\frac{p}{2}\rfloor\delta}.$$

Hence, if $p/2\in\mathbb{N}$ *is such that* $p\delta>2$, *then*

$$\sup_{x\in\mathbb{R}}\left|\widetilde{\mathbb{K}}_{b,n}(x)-\mathbb{K}(x)\right|\xrightarrow[n\to\infty]{}0\quad a.s.$$

Note that if one of the above relations holds with $p=2$, it yields the same results but only with respect to the convergence in probability. The proof of Theorem 1 and all the other proofs of the results in this paper are given in Sect. 5.

Monitoring the smoothing parameter ε_n. The smoothing parameter ϵ_n needs to be chosen suitably. A natural way to choose ϵ_n is to find equilibrium between the square of bias of statistics and the limiting variance. The order of square of bias in (9) is given by $O(\epsilon_n^2+r_b^2)$. The order of variance is listed for each case as follows.

Overlapping setting:

$$\underline{\eta\text{-dependence:}}\quad \frac{b}{n}\left[1\vee\frac{L}{\epsilon_n}\right],$$

$$\underline{\lambda\text{-dependence:}}\quad \frac{b}{n}\left[1\vee\frac{L}{\epsilon_n}\vee\frac{bL^2}{\epsilon_n^2}\right].$$

Nonoverlapping setting:

$$\underline{\eta\text{-dependence:}}\quad \frac{b}{n}\left[1\vee\frac{bL}{\epsilon_n}\right],$$

$$\underline{\lambda\text{-dependence:}}\quad \frac{b}{n}\left[1\vee\frac{bL}{\epsilon_n}\vee\frac{bL^2}{\epsilon_n^2}\right].$$

Let us consider, as an example, the η-weak dependence case. A reasonable choice of the parameter ϵ_n is then $(bL/n)^{1/3}$ for the overlapping subsampling scheme, and it is $(b^2L/n)^{1/3}$ in the nonoverlapping case, respectively.

Choice of subsampling size b. Generally, there is no unique rule to optimally select bandwidth factor (see Nordman and Lahiri 2004, Bickel et al. 1997). An optimal choice of subsampling size could be made by finding equilibrium between square of bias and the limiting variance under each of subsampling schemes. A more detailed analysis is left for future research.

Choice of procedure. An important issue is also the choice of the subsampling scheme. It is clear that the overlapping scheme for smooth subsampling is a much more "expensive" procedure in terms of the assumptions needed to be imposed on the sample size and on the bandwidth sequence.

For completeness, we give results for the rough subsampling estimator by considering successively the convergence condition (5) and the concentration condition (6).

Theorem 2 (Rough subsampler) *Assume that Condition* (C.1) *holds*. *Let* $L \equiv L(b) = \operatorname{Lip} s_b$. *Suppose moreover that respectively* the overlapping setting is used *and one of the following relations holds*:

$$\underline{\eta\text{-dependence}}: \quad \sum_{t=0}^{\infty} \eta(t)^{\frac{1}{2}} < \infty, \qquad \lim_{n\to\infty} \frac{b}{n}\left[1 \vee \frac{L}{\sqrt{b}}\right] = 0, \quad \text{or}$$

$$\underline{\lambda\text{-dependence}}: \quad \sum_{t=0}^{\infty} \lambda(t)^{\frac{2}{3}} < \infty, \qquad \lim_{n\to\infty} \frac{b}{n}\left[1 \vee \left(\frac{L^4}{b}\right)^{\frac{1}{3}} \vee \left(\frac{L}{b}\right)^{\frac{2}{3}}\right] = 0,$$

or the nonoverlapping setting is used *and*

$$\underline{\eta\text{-dependence}}: \quad \sum_{t=0}^{\infty} \eta(t)^{\frac{1}{2}} < \infty, \qquad \lim_{n\to\infty} \frac{b}{n}\left[1 \vee \sqrt{b}L\right] = 0, \quad \text{or}$$

$$\underline{\lambda\text{-dependence}}: \quad \sum_{t=0}^{\infty} \lambda(t)^{\frac{2}{3}} < \infty, \qquad \lim_{n\to\infty} \frac{b}{n}\left[1 \vee \left(bL^2\right)^{\frac{2}{3}} \vee \left(bL^2\right)^{\frac{1}{3}}\right] = 0.$$

Then $\lim_{n\to\infty} \widehat{\Delta}^{(2)}_{b,n}(x) = 0$, *and*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left|\widehat{\mathbb{K}}_{b,n}(x) - \mathbb{K}(x)\right| = 0 \quad \text{in probability.}$$

Almost sure convergence results for the rough subsampling estimator are obtained at the price of more restrictive conditions.

Theorem 3 (Rough subsampler) *Assume that Condition* (C.2) *holds*. *Let* $\delta > 0$, $p \in \mathbb{N}^*$, $L \equiv L(b) = \operatorname{Lip} s_b$, *and* $D \equiv D(b)$. *Suppose moreover that respectively* the overlapping setting is used *and one among the following relations hold*:

$$\underline{\eta\text{-dependence}}: \quad \sum_{t=0}^{\infty}(t+1)^{p-2}\eta(t)^{\frac{2+c}{1+c}} < \infty,$$

$$\frac{b}{n}\left[1 \vee \left(\frac{DL^c}{b} \vee bL^{2+c}\right)^{\frac{1}{1+c}}\right] \prec n^{-\delta}, \quad \textit{or}$$

$$\underline{\lambda\text{-dependence}}: \quad \sum_{t=0}^{\infty}(t+1)^{p-2}\lambda(t)^{\frac{1+c}{2+c}} < \infty,$$

$$\frac{b}{n}\left[1 \vee \left(\frac{(DL^c)^2}{b^{2-c}} \vee \frac{DL^c}{b^2} \vee \frac{DL^{2c}}{b^{2-c}}\right)^{\frac{1}{2+c}}\right] \prec n^{-\delta},$$

or the nonoverlapping setting is used and

$$\underline{\eta\text{-dependence}}: \quad \sum_{t=0}^{n-1}(t+1)^{p-2}\eta(t)^{\frac{2+c}{1+c}} \prec b^{p-2},$$

$$\frac{b}{n}\left[1 \vee \left(D(bL)^c \vee (bL)^{2+c}\right)^{\frac{1}{1+c}}\right] \prec n^{-\delta}, \quad \textit{or}$$

$$\underline{\lambda\text{-dependence}}: \quad \sum_{t=0}^{n-1}(t+1)^{p-2}\lambda(t)^{\frac{1+c}{2+c}} \prec b^{p-2},$$

$$\frac{b}{n}\left[1 \vee \left(\left(D(bL)^c\right)^2 \vee D(bL)^c \vee D(bL)^{2c}\right)^{\frac{1}{2+c}}\right] \prec n^{-\delta}.$$

Then

$$\widehat{\Delta}_{b,n}^{(p)}(x) \prec n^{-\lfloor \frac{p}{2} \rfloor \delta}.$$

Hence, if $p/2 \in \mathbb{N}$ *is such that* $p\delta > 2$, *then*

$$\sup_{x\in\mathbb{R}}\left|\widehat{\mathbb{K}}_{b,n}(x) - \mathbb{K}(x)\right| \xrightarrow[n\to\infty]{} 0 \quad a.s.$$

3 Subsampling the distribution of extremes

Bertail et al. (2004) studied subsampling estimators for distributions of diverging statistics but imposed that the time series is strongly mixing. We aim at adapting their results for weakly dependent sequences. Instead of considering the general case, we focus on the sample maximum because we are able to give sufficient conditions such that the normalized sample maximum converges in distribution under the weak dependence assumption. Note however that the results can easily be generalized, provided that it is possible to compute the Lipschitz coefficient of the function used to define the diverging statistics.

3.1 Convergence of the sample maximum

We first discuss conditions for convergence in distribution of the normalized sample maximum of a weakly dependent sequence.

Let $x_F = \sup\{x : F(x) < 1\}$ be the upper end point of F, and $\bar{F} := 1 - F$. We say that the stationary distribution F is in the maximum domain attraction of the generalized extreme value distribution with index γ, $-\infty < \gamma < \infty$, if there exists a positive measurable function g such that for $1 + \gamma x > 0$,

$$\lim_{u \to x_F} \bar{F}\big(u + xg(u)\big)/\bar{F}(u) = (1 + \gamma x)^{-1/\gamma}.$$

Then there exist sequences $(u_n)_{n \geq 1}$ and $(v_n)_{n \geq 1}$ such that $u_n > 0$ and

$$\lim_{n \to \infty} F^n\big(w_n(x)\big) = G_\gamma(x) := \begin{cases} \exp(-(1+\gamma x)_+^{-1/\gamma}) & \text{if } \gamma \neq 0, \\ \exp(-\exp(-x)) & \text{if } \gamma = 0, \end{cases} \tag{10}$$

where $w_n(x) = x/u_n + v_n$. Let $q(t) = F^{\leftarrow}(1 - t^{-1})$ where $F^{\leftarrow}$ is the generalized inverse of F. Then $(u_n)_{n \geq 1}$ and $(v_n)_{n \geq 1}$ can be chosen as

$$v_n = q(n),$$
$$u_n^{-1} = \begin{cases} (-\gamma)(x_F - q\,(n)) & \text{if } \gamma < 0, \\ q\,(ne) - q\,(n) & \text{if } \gamma = 0, \\ \gamma q\,(n) & \text{if } \gamma > 0. \end{cases}$$

Let us introduce the extremal dependence coefficient

$$\beta_{n,l} = \sup\big|\mathbb{P}\big(X_i \leq w_n(x), i \in A \cup B\big) - \mathbb{P}\big(X_i \leq w_n(x), i \in A\big)\mathbb{P}\big(X_i \leq w_n(x), i \in B\big)\big|,$$

where the sup bound is relative to sets A and B such that $A \subset \{1, \dots, k\}$, $B \subset \{k + l, \dots, n\}$, and $1 \leq k \leq n - l$.

O'Brien (1987) gave sufficient conditions such that the normalized sample maximum $u_n(M_n - v_n)$ converges in distribution when the stationary distribution is in the maximum domain attraction of some extreme value distributions.

Theorem 4 O'Brien (1987) *assume that F is in the maximum domain attraction of the extreme value distribution with index γ. Let (a_n) be a sequence of positive integers such that $a_n = o\,(n)$ as $n \to \infty$ and*

$$\lim_{n \to \infty} \frac{\mathbb{P}(M_{a_n} > w_n(x))}{a_n \bar{F}(w_n(x))} = \theta \in (0, 1]. \tag{11}$$

Assume that there exists a sequence (l_n) of positive integers such that

$$l_n = o\,(a_n) \quad \text{and} \quad \frac{n}{a_n}\beta_{n,l_n} \to_{n \to \infty} 0 \quad \text{as } n \to \infty. \tag{12}$$

Then

$$\lim_{n \to \infty} \mathbb{P}\big(M_n \leq w_n(x)\big) = \begin{cases} \exp(-\theta(1+\gamma x)_+^{-1/\gamma}) & \text{if } \gamma \neq 0, \\ \exp(-\theta\exp(-x)) & \text{if } \gamma = 0. \end{cases} \tag{13}$$

The constant θ is referred to as the extremal index of X (see Leadbetter et al. 1983). Note that any $l_n = o(n)$ such that $\beta_{n,l_n} \to 0$ as $n \to \infty$ can be used in constructing a sequence a_n such that (12) is satisfied by taking a_n equal to the integer part of $\max(n\beta_{n,l_n}^{1/2}, (nl_n)^{1/2})$. The condition $\beta_{n,l_n} \to 0$ as $n \to \infty$ is known as the $\mathcal{D}(w_n)$ condition (see Leadbetter 1974).

We provide an equivalent theorem when X is assumed to be either η- or λ-weakly dependent.

Theorem 5 *Assume that F is an absolutely continuous distribution in the maximum domain of attraction of the extreme value distribution with index γ such that for $1 + \gamma x > 0$,*

$$\lim_{n\to\infty} \frac{\partial}{\partial x} F^n\big(w_n(x)\big) = \frac{\partial}{\partial x} G_\gamma(x). \tag{14}$$

Let (a_n) be a sequence of positive integers such that $a_n = o(n)$ as $n \to \infty$ and (11) holds. Assume that there exists a sequence (l_n) of positive integers such that $l_n = o(a_n)$ $(n \to \infty)$. If X is η-weakly dependent and

$$\frac{n}{a_n}\big(n\eta(l_n)u_n\big)^{1/2} \to_{n\to\infty} 0,$$

or if X is λ-weakly dependent and

$$\frac{n}{a_n}\big(\big[n\lambda(l_n)u_n\big]^{1/2} \vee \big[na_n\lambda(l_n)u_n^2\big]^{1/3}\big) \to_{n\to\infty} 0,$$

then (13) holds.

The assumption that F is an absolutely continuous distribution is a strong assumption that is needed to take into account the weak dependence properties of the time series. Of course this assumption can be relaxed when only considering the convergence of the normalized sample maximum: one may consider the example of an integer-valued moving average sequence given in Hall et al. (2010), where it is shown that the limiting distribution of the normalized maxima is the same distribution as it would be obtained in the continuous case. This result has to be linked with the fact that the stationary distribution is heavy-tailed (see Andersson 1970 for further details).

3.2 Subsampling the distribution of the normalized sample maximum

Consider the sequence of extreme statistics

$$M_n = m_n(X_1, \ldots, X_n) = \max_{1\le i\le n} X_i.$$

Set $\mathbb{H}_n(x) = \mathbb{P}(M_n \le x)$. Restate the smooth subsampling estimates for nonnormalized extremes by

$$\widetilde{\mathbb{H}}_{b,n}(x) = \frac{1}{N}\sum_{i=0}^{N-1} \varphi\left(\frac{m_b(Y_{b,i}) - x}{\epsilon_n}\right). \tag{15}$$

Assume, under the assumption of Theorem 5, that (5) adapted to normalized extremes holds, i.e.,

$$r_n = \sup_{x\in\mathbb{R}} \left|\mathbb{H}_n\big(w_n(x)\big) - \mathbb{H}(x)\right| \to_{n\to\infty} 0,$$

where $\mathbb{H} = G_\gamma^\theta$.

Following the lines of Bertail et al. (2004), we have to impose conditions on the median and the distance between two quantiles of the limiting distribution in order to be able to identify it. The median of the limiting distribution to estimate is assumed to be equal to 0, and the distance between the quantiles is assumed to be equal to 1. Fix $0 < t_1 < t_2 < 1$. Then the normalizing sequences can be estimated by

$$\widetilde{v}_{b,n} = \widetilde{\mathbb{H}}_{b,n}^{\leftarrow}\left(\frac{1}{2}\right), \qquad \widetilde{u}_{b,n} = \left|\widetilde{\mathbb{H}}_{b,n}^{\leftarrow}(t_2) - \widetilde{\mathbb{H}}_{b,n}^{\leftarrow}(t_1)\right|^{-1}. \tag{16}$$

Let $C = \mathbb{H}^{\leftarrow}(t_2) - \mathbb{H}^{\leftarrow}(t_1)$. Using that $\mathrm{Lip}\, m_b = 1$, from Theorem 4 in Bertail et al. (2004) and Theorem 1 we derive the following theorem.

Theorem 6 *Assume that the conditions of Theorem 5 hold. Let $\delta > 0$ and $p \in \mathbb{N}^*$. The relation $|\mathbb{E}[\widetilde{\mathbb{H}}_{b,n}(w_b(x)) - \mathbb{E}[\widetilde{\mathbb{H}}_{b,n}(w_b(x))]]^p| \prec n^{-\lfloor \frac{p}{2} \rfloor \delta}$ holds if we assume that $\lim_{n\to\infty} \epsilon_n u_b = 0$ and respectively that*

- in the overlapping case,

$$\underline{\eta\text{-weak dependence}}: \quad \sum_{t=0}^{\infty} (t+1)^{p-2}\eta(t) < \infty, \qquad \frac{b}{n^{1-\delta}\epsilon_n} \prec 1, \quad or$$

$$\underline{\lambda\text{-weak dependence}}: \quad \sum_{t=0}^{\infty} (t+1)^{p-2}\lambda(t) < \infty, \qquad \frac{b}{n^{1-\delta}\epsilon_n^2} \prec 1,$$

- in the nonoverlapping case,

$$\underline{\eta\text{-weak dependence}}: \quad \sum_{t=0}^{n-1} (t+1)^{p-2}\eta(t) \prec b^{p-2}, \qquad \frac{b^2}{n^{1-\delta}\epsilon_n} \prec 1, \quad or$$

$$\underline{\lambda\text{-weak dependence}}: \quad \sum_{t=0}^{n-1} (t+1)^{p-2}\lambda(t) \prec b^{p-2}, \qquad \frac{b^2}{n^{1-\delta}\epsilon_n^2} \prec 1.$$

Hence, if $p/2 \in \mathbb{N}$ is such that $p\delta > 2$, then

$$\sup_{x\in\mathbb{R}} \left|\widetilde{\mathbb{H}}_{b,n}\left(\widetilde{v}_{b,n} + \frac{x}{\widetilde{u}_{b,n}}\right) - \mathbb{H}\left(\mathbb{H}^{\leftarrow}\left(\frac{1}{2}\right) + Cx\right)\right| \to_{n\to\infty} 0 \quad a.s.$$

Choice of procedure. In Robert et al. (2009) the question of the choice of the subsampling scheme for the estimator of the distribution of the sample maximum is partially addressed. The authors are interested in the estimation of the extremal index which measures the degree of clustering of extremes and consider 'disjoint' (nonoverlapping) and 'sliding' (overlapping) blocks estimators which are based on the estimation of the distribution of the sample maximum. They compare their asymptotic properties and show that the sliding blocks estimator is more efficient than the disjoint version and has a smaller asymptotic bias.

4 Simulation study

The finite sample properties of our subsampling estimators are now compared in a simulation study. We consider both rough and smooth subsampling estimators when they are computed with the overlapping or nonoverlapping schemes.

Sequences of length $n = 2000$ and $n = 5000$ have been simulated from the first-order autoregressive process of Example (1),

$$X_t = \frac{1}{r}(X_{t-1} + \varepsilon_t),$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ are iid and uniformly distributed on the set $\{0, 1, \ldots, r-1\}$, and r is equal to 3. It is well known that the asymptotic condition (13) holds with $\gamma = -1$, $\theta = r^{-1}(r-1)$, $u_n = n$, and $v_n = 1 - n^{-1}$. Following the approach presented in the previous Sect. 3.2, we have to fix conditions on the median and two quantiles of the limiting distribution. We choose $t_1 = 1/4$ and $t_2 = 3/4$. The limiting distribution becomes

$$\mathbb{K}(x) = e^{-\theta(1-(x-d)/c)}, \quad x \le c + d,$$

where $c = \theta/\ln 3$ and $d = (\ln 2 - \theta)/\ln 3$. The normalization coefficients $\bar{u}_n$ and $\bar{v}_n$ such that

$$\lim_{n\to\infty} \mathbb{P}\big(\bar{u}_n(M_n - \bar{v}_n) \le x\big) = \mathbb{K}(x)$$

are given by

$$\bar{v}_n = v_n - c^{-1}u_n^{-1}d, \qquad \bar{u}_n = cu_n. \tag{17}$$

We first simulate a sequence of length $n = 2000$ and plot the estimators of the limiting distribution in Fig. 1. As expected, smoothing estimators yield smoother curves. The differences between the estimators are small, but the smoothed versions need less strong assumptions for the convergence. Moreover note that, contrary to the last remark at the end of the previous subsection, nothing in those simulations seems to make us prefer the overlapping scheme to the nonoverlapping one.

Monte Carlo approximations to the quantiles and the means of the estimators have been then computed from 1000 simulated sequences.

The properties of our rough and smooth subsampling estimators computed with the nonoverlapping scheme are shown in the two upper graphs in Fig. 2. There are very few differences between both estimators according to their quantiles and their means. Their biases are negligible for all the values of x. The confidence intervals

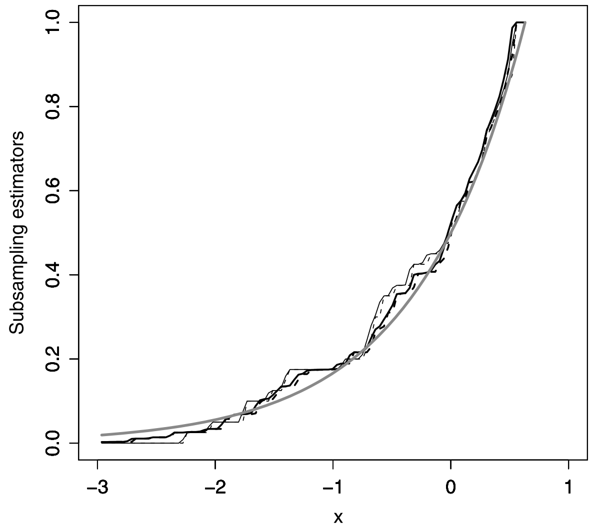


Fig. 1 AR(1) process. The rough (*dashed*) and smooth (*solid*) subsampling estimators computed with the nonoverlapping scheme (*thin*) and with the overlapping scheme (*thick*) for a sequence of length $n = 2000$, $b = 50$, $\epsilon = 0.05$. The asymptotic distribution is given by the *gray curve*

with level 90% (gray zone) vanish when x goes to 0 because 0 is the median of the empirical distribution and also the median of the asymptotic distribution. We may compare the quantiles and the means of our estimators with those obtained when the normalization coefficients given by (16) are replaced by the theoretical normalization coefficients given by (17) (see the two lower graphs in Fig. 2). First, note that the bias become negative when x is smaller than the median. Second, the confidence intervals are obviously not equal to zero for the median, but they are more narrow than the confidence intervals of our estimators when x is close to the extremal point of the asymptotic distribution, $c + d$.

The properties of our rough and smooth subsampling estimators computed with the overlapping scheme are shown in Fig. 3. We chose the same value for b as in the nonoverlapping scheme, and consequently the number of components in the definition of estimators is quite larger than in the other scheme. It follows that the empirical distribution functions given by the estimators computed with the overlapping scheme are smoother than those of the estimators computed with the nonoverlapping scheme. The confidence intervals are also a little bit more narrow.

Moreover, note that qualitatively similar results were found when the simulations were repeated with $n = 5000$, $b = 100$, $\epsilon = 0.05$.

Finally, sequences of length $n = 2000$ have also been simulated from the LARCH model with Rademacher i.i.d. inputs (see (2)) and with inputs that have a parabolic density probability function given by $x \mapsto 0.5(1 + \rho)|x|^\rho$ for $x \in [-1, 1]$. Note that the Rademacher distribution can be seen as the limit of the parabolic distribution as ρ goes to infinity. We choose $a = 0.4$. Hence the process is weakly dependent but not strong mixing when the inputs have a Rademacher distribution, and it is strong mixing when the distribution of the inputs is absolutely continuous. Neither the stationary distribution nor the extremal behavior of the processes are known. Note however that the end points of the stationary distributions are finite.

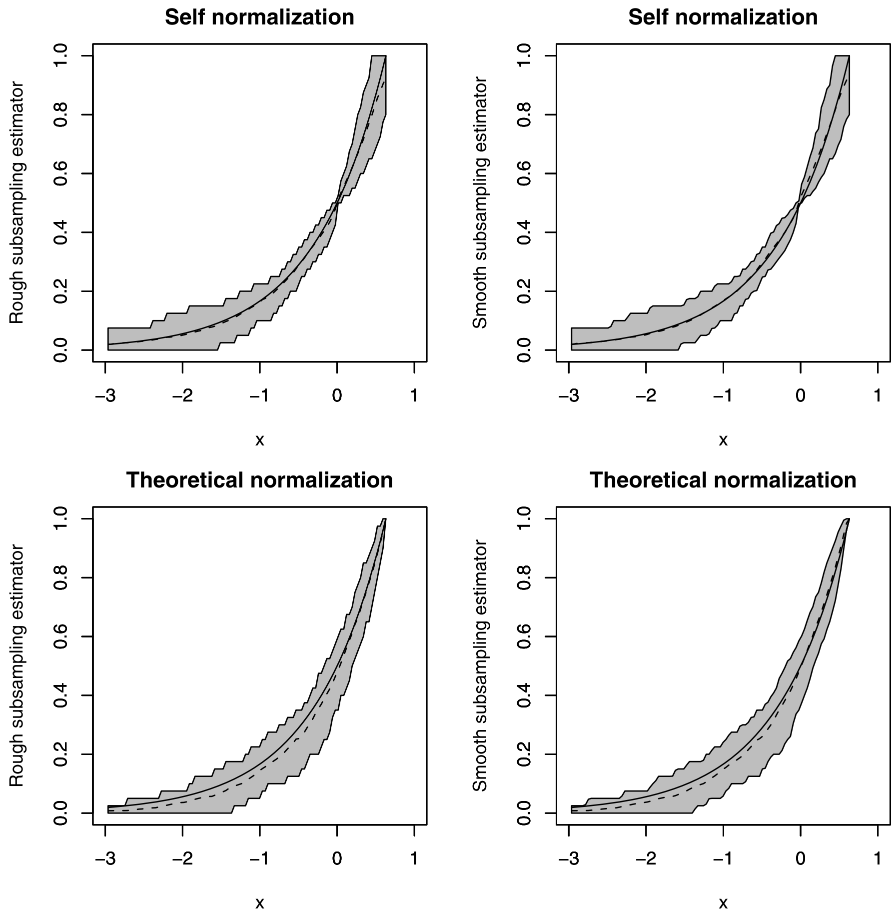


Fig. 2 AR(1) process. Monte Carlo approximations to the quantiles ($q_{0.05}$ and $q_{0.95}$) (*gray zone*) and means (*dashed line*) of the rough (*left*) and smooth (*right*) subsampling estimators computed with the nonoverlapping scheme when the normalization coefficients are given by (16) (*top*) or by (17) (*bottom*). The asymptotic distribution function, $\mathbb{K}$, is given by the solid line for a sequence of length $n = 2000$, $b = 50$, $\epsilon = 0.05$

We perform simulations and use our estimators. Results are given in Fig. 4. The shapes of the empirical distribution functions given by the estimators are different for the two processes (in particular for the large values of x). As far as we can see, the generalized extreme value distribution with a negative index could be a good choice to model the distribution of the maximum of the process with absolutely continuous inputs but not to model the distribution of the maximum of the process with Rademacher inputs. The study of the extremal behavior of these processes are intricate and left for future work.

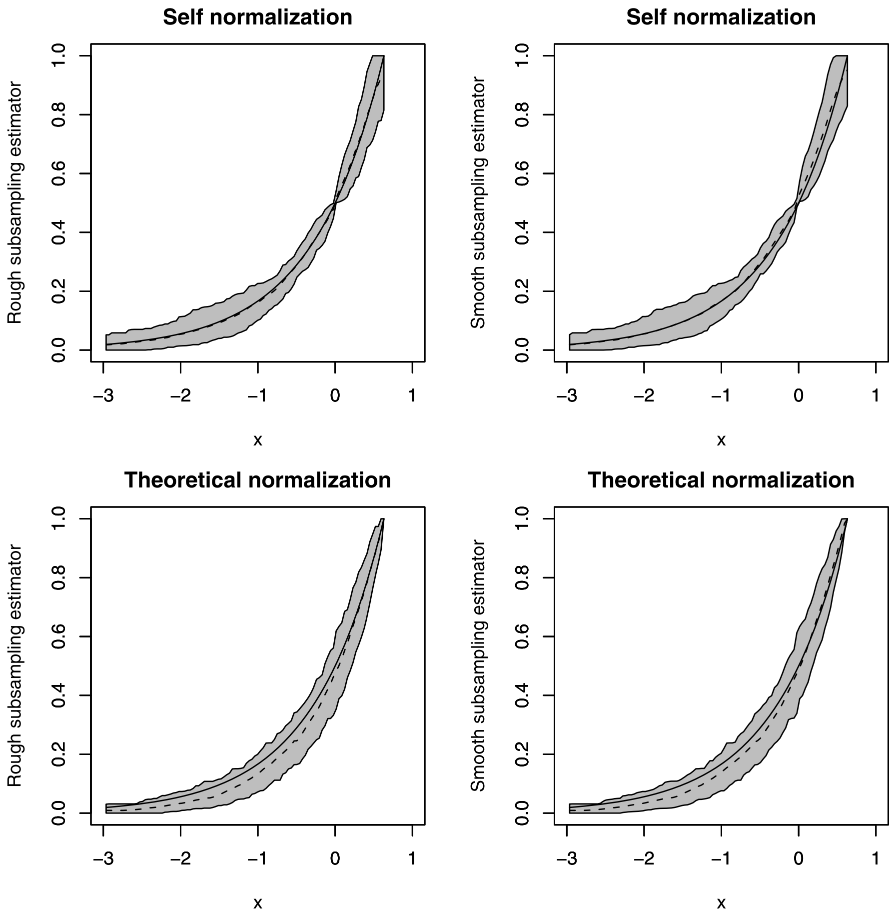


Fig. 3 AR(1) process. Monte Carlo approximations to the quantiles ($q_{0.05}$ and $q_{0.95}$) (*gray zone*) and means (*dashed line*) of the rough (*left*) and smooth (*right*) subsampling estimators computed with the overlapping scheme when the normalization coefficients are given by (16) (*top*) or by (17) (*bottom*). The asymptotic distribution function, $\mathbb{K}$, is given by the solid line for a sequence of length $n = 2000$, $b = 50$, $\epsilon = 0.05$

5 Proofs

5.1 Proofs for smooth subsampling

A bound of the expression $\widetilde{\Delta}_{b,n}^{(p)}(x)$ is closely related to the coefficients defined for $1 \leq q \leq p$ as

$$C_{b,q}(r) = \sup\left|\mathrm{Cov}(Z_{i_1} \cdots Z_{i_k}, Z_{i_{k+1}} \cdots Z_{i_q})\right|,$$

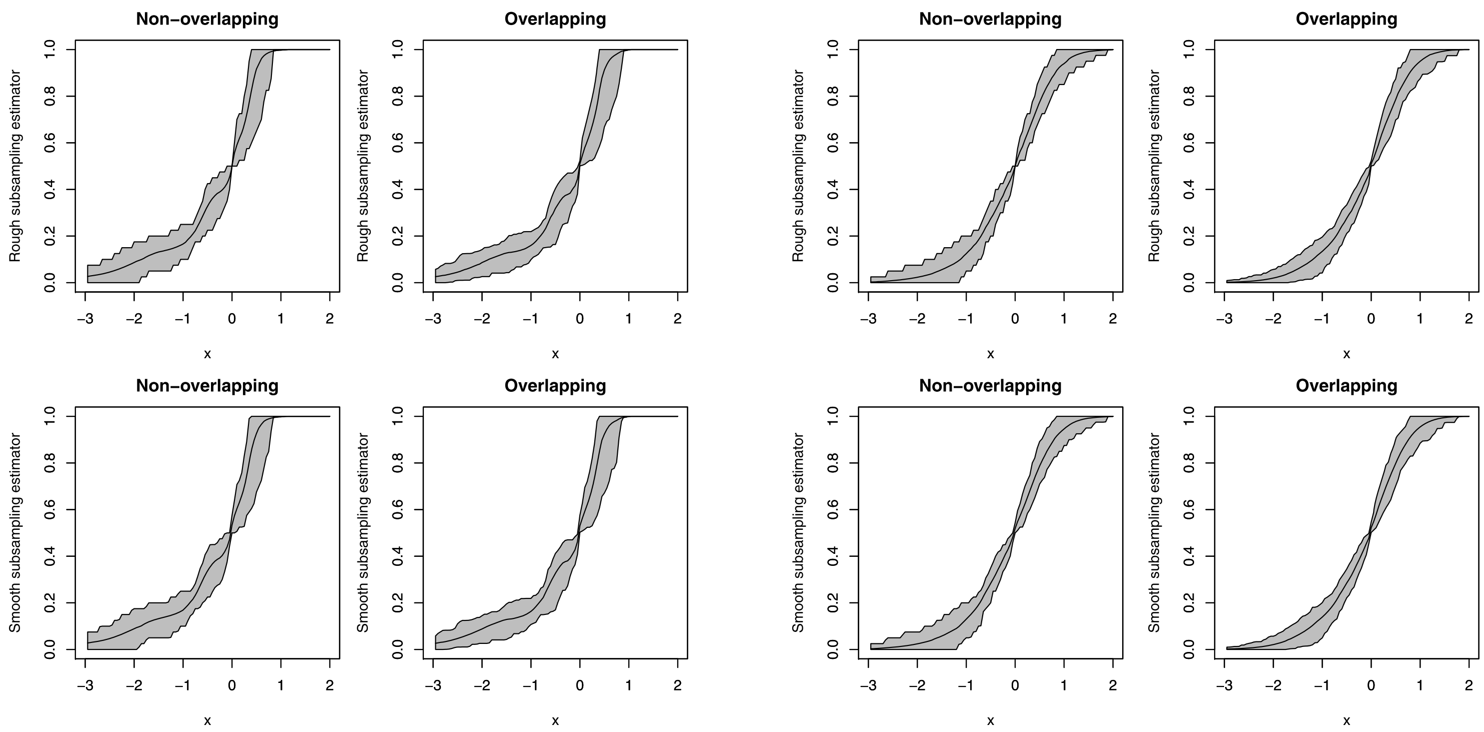


Fig. 4 LARCH processes with Rademacher inputs (the *four* graphs at *left*) and with parabolic inputs ($\rho = 4$) (the *four* graphs at *right*). Monte Carlo approximations to the quantiles ($q_{0.05}$ and $q_{0.95}$) (*gray zone*) and means (*solid line*) of the rough (*top*) and smooth (*bottom*) subsampling estimators computed with the nonoverlapping scheme (*left*) and with the overlapping scheme (*right*) for a sequence of length $n = 2000$, $b = 50$, $\epsilon = 0.05$

where the supremum refers to indices with $1 \le k < q$, $i_1 \le \cdots \le i_q$ satisfy $i_{k+1} - i_k = r$, and $Z_i = \varphi(\frac{S_{b,i}-x}{\epsilon_n}) - \mathbb{E}\varphi(\frac{S_{b,i}-x}{\epsilon_n})$ is a centered rv. Then setting

$$A_{b,q}(N) = \frac{1}{N^q} \sum_{1 \le i_1 \le \cdots \le i_q \le N} |\mathbb{E} Z_{i_1} \cdots Z_{i_q}|, \quad 2 \le q \le p,$$

Doukhan and Louhichi (1999) prove that $\widetilde{\Delta}_{b,n}^{(p)}(x) \le p! A_{b,p}(N)$. Moreover,

$$A_{b,p}(N) \le B_{b,p}(N) + \sum_{q=2}^{p-2} A_{b,q}(N) A_{b,p-q}(N),$$

$$B_{b,q}(N) = \frac{q-1}{N^{q-1}} \sum_{r=0}^{N-1} (r+1)^{q-2} C_{b,q}(r), \quad 2 \le q \le p.$$

Lemma 1 *Let p, q, b, N be integers, and $\beta(b, N) \le 1$. We assume that for all $2 \le q \le p$, there exists a constant $c_q \ge 0$ such that $B_{b,q}(N) \le c_q \beta^{\frac{q}{2}}(b, N)$. Then there exists a constant $C_p \ge 0$ only depending on p and $c_1, \ldots, c_p$ such that $A_{b,p}(N) \le C_p \beta^{\lfloor \frac{p}{2} \rfloor}(b, N)$.*

Proof of the Lemma 1 The result is the assumption if $p = 2$ because $A_{b,2}(N) \le B_{b,2}(N)$. If now the result has been proved for each $q < p$, the relation $\lfloor \frac{p}{2} \rfloor \le \lfloor \frac{q}{2} \rfloor + \lfloor \frac{p-q}{2} \rfloor$ completes the proof because $\beta(b, N) \le 1$. $\square$

The covariance $\mathrm{Cov}(f(Y_{i_1}, \ldots, Y_{i_u}), g(Y_{j_1}, \ldots, Y_{j_v}))$ writes respectively as

$$\begin{cases} \mathrm{Cov}\left(f_b\left((X_{i_h+k})_{\left\{\substack{1 \le h \le u \\ 1 \le k \le b}\right.}\right), g_b\left((X_{j_{h'}+k'})_{\left\{\substack{1 \le h' \le u \\ 1 \le k' \le b}\right.}\right)\right) \\ \mathrm{Cov}\left(f_b\left((X_{(i_h-1)b+k})_{\left\{\substack{1 \le h \le u \\ 1 \le k \le b}\right.}\right), g_b\left((X_{(j_{h'}-1)b+k'})_{\left\{\substack{1 \le h' \le u \\ 1 \le k' \le b}\right.}\right)\right) \end{cases}$$

for suitable functions f_b, g_b depending if the considered setting is the overlapping one or not. Moreover, $\mathrm{Lip}\, f_b \le \mathrm{Lip}\, f$, which proves that if the dependence coefficients relative to the sequences $X = (X_t)_{t \in \mathbb{Z}}$ are denoted by $\eta_X(r) = \eta(r)$ and $\eta_{Y_b}(r)$, then we get the following elementary lemma.

Lemma 2 (Heredity) *Assume that the stationary sequence $X = (X_t)_{t \in \mathbb{Z}}$ is weakly dependent. Then the same occurs for $Y_b = (Y_t)_{t \in \mathbb{Z}}$, and:*

- *$\eta_{Y_b}(r) \le b\eta(r-b)$ if $r \ge b$ in the overlapping case,*
- *$\lambda_{Y_b}(r) \le b^2 \lambda(r-b)$ if $r \ge b$ in the overlapping case,*
- *$\eta_{Y_b}(r) \le b\eta((r-1)b)$ if $r \ge 1$ in the nonoverlapping case,*
- *$\lambda_{Y_b}(r) \le b^2 \lambda((r-1)b)$ if $r \ge 1$ in the nonoverlapping case.*

In our setting we use the function $f(y_1, \ldots, y_b) = \varphi(\frac{s_b(y_1, \ldots, y_b) - x}{\epsilon_n})$, and the covariance inequalities write here as:

Lemma 3 *Using the overlapping and the nonoverlapping schemes, under the respective weak dependence assumptions η and λ, we respectively get*

- *in the overlapping case, $C_{b,q}(r) \prec 1$ for $r < b$, and else, respectively,*

$$C_{b,q}(r) \prec \frac{bL}{\epsilon_n}\eta(r-b), \quad or \quad C_{b,q}(r) \prec \frac{bL}{\epsilon_n}\left(1 \vee \frac{bL}{\epsilon_n}\right)\lambda(r-b);$$

- *in the nonoverlapping case, $C_{b,q}(r) \prec 1$ for $r = 0$, and else, respectively,*

$$C_{b,q}(r) \prec \frac{bL}{\epsilon_n}\eta\big((r-1)b\big), \quad or \quad C_{b,q}(r) \prec \frac{bL}{\epsilon_n}\left(1 \vee \frac{bL}{\epsilon_n}\right)\lambda\big((r-1)b\big).$$

This lemma entails the following bounds:

- *Overlapping and η-dependent case.* We obtain

$$\begin{aligned}
B_{b,q}(N) &\prec \frac{1}{N^{q-1}}\sum_{r=0}^{b-1}(r+1)^{q-2} + \frac{1}{N^{q-1}}\frac{bL}{\epsilon_n}\sum_{r=b}^{N-1}\frac{\eta(r-b)}{(r+1)^{2-q}} \\
&\prec \left(\frac{b}{N}\right)^{q-1}\left[1 + \frac{L}{\epsilon_n}\sum_{t=0}^{N-b-1}\eta(t)\right] + \frac{1}{N^{q-1}}\frac{bL}{\epsilon_n}\sum_{t=0}^{N-b-1}\frac{\eta(t)}{(t+1)^{2-q}},
\end{aligned}$$

where the second inequality follows from the change in variable $r = t + b$. We use here $N = n$. Now if we assume that $b \prec n^{1-\delta}$, we deduce that $(b/n)^{q-1} \prec n^{-\frac{q}{2}\delta}$, and if $bL \prec n^{1-\delta}\epsilon_n$, we analogously derive that $(b/N)^{q-1}(L/\epsilon_n) \prec n^{-\frac{q}{2}\delta}$. Assume now that

$$\frac{b}{n}\left[1 \vee \frac{L}{\epsilon_n}\right] \prec n^{-\delta};$$

with $\sum_{t=0}^{\infty}(t+1)^{q-2}\,\eta(t) < \infty$ this implies that $B_{b,q}(N) \prec n^{-\frac{q}{2}\delta}$.

- *Overlapping and λ-dependent case.* We obtain

$$\begin{aligned}
B_{b,q}(N) &\prec \frac{1}{N^{q-1}}\sum_{r=0}^{b-1}(r+1)^{q-2} + \frac{1}{N^{q-1}}\frac{bL}{\epsilon_n}\sum_{r=b}^{N-1}\frac{\lambda(r-b)}{(r+1)^{2-q}} \\
&\quad + \frac{1}{N^{q-1}}\frac{(bL)^2}{\epsilon_n^2}\sum_{r=b}^{N-1}\frac{\lambda(r-b)}{(r+1)^{2-q}} \\
&\prec \left(\frac{b}{N}\right)^{q-1}\left[1 + \frac{L}{\epsilon_n}\sum_{t=0}^{N-b-1}\lambda(t) + \frac{bL^2}{\epsilon_n^2}\sum_{t=0}^{N-b-1}\lambda(t)\right] \\
&\quad + \frac{1}{N^{q-1}}\left[\frac{bL}{\epsilon_n}\sum_{t=0}^{N-b-1}\frac{\lambda(t)}{(t+1)^{2-q}} + \frac{(bL)^2}{\epsilon_n^2}\sum_{t=0}^{N-b-1}\frac{\lambda(t)}{(t+1)^{2-q}}\right],
\end{aligned}$$

where the second inequality follows from the change in variable $r = t + b$. We use here $N = n$. Now if we assume that $b \prec n^{1-\delta}$, we deduce that $(b/n)^{q-1} \prec n^{-\frac{q}{2}\delta}$. Now if $(bL \vee (bL)^2) \prec n^{1-\delta}(\epsilon_n \vee \epsilon_n^2)$, we analogously derive that $(b/N)^{q-1}(L/\epsilon_n \vee bL^2/\epsilon_n^2) \prec n^{-\frac{q}{2}\delta}$. Assume now that

$$\frac{b}{n}\left[1 \vee \frac{L}{\epsilon_n} \vee \frac{bL^2}{\epsilon_n^2}\right] \prec n^{-\delta};$$

with $\sum_{t=0}^{\infty} (t+1)^{q-2}\, \lambda(t) < \infty$ this implies that $B_{b,q}(N) \prec n^{-\frac{q}{2}\delta}$.

- *Nonoverlapping and η-dependent case*. We obtain

$$\begin{aligned} B_{b,q}(N) &\prec \frac{1}{N^{q-1}} + \frac{1}{N^{q-1}} \frac{bL}{\epsilon_n} \sum_{r=1}^{N-1} \frac{\eta((r-1)b)}{(r+1)^{2-q}} \\ &\prec \frac{1}{N^{q-1}}\left[1 + \frac{bL}{\epsilon_n} \sum_{k=1}^{n-1} \eta(k)\right] + \frac{1}{N^{q-1}} \frac{b^{3-q}L}{\epsilon_n} \sum_{k=1}^{n-1} \frac{\eta(k)}{k^{2-q}}, \end{aligned}$$

where the second inequality follows from replacement of $k = b(r-1)$. We use here $N = n/b$. Now if we assume that $b \prec n^{1-\delta}$, we deduce that $(b/n)^{q-1} \prec n^{-\frac{q}{2}\delta}$, and if $b^2 L \prec n^{1-\delta}\epsilon_n$, we analogously derive that $(1/N^{q-1})(bL/\epsilon_n) \prec n^{-\frac{q}{2}\delta}$. Assume now that

$$\frac{b}{n}\left[1 \vee \frac{bL}{\epsilon_n}\right] \prec n^{-\delta};$$

with $\sum_{t=0}^{n-1}(t+1)^{q-2}\eta(t) < b^{q-2}$ this implies that $B_{b,q}(N) \prec n^{-\frac{q}{2}\delta}$.

- *Nonoverlapping and λ-dependent case*. We obtain

$$\begin{aligned} B_{b,q}(N) &\prec \frac{1}{N^{q-1}} + \frac{1}{N^{q-1}} \frac{bL}{\epsilon_n} \sum_{r=b}^{N-1} \frac{\lambda((r-1)b)}{(r+1)^{2-q}} + \frac{1}{N^{q-1}} \frac{(bL)^2}{\epsilon_n^2} \sum_{r=b}^{N-1} \frac{\lambda((r-1)b)}{(r+1)^{2-q}} \\ &\prec \left(\frac{1}{N}\right)^{q-1}\left[1 + \frac{bL}{\epsilon_n} \sum_{k=b}^{n-1} \lambda(k) + \frac{bL^2}{\epsilon_n^2} \sum_{k=b}^{n-1} \lambda(k)\right] \\ &\quad + \frac{1}{N^{q-1}}\left[\frac{b^{3-q}L}{\epsilon_n} \sum_{k=b}^{n-1} \frac{\lambda(k)}{k^{2-q}} + \frac{b^{5-q}L^2}{\epsilon_n^2} \sum_{k=b}^{n-1} \frac{\lambda(k)}{k^{2-q}}\right], \end{aligned}$$

where the second inequality follows from replacement of $k = b(r-1)$. We use here $N = n/b$. Now if we assume that $b \prec n^{1-\delta}$, we deduce that $(b/n)^{q-1} \prec n^{-\frac{q}{2}\delta}$. Now if $(b^2 L \vee (bL)^2) \prec n^{1-\delta}(\epsilon_n \vee \epsilon_n^2)$, we analogously derive that $(1/N^{q-1})(bL/\epsilon_n \vee bL^2/\epsilon_n^2) \prec n^{-\frac{q}{2}\delta}$. Assume now that

$$\frac{b}{n}\left[1 \vee \frac{bL}{\epsilon_n} \vee \frac{bL^2}{\epsilon_n^2}\right] \prec n^{-\delta};$$

with $\sum_{t=0}^{n-1}(t+1)^{q-2}\lambda(t) < (b^{q-2} \vee b^{q-4})$ this implies that $B_{b,q}(N) \prec n^{-\frac{q}{2}\delta}$.

Lemma 4 *The relation $B_{b,q}(N) \prec n^{-\frac{q}{2}\delta}$ holds in the following cases*:

- *In the overlapping case, if we have respectively*

$$\sum_{t=0}^{\infty}(t+1)^{q-2}\eta(t) < \infty, \quad \textit{and} \quad \frac{b}{n}\left[1 \vee \frac{L}{\epsilon_n}\right] \prec n^{-\delta},$$
$$\sum_{t=0}^{\infty}(t+1)^{q-2}\lambda(t) < \infty, \quad \textit{and} \quad \frac{b}{n}\left[1 \vee \frac{L}{\epsilon_n} \vee \frac{bL^2}{\epsilon_n^2}\right] \prec n^{-\delta}.$$

- *In the nonoverlapping case, if we have respectively*

$$\sum_{t=0}^{n-1}(t+1)^{q-2}\eta(t) < b^{q-2}, \quad \textit{and} \quad \frac{b}{n}\left[1 \vee \frac{bL}{\epsilon_n}\right] \prec n^{-\delta},$$
$$\sum_{t=0}^{n-1}(t+1)^{q-2}\lambda(t) < \left[b^{q-2} \vee b^{q-4}\right), \quad \textit{and} \quad \frac{b}{n}\left[1 \vee \frac{bL}{\epsilon_n} \vee \frac{bL^2}{\epsilon_n^2}\right] \prec n^{-\delta}.$$

This lemma, together with Lemma 1, yields the main theorem.

5.2 Proofs for rough subsampling

In this section we shall replace ϵ_n by some $z > 0$ to be settled later, and we set $\varphi_z(t) = \varphi(\frac{t-x}{z})$. We now set $Z_i = \mathbb{1}_{\{S_{b,i} \le x\}} - \mathbb{P}(S_{b,i} \le x)$ and $W_i = \varphi_z(S_{b,i}) - \mathbb{E}\varphi_z(S_{b,i})$. A usual trick yields:

$$\left|\mathrm{Cov}(Z_{i_1}\cdots Z_{i_k}, Z_{i_{k+1}}\cdots Z_{i_q})\right| \le \left|\mathrm{Cov}(W_{i_1}\cdots W_{i_k}, W_{i_{k+1}}\cdots W_{i_q})\right| + 2\sum_{h=1}^{p}\mathbb{E}|W_{i_h} - Z_{i_h}| = U + V$$

with $U = |\mathrm{Cov}\,(W_{i_1}\cdots W_{i_k}, W_{i_{k+1}}\cdots W_{i_q})|$ and $V = 2p\,\mathbb{P}(S_{b,i} \in [x, x+z])$.

A bound for V does not depend on the overlapping or not overlapping case, and we get

$$\begin{cases} V \le 2p(G\,\mathrm{Lip}\,S_b z + r_b) \prec Lz + r_b, & \text{under assumption (5)},\\ V \le 2pDz^c \prec Dz^c, & \text{under assumption (6)}.\end{cases}$$

Set here $A_{p,b,\epsilon} = bp\mathrm{Lip}\,S_b/z \prec bL/z$. The bound of U needs four cases (considered in Lemma 3) with

- in the overlapping case,

$$U \prec \begin{cases} \frac{bL}{z}\eta(r-b) & \text{for } r \ge b,\\ 1 & \text{for } r < b;\end{cases}$$

- in the overlapping case,

$$U \prec \begin{cases} \frac{bL}{z}(1 \vee \frac{bL}{z})\lambda(r-b) & \text{for } r \geq b, \\ 1 & \text{for } r < b; \end{cases}$$

- in the nonoverlapping case,

$$U \prec \begin{cases} \frac{bL}{z}\eta((r-1)b) & \text{for } r \geq 1, \\ 1 & \text{for } r = 0; \end{cases}$$

- in the nonoverlapping case,

$$U \prec \begin{cases} \frac{bL}{z}(1 \vee \frac{bL}{z})\lambda((r-1)b) & \text{for } r \geq 1, \\ 1 & \text{for } r = 0. \end{cases}$$

We first derive the inequality $(t+1+b)^{q-2} \leq 2^{(q-3)} \vee 1\{(t+1)^{q-2} + b^{q-2}\}$ from convexity if $q > 3$ and sublinearity else, and thus,

$$(t+1+b)^{q-2} \prec (t+1)^{q-2} + b^{q-2}.$$

The coefficients $C_{b,q}(r) \prec \sup\{U + V\}$ may thus be bounded in all the considered cases.

For simplicity, we classify the cases with couples of numbers indicating the fact overlapping (3) or not (4) setting is used and from the fact the convergence (5) or concentration (6) is assumed, which makes four different cases to consider). Consider the cases under assumption (5).

– η (3), (5) *case*. Note that

$$C_{b,q}(r) \prec L\big(b\eta(r-b)/z + z\big) + r_b \prec L\sqrt{b\eta(r-b)} + r_b$$

with the choice $z = \sqrt{b\eta(r-b)}$. This yields

$$\begin{aligned} B_{b,q}(N) &\prec \frac{1}{N^{q-1}}\sum_{r=0}^{b-1}(r+1)^{q-2} + \frac{L\sqrt{b}}{N^{q-1}}\sum_{r=b}^{N-1}\frac{\sqrt{\eta(r-b)}}{(r+1)^{2-q}} + r_b \\ &\prec \left(\frac{b}{N}\right)^{q-1}\left[1 + \frac{L}{\sqrt{b}}\sum_{t=0}^{N-b-1}\sqrt{\eta(t)}\right] + \frac{L\sqrt{b}}{N^{q-1}}\sum_{t=0}^{N-b-1}\frac{\sqrt{\eta(t)}}{(t+1)^{2-q}} + r_b, \end{aligned}$$

where the second inequality follows from the change in variable $r = t + b$. We use here $N = n$. Now if we assume that $b \prec n^{1-\delta}$, we deduce that $(b/n)^{q-1} \prec n^{-q\delta/2}$, and if $bL \prec n^{1-\delta}$, we analogously derive that $(b/N)^{q-1}(L/\sqrt{b}) \prec n^{-\frac{q}{2}\delta}$. If $\eta(t) \prec n^{-\eta}$ and $\sigma \leq \eta/2$, we assume that

$$\frac{b}{n}\left[1 \vee \frac{L}{\sqrt{b}}\right] + r_b \prec n^{-\delta}, \qquad \sum_{t=0}^{\infty}(t+1)^{q-2}\eta(t)^{1/2} < \infty.$$

– η (4), (5) *case*. Note that

$$C_{b,q}(r) \prec bL\eta\big((r-1)b\big)/z + Lz + r_b \prec L\sqrt{b\eta\big((r-1)b\big)} + r_b,$$

where we use $z = \sqrt{b\eta((r-1)b)}$. Then

$$\begin{aligned}
B_{b,q}(N) &\prec \frac{1}{N^{q-1}} + \frac{L\sqrt{b}}{N^{q-1}}\sum_{r=1}^{N-1}\frac{\sqrt{\eta((r-1)b)}}{(r+1)^{2-q}} + r_b \\
&\prec \frac{1}{N^{q-1}}\left[1 + L\sqrt{b}\sum_{k=b}^{n-1}\sqrt{\eta(k)}\right] + \frac{Lb^{\frac{5}{2}-q}}{N^{q-1}}\sum_{k=b}^{n-1}\frac{\sqrt{\eta(k)}}{k^{2-q}} + r_b,
\end{aligned}$$

where the second inequality follows from replacement of $k = b(r-1)$. We use here $N = n/b$. Let us assume that $b \prec n^{1-\delta}$; then we deduce that $(b/n)^{q-1} \prec n^{-\frac{q}{2}\delta}$, and if $b^{3/2}L \prec n^{1-\delta}$, we analogously derive that $(1/N)^{q-1}(L\sqrt{b}) \prec n^{-\frac{q}{2}\delta}$. If $\eta(t) \prec n^{-\eta}$ and $\sigma \le \eta/2$, we assume that

$$\frac{b}{n}\big[1 \vee \sqrt{b}L\big] + r_b \prec n^{-\delta}, \qquad \sum_{t=0}^{\infty}\eta(t)^{1/2} < \infty.$$

– λ (3), (5) *case*. Note that

$$\begin{aligned}
C_{b,q}(r) &\prec Lz + \big(bL/z + (bL/z)^2\big)\lambda(r-b) + r_b \\
&\prec 2\big(bL^2\big)^{\frac{2}{3}}\lambda(r-b)^{\frac{1}{3}} + \big(bL^2\big)^{\frac{1}{3}}\lambda(r-b)^{\frac{2}{3}} + r_b,
\end{aligned}$$

with the choice $z = (b^2L\lambda(r-b))^{\frac{1}{3}}$. Then

$$\begin{aligned}
B_{b,q}(N) &\prec \frac{1}{N^{q-1}}\sum_{r=0}^{b-1}(r+1)^{q-2} + \frac{2\left(bL^2\right)^{\frac{2}{3}}}{N^{q-1}}\sum_{r=b}^{N-1}\frac{\lambda(r-b)^{\frac{1}{3}}}{(r+1)^{2-q}} \\
&\quad + \frac{\left(bL^2\right)^{\frac{1}{3}}}{N^{q-1}}\sum_{r=b}^{N-1}\frac{\lambda(r-b)^{\frac{2}{3}}}{(r+1)^{2-q}} + r_b \\
&\prec \left(\frac{b}{N}\right)^{q-1}\left[1 + \left(\frac{L^4}{b}\right)^{\frac{1}{3}}\sum_{t=0}^{N-b-1}\lambda(t)^{\frac{1}{3}} + \left(\frac{L}{b}\right)^{\frac{2}{3}}\sum_{t=0}^{N-b-1}\lambda(t)^{\frac{2}{3}}\right] \\
&\quad + \frac{1}{N^{q-1}}\left[\left(bL^2\right)^{\frac{2}{3}}\sum_{t=0}^{N-b-1}\frac{\lambda(t)^{\frac{1}{3}}}{(t+1)^{2-q}} + \left(bL^2\right)^{\frac{1}{3}}\sum_{t=0}^{N-b-1}\frac{\lambda(t)^{\frac{2}{3}}}{(t+1)^{2-q}}\right] + r_b,
\end{aligned}$$

where the second inequality follows from the change in variable $r = t + b$. We use here $N = n$. Let us assume that $b \prec n^{1-\delta}$; then we deduce that $(b/n)^{q-1} \prec n^{-\frac{q}{2}\delta}$, and if $((bL^2)^{2/3} \vee (bL^2)^{1/3}) \prec n^{1-\delta}$, we analogously derive that $(b/N)^{q-1}((L^4/b)^{\frac{1}{3}} \vee$

$(L/b)^{\frac{2}{3}}) \prec n^{-\frac{q}{2}\delta}$. If $\lambda(t) \prec n^{-\lambda}$ and $\sigma \leq \lambda/2$ and $\sigma \leq 2\lambda/3$, we assume that

$$\frac{b}{n}\left[1 \vee \left(\frac{L^4}{b}\right)^{\frac{1}{3}} \vee \left(\frac{L}{b}\right)^{\frac{2}{3}}\right] + r_b \prec n^{-\delta}$$

with $\sum_{t=0}^{\infty}(t+1)^{q-2}\lambda(t)^{1/3} < \infty$ and $\sum_{t=0}^{\infty}(t+1)^{q-2}\lambda(t)^{2/3} < \infty$.

– λ (4), (5) *case*. Note that

$$\begin{aligned} C_{b,q}(r) &\prec (Lz + r_b) + \left(\frac{bL}{z} + \left(\frac{bL}{z}\right)^2\right)\lambda\big((r-1)b\big) \\ &\prec 2\big(b^2L^4\lambda\big((r-1)b\big) + bL^2\big(\lambda\big((r-1)b\big)\big)^2\big)^{\frac{1}{3}} + r_b, \end{aligned}$$

with the choice $z = (b^2L\lambda((r-1)b))^{\frac{1}{3}}$. Then we obtain

$$\begin{aligned} B_{b,q}(N) &\prec \frac{1}{N^{q-1}} + \frac{(bL^2)^{\frac{2}{3}}}{N^{q-1}}\sum_{r=1}^{N-1}(r+1)^{q-2}\lambda((r-1)b)^{\frac{1}{3}} \\ &\quad + \frac{(bL^2)^{\frac{1}{3}}}{N^{q-1}}\sum_{r=1}^{N-1}(r+1)^{q-2}\lambda\big((r-1)b\big)^{\frac{2}{3}} + r_b \\ &\prec \frac{1}{N^{q-1}}\left[1 + \big(bL^2\big)^{\frac{2}{3}}\sum_{k=1}^{n-1}\lambda(k)^{\frac{1}{3}} + \big(bL^2\big)^{\frac{1}{3}}\sum_{k=1}^{n-1}\lambda(k)^{\frac{2}{3}}\right] \\ &\quad + \frac{1}{N^{q-1}}\left[L^{\frac{4}{3}}b^{\frac{8}{3}-q}\sum_{k=1}^{n-1}\frac{\lambda(k)^{\frac{1}{3}}}{k^{2-q}} + L^{\frac{2}{3}}b^{\frac{7}{3}-q}\sum_{k=1}^{n-1}\frac{\lambda(k)^{\frac{2}{3}}}{k^{2-q}}\right] + r_b, \end{aligned}$$

where the second inequality follows from the change in variable $k = b(r-1)$. We use here $N = n/b$. Let us assume that $b \prec n^{1-\delta}$; then we deduce that $(b/n)^{q-1} \prec n^{-\frac{q}{2}\delta}$, and if $(b^{\frac{5}{3}}L^{\frac{4}{3}} \vee b^{\frac{4}{3}}L^{\frac{2}{3}}) \prec n^{1-\delta}$, then we analogously derive that $(b/n)^{q-1}(b^2L^4 \vee bL^2)^{\frac{1}{3}} \prec n^{-\frac{q}{2}\delta}$. If $\lambda(t) \prec n^{-\lambda}$ and $\sigma \leq \lambda/3$ and $\sigma \leq 2\lambda/3$, we assume that

$$\frac{b}{n}\big[\big(b^2L^4\big)^{\frac{1}{3}} \vee \big(bL^2\big)^{\frac{1}{3}}\big] + r_b \prec n^{-\delta};$$

with $\sum_{t=0}^{n-1}(t+1)^{q-2}\lambda(t)^{1/3} < \infty$ and $\sum_{t=0}^{n-1}(t+1)^{q-2}\lambda(t)^{2/3} < \infty$ that bound holds.

Consider now the cases under assumption (6).

– η (3), (6) *case*. Note that

$$\begin{aligned} C_{b,q}(r) &\prec Lb\eta(r-b)/z + Dz^c \\ &\prec \big(D\big(bL\eta(r-b)\big)^c\big)^{\frac{1}{1+c}} + \big(bL\eta(r-b)\big)^{\frac{2+c}{1+c}} \end{aligned}$$

with the choice $z = (bL\eta(r-b)/D)^{\frac{1}{c+1}}$. Then

$$\begin{aligned}
B_{b,q}(N) &\prec \frac{1}{N^{q-1}}\sum_{r=0}^{b-1}(r+1)^{q-2} + \frac{(D(bL)^c)^{\frac{1}{1+c}}}{N^{q-1}}\sum_{r=b}^{N-1}(r+1)^{q-2}\big(\eta(r-b)\big)^{\frac{c}{1+c}} \\
&\quad + \frac{(bL)^{\frac{2+c}{1+c}}}{N^{q-1}}\sum_{r=b}^{N-1}(r+1)^{q-2}\big(\eta(r-b)\big)^{\frac{2+c}{1+c}} \\
&\prec \left(\frac{b}{N}\right)^{q-1}\left[1+\left(\frac{DL^c}{b}\right)^{\frac{1}{1+c}}\sum_{t=0}^{N-b-1}\eta(t)^{\frac{c}{1+c}} + \big(bL^{2+c}\big)^{\frac{1}{1+c}}\sum_{t=0}^{N-b-1}\eta(t)^{\frac{2+c}{1+c}}\right] \\
&\quad + \frac{1}{N^{q-1}}\left[\big(D(bL)^c\big)^{\frac{1}{1+c}}\sum_{t=0}^{N-b-1}\frac{\eta(t)^{\frac{c}{1+c}}}{(t+1)^{2-q}} + (bL)^{\frac{2+c}{1+c}}\sum_{t=0}^{N-b-1}\frac{\eta(t)^{\frac{2+c}{1+c}}}{(t+1)^{2-q}}\right],
\end{aligned}$$

where the second inequality follows from the change in variable $r = t + b$. We use here $N = n$. Now if we assume that $b \prec n^{1-\delta}$, we deduce that $(b/n)^{q-1} \prec n^{-\frac{q}{2}\delta}$, and if $b((DL^c/b)^{\frac{1}{1+c}} \vee (bL^{2+c})^{\frac{2+c}{1+c}}) \prec n^{1-\delta}$, we analogously derive $(b/N)^{q-1} \times ((DL^c/b)^{\frac{1}{1+c}} \vee (bL^{2+c})^{\frac{2+c}{1+c}}) \prec n^{-\frac{q}{2}\delta}$.

If $\eta(t) \prec n^{-\eta}$ and $\sigma \le \eta \cdot \frac{c}{1+c}$ and $\sigma \le \eta \cdot \frac{2+c}{1+c}$, we assume that

$$\frac{b}{n}\left[1 \vee \left(\frac{DL^c}{b}\right)^{\frac{1}{1+c}} \vee \big(bL^{2+c}\big)^{\frac{1}{1+c}}\right] \prec n^{-\delta},$$

which implies with $\sum_{t=0}^{\infty}(t+1)^{q-2}\eta(t)^{\frac{c}{1+c}} < b^{q-2}$ and $\sum_{t=0}^{\infty}(t+1)^{q-2}\eta(t)^{\frac{2+c}{1+c}} < b^{q-2}$ that $B_{b,q}(N) \prec n^{-\frac{q}{2}\delta}$.

– η (4), (6) *case*. Note that

$$\begin{aligned}
C_{b,q}(r) &\prec Lb\eta\big((r-1)b\big)/z + Dz^c \\
&\prec \big(D\big(bL\eta\big((r-1)b\big)\big)^c\big)^{\frac{1}{1+c}} + \big(bL\eta\big((r-1)b\big)\big)^{\frac{2+c}{1+c}}
\end{aligned}$$

with the choice $z = (bL\eta((r-1)b)/D)^{\frac{1}{c+1}}$. Then

$$\begin{aligned}
B_{b,q}(N) &\prec \frac{1}{N^{q-1}} + \frac{(D(bL)^c)^{\frac{1}{1+c}}}{N^{q-1}}\sum_{r=1}^{N-1}(r+1)^{q-2}\eta\big((r-1)b\big)^{\frac{c}{1+c}} \\
&\quad + \frac{(bL)^{\frac{2+c}{1+c}}}{N^{q-1}}\sum_{r=1}^{N-1}(r+1)^{q-2}\eta\big((r-1)b\big)^{\frac{2+c}{1+c}} \\
&\prec \frac{1}{N^{q-1}}\left[1+\big(D(bL)^c\big)^{\frac{1}{1+c}}\sum_{k=b}^{n-1}\eta(k)^{\frac{c}{1+c}} + (bL)^{\frac{2+c}{1+c}}\sum_{k=b}^{n-1}\eta(k)^{\frac{2+c}{1+c}}\right]
\end{aligned}$$

$$+\frac{1}{N^{q-1}}\left[\left(DL^c\right)^{\frac{1}{1+c}}b^{\frac{2+3c}{1+c}-q}\sum_{k=b}^{n-1}\frac{\eta(k)^{\frac{c}{1+c}}}{k^{2-q}}+L^{\frac{2+c}{1+c}}b^{\frac{4+3c}{1+c}-q}\sum_{k=b}^{n-1}\frac{\eta(k)^{\frac{2+c}{1+c}}}{k^{2-q}}\right],$$

where the second inequality follows from the change in variables $k=b(r-1)$. We use here $N=n/b$. Now if we assume $b\prec n^{1-\delta}$, we deduce that $(b/n)^{q-1}\prec n^{-\frac{q}{2}\delta}$. Now if $(1\vee(D(bL)^c)^{\frac{1}{1+c}}\vee(bL)^{\frac{2+c}{1+c}})\prec n^{-\delta}$, we analogously derive $(1/N)^{q-1}(1\vee(D(bL)^c)^{\frac{1}{1+c}}\vee(bL)^{\frac{2+c}{1+c}})\prec n^{-\frac{q}{2}\delta}$. If $\eta(t)\prec n^{-\eta}$ and $\sigma\le\eta\cdot\frac{c}{1+c}$ and $\sigma\le\eta\cdot\frac{2+c}{1+c}$, we assume that

$$\frac{b}{n}\left[1\vee\left(D(bL)^c\right)^{\frac{1}{1+c}}\vee(bL)^{\frac{2+c}{1+c}}\right]\prec n^{-\delta},$$

which implies with $\sum_{t=0}^{n-1}(t+1)^{q-2}\eta(t)^{\frac{c}{1+c}}\prec b^{q-2}$ and $\sum_{t=0}^{n-1}(t+1)^{q-2}\eta(t)^{\frac{2+c}{1+c}}\prec b^{q-2}$ that $B_{b,q}(N)\prec n^{-\frac{q}{2}\delta}$.

– λ (3), (6) *case*. Note that

$$\begin{aligned}C_{b,q}(r)&\prec Dz^c+\left(bL/z+(bL/z)^2\right)\lambda(r-b),\\&\prec\left(DbL^c\right)^{\frac{2}{2+c}}\left(\lambda(r-b)\right)^{\frac{c}{2+c}}+\left(D\left(bL^c\right)^{\frac{1}{2+c}}\right)\left(\lambda(r-b)\right)^{\frac{1+c}{2+c}}\\&\quad+D\left(bL^{2c}\right)^{\frac{1}{2+c}}\left(\lambda(r-b)\right)^{\frac{1+c}{2+c}}\end{aligned}$$

with the choice $z=((bL)^2D^{-1}\lambda(r-b))^{\frac{1}{2+c}}$. Then

$$\begin{aligned}B_{b,q}(N)&\prec\frac{1}{N^{q-1}}\sum_{r=0}^{b-1}(r+1)^{q-2}+\frac{(D(bL)^c)^{\frac{2}{2+c}}}{N^{q-1}}\sum_{r=b}^{N-1}\frac{\lambda(r-b)^{\frac{c}{2+c}}}{(r+1)^{2-q}}\\&\quad+\frac{(D(bL)^c)^{\frac{1}{2+c}}}{N^{q-1}}\sum_{r=b}^{N-1}\frac{\lambda(r-b)^{\frac{1+c}{2+c}}}{(r+1)^{2-q}}+\frac{(D(bL)^{2c})^{\frac{1}{2+c}}}{N^{q-1}}\sum_{r=b}^{N-1}\frac{\lambda(r-b)^{\frac{1+c}{2+c}}}{(r+1)^{2-q}}\\&\prec\left(\frac{b}{N}\right)^{q-1}\left[1+\left(\frac{(DL^c)^2}{b^{2-c}}\right)^{\frac{1}{2+c}}\sum_{t=0}^{N-b-1}\lambda(t)^{\frac{c}{2+c}}\right.\\&\quad\left.+\left(\frac{DL^c}{b^2}\right)^{\frac{1}{2+c}}\sum_{t=0}^{N-b-1}\lambda(t)^{\frac{1+c}{2+c}}\right]\\&\quad+\left(\frac{b}{N}\right)^{q-1}\left[\left(\frac{DL^{2c}}{b^{2-c}}\right)^{\frac{1}{2+c}}\sum_{t=0}^{N-b-1}\lambda(t)^{\frac{1+c}{2+c}}\right]\\&\quad+\frac{1}{N^{q-1}}\left[\left(D(bL)^c\right)^{\frac{2}{2+c}}\sum_{t=0}^{N-b-1}\frac{\lambda(t)^{\frac{c}{2+c}}}{(t+1)^{2-q}}\right.\\&\quad\left.+\left(D(bL)^c\right)^{\frac{1}{2+c}}\sum_{t=0}^{N-b-1}\frac{\lambda(t)^{\frac{1+c}{2+c}}}{(t+1)^{2-q}}\right]\end{aligned}$$

$$+\frac{1}{N^{q-1}}\left[\left(D(bL)^{2c}\right)^{\frac{1}{2+c}}\sum_{t=0}^{N-b-1}\frac{\lambda(t)^{\frac{1+c}{2+c}}}{(t+1)^{2-q}}\right],$$

where the second inequality follows from the change in variable $r=t+b$. We use here $N=n$. Now if we assume that $b\prec n^{1-\delta}$, we deduce that $(b/n)^{q-1}\prec n^{-\frac{q}{2}\delta}$.

If $b((D^2L^{2c}/b^{2-c})^{\frac{1}{2+c}}\vee(DL^c/b^2)^{\frac{1}{2+c}}\vee(DL^{2c}/b^{2-c})^{\frac{1}{2+c}})\prec n^{1-\delta}$, we analogously derive

$$\left(\frac{b}{N}\right)^{q-1}\left(\frac{D^2L^{2c}}{b^{2-c}}\vee\frac{DL^c}{b^2}\vee\frac{DL^{2c}}{b^{2-c}}\right)^{\frac{1}{2+c}}\prec n^{-\frac{q}{2}\delta}.$$

If $\lambda(t)\prec n^{-\lambda}$ and $\sigma\le\lambda\cdot\frac{c}{1+c}$ and $\sigma\le\lambda\cdot\frac{2+c}{1+c}$, we assume that

$$\frac{b}{n}\left[1\vee\left(\frac{D^2L^{2c}}{b^{2-c}}\vee\frac{DL^c}{b^2}\vee\frac{DL^{2c}}{b^{2-c}}\right)^{\frac{1}{2+c}}\right]\prec n^{-\delta},$$

which implies with $\sum_{t=0}^{\infty}(t+1)^{q-2}\lambda(t)^{\frac{c}{1+c}}<b^{q-2}$ and $\sum_{t=0}^{\infty}(t+1)^{q-2}\lambda(t)^{\frac{2+c}{1+c}}<b^{q-2}$ that $B_{b,q}(N)\prec n^{-\frac{q}{2}\delta}$.

– λ (4), (6) *case*. Note that

$$\begin{aligned}C_{b,q}(r)&\prec Dz^c+\left(\frac{bL}{z}+\left(\frac{bL}{z}\right)^2\right)\lambda\big((r-1)b\big)\\&\prec\left(D(bL)^c\right)^{\frac{2}{2+c}}\lambda\big((r-1)b\big)^{\frac{c}{2+c}}+\left(\left(D(bL)^c\right)^{\frac{1}{2+c}}\right)\lambda\big((r-1)b\big)^{\frac{1+c}{2+c}}\\&\quad+\left(D(bL)^{2c}\right)^{\frac{1}{2+c}}\lambda\big((r-1)b\big)^{\frac{1+c}{2+c}}\end{aligned}$$

with the choice $z=((bL)^2D^{-1}\lambda((r-1)b))^{\frac{1}{2+c}}$. We obtain

$$\begin{aligned}B_{b,q}(N)&\prec\frac{1}{N^{q-1}}+\frac{(D(bL)^c)^{\frac{2}{2+c}}}{N^{q-1}}\sum_{r=1}^{N-1}(r+1)^{q-2}\lambda\big((r-1)b\big)^{\frac{c}{2+c}}\\&\quad+\frac{(D(bL)^c)^{\frac{1}{2+c}}}{N^{q-1}}\sum_{r=1}^{N-1}(r+1)^{q-2}\lambda\big((r-1)b\big)^{\frac{1+c}{2+c}}\\&\quad+\frac{(D(bL)^{2c})^{\frac{1}{2+c}}}{N^{q-1}}\sum_{r=1}^{N-1}(r+1)^{q-2}\lambda\big((r-1)b\big)^{\frac{1+c}{2+c}}\\&\prec\frac{1}{N^{q-1}}\left[1+\left(D(bL)^c\right)^{\frac{2}{2+c}}\sum_{k=1}^{n-1}\lambda(k)^{\frac{c}{2+c}}+\left(D(bL)^c\right)^{\frac{1}{2+c}}\sum_{k=1}^{n-1}\lambda(k)^{\frac{1+c}{2+c}}\right]\\&\quad+\frac{1}{N^{q-1}}\left[\left(D(bL)^{2c}\right)^{\frac{1}{2+c}}\sum_{k=1}^{n-1}\lambda(k)^{\frac{1+c}{2+c}}\right]\end{aligned}$$

$$+\frac{1}{N^{q-1}}\left[\left(DL^c\right)^{\frac{2}{2+c}}b^{\frac{4(1+c)}{2+c}-q}\sum_{k=1}^{n-1}\frac{\lambda(k)^{\frac{c}{2+c}}}{k^{2-q}}\right.$$
$$\left.+\left(DL^c\right)^{\frac{1}{2+c}}b^{\frac{4+3c}{2+c}-q}\sum_{k=1}^{n-1}\frac{\lambda(k)^{\frac{1+c}{2+c}}}{k^{2-q}}\right]$$
$$+\frac{1}{N^{q-1}}\left[\left(DL^{2c}\right)^{\frac{1}{2+c}}b^{\frac{4(1+c)}{2+c}-q}\sum_{k=1}^{n-1}\frac{\lambda(k)^{\frac{1+c}{2+c}}}{k^{2-q}}\right],$$

where the second inequality follows from the change in variable $k=b(r-1)$. We use here $N=n/b$. Now if we assume that $b\prec n^{1-\delta}$, we deduce that $(b/n)^{q-1}\prec n^{-\frac{q}{2}\delta}$.

If

$$b\Big(\big(D(bL)^c\big)^{\frac{2}{2+c}}\vee\big(D(bL)^c\big)^{\frac{1}{2+c}}\vee\big(D(bL)^{2c}\big)^{\frac{1}{2+c}}\Big)\prec n^{1-\delta},$$

we analogously derive that

$$(b/N)^{q-1}\Big(\big(D(bL)^c\big)^{\frac{2}{2+c}}\vee\big(D(bL)^c\big)^{\frac{1}{2+c}}\vee\big(D(bL)^{2c}\big)^{\frac{1}{2+c}}\Big)\prec n^{-\frac{q}{2}\delta}.$$

If $\lambda(t)\prec n^{-\lambda}$ and $\sigma\le\lambda\cdot\frac{c}{1+c}$ and $\sigma\le\lambda\cdot\frac{2+c}{1+c}$, we assume that

$$\frac{b}{n}\Big[1\vee\big(D(bL)^c\big)^2\vee D(bL)^c\vee\big(D(bL)^{2c}\big)^{\frac{1}{2+c}}\Big]\prec n^{-\delta},$$

which implies with $\sum_{t=0}^{n-1}(t+1)^{q-2}\lambda(t)^{\frac{c}{1+c}}\prec b^{q-2}$ and $\sum_{t=0}^{n-1}(t+1)^{q-2}\lambda(t)^{\frac{2+c}{1+c}}\prec b^{q-2}$ that $B_{b,q}(N)\prec n^{-\frac{q}{2}\delta}$.

Lemma 5 *The relation* $\widehat{\Delta}_{b,n}^{(2)}(x)\to_{n\to\infty}0$ *holds in the following cases under the convergence assumption* (5):

- *In the overlapping case, if we have respectively*

$$\sum_{t=0}^{\infty}\eta(t)^{1/2}<\infty,\quad \textit{and}\quad r_b+\frac{b}{n}\left[1\vee\frac{L}{\sqrt{b}}\right]\to 0,$$
$$\sum_{t=0}^{\infty}\lambda(t)^{2/3}<\infty,\quad \textit{and}\quad r_b+\frac{b}{n}\left[1\vee\left(\frac{L^4}{b}\right)^{1/3}\vee\left(\frac{L}{b}\right)^{2/3}\right]\to 0.$$

- *In the nonoverlapping case, if we have respectively*

$$\sum_{t=0}^{\infty}\eta(t)^{1/2}<\infty,\quad \textit{and}\quad r_b+\frac{b}{n}\big[1\vee\sqrt{b}L\big]\to 0,$$
$$\sum_{t=0}^{\infty}\lambda(t)^{2/3}<\infty,\quad \textit{and}\quad r_b+\frac{b}{n}\big[1\vee\big(bL^2\big)^{2/3}\vee\big(bL^2\big)^{1/3}\big]\to 0.$$

This lemma, together with Lemma 1, yields Theorem 2.

Lemma 6 *The relation $B_{b,q}(N) \prec n^{-q\delta/2}$ holds under concentration assumption* (6) *if respectively* the overlapping setting *and one of the following relations holds as* $n \to \infty$:

$$\underline{\eta\text{-}dependence}: \quad \sum_{t=0}^{\infty} (t+1)^{p-2} \eta(t)^{\frac{2+c}{1+c}} < \infty,$$

$$\frac{b}{n}\left[1 \vee \left(\frac{DL^c}{b} \vee bL^{2+c}\right)^{\frac{1}{1+c}}\right] \prec n^{-\delta}, \quad or$$

$$\underline{\lambda\text{-}dependence}: \quad \sum_{t=0}^{\infty} (t+1)^{p-2} \lambda(t)^{\frac{1+c}{2+c}} < \infty,$$

$$\frac{b}{n}\left[1 \vee \left(\frac{(DL^c)^2}{b^{2-c}} \vee \frac{DL^c}{b^2} \vee \frac{DL^{2c}}{b^{2-c}}\right)^{\frac{1}{2+c}}\right] \prec n^{-\delta},$$

or the nonoverlapping setting is used and

$$\underline{\eta\text{-}dependence}: \quad \sum_{t=0}^{n-1} (t+1)^{p-2} \eta(t)^{\frac{2+c}{1+c}} \prec b^{p-2},$$

$$\frac{b}{n}\left[1 \vee \left(D(bL)^c \vee (bL)^{2+c}\right)^{\frac{1}{1+c}}\right] \prec n^{-\delta}, \quad or$$

$$\underline{\lambda\text{-}dependence}: \quad \sum_{t=0}^{n-1} (t+1)^{p-2} \lambda(t)^{\frac{1+c}{2+c}} \prec b^{p-2},$$

$$\frac{b}{n}\left[1 \vee \left(\left(D(bL)^c\right)^2 \vee D(bL)^c \vee D(bL)^{2c}\right)^{\frac{1}{2+c}}\right] \prec n^{-\delta}.$$

This lemma, together with Lemma 1, yields Theorem 3.

5.3 Proof of Theorem 5

Put $k_n = \lfloor n/a_n \rfloor$. Partition $\{1, \ldots, n\}$ into k_n blocks of size a_n,

$$J_j = J_{j,n} = \left\{(j-1)\, a_n + 1, \ldots, j a_n\right\}, \qquad j = 1, \ldots, k_n,$$

and, in case $k_n a_n < n$, a remainder block, $J_{k_n+1} = \{k_n a_n + 1, \ldots, n\}$. Observe that

$$\mathbb{P}\big(M_n \le w_n(x)\big) = \mathbb{P}\left(\bigcap_{j=1}^{k_n+1} \left\{M(J_j) \le w_n(x)\right\}\right),$$

where $M(J_j)=\max_{i\in J_j} X_i$. Since $\mathbb{P}(M(J_j)>w_n(x))\le a_n\bar{F}(w_n(x))\to 0$ as $n\to\infty$, the remainder block can be omitted, and

$$\mathbb{P}\big(M_n\le w_n(x)\big)=\mathbb{P}\left(\bigcap_{j=1}^{k_n}\big\{M(J_j)\le w_n(x)\big\}\right)+o(1).$$

Let

$$J_j^*=J_{j,n}^*=\big\{(j-1)a_n+1,\dots,ja_n-l_n\big\},\quad j=1,\dots,k_n,$$
$$J_j'=J_{j,n}'=\{ja_n-l_n,\dots,ja_n\},\quad j=1,\dots,k_n.$$

Since $P(\bigcup_{j=1}^{k_n}M(J_j')>w_n(x))\le k_nl_n\bar{F}(w_n(x))\to 0$ as $n\to\infty$, we deduce that

$$\mathbb{P}\big(M_n\le w_n(x)\big)=\mathbb{P}\left(\bigcap_{j=1}^{k_n}\big\{M(J_j^*)\le w_n(x)\big\}\right)+o(1).$$

Let $B_j=B_{j,n}=\{M(J_j^*)\le w_n(x)\}$. We write

$$\begin{aligned}
&\mathbb{P}\left(\bigcap_{j=1}^{k_n}B_j\right)-\prod_{j=1}^{k_n}\mathbb{P}(B_j)\\
&=\sum_{i=1}^{k_n-1}\left(\mathbb{P}\left(\bigcap_{j=1}^{k_n-i+1}B_j\right)\prod_{j=k_n-i+2}^{k_n}\mathbb{P}(B_j)-\mathbb{P}\left(\bigcap_{j=1}^{k_n-i}B_j\right)\prod_{j=k_n-i+1}^{k_n}\mathbb{P}(B_j)\right)\\
&=\sum_{i=1}^{k_n-1}\left(\mathbb{P}\left(\bigcap_{j=1}^{k_n-i+1}B_j\right)-\mathbb{P}\left(\bigcap_{j=1}^{k_n-i}B_j\right)\mathbb{P}(B_{k_n-i+1})\right)\prod_{j=k_n-i+2}^{k_n}\mathbb{P}(B_j).
\end{aligned}$$

We want to bound the following quantity:

$$\left|\mathbb{P}\left(\bigcap_{j=1}^{k_n-i+1}B_j\right)-\mathbb{P}\left(\bigcap_{j=1}^{k_n-i}B_j\right)\mathbb{P}(B_{k_n-i+1})\right|.$$

Let us define $f_n^{(x)}(y)=\mathbb{I}_{\{y\le w_n(x)\}}$. Let (α_n) be a positive sequence such that $\alpha_n\to 0$ as $n\to\infty$ and put $x_n^-=x-\alpha_n$ and $x_n^+=x+\alpha_n$. We simply approximate the function $f_n^{(x)}$ by Lipschitz and bounded real functions g_n,h_n with

$$f_n^{(x_n^-)}\le g_n\le f_n^{(x)}\le h_n\le f_n^{(x_n^+)},$$

and we quote that it is easy to choose functions g_n and h_n with Lipschitz coefficient $u_n\alpha_n^{-1}$. For $I\subset\{1,\dots,n\}$, let $H_I(f_n^{(x)})=\mathbb{E}[\prod_{i\in I}f_n^{(x)}(X_i)]$. Note that

$$H_I\big(f_n^{(x_n^-)}\big)\le H_I(g_n)\le H_I\big(f_n^{(x)}\big)\le H_I(h_n)\le H_I\big(f_n^{(x_n^+)}\big).$$

Let $C_{I,J}(f_n^{(x)}) = H_{I\cup J}(f_n^{(x)}) - H_I(f_n^{(x)})H_J(f_n^{(x)})$. We have

$$C_{I,J}(g_n) - \delta_{I,J}(g_n, h_n) \le C_{I,J}\big(f_n^{(x)}\big) \le C_{I,J}(h_n) + \delta_{I,J}(g_n, h_n)$$

with

$$\delta_{I,J}(g_n, h_n) = H_I(h_n)H_J(h_n) - H_I(g_n)H_J(g_n).$$

Let $I_i = \{l : \{X_l \le w_n(x)\} \in \bigcap_{j=1}^{k_n-i} B_j\}$ and $J_i = \{l : \{X_l \le w_n(x)\} \in B_{k_n-i+1}\}$. We have

$$\begin{aligned}\big|H_{I_i}(h_n) - H_{I_i}(g_n)\big| &\le (k_n - i + 1)\, a_n\big(\bar F\big(w_n(x_n^-)\big) - \bar F\big(w_n(x_n^+)\big)\big),\\ \big|H_{J_i}(h_n) - H_{J_i}(g_n)\big| &\le a_n\big(\bar F\big(w_n(x_n^-)\big) - \bar F\big(w_n(x_n^+)\big)\big).\end{aligned}$$

Then we have

$$\big|C_{I_i,J_i}\big(f_n^{(x)}\big)\big| \le \big|C_{I_i,J_i}(h_n)\big| \vee \big|C_{I_i,J_i}(g_n)\big| + \big|\delta_{I_i,J_i}(g_n, h_n)\big|$$

and

$$\big|\delta_{I_i,J_i}(g_n, h_n)\big| \le \big|H_{I_i}(h_n) - H_{I_i}(g_n)\big| + \big|H_{J_i}(h_n) - H_{J_i}(g_n)\big|.$$

Note that, as $n \to \infty$,

$$n\big(\bar F\big(w_n(x_n^-)\big) - \bar F\big(w_n(x_n^+)\big)\big) \sim 2\alpha_n(1 + \gamma x)_+^{-1/\gamma-1}.$$

If X is η-weakly dependent, it follows that

$$\big|C_{I_i,J_i}\big(f_n^{(x)}\big)\big| \le (k_n - i + 2)\, a_n u_n \alpha_n^{-1}\eta(l_n) + 2\alpha_n(1 + \gamma x)_+^{-1/\gamma-1}\frac{(k_n - i + 2)\, a_n}{n}.$$

An optimal choice of α_n is then given by

$$\alpha_n \sim \big[n\eta(l_n)u_n\big]^{1/2},$$

and we deduce that

$$\big|C_{I_i,J_i}\big(f_n^{(x)}\big)\big| \prec \big(n\eta(l_n)u_n\big)^{1/2}.$$

It follows that

$$\left|\mathbb{P}\left(\bigcap_{j=1}^{k_n} B_j\right) - \prod_{j=1}^{k_n}\mathbb{P}(B_j)\right| \prec k_n\big(n\eta(l_n)u_n\big)^{1/2}.$$

If X is λ-weakly dependent, it follows that

$$\begin{aligned}\big|C_{I_i,J_i}\big(f_n^{(x)}\big)\big| \le{}& \big[(k_n - i + 2)\, a_n u_n \alpha_n^{-1} + (k_n - i + 1)^2\, a_n u_n^2 \alpha_n^{-2}\big]\lambda(l_n)\\ &+ 2\alpha_n(1 + \gamma x)_+^{-1/\gamma-1}\frac{(k_n - i + 2)\, a_n}{n}.\end{aligned}$$

An optimal choice of α_n is then given by

$$\alpha_n \sim \left[n\lambda(l_n)u_n\right]^{1/2} \vee \left[na_n\lambda(l_n)u_n^2\right]^{1/3},$$

and then

$$\left|C_{I_i,J_i}\left(f_n^{(x)}\right)\right| \prec \left(\left[n\lambda(l_n)u_n\right]^{1/2} \vee \left[na_n\lambda(l_n)u_n^2\right]^{1/3}\right).$$

It follows that

$$\left|\mathbb{P}\left(\bigcap_{j=1}^{k_n} B_j\right) - \prod_{j=1}^{k_n} \mathbb{P}(B_j)\right| \prec k_n\left(\left[n\lambda(l_n)u_n\right]^{1/2} \vee \left[na_n\lambda(l_n)u_n^2\right]^{1/3}\right).$$

Finally, we deduce that

$$\mathbb{P}\big(M_n \leq w_n(x)\big) = \left[\mathbb{P}\big(M_n \leq w_{a_n}(x)\big)\right]^{k_n} + o(1),$$

and the result follows.

Acknowledgements We would like first to thank the referees and an Associate Editor for their kind remarks and the co-Editors Ricardo Cao and Domingo Morales for inviting us to present this discussion paper. Then we want to thank Patrice Bertail for essential discussions during many years at ENSAE. Lastly, the first author wants to thank Paul Embrechts at ETHZ and the Swiss Banking Institute at Zurich University for their strong support.

References

Andersson CW (1970) Extreme value theory for a class of discrete distributions with applications to some stochastic processes. J Appl Probab 7:99–113

Andrews D (1984) Non strong mixing autoregressive processes. J Appl Probab 21:930–934

Ango Nze P, Doukhan P (2004) Weak dependence and applications to econometrics. Econom Theory 20:995–1045

Bertail P, Haefke C, Politis DN, White W (2004) Subsampling the distribution of diverging statistics with applications to finance. J Econom 120:295–326

Bickel P, Götze F, van Zwet W (1997) Resampling fewer than n observations: gains, losses and remedies for losses. Stat Sin 7:1–31

Carlstein E (1986) The use of subseries values for estimating the variance of a general statistic from a stationary time series. Ann Stat 14:1171–1179

Chernick M (1981) A limit theorem for the maximum of autoregressive processes with uniform marginal distribution. Ann Probab 9:145–149

Dedecker J, Doukhan P, Lang G, León JR, Louhichi S, Prieur C (2007) Weak dependence: Models, theory and applications. Lecture notes in statistics, vol 190. Springer, New York

Doukhan P (1994) Mixing: properties and examples. Lecture notes in statistics, vol 85. Springer, New York

Doukhan P, Latour A, Oraichi D (2006) A simple integer-valued bilinear time series model. Adv Appl Probab 38:559–578

Doukhan P, Louhichi S (1999) A new weak dependence condition and applications to moment inequalities. Stoch Process Appl 84:313–342

Doukhan P, Mayo N, Truquet L (2009) Weak dependence, models and some applications. Metrika 69(2–3):199–225

Doukhan P, Teyssière G, Winant P (2006) A LARCH(∞) vector valued process. Dependence in probability and statistics. Lecture notes in statist, vol 187. Springer, New York, pp. 245–258.

Doukhan P, Wintenberger O (2007) An invariance principle for weakly dependent stationary general models. Probab Math Stat 27:45–73

Doukhan P, Wintenberger O (2008) Weakly dependent chains with infinite memory. Stoch Process Appl 118:1997–2013

Hall A, Scotto MG, Cruz J (2010) Extremes of integer-valued moving average sequences. Test 19:359–374

Künsch HR (1989) The jackknife and the bootstrap for general stationary observations. Ann Stat 17:1217–1241

Leadbetter MR (1974) On extreme values in stationary sequences. Z Wahrscheinlichkeitstheor Verw Geb 28:289–303

Leadbetter MR, Lindgren G, Rootzen H (1983) Extremes and related properties of random sequences and processes. Springer series in statistics

Nordman DJ, Lahiri SN (2004) On optimal spatial subsample size for variance estimation. Ann Stat 32:1981–2027

O'Brien GL (1987) Extreme values for stationary and Markov sequences. Ann Probab 15:281–291

Politis DN, Romano JP (1994) Large sample confidence regions based on subsamples under minimal assumptions. Ann Stat 22:2031–2050

Robert C, Segers J, Ferro C (2009) A sliding blocks estimator for the extremal index. Electron J Stat 3:993–1020

Test (2011) 20:480–482
DOI 10.1007/s11749-011-0270-2

DISCUSSION

Comments on: Subsampling weakly dependent time series and application to extremes

Carlos Velasco

Published online: 1 November 2011

It has been a pleasure to read this paper that further extends the idea that subsampling methods are able to deal with a wide class of problems concerning statistical inference for time series data under minimal assumptions. In this context, subsampling has a series of advantages compared to resampling methods as it avoids the artificial construction of new time series either joining blocks of the original series or fitting approximate models that may alter the original dependence structure. Then, one of the main concerns in subsampling analysis is to allow serial dependence as general and weak as possible. This is key in the paper since, basically, only a restriction on the weak dependence is required for consistency, together with the existence of a nondegenerated limiting distribution and smoothness conditions on the statistic of interest in relation to block size. Thanks to the results obtained in this paper, the justification of subsampling is now even easier and more general than for any resampling method, covering a huge class of processes and statistics of relevance in many fields of application. Moreover, the paper considers for both converging and diverging statistics a new smooth subsampling estimator of the limiting distribution. This is compared to the rough estimator for the second class of statistics in a simulation study, even though the latter is not analyzed for the sample maxima.

Two important issues in practice are the efficiency of subsampling distribution estimators and their sensitivity to the user chosen tuning parameters, namely the block size b, the degree of overlapping and the smoothing parameter ε_n. Alternative methods, such as the bootstrap, when valid may possess higher order asymptotic properties improving over subsampling methods.

This comment refers to the invited paper available at doi:10.1007/s11749-011-0269-8.

C. Velasco (✉)
Department of Economics, Universidad Carlos III de Madrid, Madrid, Spain
e-mail: carlos.velasco@uc3m.es

In principle, the overlapping method that seems to be favored in block-bootstrap methods may have some efficiency advantages from increasing the number of subsamples. This paper illustrates that to cope with the stronger dependence of overlapping samples it is necessary to restrict the dependence strength of the original data to justify consistency, both for smooth and rough estimators. These additional restrictions on the weak dependent coefficients due to overlapping are only relevant for the a.s. bounds (since $p > 2$ in this case), though, on the other hand, overlapping apparently allows for more flexible choices of the subsample length b and/or the smoothing parameter ε_n. Even if these asymptotic results cannot be directly used to propose closed form optimal choices of b, it would be very interesting to investigate if such an extension is feasible and if the weaker asymptotic restrictions on the block size when overlapping are also relevant in terms of practical robustness to such choices.

A related question for the authors is whether they expect that a higher order asymptotic analysis as in Hall and Jing (1996) could be justified to some extent under their general weak dependence conditions and if this would shed further light on the efficiency of subsampling and block size choices, perhaps using extrapolation or other methods. This would require an extension of techniques that up to now rely mainly on strong mixing conditions, e.g., Götze and Künsch (1996).

For many applications, the full overlapping considered might be computationally very demanding, so that partial-overlapping, e.g., as in Politis and Romano (2010), could be an alternative. This would amount to the definition of partial overlapping subsamples

$$Y_{b,h,i} = (X_{(i-1)h+1}, \dots, X_{(i-1)h+b}),$$

where h is any integer in $[1, b]$. Politis and Romano showed that the same strong mixing assumption is sufficient for the consistency of both the no overlapping method, $h = b$, and the full one, $h = 1$. However, similar strong consistency results to those of the present paper might require different regularity conditions when overlapping is a proportion of the subsample length b (similar to the full overlapping case), than when overlapping is fixed (so closer to the less demanding case of nonoverlapping). Also, an asymptotically negligible degree of overlapping should be close to this last situation, while allowing simultaneously for potential efficiency improvements and perhaps for some further practical robustness on the choice of b.

When approximating the distribution of the normalized sample maximum, there can be further concerns on the choice of b in applications. This relates to the degree of clustering in the time series, especially relevant in a financial context with conditional heteroskedastic series. For example, a too small value of b will result in severely biased estimates of the maximum distribution since many subsamples would not contain any cluster of local maximum values when θ is also small. This aspect, however, is not reflected in the asymptotic analysis.

A final question is whether the results of Theorem 5 could be extended under similar weak dependence assumptions to related point processes defined on the sequence X_t, avoiding less primitive conditions on the extremal dependence coefficients. Also, the subsampling estimation of tail parameters of the distribution of X_t

could be considered to produce semiparametric estimates of the limiting distribution of the maximum valid under very general dependence conditions but with some knowledge on the marginal distribution.

Acknowledgements Financial support from the Spanish Plan Nacional de I+D+I (SEJ2007-62908) is gratefully acknowledged.

References

Götze F, Künsch HR (1996) Blockwise bootstrap for dependent observations: higher order approximations for studentized statistics. Ann Stat 24:1914–1933

Hall P, Jing B-Y (1996) On sample reuse methods for dependent data. J R Stat Soc B 58:727–737

Politis DN, Romano JP (2010) K-sample subsampling in general spaces: the case of independent time series. J Multivar Anal 101:316–326

Test (2011) 20:483–486
DOI 10.1007/s11749-011-0271-1

DISCUSSION

Comment on: Subsampling weakly dependent time series and application to extremes

Francesco Bravo

Published online: 1 November 2011

Professors Doukhan, Prohol and Robert should be congratulated for writing an interesting and mathematically elegant paper that extends the applicability of the subsampling method to η- and λ-weakly dependent time series models. This is an important generalisation that widens considerably the range of applications of subsampling, as clearly illustrated with the three examples used by the authors. The use of smoothing is also a nice addition to the literature, as it allows weaker block rate assumptions, albeit it does add an extra "tuning parameter" to be chosen, the bandwidth ε_n. I also note that because of the subsampling, the bandwidth rate of the smoother is going to be slower than the optimal $n^{-1/3}$ rate for probability distributions.

In this comment I would merely like to expand briefly on two points: (I) empirical applications and (II) choice of the subsample block size.

(I) There has been an increasing interest in a particular class of models for which the results of this paper are potentially relevant: integer-valued autoregressions. Brännäs and Hellström (2001) consider a generalisation of the integer-valued AR(1) model that encompasses some empirical features of economic time series of count data, Rudholm (2001) studies the entry dynamics of firms in an industry, Böckenholt (2003) uses an autoregressive negative binomial model for longitudinal count data to model emotional dependencies, Freeland and McCabe (2004) consider Poisson autoregression, Gouriéroux and Jasiak (2004) suggest a heterogeneous integer-valued AR(1) to model the number of claims an insurance company receives, and McCabe and Martin (2005) consider Bayesian forecasting for count data.

(II) It is well known that in general the block size b can affect the finite sample performance of subsampling for small to moderate sample sizes (see e.g. Politis

This comment refers to the invited paper available at doi:10.1007/s11749-011-0269-8.

F. Bravo (✉)
Department of Economics and Related Studies, University of York, York YO10 5DD, UK
e-mail: francesco.bravo@york.ac.uk

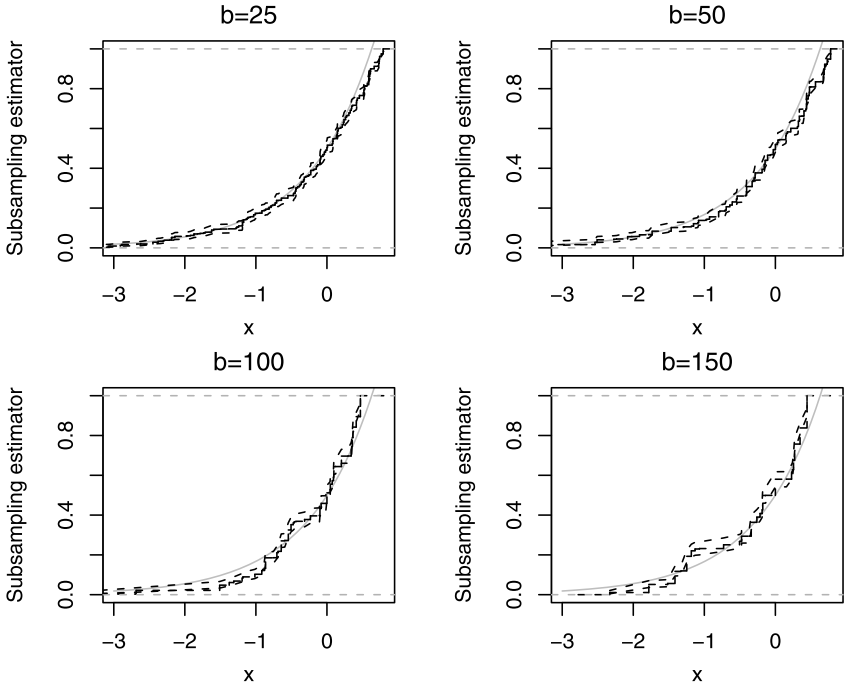


Fig. 1 AR(1) process. Rough subsampling estimator with overlapping blocks and $n = 2000$. The *dashed lines* represent 0.95 nonparametric likelihood based confidence bands calculated using Owen's (1995) method. The *grey line* represents the true asymptotic distribution

et al. 1997). I show here that the same happens (to a certain extent) in the context of one of the applications considered in the paper, namely subsampling the distribution of the normalised sample maximum of an AR(1) process with discrete innovations. Figure 1 shows the subsample rough estimator with overlapping blocks for the self-normalised sample maximum computed using the same Monte Carlo design as that used in the paper with four different block sizes, namely $b = 25$, 50, 100 and 150. In each plot a 95% confidence band for the subsampled distribution is computed using Owen's (1995) nonparametric likelihood method. Figure 1 shows that the block length has some bearings on the performances of the subsampling; in particular it seems to indicate that the block sizes $b = 25$ or $b = 50$ are more accurate.

It seems therefore of interest to consider the issue of choosing the block size. Politis et al. (1997) propose a general calibration technique that is however limited to applications where the standard bootstrap works, and thus it is not applicable to the results of this paper. The approach I suggest here borrows directly from the one suggested by Bickel and Sakov (2008) in the context of the so-called m out of n bootstrap for i.i.d. observations, and thus strictly speaking is not valid for the results of this paper. However it does seem to work (at least for the example considered here), and thus it should be interpreted as a heuristic block selection method. Its asymptotic validity is not proven, albeit we note that in general data-dependent block selection

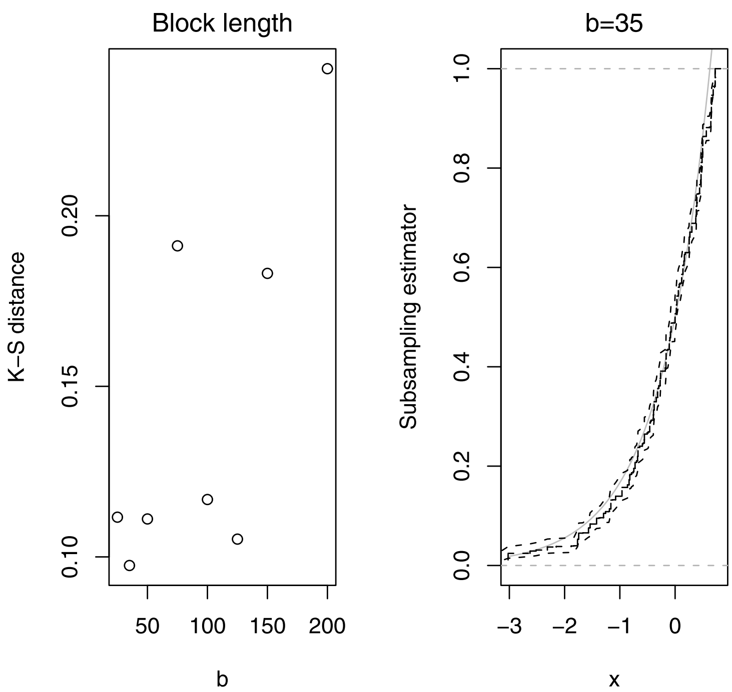


Fig. 2 AR(1) process. *Left*: Kolmogorov sup distance (K–S distance) for different block lengths b. *Right*: Rough subsampling estimator with overlapping blocks, $n = 2000$ and $b = 35$ chosen as minimum of K–S distance

methods are valid under weak conditions (see for example Theorem 2.7.1 of Politis et al. 1999). The block selection method I propose consists of the following steps: (1) choose a range of possible block sizes, say $b_j = \lfloor nq^j \rfloor$ $(j = 1, 2, \dots,\ 0 < q < 1)$, where $\lfloor\cdot\rfloor$ is the floor function, (2) consider a distance measure $D(\cdot,\cdot)$ between successive subsampling distributions $\widehat{\mathbb{K}}_{b_j,n}(x)$ and $\widehat{\mathbb{K}}_{b_{j+1},n}(x)$ (using the paper's notation), (3) choose b_j as $b_j := \arg\min_{b_j} D(\widehat{\mathbb{K}}_{b_j,n}(x), \widehat{\mathbb{K}}_{b_{j+1},n}(x))$. In this comment I use as in Bickel and Sakov (2008) the Kolmogorov sup-distance (K–S) but other distances (e.g. Cramer von Mises) could be used. Figure 2 (left panel) reports the K–S distance for $b = 15$, 25, 30, 35, 50, 75, 100, 125, 150 and 200. The minimum K–S distance was between $b = 30$ and $b = 35$ with $D(\widehat{\mathbb{K}}_{30,2000}(x), \widehat{\mathbb{K}}_{35,2000}(x)) = 0.0579$. Figure 2 (right panel) shows the resulting subsampled distribution with Owen's (1995) 95% nonparametric likelihood confidence bands. Compared with the two top panels of Fig. 1, it appears that the accuracy of this distribution is slightly superior.

In sum I think that this paper makes an important extension to the theory of subsampling with time series, which will be valuable not only to theorists but also to applied researchers in different fields. The issue of the choice of the block length is important and should be investigated in depth; in this comment I suggest a heuristic method that might give some practical guidance.

References

Bickel P, Sakov A (2008) On the choice of m in the m out of n bootstrap and confidence bounds for extrema. Stat Sin 18:967–985

Böckenholt U (2003) Analysing state dependences in emotional experiences by dynamic count data models. J R Stat Soc, Ser C 52:213–226

Brännäs K, Hellström J (2001) Generalized integer-valued autoregression. Econom Rev 20:425–443

Freeland R, McCabe B (2004) Analysis of low count time series data by Poisson autoregression. J Time Ser Anal 25:701–722

Gouriéroux C, Jasiak J (2004) Heterogeneous INAR(1) model with application to car insurance. Insur Math Econ 34:177–192

McCabe B, Martin G (2005) Bayesian predictions of low count time series. Int J Forecast 21:315–330

Owen A (1995) Nonparametric likelihood confidence bands for a distribution function. J Am Stat Assoc 90:516–521

Rudholm N (2001) Entry and the number of firms in the Swedish pharmaceuticals market. Rev Ind Organ 19:351–364

Politis D, Romano J, Wolf M (1997) Subsampling for heteroskedastic time series. J Econom 81:287–317

Politis D, Romano J, Wolf M (1999) Subsampling. Springer, Berlin

Test (2011) 20:487–490
DOI 10.1007/s11749-011-0272-0

DISCUSSION

Comments on: Subsampling weakly dependent time series and application to extremes

Patrice Bertail

Published online: 1 November 2011

The authors are to be congratulated for their paper, which pushes the potentiality of subsampling methods a step forward. Most of the results obtained for subsampling methods have been done either for i.i.d. data or for strong mixing time series or random fields. Many interesting results are summarized in the Springer book by Politis et al. (1999). Some people have also been interested in extending some results on subsampling in the framework of long-range dependent data, for which the usual bootstrap blocking techniques fail, whereas subsampling remains asymptotically valid. See Hall et al. (1998). These latter results actually escape the framework presented here but are complementary. It is well known that the strong mixing assumption also fails for a large class of times series.The potentiality of week dependence introduced by Doukhan and Louhichi (1999) is now well recognized and covers a large class of interesting processes.

In this paper the authors generalize the results of Bertail et al. (2004) especially suited for diverging statistics to times series satisfying some weak dependence conditions. They apply their results to the case of the maximum when no information on the domain of attraction is available. Even if the conditions are sometimes quite intriguing and complicated to understand, I would like to stress a few facts about their results.

First of all, the authors show that in this framework it is more interesting to smooth the subsampling distribution a little bit. This may be understood from a technical point of view by noticing that the indicator functions used in the building of the subsampling distribution are not Lipschitz. Recall that Lipschitz functions are at the heart

This comment refers to the invited paper available at doi:10.1007/s11749-011-0269-8.

P. Bertail (✉)
Crest-Insee and Modal'X, University Paris, Ouest-Nanterre-La Défense, Paris, France
e-mail: pbertail@u-paris10.fr

of the weak mixing concept; it is easy to prove that Lipschitz functions of weak-mixing times series will also satisfy weak mixing conditions. However, they also show that if one does not want to smooth and use what they call a rough subsampling distribution (the nonsmoothed one), a concentration hypothesis on the limiting distribution of the statistics of interest is needed. This somehow means that at the limit the concentration behavior implies some smoothing of the subsampling distribution: this is a very interesting phenomenon. I would be interested to have some examples for which such concentration condition fails. Then how does the rough subsampler behave in this case? From my experience a little of smoothing for the subsampling distribution may make a huge difference even if the smoothing is really small: I am glad that the authors point out this fact and give some theoretical arguments to understand this phenomenon.

The authors do not consider the problem of calibrating the subsampling size of the distribution. This is indeed a challenging task. Some works have been done in the framework of i.i.d. data using the particular U-statistics structure of the subsampling distribution. See Götze and Račkauskas (1999), Bickel and Sakov (2008), and a recent review in Bickel et al. (2010). The main idea underlying their proposition is to construct several subsampling distributions by using two different subsampling sizes, say b_n and $b_{n,2} = qb_n$ for $q \in]0, 1[$, typically $q = 1/2$. It is easy to see that when the subsampling distribution is a convergent estimator of the true distribution, then the distance d between the subsampling distribution and the true one is stochastically equivalent to

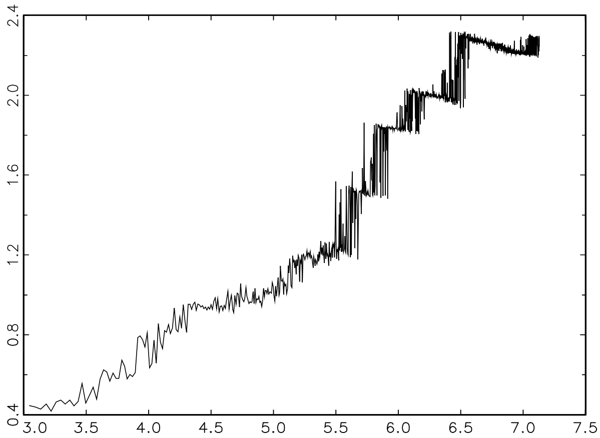
$$d(K_{b_n}, K_{qb_n}).$$

The idea is then to find the largest b_n which minimizes this quantity. Several distance (Kolmogorov distance, Wasserstein metrics, etc.) may be used. Of course, such a procedure may be difficult to analyze in the framework of dependent data because of the intricate dependence between both distributions. However, as a toy example, one may use either a double splitting trick in the spirit of Schick (2001) to estimate K_{qb} or consider nonoverlapping blocks to simplify the dependence structure. Proving that the procedure proposed by Götze and Račkauskas (1999) still works in this framework would be of interest.

Another approach to choose the subsampling size has been followed by Bertail et al. (2004) and Bertail et al. (1999) by considering log of quantile range. Indeed, some of the results of this paper may also be used to obtain convergent estimators of the rate of convergence of the statistics of interest. Most of the time the rate of convergence of the statistics say τ_n is unknown (this is precisely the reason for choosing a quantile range in the standardization of the maximum). If we denote by $K_{b,n}(x)$ either the nonnormalized smooth subsampling distribution or the rough one, based on a windows of size b_n, then it is easy to prove following the same lines as Bertail et al. (2004) and using the nice results of the authors (under their conditions) that, for any $\alpha_i \neq \alpha_j \in]0, 1[$ such that $K^{-1}(\alpha_i)$ and $K^{-1}(\alpha_j)$ are continuity points of the limiting distribution, we have in probability

$$\left|K_{b,n}^{-1}(\alpha_i) - K_{b,n}^{-1}(\alpha_j)\right| = C_{i,j}\tau_b^{-1}\left(1 + o_P(1)\right),$$

Fig. 1 Estimation of the rate of convergence, stability of the subsampling distribution



where $C_{i,j}$ is a constant depending on α_i, α_j, K, and on the underlying generating process. In particular, when $\tau_n = n^{\gamma} L(n)$ where $L()$ is a slowly varying function, then one gets a regression-type representation

$$-\log\left|K_{b,n}^{-1}(\alpha_i) - K_{b,n}^{-1}(\alpha_j)\right| = \gamma \log(b) + \log\bigl(L(b)\bigr) + \mathrm{Log}(C_{i,j}) + o_P(1).$$

It may be used to estimate the exponent γ and the slowly varying function by using different α_i, α_j and several subsampling size b_n. This boils down to a true simple regression when there is no slowly varying function. What is interesting to notice is that the rate γ generally depends on the tail index of the marginal distribution. Moreover, the dependence index θ of the time series is also directly related to the subsampling distribution of the unstandardized maximum to exceed a level u_{b_n}, which may be chosen of order $F_n^{-1}(1 - \eta b_n^{-1})$ (see Bertail et al. 2009), F_n being the empirical distribution of the observations. The subsampling methodology thus may provide some new methods for estimating the extremal index and/or the tail index in the very general framework considered by the authors.

Figure 1 illustrates why it may be interesting to draw minus the log of some quantile range of the subsampling distribution to infer both on the value of the rate of convergence and on the choice of the subsampling size. It represents minus the log range of the subsampling distribution of the unstandardized maximum for $\alpha_1 = 0.25$ and $\alpha_2 = 0.75$ as $\log(b_n)$ grows for a very simple AR times series model and $n = 5000$.

Note that the graph is quite unstable for small b_n (around 20), then stabilizes and becomes quite hieratic for moderate value of b_n (greater than 250). We thus have a very simple empirical tool for choosing the subsampling size. A natural choice for the subsampling distribution is to find the largest b_n before the break in the regression function (here leading to the choice $b \simeq 244$). The slope of the graphic in the stable part tells us the rate of convergence. It would be challenging to prove that this method is indeed valid in the framework of weak mixing time series data. An automatic choice of the subsampling size for cases where the bootstrap fails is still a challenge in any case.

References

Bertail P, Politis D, Romano J (1999) Undersampling with unknown rate of convergence. J Am Stat Assoc 94(446):569–579

Bertail P, Haeffke C, Politis D, White H (2004) A subsampling approach to estimating the distribution of diverging statistics with applications to assessing financial market risks. J Econom 120:295–326

Bertail P, Clémençon S, Tressou J (2009) Extreme values statistics for Markov chains via the pseudo-regenerative method. Extremes 12:327–360

Bickel PJ, Sakov A (2008) On the choice of the *m* out *n* bootstrap and confidence bounds for extrema. Stat Sin 18:967–985

Bickel PJ, Boley N, Brown JB, Huang H, Zhang NR (2010) Subsampling methods for genomic inference. Ann Appl Stat 4:1660–1697

Doukhan P, Louhichi S (1999) A new weak dependence condition and applications to moment inequalities. Stoch Process Appl 84:313–342

Götze F, Račkauskas A (1999) Adaptive choice of bootstrap sample sizes. In: de Gunst M, Klaassen C, van der Vaart A (eds) State of the art in probability and statistics: Festschrift for Willem R van Zwet, IMS lecture notes, monograph series, pp 286–309

Hall P, Jing BY, Lahiri SN (1998) On the sampling window method for long-range dependent data. Stat Sin 8:1189–1204

Politis D, Romano JP, Wolf M (1999) Subsampling. Springer, Berlin

Schick (2001) Sample splitting with Markov chains. Bernoulli 7:33–61

Test (2011) 20:491–496
DOI 10.1007/s11749-011-0273-z

DISCUSSION

Comments on: Subsampling weakly dependent time series and application to extremes

S.N. Lahiri · S. Mukhopadhyay

Published online: 1 November 2011

1 Introduction

Professors Doukhan, Prohl, and Robert are to be congratulated for their work on extending the validity of the subsampling method to a much wider class of processes compared to the existing literature that typically requires the processes to be strongly mixing (cf. Politis et al. 1999). As described in Sects. 1 and 2, many common time series models, including the ARMA models, often fail to satisfy the strong mixing condition but they typically satisfy the η- and λ-weak dependence conditions of Doukhan and Louhichi (1999) considered in this paper. As a result, extending the validity of the subsampling method under suitable η- and λ-weak dependence conditions is an important contribution. Expectedly, the smooth version of the subsampling estimator is especially suited to the form of the η- and λ-weak dependence conditions which give covariance bounds for smooth functions of the observations. This is one reason why the validity of the smooth subsampling estimator in Theorem 1 holds under *weaker* conditions than those for the rough subsampling estimator in Theorem 3. However, from the applications point of view, it is worth noting that while smoothing is known to play an important role in the resampling methodology in certain inference problems (e.g., inference on quantiles), caution must be exercised in cases where the limit distribution of the (unbootstrapped) statistic has points of discontinuity.

The authors also prove validity of the (smooth) subsampling estimator for the sample maximum under different sets of η- and λ-weak dependence conditions, for both the overlapping and the non-overlapping cases. This is an important problem where its natural competitor, namely, the block bootstrap method (cf. Künsch 1989

This comment refers to the invited paper available at doi:10.1007/s11749-011-0269-8.

S.N. Lahiri (✉) · S. Mukhopadhyay
Texas A & M University, College Station, USA
e-mail: snlahiri@stat.tamu.edu

and Liu and Singh 1992) does not always provide the correct answer. More precisely, it is well known (cf. Athreya et al. 1999; Lahiri 2003) that even under stronger strong mixing conditions, the block bootstrap methods do not provide a valid approximation to the distribution of the sample maximum when the resample size equals the sample size. Like in many other similar problems (cf. Bickel et al. 1997), a consistent block bootstrap approximation can be generated by choosing a resample size m that grows at a slower rate than the sample size n, i.e., when

$$m = o(n),$$

which is referred to as the 'm out of n' block bootstrap. In the rest of this note, we compare performance of the subsampling and the 'm out of n' block bootstrap methods for the sample maximum under the η-mixing condition.

2 Theoretical properties

For completeness, we briefly describe the 'm out of n' (overlapping or moving) block bootstrap (MBB) method using the notation of the main paper. Given observations $X_1, \ldots, X_n$ from a stationary time series, let $\{Y_{b,i} : i = 0, \ldots, N\}$ denote the overlapping blocks of size b, as defined in (3), where $N = n - b$. The 'm out of n' MBB resamples $k \geq 1$ blocks with replacement from this collection to generate a bootstrap sample of size $m = bk$, which we shall denote as $X_1^*, \ldots, X_m^*$. Then, the 'm out of n' MBB estimator of the distribution of a statistic $R_n = r_n(X_1, \ldots, X_n)$ is given by the conditional distribution of $R_{m,n}^* \equiv r_m(X_1^*, \ldots, X_m^*)$, given the X_i's. In particular, the 'm out of n' MBB estimator of $\mathbb{H}_n(x) \equiv P([M_n - v_n]/u_n \leq x)$ is given by

$$\widehat{\mathbb{H}}_{m,n}(x) \equiv P_*\big([M_{m,n}^* - \tilde{v}_m]/\tilde{u}_m \leq x\big),$$

where $M_{m,n}^* = \max\{X_1^*, \ldots, X_m^*\}$ is the bootstrap version of $M_n = \max\{X_1, \ldots, X_n\}$, P_* denotes the conditional probability given the X_i's, and where $\tilde{v}_m$ and $\tilde{u}_m$ are analogs of the centering and scaling constants v_n and u_n, respectively. A possible choice of $\tilde{v}_m$ and $\tilde{u}_m$ that leads to a valid approximation is given by $\tilde{v}_m = v_m$ and $\tilde{u}_m = u_m$, which normalizes $M_{m,n}^*$ at the level of the subsamples (that are of size m). A more standard choice, especially when these constants are unknown, are given by replacing F and n in the definitions of v_n and u_n (given right after (10)) with the empirical distribution function F_n (say) of $X_1, \ldots, X_n$ and with m, respectively. Finally, when $k = 1$, the 'm out of n' MBB reduces to the *rough* subsampling estimator of $\mathbb{H}_n(\cdot)$. In this case, the standard choice of $\tilde{v}_m$ is given by $\tilde{v}_m = F_n^{\leftarrow}(1 - n^{-1})$ while $\tilde{u}_m$ is chosen as in the last case. For all these variants, the key requirement for the 'm out of n' MBB to work is that

$$u_m^{-1}\big[|\tilde{v}_m - v_m| + |\tilde{u}_m - u_m|\big] \to_p 0 \quad \text{as } n \to \infty, \tag{1}$$

which we shall assume for the rest of this discussion. We, however, point out that for the data-based choices of $\tilde{v}_m$ and $\tilde{u}_m$ mentioned above, (1) can be proved using the arguments developed in Fukuchi (1994). Then, we have the following result:

Theorem *Suppose that* (1) *holds and that the conditions of Theorem* 5 *hold for the* η*-weakly dependent case. Also suppose that* $b = b_n$, $k = k_n$, *and* $m = m_n \equiv b_n k_n$ *are* *such that* (i) $b_n^{-1} + n^{-1}m_n = o(1)$; (ii) $\lim_{p\to\infty} \limsup_{n\to\infty} k[mu_m\eta(pb)]^{1/2} = 0$; (iii) $kP(M_b > w_m(x)) = O(1)$ *for each* $x \in \mathbb{R}$. *Then,*

$$\sup_{x\in\mathbb{R}} \left|\hat{\mathbb{H}}_{m,n}(x) - \mathbb{H}_n(x)\right| \to_p 0 \quad \text{as } n \to \infty.$$

A sketch of the proof of the theorem is given in the Appendix. To briefly comment on the conditions, (i) is a very standard condition on the block size b in the block bootstrap literature, (ii) is similar to the η-weak dependence condition in Theorem 5, and (iii) is implied by $P(M_b > w_m(x)) \le b\bar{F}(w_m(x))$ for each $x \in \mathbb{R}$, which is akin to (11). Note that the theorem requires the block size b to grow to infinity with n and the resample size m to grow at a rate slower than n, but it does not otherwise impose any conditions on the number of resamples k. In particular, the consistency holds for the case $k = 1$, i.e., for the rough subsampling estimator and, more generally, for the 'm out of n' MBB under the conditions of the Theorem. This extends the results of Athreya et al. (1999) to the η-dependence case. Similar result also holds for the nonoverlapping version of the 'm out of n' block bootstrap, which we omit to save space.

3 Numerical results

We now report the results from a simulation study comparing the performance of the 'm out of n' MBB and the (rough) subsampling methods for the sample maximum. As in the main paper, we considered the model given by (1) and drew 1000 Monte Carlo samples with sample sizes $n = 40$, 200, and 2000. For each sample $(X_1, X_2, \ldots, X_n)$, we resampled the data randomly 1000 times for constructing the block bootstrap estimator. The block lengths we have considered here are (closest integers to) $b_1 = 2n^{1/3}$ and $b_2 = \sqrt{(1.25)n}$; The second block length b_2 equals 50 for $n = 2000$, matching the choice of b in the main paper. The following table reports the (scaled) global error (using a version of the Cramer–von Mises distance) of approximating the true distribution $H_n(\cdot)$ by the 'm out of n' MBB and the subsampling methods.

Results in Table 1 show that the block bootstrap produces more accurate global approximation to the true distribution than the subsampling method for all combinations of n and b. Further, for both methods, the overlapping versions have slightly

Table 1 Global errors (in %) for approximating the true distribution $\mathbb{H}_n$ by the 'm out of n' MBB and the subsampling methods (denoted as MBB and SS, respectively, in the table)

		Overlapping			Non-overlapping		
		$n = 50$	$n = 200$	$n = 2000$	$n = 50$	$n = 200$	$n = 2000$
MBB	b_1	1.03	0.84	0.37	0.9	0.78	0.36
MBB	b_2	0.3	0.2	0.15	0.22	0.18	0.14
SS	b_1	4.66	2.89	0.9	4.86	2.7	0.88
SS	b_2	3.4	1.55	0.4	3.42	1.48	0.42

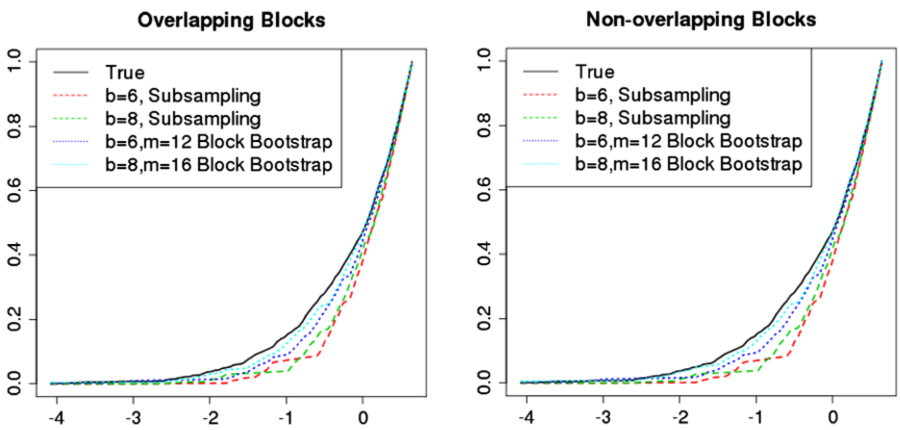


Fig. 1 Block bootstrap and subsampling estimators of the true distribution $\mathbb{H}_n$ for $n = 50$

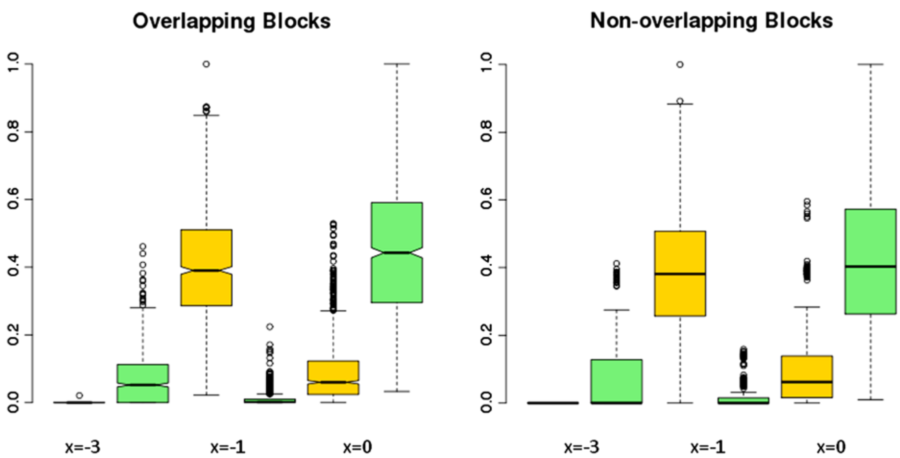


Fig. 2 Box-plots of the block bootstrap (even numbered) and subsampling estimators (odd numbered) of $\mathbb{H}_n(x)$ at $x = -3, -1, 0$ for $n = 50$

better performance than the non-overlapping versions. This is also evident from Fig. 1 which gives the mean CDF curves for the two methods for $n = 50$, based on 1000 simulation runs. We also considered local performance of the two methods, comparing the estimators of the CDF $\mathbb{H}_n(x)$ at different values of x. Figure 2 gives the box-plots of the estimators based on 1000 simulation runs for $n = 50$.

From Fig. 2, it follows that the estimators based on the non-overlapping versions of each method have higher variability than the corresponding overlapping versions, like in the case of smooth functions of means (cf. Lahiri 1999). Further, the variabilities of the subsampling estimators are much smaller than their block bootstrap counterparts at the tails ($x = -3, 0$), but the pattern reverses in the middle ($x = -1$). We observed a similar behavior also for $n = 200$ and $n = 2000$ (not shown here).

Acknowledgement Research partially supported by NSF grant number DMS 1007703 and NSA grant number H98230-11-1-0130.

Appendix

Here we provide an outline of the proof of the Theorem. First, note that by the continuity of the limit law $\mathbb{H}(\cdot)$ and by a subsequence argument, it is enough to show that for each fixed x,

$$\hat{\mathbb{H}}_{m,n}(x) - \mathbb{H}(x) \to_p 0.$$

Fix $x \in \mathbb{R}$. Write $M_{i,b} = \max Y_{i,b}$ and $M_b^* = \max\{X_1^*, \dots, X_b^*\}$, the maximum over a single resampled block. Also, let $\mathbb{H}_b^\dagger(x) = P_*([M_b^* - v_m]/u_m \le x)$. Then, by (1) and the independence of the resampled blocks, it is easy to check that

$$\hat{\mathbb{H}}_{m,n}(x) = \big[\mathbb{H}_b^\dagger(x)\big]^k + o_p(1) = \left(1 + \frac{[k(1-\mathbb{H}_b^\dagger(x)]}{k}\right)^k + o_p(1),$$

which, by Theorem 5 and Condition (i), converges to $\mathbb{H}(x)$ in probability provided $k^2 \operatorname{Var}(\mathbb{H}_b^\dagger(x)) \to 0$. Note that, by stationarity, with $w_m(x) = u_m^{-1}x + v_m$,

$$\begin{aligned}
&k^2 \operatorname{Var}\big(\mathbb{H}_b^\dagger(x)\big) \\
&\quad \le C_1 N^{-1} k^2 \sum_{j=0}^{N} \big|\operatorname{Cov}\big(\boldsymbol{I}\big(M_{0,b} > w_m(x)\big), \boldsymbol{I}\big(M_{j,b} > w_m(x)\big)\big)\big| \\
&\quad \le C_2 N^{-1} k^2 \Bigg[pbP\big(M_{0,b} > w_m(x)\big) \\
&\qquad + \sum_{j=(p+1)b}^{N} \big|\operatorname{Cov}\big(\boldsymbol{I}\big(M_{0,b} > w_m(x)\big), \boldsymbol{I}\big(M_{j,b} > w_m(x)\big)\big)\big| \Bigg] \\
&\quad \equiv I_{1n}(p) + I_{2n}(p) \quad \text{(say)},
\end{aligned}$$

where $C_1, C_2, \ldots \in (0, \infty)$ are constants. By Conditions (i) and (iii), it follows that $I_{1n}(p) = O(m/n) = o(1)$ for every fixed $p \ge 1$. And, by retracing the arguments in the proof of Theorem 5, one can show that for any $j \ge pb$, $p > 1$, and $\alpha_{jn} > 0$,

$$\begin{aligned}
&\big|\operatorname{Cov}\big(\boldsymbol{I}\big(M_0 > w_m(x)\big), \boldsymbol{I}\big(M_j > w_m(x)\big)\big)\big| \\
&\quad \le 2b\big(u_m \alpha_{jn}^{-1} \eta(j-b) + 2\alpha_{jn}\big[(1+\gamma x)_+^{-\frac{1}{\gamma}-1}/m\big]\big).
\end{aligned}$$

Setting $\alpha_{jn} = [m u_m \eta(j-b)]^{1/2}$ and noting that $m = kb$, we have

$$I_{2n}(p) \le C_3 N^{-1} k^2 \sum_{j=[p+1]b}^{N} \big[b m^{-1/2} u_m^{1/2} \eta(j-b)^{1/2}\big] \le C_3 k m^{1/2} u_m^{1/2} \eta(pb)^{1/2},$$

which, by Condition (iii), goes to zero by first letting $n \to \infty$ and then $p \to \infty$. This completes the proof of the theorem.

References

Athreya KB, Fukuchi J-I, Lahiri SN (1999) On the bootstrap and the moving block bootstrap for the maximum of a stationary process. J Stat Plan Inference 76:1–17

Bickel PJ, Götze F, van Zwet WR (1997) Resampling fewer than nn observations: gains, losses, and remedies for losses. Stat Sin 7:1–31

Doukhan P, Louhichi S (1999) A new weak dependence condition and applications to moment inequalities. Stoch Process Appl 84:313–342

Fukuchi J-I (1994) Bootstrapping extremes of random variables. PhD Thesis, Iowa State University, IA

Künsch HR (1989) The jackknife and the bootstrap for general stationary observations. Ann Stat 17:1217–1241

Lahiri SN (1999) Theoretical comparison of block bootstrap methods. Ann Stat 27:386–404

Lahiri SN (2003) Resampling methods for dependent data. Springer, New York

Liu RY, Singh K (1992) Moving blocks jackknife and bootstrap capture weak dependence. In: Lepage R, Billard L (eds) Exploring the limits of bootstrap. Wiley, New York, pp 225–248

Politis DN, Romano JP, Wolf M (1999) Subsampling. Springer, New York

Test (2011) 20:497–498
DOI 10.1007/s11749-011-0274-y

DISCUSSION

Comments on: Subsampling weakly dependent time series and application to extremes

Efstathios Paparoditis

Published online: 1 November 2011

Doukhan, Prohl and Robert should be congratulated for a nice piece of work. They succeeded in extending the validity of subsampling procedures with overlapping or non-overlapping blocks to a wider class of processes than the one commonly considered in the literature, following the invention of this resampling procedure by Politis and Romano (1994). In fact, they weakened the basic strong mixing assumption to a weak dependence condition introduced by Doukhan and Louhichi (1999). It is known that the class of processes obeying this kind of dependence structure is much broader than the strong mixing class and includes, for instance, Bernoulli shifts with discrete distribution of the innovations, which turns out to be quite important for many statistical applications.

One concern that I have refers to the technical Assumption C2 which is called a concentration condition. How easy is it to verify that this condition is satisfied in a particular application of the subsampling procedure, taking into account that it refers to the behavior of the (finite sample) distribution of the statistic of interest? The authors refer to some examples given in the literature, but I think it is worth elaborating more on this important point. Another issue refers to the subsampling estimator (7) which the author introduce as more suitable for the weakly depended case considered. It should be made more clear why this smoothed estimator is more suitable, taking into account that it requires the choice of additional (smoothing) parameters.

This comment refers to the invited paper available at doi:10.1007/s11749-011-0269-8.

E. Paparoditis (✉)
University of Cyprus, Department of Mathematics and Statistics, 1678 Nicosia, Cyprus
e-mail: stathisp@ucy.ac.cy

References

Doukhan P, Louhichi S (1999) A new weak dependence condition and applications to moment inequalities. Stoch Process Appl 84:313–342

Politis DN, Romano JP (1994) The stationary bootstrap. J Am Stat Assoc 89:1303–1313

Rejoinder on: Subsampling weakly dependent time series and application to extremes

Paul Doukhan, Silika Prohl & Christian Y. Robert

TEST
An Official Journal of the Spanish Society of Statistics and Operations Research

ISSN 1133-0686
Volume 20
Number 3

TEST (2011) 20:499-502
DOI 10.1007/s11749-011-0275-x

Test (2011) 20:499–502
DOI 10.1007/s11749-011-0275-x

DISCUSSION

Rejoinder on: Subsampling weakly dependent time series and application to extremes

Paul Doukhan · Silika Prohl · Christian Y. Robert

Published online: 1 November 2011

We are very grateful to all discussants for their valuable and insightful comments. We sincerely appreciate that they were willing to take the time to provide such valuable feedback. We also want to thank the co-editors Ricardo Cao and Dominigo Morales for initiation of this discussion.

As Prof. Velasco points out, among the main important issues in practice using the smooth and/or the rough subsampling estimators are the choices of the block size b, the smoothing parameter ε and the degree of overlapping.

– Two ideas have emerged from all the comments to find an appropriate choice for the block size.

  Firstly Prof. Bertail and Bravo propose to extend the method introduced by Bickel and Sakov (2008) in the context of the so-called 'm out of n' bootstrap for

This rejoinder refers to the comments available at doi:10.1007/s11749-011-0270-2, doi:10.1007/s11749-011-0271-1, doi:10.1007/s11749-011-0272-0, doi:10.1007/s11749-011-0273-z, doi:10.1007/s11749-011-0274-y.

P. Doukhan (✉)
French Universitary Institute, IUF University Cergy Pontoise, UFR Sciences-Techniques, Saint-Martin, 2, Avenue Adolphe-Chauvin, Pontoise, B.P. 222, 95302 Cergy-Pontoise cedex, France
e-mail: doukhan@u-cergy.fr

S. Prohl
Department of Operations Research and Financial Engineering, Princeton University, Sherrerd Hall, Charlton Street, 08544 Princeton, NJ, USA
e-mail: sprohl@princeton.edu

C.Y. Robert
Institut de Science Financière et d'Assurances, Université de Lyon, Université Lyon 1, 50 Avenue Tony Garnier, 69007 Lyon, France
e-mail: christian.robert@univ-lyon1.fr

i.i.d. observations to dependent data and subsampling estimators. Prof. Bravo provides moreover a very interesting simulation study that suggests that the method seems to work very well. It is clearly an essential approach which will be considered. It still needs cumbersome calculations to justify it theoretically so that we prefer to report it to a careful separate study. Certainly coupling arguments will be needed to go back with arguments used under independence: such arguments are already developed for β-mixing, see Doukhan (1994), and for τ-dependence, see Dedecker and Prieur (2004).

We should underline that this approach aims at minimizing a distance between the subsampling distribution and the distribution of the statistics for the overall sample but not between the subsampling distribution and the asymptotic distribution. It raises the question about considering or not the asymptotic distribution when choosing the subsampling size. Prof. Velasco sheds light on this question for the particular example of the distribution of the maximum for which clusters of extreme values may have strong impact on bias for a too small value of the block size. We agree with this remark. However, one has to be able to have an accurate idea of the bias between the distribution of the statistics for the overall sample and the asymptotic distribution, which is rarely the case in general without imposing additional assumptions.

Secondly, Prof. Bertail proposes to consider the rate of convergence of the logarithm of some quantile ranges of the statistics of interest and to perform a regression analysis as is done in Bertail et al. (2004). A break in the regression slope gives then an idea for the choice of the block size. It is a quite promising approach when the rate of convergence of the statistics is given by a regularly varying function. As for the previous approach, it still needs a detailed theoretical analysis in the framework of weak mixing time series data.

Prof. Velasco raises the question whether it is possible to have a higher order asymptotic analysis of the estimator because this could be helpful for the choice of the block size. This is a really hard issue that has been considered e.g. in Götze and Hipp (1994), but we think that deep-going work remains to be handled in our setting.

- The choice of the smoothing parameter is intricate and should be done at the same time as the block size. As underlined by Prof. Bertail, the sensitivity of the subsampling distribution to smoothing can be very important in practice. We certainly agree that this is a problem that should be carefully addressed.
- The question of the degree of overlapping is a difficult theoretical one. However, as can be seen from the numerical results presented by Prof. Lahiri and Mukhopadhyay and our simulation study, the overlapping estimators have slightly better performance than the non-overlapping version. This confirms what we explained in our remark on the choice of procedure before Sect. 4 of our paper. Prof. Velasco makes mention of a partial overlapping scheme as suggested in Politis and Romano (2010). It is an interesting suggestion, which, however, leads to one introduce a new parameter h that is used to balance between the two schemes (overlapping/non-overlapping). It is also not clear that it will lead to a larger efficiency in practice for small to moderate sample size.

Prof. Bravo brings an important point regarding the integer-valued time series models and proposes to carry on the analysis of our estimators in the case of such time series models. This is a good suggestion since such integer-valued models have an important development nowadays, for application's sake. Conferences now develop such preoccupations: for example, the Second international workshop on integer-valued time series was organized in Protaras, Cyprus, 19–21 June 2011. Many papers also appeared, and we only cite some few of them: Ferland et al. (2006), Doukhan et al. (2006), Fokianos and Tjostheim (2011), Fokianos (2011), Doukhan et al. (2011).

Prof. Lahiri and Mukhopadhyay compare the rough subsampling estimator with a natural competitor, the 'm out of n' moving block bootstrap estimator, which is based on a resampling method. As underlined by Prof. Velasco, subsampling has the advantage to avoid the construction of new time series joining blocks of the original series that may alter the original dependence structure and that need to specify the choice of the length of the new time series. However, a simulation study seems to show that the moving block bootstrap estimator has interesting properties for estimating the sample maximum distribution when the theoretical normalization coefficients are used. We are particularly intrigued by such results and we will conduct our own simulation study, applying to other statistics and time series models to understand the differences between the two methods.

Prof. Paparoditis and Bertail both stressed the fact that a concentration assumption is needed to derive a.s. sure convergence of the subsampling procedures. This condition is used (see e.g. (4.1.1), in Lemma 4.1, p. 84, Dedecker et al. 2007) to provide weak dependence from dependence bounds on indicators. Such a condition seems indeed necessary in order to get higher order moments and corresponding faster decay rates necessary to use the Borel–Cantelli argument. The condition C.2 holds if a concentration condition holds for the finite repartitions $(X_1, \ldots, X_b)$ and if the function $s_b : \mathbb{R}^b \to \mathbb{R}$ is regular enough. As mentioned in the paper after (6), sufficient conditions for this result are already given in Doukhan and Wintenberger (2008) for infinite memory models; some arguments are also given in Lemma 3.7, p. 67, Dedecker et al. (2007) for the case of LARCH(∞) models. The case of infinite moving averages is also of interest; Doukhan et al. (2002) prove the regularity of marginal distributions if the independent innovations satisfy $|\mathbb{E}(e^{i\lambda\zeta_0})| \le C(1+|\lambda|)^{-\epsilon}$ for arbitrarily small $\epsilon > 0$. Finally the simplest situation for which such a condition holds is the case of noisy data $X_t = Y_t + \zeta_t$ for strong white noise $(\zeta_t)_{t\ge 0}$ already satisfying C.2 and independent of $(Y_t)_{t\ge 0}$.

References

Bertail P, Haëfke C, Politis D, White H (2004) A subsampling approach to estimating the distribution of diverging statistics with applications to assessing financial market risks. J Econom 120:295–326

Bickel PJ, Sakov A (2008) On the choice of the m out n bootstrap and confidence bounds for extrema. Stat Sin 18:967–985

Dedecker J, Prieur C (2004) Coupling for τ-dependent sequences and applications. J Theor Probab 17:855–861

Dedecker J, Doukhan P, Lang G, León JR, Louhichi S, Prieur C (2007) Weak dependence: models, theory and applications. Lecture notes in statistics, vol 190. Springer, New York

Doukhan P (1994) Mixing properties and examples. Lectures notes in statistics, vol 85. Springer, New York

Doukhan P, Wintenberger O (2008) Weakly dependent chains with infinite memory. Stoch Process Appl 118:1997–2013

Doukhan P, Fokianos K, Tjostheim D (2011) On weak dependence conditions for Poisson autoregressions (submitted for publication)
Doukhan P, Lang D, Surgailis D (2002) Functional CLTs for short or long memory linear sequences. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov. Ann Inst Henri Poincaré B 38(6):879–896
Doukhan P, Latour A, Oraichi D (2006) A simple integer-valued bilinear times series models. Adv Appl Probab 38:1–20
Ferland R, Latour A, Oraichi D (2006) Integer-valued GARCH processes. J Time Ser Anal 27:923–942
Fokianos K (2011) Count time series (submitted for publication in The Handbook of Time Series Analysis)
Fokianos K, Tjostheim D (2011) Loglinear Poisson autoregression. J Multivar Anal 102:563–578
Götze F, Hipp C (1994) Asymptotic expansions for statistics in time series. Ann Stat 22:211–239
Politis DN, Romano JP (2010) K-sample subsampling in general spaces: the case of independent time series. J Multivar Anal 101:316–326