Extremes (2013) 16:203–239
DOI 10.1007/s10687-012-0159-9

Estimation of limiting conditional distributions for the heavy tailed long memory stochastic volatility process

Rafał Kulik · Philippe Soulier

Received: 14 February 2012 / Revised: 18 August 2012 /
Accepted: 18 September 2012 / Published online: 2 October 2012

Abstract We consider Stochastic Volatility processes with heavy tails and possible long memory in volatility. We study the limiting conditional distribution of future events given that some present or past event was extreme (i.e. above a level which tends to infinity). Even though extremes of stochastic volatility processes are asymptotically independent (in the sense of extreme value theory), these limiting conditional distributions differ from the i.i.d. case. We introduce estimators of these limiting conditional distributions and study their asymptotic properties. If volatility has long memory, then the rate of convergence and the limiting distribution of the centered estimators can depend on the long memory parameter (Hurst index).

Keywords Stochastic volatility · Heavy tails · Long memory · Regular variation

AMS 2000 Subject Classifications 60F05 · 60G70

1 Introduction

One of the empirical features of financial data is that log-returns are uncorrelated, but their squares, or absolute values, are dependent, possibly with long memory. Another important feature is that log-returns are heavy-tailed. There are two common classes

The research of the first author was supported by the NSERC Discovery grant.

The research of the second author was partially supported by the ANR grant ANR-08-BLAN-0314-02.

R. Kulik (✉)
University of Ottawa, Ottawa, Canada
e-mail: rkulik@uottawa.ca

P. Soulier
Université de Paris-Ouest Nanterre, Paris, France
e-mail: philippe.soulier@u-paris10.fr

of processes to model such behaviour: the generalized autoregressive conditional heteroscedastic (GARCH) process and the stochastic volatility (SV) process; the latter introduced by Breidt et al. (1998) and Harvey (1998). The former class of models rules out long memory in the squares, while the latter allows for it. We will therefore concentrate in this paper on the class of SV processes, which we define now.

Throughout the paper, we will assume that the observed process $\{Y_j, j \in \mathbb{Z}\}$ can be expressed as

$$Y_j = \sigma(X_j)Z_j = \sigma_j Z_j \tag{1}$$

where σ is some (possibly unknown) positive function, $\{Z_j, j \in \mathbb{Z}\}$ is an i.i.d. sequence and $\{X_j, j \in \mathbb{Z}\}$ is a stationary Gaussian process with mean zero, unit variance, autocovariance function $\{\gamma_n\}$, and independent from the i.i.d. sequence. The sequence $\{\sigma(X_j), j \in \mathbb{Z}\}$ can be seen as a proxy for the volatility. We will assume that either $\{X_j, j \in \mathbb{Z}\}$ is weakly dependent in the sense that

$$\sum_{n=1}^{\infty} |\gamma_n| < \infty, \tag{2}$$

or that it has long memory with Hurst index $H \in (1/2, 1)$, i.e.

$$\gamma_n = \mathrm{cov}(X_0, X_n) = n^{2H-2}\ell(n) \tag{3}$$

where ℓ is a slowly varying function.

It will also always be assumed that the marginal distribution F_Z of the i.i.d. sequence $\{Z_j\}$ has a regularly varying tail with index $\alpha > 0$:

$$\lim_{x\to\infty} \frac{\mathbb{P}(Z > x)}{x^{-\alpha}L(x)} = \beta, \quad \lim_{x\to\infty} \frac{\mathbb{P}(Z < -x)}{x^{-\alpha}L(x)} = (1-\beta), \tag{4}$$

where $L(\cdot)$ is slowly varying at infinity and $\beta \in [0, 1]$. Examples of heavy tailed distributions include the stable distributions with index $\alpha \in (0, 2)$, the t distribution with α degrees of freedom, and the Pareto distribution with index α.

By Breiman's lemma (Breiman 1965; Resnick 2007), if

$$\mathbb{E}[\sigma^{\alpha+\epsilon}(X)] < \infty \tag{5}$$

for some $\epsilon > 0$, then the marginal distribution of $\{Y_j\}$ also has a regularly varying right tail with index α and

$$\lim_{x\to\infty} \frac{\mathbb{P}(Y > xy)}{\mathbb{P}(Z > x)} = \mathbb{E}[\sigma^{\alpha}(X)]y^{-\alpha}, \tag{6}$$

where X, Y and Z denote random variables with the same joint distribution as X_0, Y_0 and Z_0. This one-dimensional result can be extended to a multivariate setting. The finite dimensional marginal distributions of the SV process are multivariate regularly varying with spectral measure concentrated on the axes; see Proposition 1 for details.

Estimation and test of the possible long memory of such processes has been studied by Hurvich et al. (2005). Estimation of the tail of the marginal distribution by the Hill estimator has been studied in Kulik and Soulier (2011).

In this paper we are concerned with certain extremal properties of the finite dimensional joint distributions of the process $\{Y_j\}$ when Z is heavy tailed and the Gaussian process $\{X_j\}$ possibly has long memory.

From the extreme value point of view, there is a significant distinction between GARCH and SV models. In the first one, exceedances over a large threshold are asymptotically dependent and extremes do cluster. In the SV model, it follows from the multivariate regular variation result (Proposition 1) that exceedances are asymptotically independent. More precisely, for any positive integer m, and positive real numbers x, y,

$$\lim_{t\to\infty} t\mathbb{P}(Y_0 > a(t)x\ ,\ Y_m > a(t)y) = 0\ , \tag{7}$$

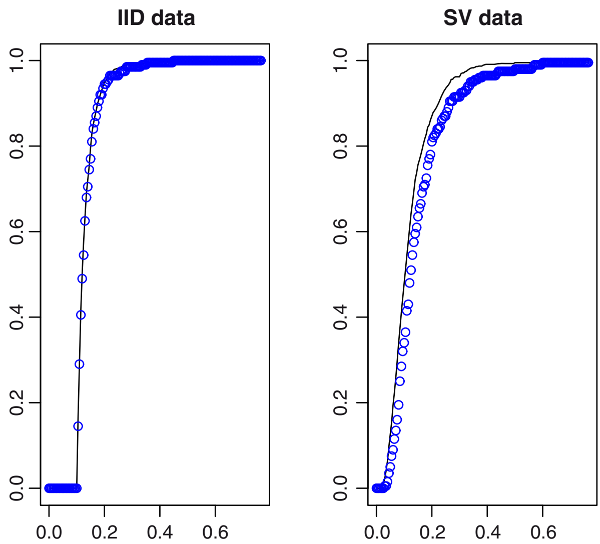
where $a(t) = F_Z^{\leftarrow}(1 - 1/t)$ and $F_Z^{\leftarrow}$ is the left continuous inverse of F_Z.

The above observations may lead to the incorrect conclusion that, for the SV process, there is no spillover from past extreme observations onto future values and from the extremal behaviour point of view we can treat the SV process as an i.i.d. sequence. However, under the assumptions stated previously, it holds that

$$\lim_{t\to\infty} \mathbb{P}(Y_m \le y \mid Y_0 > t) = \frac{\mathbb{E}[\sigma^\alpha(X_0)F_Z(y/\sigma(X_m))]}{\mathbb{E}[\sigma^\alpha(X_0)]}. \tag{8}$$

See Lemma 11 in Section 4 for a proof. Therefore, the limiting conditional distribution is influenced by the dependence structure of the time series. To illustrate this, we show in Fig. 1 estimates of the standard distribution function and of the conditional

Fig. 1 Empirical conditional distribution (*points*) and empirical distribution (*solid line*) for SV model (*right panel*) and i.i.d. data (*left panel*)



distribution for a simulated SV process. Clearly, the two estimated distributions are different, as suggested by Eq. 8. For a comparison, we also plot the corresponding estimates for i.i.d. data. Other kind of extremal events can be considered, for instance, we may be interested in the conditional distribution of some future values given that a linear combination (portfolio) of past values is extremely large, or that two consecutive values are large. As in Eq. 8, in each of these cases, a proper limiting distribution can be obtained. To give a general framework for these conditional distributions, we introduce a modified version of the extremogram of Davis and Mikosch (2009). For fixed positive integers $h < m$ and $h' \geq 0$, Borel sets $A \subset \mathbb{R}^h$ and $B \subset \mathbb{R}^{h'+1}$, we are interested in the limit denoted by $\rho(A, B, m)$, if it exists:

$$\rho(A, B, m) = \lim_{t\to\infty} \mathbb{P}((Y_m, \dots, Y_{m+h'}) \in B \mid (Y_1, \dots, Y_h) \in tA). \tag{9}$$

The set A represents the type of events considered. For instance, if we choose $A = \{(x, y, z) \in [0, \infty)^3 \mid x + y + z > 1\}$, then for large t, $\{(Y_{-2}, Y_{-1}, Y_0) \in tA\}$ is the event that the sum of last three observations was extremely large. The set B represents the type of future events of interest.

In the original definition of the extremogram of Davis and Mikosch (2009), the set B is also dilated by t. This is well suited to the context of asymptotic dependence, as arises in GARCH processes. But in the context of asymptotic independence, this would yield a degenerate limit: if $h < m$, then for most sets A and B,

$$\lim_{t\to\infty} \mathbb{P}((Y_m, \dots, Y_{m+h'}) \in tB \mid (Y_1, \dots, Y_h) \in tA) = 0.$$

The general aim of this paper is to investigate the existence of these limiting conditional distributions appearing in Eq. 9 and their statistical estimation. The paper is the first step towards understanding conditional laws for stochastic volatility models. Although we provide theoretical properties of estimators, their practical use should be investigated in conjunction with resampling techniques. This is a topic of the authors' current research.

The paper is structured as follows. In Section 2, we present a general framework that enables to treat various examples in a unified way. In Section 3 we present the estimation procedure with appropriate limiting results. The proofs are given in Section 4. In the Appendix we collect relevant results on second order regular variation, (long memory) Gaussian processes, and criteria for tightness.

1.1 Notation

We conclude this introduction by gathering some notation that will be used throughout the paper. We denote convergence in probability by $\to_P$, weak convergences of sequences of random variables or vectors by $\to_d$ and weak convergence in the Skorokhod space $\mathcal{D}(\mathbb{R}^q)$ of cadlag functions defined on $\mathbb{R}^q$ endowed with the J_1 topology by $\Rightarrow$.

Boldface letters denote vectors. Product of vectors and inequalities between vectors are taken componentwise: $\mathbf{u} \cdot \mathbf{v} = (u_1 v_1, \dots, u_d v_d)$; $\mathbf{x} \leq \mathbf{y}$ if and only if $x_i \leq y_i$

for all $i = 1, \ldots, d$. The (multivariate) interval $(-\infty, \mathbf{y}]$ is defined accordingly: $(-\infty, \mathbf{y}] = \prod_{i=1}^{d}(-\infty, y_i]$.

For any univariate process $\{\xi_j\}$ and any integers $h \le h'$, let $\boldsymbol{\xi}_{h,h'}$ denote the $(h' - h + 1)$-dimensional vector $(\xi_h, \ldots, \xi_{h'})$.

If $\boldsymbol{\xi}_{h,h'} = (\xi_h, \ldots, \xi_{h'})$ is a random vector and $\sigma : \mathbb{R} \to \mathbb{R}$ is a deterministic function, then $\boldsymbol{\sigma}(\boldsymbol{\xi}_{h,h'})$ denotes a vector

$$\boldsymbol{\sigma}(\boldsymbol{\xi}_{h,h'}) = (\sigma(\xi_h), \ldots, \sigma(\xi_{h'})).$$

For $A \subset \mathbb{R}^d$ and $\mathbf{u} \in (0, \infty)^d$, $\mathbf{u}^{-1} \cdot A = \{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{u} \cdot \mathbf{x} \in A\}$.

If $\mathbf{X}$ is a random vector, we denote by $L^p(\mathbf{X})$ the set of measurable functions f such that $\mathbb{E}[|f(\mathbf{X})|^p] < \infty$.

The σ-field generated by the process $\{X_j\}$ is denoted by $\mathcal{X}$.

2 Regular variation on subcones

In this section, we will present our general framework. A crucial property of the SV processes is that the finite dimensional marginal distributions are multivariate regularly varying and are asymptotically independent (in the sense of extreme value theory). For the sake of completeness, we state and prove this fact formally. Recall that a measure ν on the Borel sets of $[-\infty, \infty]^h \setminus \{\mathbf{0}\}$ is said to be a Radon measure if $\nu(A) < \infty$ for each relatively compact set A, i.e. for each set A bounded away from $\mathbf{0}$. A sequence of Radon measures ν_n on $\mathbb{R}^h$ is said to converge vaguely to a Radon measure ν, which will be denoted by $\nu_n \to_v \nu$ if $\nu_n(A) \to \nu(A)$ for all relatively compact set A of $[-\infty, \infty]^h \setminus \{\mathbf{0}\}$. Recall that $a(t) = F_Z^{\leftarrow}(1 - 1/t)$.

Proposition 1 *The finite dimensional distributions of the process $\{Y_j, j \in \mathbb{Z}\}$ are multivariate regularly varying and for each fixed integer h*

$$\lim_{t\to\infty} t\mathbb{P}((Y_1, \ldots, Y_h) \in a(t)\cdot) \to_v \mathbb{E}[\sigma^\alpha(X_1)]\boldsymbol{\nu}(\cdot) \tag{10}$$

where the measure $\boldsymbol{\nu}$ is characterized by

$$\boldsymbol{\nu}([\mathbf{x}, \mathbf{y}]^c) = (1 - \beta)\sum_{i=1}^{h}(-x_i)^{-\alpha} + \beta\sum_{i=1}^{h} y_i^{-\alpha},$$

for $\mathbf{x} = (x_1, \ldots, x_h) \in [-\infty, 0)^h$ and $\mathbf{y} = (y_1, \ldots, y_k) \in (0, \infty]^h$, and β is defined in Eq. 4.

The special form of the measure $\boldsymbol{\nu}$ which is concentrated on the axes is due to the asymptotic independence (in the sense of extreme value theory) of the bivariate distributions of the process $\{Y_j, j \in \mathbb{Z}\}$, regardless of the memory of the volatility process $\{\sigma(X_j), j \in \mathbb{Z}\}$. In fact, as will be clear from the proof in Section 4, a particular structure for the volatility process is not needed.

Let us now describe the conditional distributions we will consider. Since we consider dilated sets $tA = \{tx : x \in A\}$, where $A \subset \mathbb{R}^h$ for some integer $h > 0$ and $t > 0$, it is natural to consider cones, that is subsets $\mathcal{C}$ of $[0, \infty]^h$ such that $tx \in \mathcal{C}$ for all $x \in \mathcal{C}$ and $t > 0$. Hence, our discussion in this section is related to the concept of regular variation on cones or hidden regular variation (see Das and Resnick 2011; Mitra and Resnick 2011; Resnick 2008). We endow $\mathbb{R}^h$ with the topology induced by any norm and $[0, \infty]^h$ is the compactification of $[0, \infty)^h$. A subset A of $[0, \infty]^h \setminus \{0\}$ is relatively compact if its closure is compact. See Resnick (2007, Chapter 6) for more details.

We are interested in cones $\mathcal{C}$ such that there exists an integer $\beta_{\mathcal{C}}$ and a Radon measure $\nu_{\mathcal{C}}$ on $\mathcal{C}$ such that, for all relatively compact subsets A of $\mathcal{C}$ with $\nu_{\mathcal{C}}(\partial A) = 0$,

$$\lim_{t\to\infty} \frac{\mathbb{P}((Z_1, \dots, Z_h) \in tA)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} = \nu_{\mathcal{C}}(A). \tag{11}$$

Intuitively, the number $\beta_{\mathcal{C}}$ corresponds to the number of components of a point of a relatively compact subset A of the cone $\mathcal{C}$ that must be separated from zero. For the simplicity and clarity of exposition, we will restrict our considerations to the following type of cones. Let $k \geq 1$ and $P_1, \dots, P_k$ be nonempty subsets of $\{1, \dots, h\}$ such that $P_i \not\subset P_j$ for any pair i, j, though it is not assumed that $P_u \cap P_v = \emptyset$ for $u, v \geq 1$. To avoid trivialities, we also assume that $h \in \cup_{u=1}^k P_u$. Let then $\mathcal{C}$ be the cone defined by

$$\mathcal{C} = \left\{ z \in [0, \infty]^h \mid \prod_{u=1}^{k} \left(\sum_{i\in P_u} z_i \right) > 0 \right\}. \tag{12}$$

In words, a vector $\mathbf{z} = (z_1, \dots, z_h)$ belongs to $\mathcal{C}$ if in each set P_u, $1 \leq u \leq k$, we can find at least one index $i \in P_u$ such that $z_i > 0$. This class of cones is of interest for several reasons. First, it will allow to deal with practical examples. From a theoretical point of view, it is noteworthy that this class is stable by intersection, and relative compactness in such a cone $\mathcal{C}$ is easily characterized: a subset A is relatively compact in $\mathcal{C}$ if and only if there exists $\eta > 0$ such that $\sum_{i\in P_u} z_i > \eta$ for all $u = 1, \dots, k$. Examples of such cones are $\mathcal{C}_1 = \{z_1 > 0, z_2 > 0, z_3 + z_4 > 0\}$ in $[0, \infty]^4$ and $\mathcal{C}_2 = \{z_1 + z_2 > 0, z_2 + z_3 > 0, z_3 + z_4 > 0, z_4 + z_5 > 0\}$ in $[0, \infty]^5$; for $a, b, c > 0$, $\{z_1 > a, z_2 > b, z_3 > c\}$ is relatively compact in $\mathcal{C}_1$ and $\{z_2 > a, z_4 > b\}$ is relatively compact in $\mathcal{C}_2$. More detailed examples will be given in Section 2.1.

Proposition 2 *Assume that there exists* $\epsilon > 0$ *such that*

$$\mathbb{E}[\sigma^{2h\alpha+\epsilon}(X_0)] < \infty. \tag{13}$$

Let $\mathcal{C}$ *be one of the cones defined in Eq. 12. Then there exists an integer* $\beta_{\mathcal{C}}$ *and a Radon measure* $\nu_{\mathcal{C}}$ *on* $\mathcal{C}$ *such that Eq. 11 holds and for all relatively compact sets*

$A \subset \mathcal{C}$ with $\nu_{\mathcal{C}}(\partial A) = 0$, for $m > h$ and $h' \geq 0$, and for any Borel measurable set $B \subset \mathbb{R}^{h'+1}$, we have

$$\lim_{t\to\infty} \frac{\mathbb{P}(\mathbf{Y}_{1,h} \in tA, \mathbf{Y}_{m,m+h'} \in B)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} = \mathbb{E}\left[\nu_{\mathcal{C}}(\boldsymbol{\sigma}(\mathbf{X}_{1,h})^{-1} \cdot A)\mathbb{P}(\mathbf{Y}_{m,m+h'} \in B \mid \mathcal{X})\right]. \tag{14}$$

Furthermore, for $r = 1, \dots, h$, there exist functions $\mathcal{L}_r$ such that for all $s, s' \geq 1$, $\mathbf{u}, \mathbf{v} \in (0, \infty)^h$,

$$\lim_{t\to\infty} \frac{\mathbb{P}(\mathbf{u} \cdot \mathbf{Z}_{1,h} \in tsA, \mathbf{v} \cdot \mathbf{Z}_{r,r+h-1} \in ts'A)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} = \mathcal{L}_r(A, \mathbf{u}, \mathbf{v}, s, s'). \tag{15}$$

Some comments are in order. Note first that we assume that $h < m$. Otherwise, if $m < h$, then vectors $\mathbf{Y}_{m,m+h'}$ and $\mathbf{Y}_{1,h}$ may be asymptotically dependent. For example, if $\{Z_j\}$ is i.i.d with the tail distribution as in Eq. 4, then $\mathbb{P}(Z_2 + Z_3 > t \mid Z_1 + Z_2 > t) \to 1/2$. We do not think that this is of particular interest, since one is primary interested in estimating distribution of *future* vector $\mathbf{Y}_{m,m+h'}$ based on the *past* observations $\mathbf{Y}_{1,h}$. Condition 13 is sufficient to deal with any of the cones $\mathcal{C}$. For a given cone $\mathcal{C}$, it might be relaxed. However, it holds if $\sigma(x) = \mathrm{e}^x$, which is a common choice in the econometric literature. It is easily seen that the coefficient $\beta_{\mathcal{C}}$ is the smallest integer ℓ for which there exists $i_1, \dots, i_\ell \in \{1, \dots, h\}$ such that $z_{i_1} > 0, \dots, z_{i_\ell} > 0$ implies $\prod_{u=1}^{k}(\sum_{i\in P_u} z_i) > 0$. The measure $\nu_{\mathcal{C}}$ is cumbersome to write precisely in general, but is easily obtained in each example. See Eq. 36 in the proof of Proposition 2. The condition 13 obviously holds if $\sigma(x) = \mathrm{e}^x$ or if σ is a polynomial. For $B = \mathbb{R}^{h'+1}$, Eq. 14 specializes to

$$\lim_{t\to\infty} \frac{\mathbb{P}(\boldsymbol{\sigma}(\mathbf{X}_{1,h}) \cdot \mathbf{Z}_{1,h} \in tA)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} = \mathbb{E}[\nu_{\mathcal{C}}(\boldsymbol{\sigma}(\mathbf{X}_{1,h})^{-1} \cdot A)].$$

If $\nu_{\mathcal{C}}(A) > 0$, then $\mathbb{E}[\nu_{\mathcal{C}}(\boldsymbol{\sigma}(\mathbf{X}_{1,h})^{-1} \cdot A)] > 0$ and Eq. 14 implies that the extremogram defined in Eq. 9 can be expressed as

$$\rho(A, B, m) = \frac{\mathbb{E}\left[\nu_{\mathcal{C}}(\boldsymbol{\sigma}(\mathbf{X}_{1,h})^{-1}A)\mathbb{P}(\mathbf{Y}_{m,m+h'} \in B \mid \mathcal{X})\right]}{\mathbb{E}[\nu_{\mathcal{C}}(\boldsymbol{\sigma}(\mathbf{X}_{1,h})^{-1} \cdot A)]}. \tag{16}$$

It may happen that $\mathcal{L}_r(A, \cdot) \equiv 0$ for $r = 2, \dots, h$. Intuitively, this happens if $\mathbf{u} \cdot \mathbf{Z}_{1,h}$ and $\mathbf{v} \cdot \mathbf{Z}_{r,r+h-1}$ belonging simultaneously to tA implies that at least $\beta_{\mathcal{C}} + 1$ coordinates of $\mathbf{Z}_{1,r+h-1}$ are large. This is the case for instance in Examples 1, 2 and 4. Otherwise, $\mathcal{L}_r$ may have quite a complicated form, as in Example 3.

Let us finally note an important fact. In practice, the conditioning set A is given, not the cone $\mathcal{C}$. So it is important to know if the choice of the cone has any effect on the quantities that will appear in the inference theory. The following lemma shows that fortunately this is not the case.

Lemma 3 *Let A be a subset of $[0,\infty]^h \setminus \{\mathbf{0}\}$. If there exists two cones $\mathcal{C}$ and $\mathcal{C}'$ such that Eq. 11 hold, A is relatively compact in both $\mathcal{C}$ and $\mathcal{C}'$, $\nu_{\mathcal{C}}(A) > 0$ and $\nu_{\mathcal{C}'}(A) > 0$, then $\beta_{\mathcal{C}} = \beta_{\mathcal{C}'}$ and for all $\mathbf{u} \in (0,\infty]^h$, $\nu_{\mathcal{C}}(\mathbf{u}\cdot A) = \nu_{\mathcal{C}'}(\mathbf{u}\cdot A)$.*

However, introducing the cone $\mathcal{C}$ is not superfluous, since all the constants in the limiting distributions will be expressed in terms of the measure $\nu_{\mathcal{C}}$ and cannot be expressed in terms of A only.

2.1 Examples

Example 1 Fix some positive integer h and consider the cone $\mathcal{C} = (0,\infty]^h$. Then $\beta_{\mathcal{C}} = h$ and the measure $\nu_{\mathcal{C}}$ is defined by

$$\nu_{\mathcal{C}}(\mathrm{d}z_1,\ldots,\mathrm{d}z_h) = \alpha^h \prod_{i=1}^{h} z_i^{-\alpha-1}\mathrm{d}z_i .$$

Consider the set A defined by $A = \{(z_1,\ldots,z_h) \in \mathbb{R}_+^h \mid z_1 > 1,\ldots,z_h > 1\}$. If Eq. 13 holds, then for $m > h$, and $B \in \mathbb{R}^{h'+1}$, Proposition 2 yields

$$\lim_{t\to\infty} \mathbb{P}(\mathbf{Y}_{m,m+h'} \in B \mid Y_1 > t,\ldots,Y_h > t) = \frac{\mathbb{E}\left[\prod_{i=1}^{h}\sigma^\alpha(X_i)\mathbb{P}(\mathbf{Y}_{m,m+h'} \in B \mid \mathcal{X})\right]}{\mathbb{E}\left[\prod_{i=1}^{h}\sigma^\alpha(X_i)\right]}.$$

In particular, setting $B = (-\infty, y]$ and $h' = 0$, the limiting conditional distribution of Y_m given that $Y_1,\ldots,Y_h$ are simultaneously large is given by

$$\Psi_h(y) = \lim_{t\to\infty} \mathbb{P}(Y_m \le y \mid Y_1 > t,\ldots,Y_h > t) = \frac{\mathbb{E}\left[\prod_{i=1}^{h}\sigma^\alpha(X_i)F_Z(y/\sigma(X_m))\right]}{\mathbb{E}\left[\prod_{i=1}^{h}\sigma^\alpha(X_i)\right]}. \tag{17}$$

Finally, note that the function $\mathcal{L}_r$ defined in Eq. 15 vanish for $r = 2,\ldots,h$.

Example 2 Consider now $\mathcal{C} = (0,\infty]$. Another quantity of interest is the limiting distribution of the sum of h' consecutive values, given that past values are extreme. To keep notation simple, consider $h' = 1$ and, for $m > 1$,

$$\Psi^*(y) = \lim_{t\to\infty} \mathbb{P}(Y_m + Y_{m+1} \le y \mid Y_1 > t) = \frac{\mathbb{E}[\sigma^\alpha(X_1)\mathbb{P}(Y_m + Y_{m+1} \le y \mid \mathcal{X})]}{\mathbb{E}[\sigma^\alpha(X_1)]}.$$

Estimating this distribution yields for instance empirical quantiles of the sum of future returns, given the present one is large.

Example 3 Consider $\mathcal{C} = [0, \infty] \times [0, \infty] \setminus \{\mathbf{0}\}$. Then $\beta_{\mathcal{C}} = 1$ and the measure $\nu_{\mathcal{C}}$ is defined by

$$\nu_{\mathcal{C}}(\mathrm{d}\mathbf{z}) = \alpha\{z_1^{-\alpha-1}\,\mathrm{d}_1\delta_0(\mathrm{d}z_2) + \delta_0(\mathrm{d}z_1)z_2^{-\alpha-1}\mathrm{d}z_2\},$$

where δ_0 is the Dirac point mass at 0. Consider the set A defined by $A = \{(z_1, z_2) \in \mathbb{R}_+^2 \mid z_1 + z_2 > 1\}$. If $\mathbb{E}[\sigma^{\alpha+\epsilon}(X_1)] < \infty$ for some $\epsilon > 0$, then Proposition 2 yields

$$\lim_{t\to\infty} \mathbb{P}(\mathbf{Y}_{m,m+h'} \in B \mid Y_1 + Y_2 > t)$$
$$= \frac{\mathbb{E}\left[\mathbb{P}(\mathbf{Y}_{m,m+h'} \in B \mid \mathcal{X})(\sigma^\alpha(X_1) + \sigma^\alpha(X_2))\right]}{\mathbb{E}[\sigma^\alpha(X_1)] + \mathbb{E}[\sigma^\alpha(X_2)]}.$$

In particular, take $B = (-\infty, y]$ and $h' = 0$. The limiting conditional distribution of Y_m given $Y_1 + Y_2$ is large is defined by

$$\Lambda(y) = \lim_{t\to\infty} \mathbb{P}(Y_m \le y \mid Y_1 + Y_2 > t) = \frac{\mathbb{E}[\{\sigma^\alpha(X_1) + \sigma^\alpha(X_2)\}F_Z(y/\sigma(X_m)]}{\mathbb{E}[\sigma^\alpha(X_1) + \sigma^\alpha(X_2)]}.$$

Finally, the function $\mathcal{L}_2$ equals

$$\mathcal{L}_2(A, u_1, u_2, v_1, v_2, s, s') = \left(\frac{1+s}{u_2} \vee \frac{1+s'}{v_1}\right)^{-\alpha}.$$

Example 4 Consider the cone $\mathcal{C} = ([0, \infty]^2 \setminus \{\mathbf{0}\}) \times ([0, \infty]^2 \setminus \{\mathbf{0}\}) \times (0, \infty]$. Then $\beta_{\mathcal{C}} = 2$ and

$$\nu_{\mathcal{C}}(\mathrm{d}\mathbf{z}) = \alpha^3\{z_1^{-\alpha-1}\mathrm{d}z_1\delta_0(\mathrm{d}z_2) + \delta_0(\mathrm{d}z_1)z_2^{-\alpha-1}\,\mathrm{d}z_2)\}$$
$$\times \{z_3^{-\alpha-1}\mathrm{d}z_3\delta_0(\mathrm{d}z_4) + \delta_0(\mathrm{d}z_3)z_4^{-\alpha-1}\mathrm{d}z_4\}z_5^{-\alpha-1}\mathrm{d}z_5.$$

Consider $A = \{(z_1, z_2, z_3, z_4, z_5) \in \mathbb{R}_+^5 | z_1 + z_2 > 1, z_3 + z_4 > 1, z_5 > 0\}$. If $\mathbb{E}[\sigma^{3\alpha+\epsilon}(X_0)] < \infty$ for some $\epsilon > 0$, then we obtain, for $m > 3$,

$$\lim_{t\to\infty} \mathbb{P}(\mathbf{Y}_{m,m+h'} \in B \mid Y_1 + Y_2 > t, Y_3 + Y_4 > t, Y_5 > t)$$
$$= \frac{\mathbb{E}\left[\mathbb{P}(\mathbf{Y}_{m,m+h'} \in B \mid \mathcal{X})\{\sigma^\alpha(X_1) + \sigma^\alpha(X_2)\}\{\sigma^\alpha(X_3) + \sigma^\alpha(X_4)\}\sigma^\alpha(X_5)\right]}{\mathbb{E}[\{\sigma^\alpha(x_1) + \sigma^\alpha(X_2)\}\{\sigma^\alpha(X_3) + \sigma^\alpha(X_4)\}\sigma^\alpha(X_5)]}.$$

Here the functions $\mathcal{L}_r$ vanish for $r \ge 2$.

Example 5 In this example, we illustrate Lemma 3. Let $h = 4$ and $A = \{z_1 > a, z_3 + z_4 > b\}$. Then A is relatively compact in $\mathcal{C}_1 = ([0, \infty]^2 \setminus \{0\}) \times ([0, \infty]^2 \setminus \{0\})$,

$\mathcal{C}_2 = (0,\infty] \times ([0,\infty]^3 \setminus \{0\})$ and $\mathcal{C}_3 = [0,\infty]^4 \setminus \{0\}$. Then it is easily seen that $\beta_{\mathcal{C}_1} = \beta_{\mathcal{C}_2} = 2$ and $\beta_{\mathcal{C}_2} = 1$, and for all $\mathbf{u} \in [(0,\infty]^4$, $\nu_{\mathcal{C}_1}(\mathbf{u}^{-1} \cdot A) = \nu_{\mathcal{C}_2}(\mathbf{u}^{-1} \cdot A) = u_1^\alpha(u_3^\alpha + u_4^\alpha)a^{-\alpha}b^{-\alpha}$ and $\nu_{\mathcal{C}_3}(A) = 0$.

3 Estimation

Let $\mathcal{C}$ be a cone defined in Eq. 12 and let A be a relatively compact subset of $\mathcal{C}$ such that $\nu_{\mathcal{C}}(A) > 0$. To simplify the notation, assume that we observe $Y_1,\dots,Y_{n+m+h'}$. An estimator $\hat\rho_n(A,B,m)$ is naturally defined by

$$\hat\rho_n(A,B,m) = \frac{\sum_{r=1}^n \mathbf{1}_{\{\mathbf{Y}_{r,r+h-1} \in Y_{(n:n-k)}A\}} \mathbf{1}_{\{\mathbf{Y}_{r+m,r+m+h'} \in B\}}}{\sum_{r=1}^n \mathbf{1}_{\{\mathbf{Y}_{r,r+h-1} \in Y_{(n:n-k)}A\}}},$$

where k is a user chosen threshold and $Y_{(n:1)} \le \cdots \le Y_{(n:n)}$ are the increasing order statistics of the observations $Y_1,\dots,Y_n$. We will also consider the case $B = (-\infty, \mathbf{y}]$, i.e. the case of the limiting conditional distribution of $\mathbf{Y}_{m,m+h'}$ given $\mathbf{Y}_{1,h} \in tA$, that means

$$\begin{aligned}\Psi_{A,m,h'}(\mathbf{y}) &= \lim_{t\to\infty} \mathbb{P}(\mathbf{Y}_{m,m+h'} \le \mathbf{y} \mid \mathbf{Y}_{1,h} \in tA) \\ &= \rho(A,(\infty,\mathbf{y}],m) \\ &= \frac{\mathbb{E}[\nu_{\mathcal{C}}(\boldsymbol{\sigma}(\mathbf{X}_{1,h})^{-1} \cdot A) \prod_{i=1}^{h'} F(y_i/\sigma(X_{m+i}))]}{\mathbb{E}[\nu_{\mathcal{C}}(\boldsymbol{\sigma}(\mathbf{X}_{1,h})^{-1} \cdot A)]}. \qquad (18)\end{aligned}$$

An estimator $\hat\Psi_{n,A,m,h'}$ of $\Psi_{A,m,h'}$ is defined on $\mathbb{R}^{h'+1}$ by

$$\hat\Psi_{n,A,m,h'}(\mathbf{y}) = \frac{\sum_{r=1}^n \mathbf{1}_{\{\mathbf{Y}_{r,r+h-1} \in Y_{(n:n-k)}A\}} \mathbf{1}_{\{\mathbf{Y}_{r+m,r+m+h'} \le \mathbf{y}\}}}{\sum_{r=1}^n \mathbf{1}_{\{\mathbf{Y}_{r,r+h-1} \in Y_{(n:n-k)}A\}}}. \qquad (19)$$

As usual, the bias of the estimators will be bounded by a second order type condition. Let k be a non decreasing sequence of integers, let F_Y denote the distribution of Y and let $u_n = (1/\bar F_Y)^{\leftarrow}(n/k)$. Consider the measure defined on $\mathcal{C}$ by

$$\mu_{\mathcal{C}}(A) = \frac{\mathbb{E}[\nu_{\mathcal{C}}(\boldsymbol{\sigma}(\mathbf{X}_{1,h})^{-1} \cdot A)]}{(\mathbb{E}[\sigma^\alpha(X)])^{\beta_{\mathcal{C}}}}. \qquad (20)$$

We introduce the following quantity which will be used as a bound for the bias of the estimators and thus will also determine their rate of convergence.

$$v_n(A) = \mathbb{E}\left[\sup_{s\ge1}\left|\frac{\mathbb{P}(\mathbf{Y}_{1,h} \in u_n sA \mid \mathcal{X})}{(k/n)^{\beta_{\mathcal{C}}}} - s^{-\alpha\beta_{\mathcal{C}}}\mu_{\mathcal{C}}(A)\right|\right]. \qquad (21)$$

Lemma 4 *If Eqs. 4 and 13 hold, then* $\lim_{n\to\infty} v_n(A) = 0$.

We need also the following quantities, which are well defined when Eq. 13 holds. For $r = 2, \ldots, h$ and measurable subsets B, B' of $\mathbb{R}^{h'+1}$, define

$$\mathcal{R}_r(A, B, B') = \frac{\mathbb{E}\big[\mathcal{L}(A, \boldsymbol{\sigma}(\mathbf{X}_{1,h}), \boldsymbol{\sigma}(\mathbf{X}_{k,k+h-1}), 1, 1) \times \mathbb{P}(\mathbf{Y}_{m,m+h'} \in B, \mathbf{Y}_{m+k-1,m+h'+k-1} \in B' \mid \mathcal{X})\big]}{\mu_{\mathcal{C}}(A)} + \frac{\mathbb{E}\big[\mathcal{L}(A, \boldsymbol{\sigma}(\mathbf{X}_{1,h}), \boldsymbol{\sigma}(\mathbf{X}_{k,k+h-1}), 1, 1) \times \mathbb{P}(\mathbf{Y}_{m,m+h'} \in B', \mathbf{Y}_{m+k-1,m+h'+k-1} \in B \mid \mathcal{X})\big]}{\mu_{\mathcal{C}}(A)}. \tag{22}$$

For brevity, denote $\mathcal{R}_r(A, B) = \mathcal{R}_r(A, B, B)$.

3.1 General result: weak dependence

We can now state our main result in the weak dependence setting, i.e. when absolute summability (Eq. 2) of the autocovariance function of the process $\{X_j\}$ holds. In order to simplify the proof, and without loss of meaningful generality we will hereafter assume that the set A is itself a cone. This assumption is satisfied by all reasonable examples.

Theorem 5 *Let Eqs. 2, 4 and 13 hold. Assume moreover that A is a relatively compact subcone of $\mathcal{C}$ such that $\mu_{\mathcal{C}}(A) > 0$, that $k/n \to 0$, $n(k/n)^{\beta_{\mathcal{C}}} \to \infty$ and*

$$\lim_{n\to\infty} \sqrt{n(k/n)^{\beta_{\mathcal{C}}}}\ v_n(A) = 0. \tag{23}$$

Then

$$\sqrt{n(k/n)^{\beta_{\mathcal{C}}}\mu_{\mathcal{C}}(A)}\{\hat{\rho}_n(A, B, m) - \rho(A, B, m)\}$$

converges weakly to a centered Gaussian distribution with variance

$$\rho(A, B, m)\{1 - \rho(A, B, m)\} + \sum_{r=2}^{h\wedge(m-h)} \big\{\mathcal{R}_r(A, B) - 2\rho(A, B, m)\mathcal{R}_r(A, B, \mathbb{R}^{h'+1}) + \rho^2(A, B, m)\mathcal{R}_r(A, \mathbb{R}^{h'+1})\big\}. \tag{24}$$

If $h = 1$ or if the functions $\mathcal{L}_r$ defined in Eq. 15 are identically zero for $r \geq 2$, then the limiting covariance in Eq. 24 is simply $\rho(A, B, m)\{1 - \rho(A, B, m)\}$. Otherwise, the additional terms can be canceled by modifying the estimator of $\hat{\rho}_n(A, B, m)$. Assuming we have $nh + m + h' + 1$ observations, we can define

$$\tilde{\rho}_n(A, B, m) = \frac{\sum_{r=1}^{n} \mathbf{1}_{\{\mathbf{Y}_{(r-1)h+1,rh} \in Y_{(n:n-k)}A\}} \mathbf{1}_{\{\mathbf{Y}_{(r-1)h+m,(r-1)h+m+h'} \in B\}}}{\sum_{r=1}^{n} \mathbf{1}_{\{\mathbf{Y}_{(r-1)h+1,rh} \in Y_{(n:n-k)}A\}}}$$

Noting that the events $\{\mathbf{Y}_{r,r+h-1} \in A\}$ are h-dependent conditionally on $\mathcal{X}$, the proof of Theorem 5 can be easily adapted to show that the limiting variance of $\sqrt{n(k/n)^{\beta_\mathcal{C}}}\{\tilde{\rho}_n(A, B, m) - \rho(A, B, m)\}$ is the same as in the case where $\mathcal{L}_r \equiv 0$ for $r = 2, \dots, h$. But this is of course at the cost of an increase of the asymptotic variance, due to a different sample size.

We can also obtain the functional convergence of the estimator $\hat{\Psi}_{n,A,m,h'}$ of the limiting conditional distribution function $\Psi_{A,m,h'}$, defined respectively in Eqs. 19 and 18.

Corollary 6 *Under the Assumptions of Theorem* 5*, and if moreover the distribution* $\Psi_{A,m,h'}$ *is continuous, then*

$$\sqrt{n(k/n)^{\beta_\mathcal{C}}\mu_\mathcal{C}(A)}\{\hat{\Psi}_{n,A,m,h'} - \Psi_{A,m,h'}\}$$

converges in $\mathcal{D}(\mathbb{R}^{h'+1})$ *to a Gaussian process. If* $h = 1$ *or if the functions* $\mathcal{L}_r$ *are identically zero for* $r = 2, \dots, h$*, then the limiting process can be expressed as* $\mathbb{B} \circ \Psi_{A,m,h'}$*, where* $\mathbb{B}$ *is the standard Brownian bridge.*

Note that a sufficient condition for $\Psi_{A,m,h'}$ to be continuous is that F_Z is continuous.

3.2 General result: long memory

We now state our results in the framework of long memory. This requires several additional notions, such as multivariate Hermite expansion and Hermite ranks which are recalled in Appendix B.

Define the functions G_n and G for $(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^h \times \mathbb{R}^{h'+1}$ and $s \geq 1$ by

$$G_n(A, B, s, \mathbf{x}, \mathbf{x}') = \frac{\mathbb{P}(\boldsymbol{\sigma}(\mathbf{x}) \cdot \mathbf{Z}_{1,h} \in u_n s A)}{(k/n)^{\beta_\mathcal{C}}} \mathbb{P}(\boldsymbol{\sigma}(\mathbf{x}') \cdot \mathbf{Z}_{m,m+h'} \in B) \tag{25}$$

$$\begin{aligned} G(A, B, \mathbf{x}, \mathbf{x}') &= \lim_{n\to\infty} G_n(A, B, 1, \mathbf{x}, \mathbf{x}') \\ &= \frac{\nu_\mathcal{C}(\boldsymbol{\sigma}(\mathbf{x})^{-1} \cdot A)}{\mathbb{E}[\sigma^\alpha(X_1)])^{\beta_\mathcal{C}}} \mathbb{P}(\boldsymbol{\sigma}(\mathbf{x}') \cdot \mathbf{Z}_{m,m+h'} \in B). \end{aligned} \tag{26}$$

Let $\tau_n(A, B, s)$ and $\tau(A, B)$ be the Hermite ranks with respect to $(\mathbf{X}_{1,h}, \mathbf{X}_{m,m+h'})$ of the functions $G_n(A, B, s, \cdot, \cdot)$ and $G(A, B, \cdot, \cdot)$, respectively. Define $\tau(A) = \tau(A, \mathbb{R}^d)$.

Assumption 1 *For large* n*,* $\inf_s \tau_n(A, B, s) = \tau(A, B)$ *and* $\tau(A, B) \leq \tau(A)$.

This assumption is fulfilled for example when $\sigma(x) = \exp(x)$, in which case all the considered Hermite ranks are equal to one, or if σ is an even function with Hermite rank 2 (such as $\sigma(x) = x^2$), in which case they are equal to two. The modification of Theorem 5 reads as follows.

Theorem 7 *Assume that $\{X_j\}$ is the long memory Gaussian sequence with covariance given by Eq. 3. Assume that A is a relatively compact subcone of $\mathcal{C}$ such that $\nu_{\mathcal{C}}(A) > 0$. Let Assumption 1 and Eq. 13 hold, and $k/n \to 0$, $n(k/n)^{\beta_{\mathcal{C}}} \to \infty$ and*

$$\lim_{n\to\infty} \left\{ \sqrt{n(k/n)^{\beta_{\mathcal{C}}}} \wedge \gamma_n^{-\tau(A,B)/2} \right\} v_n(A) = 0. \tag{27}$$

(i) *If $n(k/n)^{\beta_{\mathcal{C}}} \gamma_n^{\tau(A,B)} \to 0$, then*

$$\sqrt{n(k/n)^{\beta_{\mathcal{C}}} \mu_{\mathcal{C}}(A)} \{\hat{\rho}_n(A, B, m) - \rho(A, B, m)\}$$

converges to a centered Gaussian distribution with variance given in Eq. 24.

(ii) *If $n(k/n)^{\beta_{\mathcal{C}}} \gamma_n^{\tau(A,B)} \to \infty$, then $\gamma_n^{-\tau(A,B)/2} \{\hat{\rho}_n(A, B, m) - \rho(A, B, m)\}$ converges weakly to a distribution which is non-Gaussian except if $\tau(A, B) = 1$.*

The exact definition of the limiting distribution will be given in Section 4. It suffices to mention here that this distribution depends on H and $\tau(A, B)$. The meaning of the above result is the following. In the long memory setting, it is still possible to obtain the same limit as in the weakly dependent case, if k (i.e., the number of high order statistics used in the definition of the estimators) is not too large, so that both the bias and the long memory effect are canceled.

Define a new Hermite rank $\tau^*(A) = \inf_{y\in\mathbb{R}^{h'+1}} \tau(A, (\infty, \mathbf{y}])$.

Corollary 8 *Under the assumptions of Theorem 7, if the distribution function $\Psi_{A,m,h'}$ is continuous and if $\tau^*(A) \le \tau(A)$, then*

- *If $n(k/n)^{\beta_{\mathcal{C}}} \gamma_n^{\tau^*(A)} \to 0$, then*

$$\sqrt{n(k/n)^{\beta_{\mathcal{C}}} \mu_{\mathcal{C}}(A)} \{\hat{\Psi}_{n,A,m,h'} - \Psi_{A,m,h'}\}$$

  converges in $\mathcal{D}((-\infty, +\infty)^{h'+1}$ to a Gaussian process. If $h = 1$ or if the functions $\mathcal{L}_r$ are identically zero for $r = 2, \ldots, h$, then the limiting process can be expressed as $\mathbb{B} \circ \Psi_{A,m,h'}$, where $\mathbb{B}$ is the standard Brownian bridge.
- *If $n(k/n)^{\beta_{\mathcal{C}}} \gamma_n^{\tau^*(A)} \to \infty$, then $\gamma_n^{-\tau^*(A)/2} \{\hat{\Psi}_{n,A,m,h'} - \Psi_{A,m,h'}\}$ converges in $\mathcal{D}((-\infty, +\infty)^{h'+1}$ to a process which can be expressed as $J_{A,m,h'} \cdot \aleph$ where $J_{A,m,h'}$ is a deterministic function and $\aleph$ is a random variable, which is non Gaussian except if $\tau^*(A) = 1$.*

The exact definition of the function $J_{A,m,h'}$ and of the random variable $\aleph$ will be given in Section 4. Anyhow, they are not of much practical interest. In practice, the main goal will be to choose the number k of order statistics used in the estimation procedure so that both the bias and the long memory effect are canceled, and the limiting distribution of the weakly dependent case can be used in the inference.

3.3 Examples

We now discuss the Examples introduced in Section 2.1. In order to evaluate the bias term (Eq. 21), it is necessary to introduce a second order regular variation condition. We follow here Drees (1998). This assumption is referred to as second order regular variation (SO).

Assumption 2 (SO) *There exists a bounded non increasing function* η^* *on* $[0, \infty)$*, regularly varying at infinity with index* $-\alpha\zeta$ *for some* $\zeta \geq 0$*, and such that* $\lim_{t\to\infty} \eta^*(t) = 0$ *and there exists a measurable function* η *such that for* $z > 0$*,*

$$\mathbb{P}(Z > z) = cz^{-\alpha} \exp\left(\int_1^z \frac{\eta(s)}{s}\,\mathrm{d}s\right),$$
$$\exists C > 0\,, \quad \forall s \geq 0, \quad |\eta(s)| \leq C\eta^*(s).$$

The most commonly used second order assumption is that $\eta^*(s) = O(s^{-\alpha\zeta})$. Then

$$\mathbb{P}(Z > z) = cz^{-\alpha}(1 + O(z^{-\alpha\zeta})) \quad \text{as } z \to \infty, \tag{28}$$

for some constant $c > 0$.

On account of Breiman's lemma, if the tail of Z is regularly varying with index $-\alpha$, then the same holds for $Y = \sigma(X)Z$, as long as X and Z are independent, and $\mathbb{E}[\sigma^\alpha(X)] < \infty$. Also, (SO) property is transferred from the tail of Z to Y; see Kulik and Soulier (2011, Proposition 2.1).

For the sake of simplicity and clarity of exposition, we will make in this section the usual assumption that $\sigma(x) = \exp(x)$, so that the Hermite rank of σ is 1 and Assumption 1 is fulfilled with the Hermite rank equal to one. This will avoid to define many auxiliary functions and Hermite ranks. But the examples can of course be treated in a more general framework. For the exponential function, Eq. 13 obviously holds for any h. Also, we will only state the convergence results under the conditions which imply that the limiting distribution is the same as in the weak dependence case, since this is the case of practical interest. We only treat Examples 1 and 3 since they exhibit two different limiting distributions. The computations for the other examples are straightforward.

3.3.1 Example 1 *continued*

Fix integers $h \geq 1$ and $m > h$. Recall the formula 17 for the conditional distribution of Y_m given that $Y_1, \ldots, Y_h$ are simultaneously large. Its estimator $\hat{\Psi}_{n,h}$ is defined by

$$\hat{\Psi}_{n,h}(y) = \frac{\sum_{r=1}^n \mathbf{1}_{\{Y_r > Y_{(n:n-k)}, \ldots, Y_{r+h-1} > Y_{(n:n-k)}, Y_{r+m} \leq y\}}}{\sum_{r=1}^n \mathbf{1}_{\{Y_r > Y_{(n:n-k)}, \ldots, Y_{r+h-1} > Y_{(n:n-k)}\}}}$$

with a user chosen k.

In this case, if Eq. 13 holds, then the functions $\mathcal{L}_r(A,\cdot)$ vanish for $r = 2,\dots,h$. Assumption 2 and Kulik and Soulier (2011, Proposition 2.8) imply that a bound for $v_n(A)$ is then given by

$$v_n(A) = O(\eta^*(u_n)). \tag{29}$$

Corollary 9 *Assume that* $\sigma(x) = \exp(x)$. *Let Assumption* 2 *hold. Let* k *be such that* $k/n \to 0$, $n(k/n)^h \to \infty$, *and*

$$\lim_{n\to\infty} (n(k/n)^h)^{1/2}\eta^*(u_n) = 0. \tag{30}$$

In the weakly dependent case (Eq. 2*) or in the long memory case (Eq.* 3*) if moreover* $n(k/n)^h\gamma_n \to 0$, *then*

$$\sqrt{n(k/n)^h}(\hat{\Psi}_{n,h} - \Psi_h) \Rightarrow \left(\frac{\mathbb{E}[\sigma^\alpha(X_1)\cdots\sigma^\alpha(X_h)]}{\mathbb{E}^h[\sigma^\alpha(X_1)]}\right)^{-1/2} \mathbb{B}\circ\Psi_h \tag{31}$$

weakly in $\mathcal{D}((-\infty,\infty))$, *where* $\mathbb{B}$ *is the standard Brownian bridge.*

We note that the rate of convergence $\sqrt{n(k/n)^h}$ depends explicitly on the exponent $\beta_{\mathcal{C}} = h$. Let us have a closer look at the bias condition 30. Assume that in Assumption 2 we have $\eta^*(s) = O(s^{-\alpha\zeta})$ so that $\bar{F}_Z(z) = cz^{-\alpha}(1 + O(z^{-\alpha\zeta}))$ and $\eta^*(u_n) = O((k/n)^\zeta)$. Hence, Eq. 30 becomes

$$\lim_{n\to\infty} n(k/n)^{h+2\zeta} = 0.$$

Consequently, the largest allowed number of extremes k used in the construction of the estimator $\hat{\Psi}_{n,h}$ is proportional to $n^{(h-1+2\zeta)/(h+2\zeta)-\epsilon}$ with arbitrary small $\epsilon > 0$ and the best possible rate of convergence in Eq. 31 is $n^{\zeta/(2\zeta+h)-\epsilon}$.

3.3.2 Example 3 *continued*

Consider the estimation of

$$\Lambda(y) = \lim_{t\to\infty}\mathbb{P}(Y_m \le y \mid Y_1 + Y_2 > t) = \frac{\mathbb{E}[\{\sigma^\alpha(X_1) + \sigma^\alpha(X_2)\}F_Z(y/\sigma(X_m))]}{\mathbb{E}[\sigma^\alpha(X_1) + \sigma^\alpha(X_2)]}.$$

An estimator if defined by

$$\hat{\Lambda}_n(y) = \frac{\sum_{r=1}^n \mathbf{1}_{\{Y_r+Y_{r+1}>Y_{(n:n-k)}\}}\mathbf{1}_{\{Y_{r+m}\le y\}}}{\sum_{r=1}^n \mathbf{1}_{\{Y_r+Y_{r+1}>Y_{(n:n-k)}\}}}.$$

As argued before, if Eq. 13 holds, then the function $\mathcal{L}_2$ is equal to

$$\mathcal{L}_2(A, u_1, u_2, v_1, v_2, s, s') = \left(\frac{1+s}{u_2}\vee\frac{1+s'}{v_1}\right)^{-\alpha}.$$

Applying Lemma A.1, we obtain a bound for $v_n(A)$:

$$v_n(A) = O\left(\eta^*(u_n) + u_n^{-1}\int_0^{u_n} \bar{F}_Z(s)\,\mathrm{d}s\right). \tag{32}$$

Corollary 10 *Assume that* $\sigma(x) = \exp(x)$. *Let Assumption* 2 *and holds. Let* k *be such that* $k \to \infty$, $k/n \to 0$ *and*

$$\lim_{n\to\infty} k^{1/2}\left(\eta^*(u_n) + u_n^{-1}\int_0^{u_n} \bar{F}_Z(s)\,\mathrm{d}s\right) = 0. \tag{33}$$

In the weakly dependent case (Eq. 2) or in the long memory case (Eq. 3) if moreover $k\gamma_n \to 0$, *then*

$$k^{1/2}(\hat{\Lambda}_n - \Lambda) \Rightarrow \left(\frac{\mathbb{E}[\sigma^\alpha(X_1) + \sigma^\alpha(X_2)]}{\mathbb{E}[\sigma^\alpha(X_1)]}\right)^{-1/2} \mathbb{W}$$

weakly in $\mathcal{D}((-\infty,\infty))$, *where* $\mathbb{W}$ *is a Gaussian process with covariance*

$$\begin{aligned}
&\mathrm{cov}(\mathbb{W}(y), \mathbb{W}(y')) \\
&\quad = \Lambda(y \wedge y') - 2\Lambda(y)\Lambda(y') \\
&\qquad + \frac{\mathbb{E}[\sigma^\alpha(X_2)\{F_Z(y/\sigma(X_m))F_Z(y'/\sigma(X_{m+1})) + F_Z(y/\sigma(X_m))F_Z(y'/\sigma(X_{m+1}))\}]}{\mathbb{E}[\sigma^\alpha(X_1) + \sigma^\alpha(X_2)]}.
\end{aligned}$$

Consider as in the previous example that $\eta^*(s) = O(s^{-\alpha\zeta})$ and $\alpha > 1$ so that Eq. 32 becomes

$$v_n(A) = O\left((k/n)^\zeta + k/n\right).$$

Thus, the vanishing bias condition 33 implies that the maximum allowed number of extremes is $n^{2\zeta/(2\zeta+1)-\epsilon}$ if $\zeta < 1$ and $n^{2/3-\epsilon}$ otherwise, for some arbitrarily small $\epsilon > 0$. If $\alpha < 1$, then the rate of convergence depends on α, since Eq. 32 becomes

$$v_n(A) = O\left((k/n)^\zeta + (k/n)^\alpha\right).$$

Thus the maximum allowed number of extremes is $n^{2\zeta/(2\zeta+1)-\epsilon}$ if $\zeta < \alpha$ and $n^{2\alpha/2\alpha+1-\epsilon}$ otherwise, for some arbitrarily small $\epsilon > 0$. Summarizing, the best possible rate of convergence is $n^{\rho/(2\rho+1)}$ with $\rho = \alpha \wedge \zeta \wedge 1$.

Remark 1 If the estimator is modified by taking only every other observation,

$$\tilde{\Lambda}_n(y) = \frac{\sum_{r=1}^n \mathbf{1}_{\{Y_{2(r-1)+1}+Y_{2r} > Y_{(n:n-k)}\}}\mathbf{1}_{\{Y_{r+m}\le y\}}}{\sum_{r=1}^n \mathbf{1}_{\{Y_{2(r-1)+}+Y_{2r} > Y_{(n:n-k)}\}}}.$$

then $\sqrt{k}(\tilde{\Lambda}_n - \Lambda)$ converges weakly to $2\mathbb{B} \circ \Lambda$ where $\mathbb{B}$ is the standard Brownian bridge. Indeed, the random vectors $\{Y_1, Y_2\}$, $\{Y_3, Y_4\}, \ldots$ are conditionally independent given $\mathcal{X}$ but the price to be paid is the bigger variance.

4 Proofs

We start by proving the limiting behaviour of the conditional distribution (Eq. 8).

Lemma 11 *Suppose that assumptions of Proposition* 1 *are fulfilled. Then*

$$\lim_{t\to\infty} \mathbb{P}(Y_m \le y \mid Y_0 > t) = \frac{\mathbb{E}[\sigma^\alpha(X_0)F_Z(y/\sigma(X_m))]}{\mathbb{E}[\sigma^\alpha(X_0)]}. \tag{34}$$

Proof Conditioning on the sigma-field $\mathcal{X}$ yields

$$\begin{aligned} \mathbb{P}(Y_m \le y, Y_0 > t) &= \mathbb{E}[\mathbb{P}(\sigma_0 Z_0 \le y, \sigma_m Z_m > t)|\mathcal{X}] \\ &= \mathbb{E}[\mathbb{P}(\sigma_0 Z_0 \le y|\mathcal{X})\mathbb{P}(\sigma_m Z_m > t|\mathcal{X})] \\ &= \mathbb{E}[F_Z(y/\sigma_0)\bar{F}(t/\sigma_0)] \end{aligned}$$

Applying Potter's bound (see Bingham et al. 1989, Theorem 1.5.6), yields, for some constant C and $\epsilon > 0$

$$\frac{F_Z(y/\sigma_0)\bar{F}(t/\sigma_0)}{\bar{F}(t)} \le C(\sigma_0 \vee 1)^{\alpha+\epsilon}.$$

Thus, the assumption $\mathbb{E}[\sigma^{\alpha+\epsilon}(X_m)] < \infty$ and the bounded convergence theorem imply that

$$\lim_{t\to\infty} \frac{\mathbb{P}(Y_m \le y, Y_0 > t)}{\bar{F}(t)} = \mathbb{E}\left[\lim_{t\to\infty} \frac{F(y/\sigma_m)\bar{F}(t/\sigma_0)}{\bar{F}(t)}\right] = \mathbb{E}[\sigma^\alpha(X_0)F_Z(y/\sigma(X_m))].$$

Finally, noting that by Eq. 6 we have $\mathbb{P}(Y_0 > t) \sim \mathbb{E}[\sigma^\alpha(X_0)]\bar{F}(t)$ as $t \to \infty$ yields Eq. 8. □

Proof of Proposition 1 Since the random variables $Z_1, \ldots, Z_h$ are i.i.d., for each $(u_1, \ldots, u_h) \in [0, \infty]^h$, $x \in [-\infty, 0)^h$ and $y \in (0, \infty]^h$, it holds that

$$\begin{aligned} &\lim_{t\to\infty} t\mathbb{P}(a^{-1}(t)(u_1 Z_1, \ldots, u_h Z_h) \in [\mathbf{x}, \mathbf{y}]^c) \\ &= (1-\beta)\sum_{i=1}^h u_i^\alpha |x_i|^{-\alpha} + \beta \sum_{i=1}^h u_i^\alpha y_i^{-\alpha} \\ &= \sum_{i=1}^h u_i^\alpha \nu_{\alpha,\beta}([x_i, y_i]^c), \end{aligned}$$

where $\nu_{\alpha,\beta}$ is the Radon measure on $[-\infty, \infty] \setminus \{0\}$ defined by

$$\nu_{\alpha,\beta}(\mathrm{d}x) = \alpha\{(1-\beta)(-x)^{-\alpha-1}\mathbf{1}_{\{x<0\}} + \beta x^{-\alpha-1}\mathbf{1}_{\{x>0\}}\}\,\mathrm{d}x.$$

Moreover, by Potter's bound, for any $\epsilon > 0$, there exists a constant C (which also depends on $\mathbf{x}$ and $\mathbf{y}$) such that

$$t\mathbb{P}(a^{-1}(t)(u_1 Z_1, \ldots, u_k Z_k) \in [\mathbf{x}, \mathbf{y}]^c) \le C \sum_{i=1}^{k} (u_i \vee 1)^{\alpha+\epsilon}. \tag{35}$$

Thus, we can apply the bounded convergence theorem and obtain

$$\begin{aligned}
&\lim_{t\to\infty} t\mathbb{P}(a^{-1}(t)(\sigma(X_1)Z_1, \ldots, \sigma(X_k)Z_k) \in [\mathbf{x}, \mathbf{y}]^c) \\
&\quad = \mathbb{E}[\lim_{t\to\infty} t\mathbb{P}(a^{-1}(t)(\sigma(x_1)Z_1, \ldots, \sigma(X_k)Z_k) \in [\mathbf{x}, \mathbf{y}]^c \mid \mathcal{X})] \\
&\quad = \mathbb{E}[\sigma^\alpha(X_0)] \sum_{i=1}^{k} \nu_{\alpha,\beta}([x_i, y_i]^c).
\end{aligned}$$

□

Proof of Proposition 2 Let $\mathcal{C}$ be a cone of type 12 and let $\beta_{\mathcal{C}}$ be the smallest integer ℓ for which there exists $i_1 < \cdots < i_\ell \in \{1, \ldots, h\}$ such that $z_{i_1} > 0, \ldots, z_{i_\ell} > 0$ implies $\prod_{i=1}^{u}(\sum_{i\in P_u} z_i) > 0$. Such an integer exists since obviously $z_i > 0$ for all $i \in \{1, \ldots, h\}$ implies that $\prod_{i=1}^{u}(\sum_{i\in P_u} z_i) > 0$. Moreover, since $\beta_{\mathcal{C}}$ is the smallest such integer, then it clearly holds conversely that $\prod_{i=1}^{u}(\sum_{i\in P_u} z_i) > 0$ implies that at least $\beta_{\mathcal{C}}$ among z_i, $i = 1, \ldots, h$, are positive. Let P^* be the sets of $\beta_{\mathcal{C}}$-tuples $\mathbf{i} = (i_1, \ldots, i_{\beta_{\mathcal{C}}}) \in \{1, \ldots, h\}^{\beta_{\mathcal{C}}}$ such that $z_{i_q} > 0$ for $q = 1, \ldots, \beta_{\mathcal{C}}$ implies that $\mathbf{z} \in \mathcal{C}$. We now prove Eq. 11. It suffices to prove it for sets A of the form $A = \{\mathbf{z} \in [0, \infty]^h \mid \sum_{i\in P_u} z_i \ge a_u,\ u = 1, \ldots, k\}$, where $a_u > 0$, $u = 1, \ldots, k$. By relative compactness, there exist $\eta > 0$ and $\mathbf{i} = (i_1, \ldots, i_{\beta_{\mathcal{C}}}) \in P^*$ such that $z_{i_j} > \eta$, $1 \le j \le \beta_{\mathcal{C}}$. Moreover, by independence, asymptotically there is only one such $\mathbf{i} \in P^*$, i.e.

$$\frac{\mathbb{P}(\mathbf{Z}_{1,h} \in tA)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} \sim \sum_{\mathbf{i}\in P^*} \frac{\mathbb{P}(\mathbf{Z}_{1,h} \in tA, Z_{i_1} > t\eta, \ldots, Z_{i_{\beta_{\mathcal{C}}}} > t\eta)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}},$$

since by independence it holds that for $\mathbf{i} \neq \mathbf{i}' \in P^*$,

$$\lim_{t\to\infty} \frac{\mathbb{P}(Z_{i_1} > t\eta, \ldots, Z_{i_{\beta_{\mathcal{C}}}} > t\eta,\ Z_{i'_1} > t\eta, \ldots, Z_{i'_{\beta_{\mathcal{C}}}} > t\eta)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} = 0.$$

Let us now consider one arbitrary $\mathbf{i} \in P^*$, and for clarity assume that $\mathbf{i} = (1, \ldots, \beta_{\mathcal{C}})$. For any $\epsilon > 0$, again by independence, it holds that

$$\frac{\mathbb{P}(\mathbf{Z}_{1,h} \in tA, Z_1 > t\eta, \ldots, Z_{\beta_{\mathcal{C}}} > t\eta)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} \sim \frac{\mathbb{P}(\mathbf{Z}_{1,h} \in tA\ ,\ Z_1 > t\eta, \ldots, Z_{\beta_{\mathcal{C}}} > t\eta, Z_{\beta_{\mathcal{C}}+1} \le t\epsilon, \ldots, Z_h \le t\epsilon)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}}.$$

Fix some arbitrary $\zeta > 0$, $\zeta < \inf_{u=1}^k a_u$. Then ϵ can be chosen small enough, so that the last term is less than $\mathbb{P}((Z_1, \ldots, Z_{\beta_{\mathcal{C}}}) \in A_\zeta)$ where

$$A_\zeta = \{z_1, \ldots, z_{\beta_{\mathcal{C}}} \mid \sum_{i \in P_u \cap \{1, \ldots, \beta_{\mathcal{C}}\}} z_i \ge a_u - \zeta,\ u = 1, \ldots, k\}.$$

Thus, we obtain

$$\begin{aligned}
\limsup_{t\to\infty} & \frac{\mathbb{P}(\mathbf{Z}_{1,h} \in tA, Z_1 > t\eta, \ldots, Z_{\beta_{\mathcal{C}}} > t\eta)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} \\
&\le \limsup_{t\to\infty} \frac{\mathbb{P}((Z_1, \ldots, Z_{\beta_{\mathcal{C}}}) \in tA_\zeta)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} \\
&= \alpha^{\beta_{\mathcal{C}}} \int_{A_\zeta} \prod_{i=1}^{\beta_{\mathcal{C}}} z_i^{-\alpha-1}\, \mathrm{d}z_i.
\end{aligned}$$

Moreover, $\lim_{\zeta\to 0} \int_{A_\zeta} \prod_{i=1}^{\beta_{\mathcal{C}}} z_i^{-\alpha-1}\, \mathrm{d}z_i = \int_{A_0} \prod_{i=1}^{\beta_{\mathcal{C}}} z_i^{-\alpha-1}\, \mathrm{d}z_i$, thus it actually holds that

$$\limsup_{t\to\infty} \frac{\mathbb{P}(\mathbf{Z}_{1,h} \in tA, Z_1 > t\eta, \ldots, Z_{\beta_{\mathcal{C}}} > t\eta)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} \le \alpha^{\beta_{\mathcal{C}}} \int_{A_0} \prod_{i=1}^{\beta_{\mathcal{C}}} z_i^{-\alpha-1}\, \mathrm{d}z_i.$$

Conversely, for the lower bound, it obviously holds that

$$\begin{aligned}
\frac{\mathbb{P}(\mathbf{Z}_{1,h} \in tA, Z_1 > t\eta, \ldots, Z_{\beta_{\mathcal{C}}} > t\eta)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} &\ge \frac{\mathbb{P}((Z_1, \ldots, Z_{\beta_{\mathcal{C}}}) \in tA_0)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} \\
&\to \alpha^{\beta_{\mathcal{C}}} \int_{A_0} \prod_{i=1}^{\beta_{\mathcal{C}}} z_i^{-\alpha-1}\, \mathrm{d}z_i.
\end{aligned}$$

Comparing the lower bound and the upper bound and summing over $\mathbf{i} \in P^*$ yields

$$\lim_{t\to\infty} \frac{\mathbb{P}(\mathbf{Z}_{1,h} \in tA)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} = \alpha^{\beta_{\mathcal{C}}} \sum_{\mathbf{i} \in P^*} \int_{A_0(\mathbf{i})} \prod_{q=1}^{\beta_{\mathcal{C}}} z_{i_q}^{-\alpha-1}\, \mathrm{d}z_{i_q},$$

where $\mathbf{i} = (i_1, \dots, i_{\beta_\mathcal{C}})$, $A_0(\mathbf{i}) = \{(z_{i_1}, \dots, z_{i_{\beta_\mathcal{C}}}) \mid \sum_{i \in P_u \cap \{i_1,\dots,i_{\beta_\mathcal{C}}\}} z_i \geq a_u, \ u = 1, \dots, k\}$. This proves that the measure $\nu_\mathcal{C}$ has the following expression

$$\nu_\mathcal{C}(\mathrm{d}\mathbf{z}) = \alpha^{\beta_\mathcal{C}} \sum_{\mathbf{i} \in P^*} \prod_{q=1}^{\beta_\mathcal{C}} z_{i_q}^{-\alpha-1} \mathrm{d}z_{i_q} \prod_{i \notin \{i_1,\dots,i_{\beta_\mathcal{C}}\}} \delta_0(\mathrm{d}z_i). \tag{36}$$

where δ_0 denotes the Dirac point measure at 0.

We now prove Eq. 14. By the characterization of relatively compact sets given above, if A is relatively compact in $\mathcal{C}$, then for $\mathbf{u} \in (0, \infty)^h$, $\mathbf{u}^{-1} \cdot A$ is also relatively compact in $\mathcal{C}$. Thus Eq. 11 implies that

$$\lim_{t \to \infty} \frac{\mathbb{P}(\mathbf{u} \cdot \mathbf{Z}_{1,h} \in tA)}{(\bar{F}_Z(t))^{\beta_\mathcal{C}}} = \nu_\mathcal{C}(\mathbf{u}^{-1} \cdot A). \tag{37}$$

It follows from Potter's bound and the characterization of a relatively compact set A of $\mathcal{C}$, that for any $\epsilon > 0$, there exists a constant C (which depends on A and ϵ) such that, for all $\mathbf{u} \in (0, \infty)^h$, all $t \geq 1$,

$$\begin{aligned} \frac{\mathbb{P}(\mathbf{u} \cdot \mathbf{Z}_{1,h} \in tA)}{\left(\bar{F}_Z(t)\right)^{\beta_\mathcal{C}}} &\leq \frac{\mathbb{P}(\exists i_1, \dots, i_{\beta_\mathcal{C}} \in \{1, \dots, h\}, \ u_{i_j} Z_{i_j} > \eta)}{\left(\bar{F}_Z(t)\right)^{\beta_\mathcal{C}}} \\ &\leq C \sum_{1 \leq i_1 < \dots < i_{\beta_\mathcal{C}} \leq h} \prod_{q=1}^{\beta_\mathcal{C}} (u_{i_q} \vee 1)^{\alpha+\epsilon}. \end{aligned}$$

Thus, denoting $M(u) = \prod_{i=1}^h (u_i \vee 1)^{\alpha+\epsilon}$, we obtain that there exists a constant C (which depends on A and ϵ) such that

$$\sup_{t \geq 1} \frac{\mathbb{P}(\mathbf{u} \cdot \mathbf{Z}_{1,h} \in tA)}{(\bar{F}_Z(t))^{\beta_\mathcal{C}}} \leq C M(\mathbf{u}). \tag{38}$$

Assumption 13 implies that $\mathbb{E}[M(\boldsymbol{\sigma}(\mathbf{X}_{1,h}))] < \infty$. Then Eqs. 37, 38 and bounded convergence yield (Eq. 14). We now prove Eq. 15. For $r \geq 2$, let $\mathcal{C}_r$ be the subcone of $[0, \infty]^{h+r-1}$ defined by

$$(z_1, \dots, z_{h+r-1}) \in \mathcal{C}_r \iff (z_1, \dots, z_h) \in \mathcal{C}, \ (z_r, \dots, z_{h+r-1}) \in \mathcal{C}.$$

For $u = 0, \dots, k$, define $P_u^r = r - 1 + P_u$, i.e. $i \in P_u^r$ if and only if $i - r + 1 \in P_u$ (which implies that $i \geq r$). Then

$$(z_1, \dots, z_{h+r-1}) \in \mathcal{C}_r \iff \prod_{u=1}^k \left(\sum_{i \in P_u} z_i \right) \prod_{u=1}^k \left(\sum_{i \in P_u^r} z_i \right) > 0.$$

The sum over the sets P_u^r which include one of the sets P_v can be removed from the second product, and thus we see that $\mathcal{C}_r$ is of the form 12, and Eq. 11 holds. Necessarily, it holds that $\beta_{\mathcal{C}_r} \geq \beta_\mathcal{C}$. Indeed, if there exists only $\ell < \beta_\mathcal{C}$ indices $i_1, \dots, i_\ell$ such that $z_i > 0$, then the first product above is zero, hence $(z_1, \ldots, z_{h+r-1}) \notin \mathcal{C}_r$. Let now $A_r(s, s')$ be the subset of $[0, \infty]^{h+r-1}$ such that $(z_1, \ldots, z_{h+r-1}) \in A_r(s, s')$ if and only if $(z_1, \ldots, z_h) \in sA$ and $(z_r, \ldots, z_{h+r-1}) \in s'A$. If A is relatively compact in $\mathcal{C}$, then $A_r(s, s')$ is also relatively compact in $\mathcal{C}_r$ and thus, it holds that

$$\lim_{t\to\infty} \frac{\mathbb{P}(\mathbf{Z}_{1,h} \in tsA,\ \mathbf{Z}_{r,r+h-1} \in ts'A)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}_r}}} = \lim_{t\to\infty} \frac{\mathbb{P}(\mathbf{Z}_{1,r+h-1} \in tA_r(s, s'))}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}_r}}}$$
$$= \nu_{\mathcal{C}_r}(A_r(s, s')),$$

and Eq. 15 follows straightforwardly, with $\mathcal{L}_r \equiv 0$ if $\beta_{\mathcal{C}_r} > \beta_{\mathcal{C}}$. □

Proof of Lemma 3 By assumption, we have

$$\lim_{t\to\infty} \frac{\mathbb{P}((Z_1, \ldots, Z_h) \in tA)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}}}} = \nu_{\mathcal{C}}(A)\ , \quad \lim_{t\to\infty} \frac{\mathbb{P}((Z_1, \ldots, Z_h) \in tA)}{(\bar{F}_Z(t))^{\beta_{\mathcal{C}'}}} = \nu_{\mathcal{C}'}(A),$$

with $\nu_{\mathcal{C}}(A) \in (0, \infty)$ and $\nu_{\mathcal{C}'}(A) \in (0, \infty)$. This implies that $\beta_{\mathcal{C}} = \beta_{\mathcal{C}'}$ and $\nu_{\mathcal{C}}(A) = \nu_{\mathcal{C}'}(A)$. It easily follows that for all $\mathbf{u} \in (0, \infty]^h$, $\nu_{\mathcal{C}}(\mathbf{u} \cdot A) = \nu_{\mathcal{C}'}(\mathbf{u} \cdot A)$. □

We now prove the results of Section 3. For clarity of notation, denote $\sigma_i = \sigma(X_i)$, $\nu_{\mathcal{C}} = \nu$, $\mu_{\mathcal{C}} = \mu$ and define $g(t) = t^{\beta_{\mathcal{C}}}$ and $T(s) = s^{-\alpha\beta_{\mathcal{C}}}$. Recall that F_Y denotes the distribution function of Y and $u_n = (1/\bar{F}_Y)^{\leftarrow}(n/k)$. By Eqs. 4 and 13, Breiman's Lemma applies and thus it holds that $\bar{F}_Y(u_n) \sim \mathbb{E}[\sigma_0^\alpha]\bar{F}_Z(u_n)$ and

$$\lim_{n\to\infty} \frac{g(k/n)}{g(\bar{F}_Z(u_n))} = (\mathbb{E}[\sigma_0^\alpha])^{\beta_{\mathcal{C}}}.$$

Whenever there is no risk of confusion, we omit dependence on h, m, h' and A in the notation. For $r = 1, \ldots, n$, define the following random variables

$$W_{r,n}(s) = \mathbf{1}_{\{\mathbf{Y}_{r,r+h-1} \in u_n sA\}},\ s \geq 1, \quad V_r(B) = \mathbf{1}_{\{\mathbf{Y}_{r+m,r+m+h'} \in B\}}. \tag{39}$$

The choice of u_n implies that (recall the definitions (Eqs. 16 and 20) of $\rho(A, B, m)$ and $\mu(A)$),

$$\lim_{n\to\infty} \frac{\mathbb{E}[W_{r,n}(s)]}{g(k/n)} = T(s)\mu(A), \tag{40}$$

$$\lim_{n\to\infty} \frac{\mathbb{E}[W_{r,n}(s)V_r(B)]}{g(k/n)} = T(s)\mu(A)\rho(A, B, m). \tag{41}$$

Recall the definition (Eq. 25) of the function G_n:

$$G_n(A, B, s, \mathbf{x}, \mathbf{x}') = \frac{\mathbb{P}(\boldsymbol{\sigma}(\mathbf{x}) \cdot \mathbf{Z}_{1,h} \in u_n sA)}{g_{\mathbf{j}}(k/n)}\, \mathbb{P}(\boldsymbol{\sigma}(\mathbf{x}') \cdot \mathbf{Z}_{m,m+h'} \in B)$$

Also, define, for $s \geq 1$ and $\mathbf{x} \in \mathbb{R}^h$ and $\mathbf{x}' \in \mathbb{R}^{h'+1}$, the function L_n by

$$L_n(s, \mathbf{x}) = \frac{\mathbb{P}(\boldsymbol{\sigma}(\mathbf{x}) \cdot \mathbf{Z}_{1,h} \in u_n s A)}{g(k/n)} , \tag{42}$$

With these notations, we have,

$$\begin{aligned} L_n(s, \mathbf{X}_{r,r+h-1}) &= \frac{\mathbb{E}[W_{r,n}(s) \mid \mathcal{X}]}{g(k/n)}, \\ G_n(A, B, s, \mathbf{X}_{r,r+h-1}, \mathbf{X}_{r+m,r+m+h'}) &= \frac{\mathbb{E}[W_{r,n}(s) V_r(B) \mid \mathcal{X}]}{g(k/n)}. \end{aligned}$$

For $\mathbf{x} \in \mathbb{R}^h$, denote

$$L(\mathbf{x}) = \frac{\nu(\boldsymbol{\sigma}(\mathbf{x})^{-1} \cdot A)}{(\mathbb{E}[\sigma^\alpha(X)])^\delta}, \tag{43}$$

so that $\mathbb{E}[L(\mathbf{X}_{1,h})] = \mu(A)$.

Proof of Lemma 4 Write

$$\begin{aligned} &L_n(s, \mathbf{x}) - T(s) L(\mathbf{x}) \\ &\quad = \left\{ \frac{g(\bar{F}_Z(u_n s))}{g(k/n)} - (\mathbb{E}[\sigma^\alpha(X)])^{-\delta}\, T(s) \right\} \frac{\mathbb{P}(\sigma(\mathbf{x}) \cdot \mathbf{Z}_{1,h} \in u_n s A)}{g(\bar{F}_Z(u_n s))} \\ &\qquad + (\mathbb{E}[\sigma^\alpha(X)])^{-\delta} T(s) \left\{ \frac{\mathbb{P}(\sigma(\mathbf{x}) \cdot \mathbf{Z}_{1,h} \in u_n s A)}{g(\bar{F}_Z(u_n s))} - (\mathbb{E}[\sigma^\alpha(X)])^\delta L(\mathbf{x}) \right\}. \end{aligned}$$

Thus, recalling the definition of v_n from Eq. 21, we have

$$\begin{aligned} v_n(A) \leq \sup_{s \geq 1} &\left| \frac{g(\bar{F}_Z(u_n s))}{g(k/n)} - (\mathbb{E}[\sigma^\alpha(X)])^{-\delta}\, T(s) \right| \mathbb{E}[M(\boldsymbol{\sigma}(\mathbf{X}_{1,h}))] + (\mathbb{E}[\sigma^\alpha(X)])^{-\delta} \\ &\times \mathbb{E}\left[\sup_{s \geq 1} \left| \frac{\mathbb{P}(\sigma(\mathbf{X}_{1,h}) \cdot \mathbf{Z}_{1,h} \in u_n s A \mid \mathcal{X})}{g(\bar{F}_Z(u_n s))} - (\mathbb{E}[\sigma^\alpha(X)])^\delta L(\mathbf{X}_{1,h}) \right| \right] \end{aligned}$$

For all $\mathbf{x} \in \mathbb{R}^h$, we have

$$\lim_{n \to \infty} \sup_{s \geq 1} \left| \frac{\mathbb{P}(\sigma(\mathbf{x}) \cdot \mathbf{Z}_{1,h} \in u_n s A)}{g(\bar{F}_Z(u_n s))} - (\mathbb{E}[\sigma^\alpha(X)])^\delta L(\mathbf{x}) \right| = 0.$$

Moreover, by Eq. 38,

$$\sup_{s \geq 1} \left| \frac{\mathbb{P}(\sigma(\mathbf{x}) \cdot \mathbf{Z}_{1,h} \in u_n s A)}{g(\bar{F}_Z(u_n s))} - (\mathbb{E}[\sigma^\alpha(X)])^\delta L(\mathbf{x}) \right| \leq C M(\boldsymbol{\sigma}(\mathbf{x})).$$

Thus, by Eq. 13 and bounded convergence,

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{s\geq 1}\left|\frac{\mathbb{P}(\sigma(\mathbf{X}_{1,h})\cdot\mathbf{Z}_{1,h}\in u_n sA \mid \mathcal{X})}{g(\bar{F}_Z(u_n s))} - (\mathbb{E}[\sigma^\alpha(X)])^\delta L(\mathbf{X}_{1,h})\right|\right] = 0.$$

Since $g\circ\bar{F}$ is regularly varying at infinity with negative index, by Bingham et al. (1989, Theorem 1.5.2), the convergence of $g(\bar{F}_Z(u_n s))/g(k/n)$ to $(\mathbb{E}[\sigma^\alpha(X)])^{-\delta}T(s)$ is uniform on $[1,\infty)$. Thus we have proved that $v_n(A)\to 0$. □

Proof of Theorem 5 For $s\geq 1$, define

$$K(B,s) = T(s)\mu(A)\rho(A,B,m),$$

$$\tilde{K}_n(B,s) = \frac{1}{ng(k/n)}\sum_{r=1}^{n} W_{r,n}(s)V_r(B),$$

$$\tilde{e}_n(s) = \tilde{K}_n(\mathbb{R}^{h'+1},s) = \frac{1}{ng(k/n)}\sum_{r=1}^{n} W_{r,n}(s),$$

$$\xi_n = \frac{Y_{(n:n-k)}}{u_n}.$$

With this notation, we have

$$\hat{\rho}_n(A,B,m) = \frac{\tilde{K}_n(B,\xi_n)}{\tilde{e}_n(\xi_n)}$$

Equations 41 and 40 imply, respectively, that

$$\lim_{n\to\infty}\mathbb{E}[\tilde{K}_n(B,s)] = K(B,s) \qquad \lim_{n\to\infty}\mathbb{E}[\tilde{e}_n(s)] = T(s)\mu(A).$$

With this in mind, we split

$$\hat{\rho}_n(A,B,m) - \rho(A,B,m) = \frac{\tilde{K}_n(B,\xi_n) - K(B,\xi_n)}{\tilde{e}_n(\xi_n)} - \frac{\rho(A,B,m)}{\tilde{e}_n(\xi_n)}\{\tilde{e}_n(\xi_n) - \mu(A)T(\xi_n)\}. \tag{44}$$

We note that the term $\tilde{e}_n(\xi_n) - \mu(A)T(\xi_n)$ that appears in the second part of Eq. 44 has the form $\tilde{K}_n(B,\xi_n) - K(B,\xi_n)$ with $B=\mathbb{R}^{h'+1}$. Thus, we only need to find the correct norming sequence w_n and asymptotic distribution in $\mathcal{D}([a,b])$ for any $0<a<b$ of the sequence of processes $w_n\{\tilde{K}_n(B,\cdot) - K(B,\cdot)\}$. To do this, define further

$$K_n(B,s) = \mathbb{E}[\tilde{K}_n(B,s)]. \tag{45}$$

Then

$$\tilde{K}_n(B,s) - K(B,s) = \tilde{K}_n(B,s) - K_n(B,s) + K_n(B,s) - K(B,s).$$

The term $K_n(B,s) - K(B,s)$ is a deterministic bias term that will be dealt with by the second order condition 23. Write $\tilde{K}_n - K_n = (ng(k/n))^{-1/2}E_{n,1} + E_{n,2}$ with

$$E_{n,1}(B,s) = \frac{1}{\sqrt{ng(k/n)}}\sum_{r=1}^{n}\{W_{r,n}(s)V_r(B) - \mathbb{E}[W_{r,n}(s)V_r(B) \mid \mathcal{X}]\}, \tag{46}$$

$$\begin{aligned} E_{n,2}(B,s) &= \frac{1}{ng(k/n)}\sum_{r=1}^{n}\mathbb{E}[W_{r,n}(s)V_r(B) \mid \mathcal{X}] - K_n(B,s) \\ &= \frac{1}{n}\sum_{r=1}^{n}\{G_n(A,B,s,\mathbf{X}_{r,r+h-1},\mathbf{X}_{r+m,r+m+h'}) - K_n(B,s)\}. \end{aligned} \tag{47}$$

We note that $K_n(B,s)$ is the proper centering in Eq. 47 since, using the definition of K_n, Eqs. 42 and 25 we have

$$\begin{aligned} K_n(B,s) &= \mathbb{E}[G_n(A,B,s,\mathbf{X}_{1,h},\mathbf{X}_{m,m+h'})] \\ &= \mathbb{E}[L_n(s,\mathbf{X}_{1,h})\mathbb{P}(\sigma(\mathbf{X}_{m,m+h'})\cdot \mathbf{Z}_{m,m+h'} \in B \mid \mathcal{X})]. \end{aligned} \tag{48}$$

The term in Eq. 46 will be called the i.i.d. term. It is a sum of conditionally independent random variables. The term in Eq. 47 will be called the dependent term. It is a function of the dependent vectors $(\mathbf{X}_{r,r+h-1},\mathbf{X}_{r+m,r+m+h'})$.

We now state some claims whose proofs are postponed to the end of this section. The implication of Claims 1 and 3 is, in particular, that in the weakly dependent case only the i.i.d. part contributes to the limit.

Claim 1 *The process $E_{n,1}$ converges in the sense of finite-dimensional distributions to a Gaussian process W with covariance*

$$\begin{aligned} &(\mathbb{E}[\sigma^{\alpha}(X_1)])^{\delta}\mathrm{cov}(W(B,s),W(B',s')) \\ &= \mathbb{E}\Big[\mathcal{L}_1(A,\boldsymbol{\sigma}(\mathbf{X}_{1,h}),\boldsymbol{\sigma}(\mathbf{X}_{1,h}),s,s')\times \mathbb{P}(\mathbf{Y}_{m,m+h'} \in B, \mathbf{Y}_{m,m+h'} \in B' \mid \mathcal{X})\Big] \\ &\quad + \sum_{r=2}^{h\wedge(m-h)} \mathbb{E}\Big[\mathcal{L}_r(A,\boldsymbol{\sigma}(\mathbf{X}_{1,h}),\boldsymbol{\sigma}(\mathbf{X}_{r,r+h-1}),s,s') \\ &\qquad\qquad \times \{\mathbb{P}(\mathbf{Y}_{m,m+h'} \in B, \mathbf{Y}_{m+r-1,m+h'+r-1} \in B' \mid \mathcal{X}) \\ &\qquad\qquad + \mathbb{P}(\mathbf{Y}_{m,m+h'} \in B', \mathbf{Y}_{m+r-1,m+h'+r-1} \in B \mid \mathcal{X})\}\Big], \end{aligned} \tag{49}$$

where the functions $\mathcal{L}_r$ are defined in Eq. 15.

Claim 2 *For each fixed B, $E_{n,1}(B,\cdot)$ is tight in $\mathcal{D}([a,b])$ for each $0 < a < b$.*

This claim is proved in Lemma C.3.

The previous two statements are valid in both weakly dependent and long memory case. The next one may not be valid in the long memory case. See Section 3.2.

Claim 3 *In the weakly dependent case* $E_{n,2}(B,\cdot) = O_P(\sqrt{n})$*, uniformly with respect to* $s \in [a,b]$ *for any* $0 < a < b$.

The next claim is proved in Kulik and Soulier (2011, Corollary 2.4).

Claim 4 $\xi_n - 1 = o_P(1)$.

The last thing we need is the negligibility of the bias term.

Claim 5 *For any* $a > 0$, $\sup_{s \geq a} \sup_B |K_n(B,s) - K(B,s)| = O(v_n(A))$.

Therefore if $ng(k/n) \to \infty$ and Eq. 23 holds (i.e. $ng(k/n)v_n(A) \to 0$), then

$$\sqrt{ng(k/n)}\{\tilde{K}_n(B,\cdot) - K(B,\cdot), \tilde{e}_n(\cdot) - K(\mathbb{R}^d,\cdot)\} \Rightarrow (W(B,\cdot), W(\mathbb{R}^{h'+1},\cdot)).$$

This convergence and the decomposition 44 imply

$$\begin{aligned}&\sqrt{ng(k/n)\mu(A)}\{\hat{\rho}_n(A,B,m) - \rho(A,B,m)\}\\ &\quad \to_d W(B,1) - \rho(A,B,m)W(\mathbb{R}^{h'+1},1).\end{aligned}$$

This distribution is Gaussian. Applying Eq. 49 and the fact that $\rho(A,\mathbb{R}^{h'+1},m) = 1$, it is easily checked that its variance is given by Eq. 24. This concludes the proof of Theorem 5. □

We now prove the claims.

Proof of Claim 1 For $r = 1,\dots,n$, denote

$$\zeta_{n,r}(B,s) = \frac{1}{\sqrt{ng(k/n)}} W_{r,n}(s)V_r(B).$$

In order to prove our claim, we apply the central limit theorem for m-dependent random variables, see Orey (1958). Let $C(B,B',s,s')$ denote the quantity in the right hand side of Eq. 49. We need to check that

$$\operatorname{cov}\left(\sum_{r=1}^n \zeta_{n,r}(B,s), \sum_{r=1}^n \zeta_{n,r}(B',s') \mid \mathcal{X}\right) \to_P C(B,B',s,s'), \tag{50}$$

$$\sum_{r=1}^n \mathbb{E}[\zeta_{n,r}^4(B,s) \mid \mathcal{X}] \to_P 0. \tag{51}$$

By standard Lindeberg–Feller type arguments, this proves the one-dimensional convergence. The finite-dimensional convergence is proved by similar arguments and by computing the asymptotic covariances. We now prove Eqs. 50 and 51.

For $u \geq 1$, $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^h$, denote

$$\mathcal{L}_{n,u}(A, \mathbf{x}, \mathbf{x}', s, s') = \frac{\mathbb{P}(\boldsymbol{\sigma}(\mathbf{x}) \cdot \mathbf{Z}_{1,h} \in u_n s A, \boldsymbol{\sigma}(\mathbf{x}') \cdot \mathbf{Z}_{u,u+h-1} \in u_n s' A)}{g(\bar{F}_Z(u_n))}.$$

The functions $\mathcal{L}_{n,u}$ converge in $L^1(\mathbf{X}_{1,h}, \mathbf{X}_{u,u+h-1})$ to the functions $\mathcal{L}_u$ defined Eq. 15. For $u > h$, $\mathbf{Z}_{1,h}$ and $\mathbf{Z}_{u,u+h-1}$ are independent, so $\mathcal{L}_{n,u}$ converges a.s. and in $L^1(\mathbf{X}_{1,h}, \mathbf{X}_{u,u+h-1})$ to 0.

The random variables $\zeta_{n,r}$ are $m + h'$ dependent. Thus,

$$\begin{aligned}
&\operatorname{cov}\left(\sum_{r=1}^{n} \zeta_{n,r}(B, s), \sum_{r=1}^{n} \zeta_{n,r}(B', s') \mid \mathcal{X}\right) \\
&= \sum_{r=1}^{n} \operatorname{cov}(\zeta_{n,r}(B, s), \zeta_{n,r}(B', s') \mid \mathcal{X}) \\
&\quad + \sum_{r=1}^{n} \sum_{u=1}^{m+h'} \operatorname{cov}(\zeta_{n,r}(B, s), \zeta_{n,j+u}(B', s') \mid \mathcal{X}) \qquad (52)\\
&\quad + \sum_{r=1}^{n} \sum_{u=1}^{m+h'} \operatorname{cov}(\zeta_{n,j+u}(B, s), \zeta_{n,r}(B', s') \mid \mathcal{X}). \qquad (53)
\end{aligned}$$

For $u = 1, \ldots, h \wedge (m - h)$ it is easily seen that

$$\begin{aligned}
&\sum_{r=1}^{n} \operatorname{cov}(\zeta_{n,r}(B, s), \zeta_{n,r+u}(B', s') \mid \mathcal{X}) \\
&\sim \frac{g(\bar{F}_Z(u_n))}{ng(k/n)} \sum_{r=1}^{n} \mathcal{L}_{n,u}(\mathbf{X}_{r,r+h-1}, \mathbf{X}_{r+u,r+u+h-1}, s, s') \\
&\quad \times \mathbb{P}(\mathbf{Y}_{r+m,r+m+h} \in B, \mathbf{Y}_{r+u+m,r+u+m+h'} \in B' \mid \mathcal{X}) \\
&\to_P \frac{\mathbb{E}\left[\mathcal{L}_u(A, \mathbf{X}_{1,h}, \mathbf{X}_{u,h+u-1}, s, s')\mathbb{P}(\mathbf{Y}_{m,m+h} \in B, \mathbf{Y}_{u+m,u+m+h'} \in B' \mid \mathcal{X})\right]}{(\mathbb{E}[\sigma^\alpha(X)])^\delta}.
\end{aligned}$$

This yields the right-hand side of Eq. 49, so we must prove that the terms in Eqs. 52 and 53 are negligible. If $h > m - h$, then for large n and $m - h < u \leq h$, we have $(u_n s' A) \cap B = 0$, so, for all $r = 1 \ldots, n$,

$$\begin{aligned}
&\mathbb{P}\big(\mathbf{Y}_{r,r+h-1} \in u_n s A, \mathbf{Y}_{r+u,r+u+h-1} \in u_n s' A, \\
&\mathbf{Y}_{r+m,r+m+h} \in B, \mathbf{Y}_{r+u+m,r+u+m+h'} \in B' \mid \mathcal{X}\big) = 0.
\end{aligned}$$

For $u > h$, then as mentioned above, $\mathcal{L}_u(A, \cdot, \cdot, s, s')$ converges to 0 in $L^1(\mathbf{X}_{1,h}, \mathbf{X}_{u,u+h-1})$ so

$$\sum_{r=1}^{n} \operatorname{cov}(\zeta_{n,r}(B, s), \zeta_{n,r+u}(B', s') \mid \mathcal{X}) \to_P 0.$$

This proves Eq. 50. Next, since $\zeta_{n,r}$ are indicators and applying Eq. 41

$$\sum_{r=1}^{n} \mathbb{E}[\zeta_{n,r}^4(B, s)] \leq C \frac{\mathbb{E}[W_{1,n}(s, A) V_1(B)]}{ng(k/n)} \to 0.$$

This proves Eq. 51 and the weak convergence of finite dimensional distributions. □

Proof of Claim 3 By definition of the functions L_n and G_n (cf. Eqs. 42 and 25), it clearly holds that

$$|G_n(A, B, s, \mathbf{X}_{r,r+h-1}, \mathbf{X}_{r+m,r+m+h'})| \leq L_n(s, \mathbf{X}_{r,r+h-1}).$$

We apply the variance inequality B.3 in the weak dependence case to get

$$\operatorname{var}(E_{n,2}(B, s)) \leq \frac{C}{n} \operatorname{var}(G_n(A, B, s, \mathbf{X}_{1,h}, \mathbf{X}_{1+m,1+m+h'})) \leq \frac{C}{n} \mathbb{E}[L_n^2(s, \mathbf{X}_{1,h})].$$

By Eq. 38, $L_n(s, \mathbf{x}) \leq CM(\boldsymbol{\sigma}(\mathbf{x}))$. Thus, by Eq. 13, the right hand side is uniformly bounded, thus $\operatorname{var}(E_{n,2}(B, s)) = O(1/n)$ and for any fixed $s > 0$, $\sqrt{n} E_{n,2}(B, s) = O_P(1)$. Tightness follows from Lemma C.4, thus $E_{n,2}(B, \cdot)$ converges uniformly to 0 on any compact set of $(0, \infty]$. □

Proof of Claim 5 Consider now the bias term $K_n - K$. Recall that (see Eqs. 45 and 41)

$$K_n(B, s) = \mathbb{E}[\bar{K}_n(B, s)] \to T(s)\mu(A)\rho(A, B, m) = K(B, s)$$

Therefore, $K_n(B, s)$ converges pointwise to $K(B, s)$. The goal here is to show that this convergence is uniform. Using Eq. 48 and recalling the formula for $\rho(A, B, m)$ (see Eq. 16), we have

$$K(B, s) = T(s)\mathbb{E}[L(\mathbf{X}_{1,h})\mathbb{P}(\sigma(\mathbf{X}_{m,m+h'}) \cdot \mathbf{Z}_{m,m+h'} \in B \mid \mathcal{X})].$$

Therefore, recalling the definition (Eq. 21) of $v_n(A)$, we obtain that

$$|K_n(B, s) - K(B, s)| \leq \mathbb{E}\left[\sup_{s \geq 1} |L_n(s, \mathbf{X}_{1,h}) - T(s)L(\mathbf{X}_{1,h})|\right] = v_n(A).$$

□

Proof of Corollary 6 In the following, $\mathbf{y}$ stands for the set $(-\infty, \mathbf{y}]$ in the previous notation. For $\mathbf{y} \in \mathbb{R}^{h'+1}$, rewrite the decomposition 44 in the present context to get

$$\hat{\Psi}_n(\mathbf{y}) - \Psi(\mathbf{y}) = \frac{\tilde{K}_n(\mathbf{y}, \xi_n) - K(\mathbf{y}, \xi_n)}{\tilde{e}_n(\xi_n)} - \frac{\Psi(\mathbf{y})}{\tilde{e}_n(\xi_n)} \{\tilde{e}_n(\xi_n) - \mu(A)\, T(\xi_n)\}.$$

Thus we need only prove that the sequence of suitably normalized processes $\tilde{K}_n(s, \mathbf{y}) - K_n(\mathbf{y}, s)$ converge weakly to the claimed limit. The convergence of finite dimensional distributions follows from Theorem 5 and the tightness follows from Lemmas C.3 and C.4. □

Proof of Theorem 7 Claims 1, 2, 4 and 5 hold under the assumptions of Theorem 7. Thus, the result will follow if we prove a modified version of Claim 3.

Claim 6 *If* $2\tau(A, B)(1 - H) < 1$, *then* $\gamma_n^{-\tau(A,B)/2} E_{n,2}(A, B, \cdot)$ *converges weakly uniformly on compact sets of* $(0, \infty]$ *to a process* $T \cdot Z(A, B)$ *where the random variable* $Z(A, B)$ *is in a Gaussian chaos of order* $\tau(A, B)$ *and its distribution depends only on the Gaussian process* $\{X_n\}$.

For any $d \in \mathbb{N}^*$, $\mathbf{q} \in \mathbb{N}^d$ and $\mathbf{x} \in \mathbb{R}^d$, denote

$$\mathbf{H}_{\mathbf{q}}(\mathbf{x}) = \prod_{i=1}^{d} H_{q_i}(x_i) \ .$$

Define $\mathbb{X}_j = (X_{j+1}, \dots, X_{j+h}, X_{j+m}, \dots, X_{j+m+h'})$. The Hermite coefficients of $G_n(A, B, s, \cdot)$ and G with respect to $\mathbb{X}_0$ can be expressed, for $\mathbf{q} \in \mathbb{N}^{h+h'+1}$, as

$$J_n(\mathbf{q}, s) = \mathbb{E}[\mathbf{H}_{\mathbf{q}}(\mathbb{X}_0) G_n(A, B, s, \mathbb{X}_0)], \quad J(\mathbf{q}) = \mathbb{E}[\mathbf{H}_{\mathbf{q}}(\mathbb{X}_0) G(\mathbb{X}_0)].$$

Since $G_n(A, B, s, \cdot)$ converges to $T(s)G(\cdot)$ in $L^p(\mathbb{X}_0)$ for some $p > 1$, $J_n(\mathbf{q}, s)$ converges to $s^{-\alpha\delta} J(\mathbf{q})$. Let U be an $(h + h' + 1) \times (h + h' + 1)$ matrix such that UU' is equal to the inverse of the covariance matrix of $\mathbb{X}_0$. Define $J_n^*(\mathbf{q}, s) = \mathbb{E}[\mathbf{H}_{\mathbf{q}}(U\mathbb{X}_0) G_n(A, B, s, U\mathbb{X}_0)]$ and $J^*(q) = \mathbb{E}[\mathbf{H}_{\mathbf{q}}(U\mathbb{X}_0) G(\mathbb{X}_0)]$. Under Assumption 1, the function G_n can be expanded for $\mathbf{x} \in \mathbb{R}^{h+h'+1}$ as

$$G_n(A, B, s, \mathbf{x}) - \mathbb{E}[G_n(A, B, s, \mathbb{X}_0)] = \sum_{|\mathbf{q}|=\tau(A,B)} \frac{J_n^*(\mathbf{q}, s)}{\mathbf{q}!} \mathbf{H}_{\mathbf{q}}(U\mathbf{x}) + r_n(s, \mathbf{x}),$$

where r_n is implicitly defined and has Hermite rank at least $\tau(A, B) + 1$ with respect to $U\mathbb{X}_0$. Denote $R_n(s) = n^{-1} \sum_{r=1}^{n} r_n(s, \mathbb{X}_j)$. Applying Eq. B.3, we have

$$\begin{aligned} \operatorname{var}(R_n(s)) &\le C \left(\gamma_n^{\tau(A,B)+1} \vee \frac{1}{n} \right) \operatorname{var}(G_n(A, B, s, \mathbb{X}_0)) \\ &\le C \left(\gamma_n^{\tau(A,B)+1} \vee \frac{1}{n} \right) \mathbb{E}[L_n^2(s, \mathbf{X}_{1,h})]. \end{aligned}$$

By Assumption 13, $\mathbb{E}[L_n^2(s, \mathbf{X}_{1,h})]$ is uniformly bounded, thus $\mathrm{var}(R_n(s)) = o(\gamma_n^{\tau(A,B)})$ and $\gamma_n^{-\tau(A,B)} R_n(s)$ converges weakly to zero. The convergence is uniform by an application of Lemma C.1.

Thus, the asymptotic behaviour of $\gamma_n^{-\tau(A,B)/2} E_{n,2}$ is the same as that of

$$Z_n(s) = \sum_{|\mathbf{q}|=\tau(A,B)} \frac{J_n^*(\mathbf{q}, s)n^{-1}}{\mathbf{q}!} \gamma_n^{-\tau(A,B)/2} \sum_{r=1}^{n} \mathbf{H}_{\mathbf{q}}(U\mathbb{X}_j).$$

By Arcones (1994, Theorem 6), there exist random variables $\aleph^*(\mathbf{q})$ such that $Z_n(s)$ converges to

$$T(s) \sum_{|\mathbf{q}|=\tau(A,B)} \frac{J^*(\mathbf{q})}{\mathbf{q}!} \aleph^*(\mathbf{q})$$

for each $s \geq 0$. To prove that the convergence is uniform, we only need to prove that $J_n^*(\mathbf{q}, s)$ converges uniformly to $T(s)J^*(\mathbf{q})$ for each $\mathbf{q}$ such that $|\mathbf{q}| = \tau(A)$. Since the coefficients J_n^* can be expressed linearly in terms of the coefficients J_n, it suffices to prove uniform convergence of the coefficients J_n. Applying Hölder inequality, we obtain, for $p > 1$ and for any $a > 0$,

$$\sup_{s \geq a} |J_n(\mathbf{q}, s) - T(s)J(\mathbf{q})| \leq C\mathbb{E}\left[\sup_{s \geq a} \left|L_n(s, \mathbf{X}_{1,h}) - T(s)L(\mathbf{X}_{1,h})\right|^p\right].$$

As already shown in the proof of Lemma 4, this last quantity converges to 0 for $p = 2$. □

Appendix A: Second order regular variation of convolutions

Lemma A.1 *Let Z_1 and Z_2 be i.i.d. non negative random variables with common distribution function F that satisfies Assumption 2. Then*

$$\left|\mathbb{P}(a_1 Z_1 + a_2 Z_2 > t) - \bar{F}(t/a_1) - \bar{F}(t/a_2)\right| \leq C(a_1 \vee 1)^{\alpha+\epsilon}(a_2 \vee 1)^{\alpha+\epsilon} t^{-1}\bar{F}(t) \int_0^t \bar{F}(v)\,\mathrm{d}v.$$

Proof of Theorem 7 Obviously, we have

$$\begin{aligned}
\mathbb{P}(a_1 Z_1 + a_2 Z_2 > t) = {} & \bar{F}(t/a_1) + \bar{F}(t/a_2) - \bar{F}(t/a_1)\bar{F}(t/a_2) \\
& + \mathbb{P}(t/2 < a_1 Z_1 \leq t)\mathbb{P}(t/2 < a_2 Z_2 \leq t) \\
& + \mathbb{P}(a_1 Z_1 \leq t/2, a_2 Z_2 \leq t, a_1 Z_1 + a_2 Z_2 > t) \\
& + \mathbb{P}(a_2 Z_2 \leq t/2, a_1 Z_1 \leq t, a_1 Z_1 + a_2 Z_2 > t).
\end{aligned}$$

Thus, we obtain

$$\left|\mathbb{P}(a_1 Z_1 + a_2 Z_2 > t) - \bar{F}(t/a_1) - \bar{F}(t/a_2)\right| \\ \leq 2\bar{F}(t/2a_1)\bar{F}(t/2a_2) + \bar{F}(t/a_2) I_1 + \bar{F}(t/a_1) I_2.$$

with

$$I_1 = \mathbb{E}\left[\mathbf{1}_{\{a_1 Z_1 \leq t/2\}} \left\{\frac{\bar{F}(t(1 - a_1 Z_1/t)/a_2)}{\bar{F}(t/a_2)} - 1\right\}\right],$$

$$I_2 = \mathbb{E}\left[\mathbf{1}_{\{a_2 Z_2 \leq t/2\}} \left\{\frac{\bar{F}(t(1 - a_2 Z_1/t)/a_1)}{\bar{F}(t/a_1)} - 1\right\}\right].$$

Since F satisfies Assumption 2, we have, for $u \in [1/2, 1]$ and $s > 0$,

$$0 \leq \frac{\bar{F}(us)}{\bar{F}(s)} - 1 = u^{-\alpha} e^{\int_1^u \frac{\eta(sv)}{v}\, \mathrm{d}v} - 1 = \{u^{-\alpha} - 1\} e^{\int_1^u \frac{\eta(sv)}{v}\, \mathrm{d}v} + e^{\int_1^u \frac{\eta(sv)}{v}\, \mathrm{d}v} - 1 \\ \leq |u^{-\alpha} - 1| e^{\int_{1/2}^1 \frac{\eta^*(sv)}{v}\, \mathrm{d}v} + e^{\int_{1/2}^1 \frac{\eta^*(sv)}{v}\, \mathrm{d}v} \int_u^1 \frac{\eta^*(sv)}{v}\, \mathrm{d}v.$$

Since η^* is decreasing, we have, for all $u \in [1/2, 1]$,

$$0 \leq \frac{\bar{F}(us)}{\bar{F}(s)} - 1 \leq C\{|u^{-\alpha} - 1| + \log(u)\} \leq C(1 - u).$$

Applying this inequality with $s = t/a_2$ and $1 - u = a_1 Z_1/t$ on the event $a_1 Z_1 \leq t/2$ yields

$$I_1 \leq C a_1 t^{-1} \mathbb{E}\left[Z_1 \mathbf{1}_{\{a_1 Z_1 \leq t\}}\right] \leq C t^{-1} \int_0^t \bar{F}(v/a_1)\, \mathrm{d}v \\ \leq C(a_1 \vee 1)^{\alpha+\epsilon} t^{-1} \int_0^t \bar{F}(v)\, \mathrm{d}v.$$

where the last bound is obtained by applying Potter's bound for some $\epsilon > 0$.

This yields the desired bounds for the term I_1. The bound for the term I_2 is obtained similarly. To conclude, note that $\bar{F}^2(t) = O(t^{-1}\bar{F}(t) \int_0^t \bar{F}(v)\, \mathrm{d}v)$ if $\alpha < 1$ and $\bar{F}^2(t) = o(t^{-1}\bar{F}(t) \int_0^t \bar{F}(v)\, \mathrm{d}v)$ if $\alpha \geq 1$. □

Remark 2 By induction, we can obtain the bound

$$\left|\mathbb{P}(Z_1 + \cdots + Z_n > t) - n\bar{F}(t)\right| \leq C\, t^{-1} \bar{F}(t) \int_0^t \bar{F}(v)\, \mathrm{d}v,$$

and we can also recover a particular case of a result of Omey and Willekens (1987) in a slightly different form. For $\alpha \geq 1$ and $\mathbb{E}[Z_1] < \infty$,

$$\lim_{t\to\infty} t\left\{\frac{\mathbb{P}(Z_1 + \cdots + Z_n > t)}{\mathbb{P}(Z_1 > t)} - n\right\} = \frac{n(n-1)}{2}\mathbb{E}[Z_1].$$

We apply Lemma A.1 to obtain the bound 32 for the bias term in Example 3 considered in Section 3.3.2. Recall that in this case we have $\beta_{\mathcal{C}} = 1$, $\mu_{\mathcal{C}}(A) = 2$, and that it always holds that $k/n = \bar{F}_Y(u_n) \sim \bar{F}(u_n)\mathbb{E}[\sigma^\alpha]$. Then $v_n(A)$ becomes

$$v_n(A) = \mathbb{E}\left[\sup_{s\geq 1}\left|\frac{\mathbb{P}(\sigma_1 Z_1 + \sigma_2 Z_2 > u_n s \mid \mathcal{X})}{\bar{F}_Y(u_n)} - 2s^{-\alpha}\right|\right]$$

$$\leq \frac{\bar{F}(u_n)}{\bar{F}_Y(u_n)}(A_1 + 2A_2) + 2A_3,$$

with

$$A_1 = \mathbb{E}\left[\sup_{s\geq 1}\left|\frac{\mathbb{P}(\sigma_1 Z_1 + \sigma_2 Z_2 > u_n s \mid \mathcal{X})}{\bar{F}(u_n)} - \frac{\mathbb{P}(\sigma_1 Z_1 > u_n s \mid \mathcal{X})}{\bar{F}(u_n)} - \frac{\mathbb{P}(\sigma_2 Z_2 > u_n s \mid \mathcal{X})}{\bar{F}(u_n)}\right|\right],$$

$$A_2 = \mathbb{E}\left[\sup_{s\geq 1}\left|\frac{\mathbb{P}(\sigma_1 Z_1 > u_n s \mid \mathcal{X})}{\bar{F}(u_n)} - s^{-\alpha}\sigma_1^\alpha\right|\right], \quad A_3 = \left|\frac{\mathbb{E}[\sigma_1^\alpha]\bar{F}(u_n)}{\bar{F}_Y(u_n)} - 1\right|.$$

By Kulik and Soulier (2011, Proposition 2.8), the terms A_2 and A_3 are of order $O(\eta^*(u_n))$. Next, applying Lemma A.1 with $t = u_n$ and $a_i = \sigma_i/s$ for $s \geq 1$ yields $A_1 \leq C u_n^{-1} \int_0^{u_n} \bar{F}(t)\,\mathrm{d}t$. Altogether, we obtain the bound 32.

Appendix B: Gaussian long memory sequences

For the sake of completeness, we recall in this appendix the main definitions and results pertaining to Hermite coefficients and expansions of square integrable functions with respect to a possibly non standard multivariate Gaussian distribution. Expansions with respect to the multivariate standard Gaussian distribution are easy to obtain and describe. The theory for non standard Gaussian vectors is more cumbersome. The main reference is Arcones (1994).

B.1 Hermite coefficients and rank

Let G be a function defined on $\mathbb{R}^k$ and $\mathbf{X} = (X^{(1)}, \ldots, X^{(k)})$ be a k-dimensional centered Gaussian vector with covariance matrix Γ. The Hermite coefficients of G with respect to X are defined as

$$J(G, \mathbf{X}, \mathbf{q}) = \mathbb{E}\left[G(\mathbf{X})\prod_{j=1}^{k} H_{q_j}(X^{(j)})\right],$$

where $\mathbf{q} = (q_1, \ldots, q_k) \in \mathbb{N}^k$. If Γ is the $k \times k$ identity matrix (denoted by I_k), i.e. the component of $\mathbf{X}$ are i.i.d. standard Gaussian, then the corresponding Hermite coefficients are denoted by $J^*(G, \mathbf{q})$. The Hermite rank of G with respect to $\mathbf{X}$, is the smallest integer τ such that

$$J(G, \mathbf{X}, \mathbf{q}) = 0 \text{ for all } \mathbf{q} \text{ such that } 0 < |q_1 + \cdots + q_k| < \tau.$$

B.2 Variance inequalities

Consider now a k-dimensional stationary centered Gaussian process $\{\mathbf{X}_i, i \geq 0\}$ with covariance function $\gamma_n(i, j) = \mathbb{E}[X_0^{(i)} X_n^{(j)}]$ and assume either

$$\forall 1 \leq i, j \leq k \ , \ \sum_{n=0}^{\infty} |\gamma_n(i, j)| < \infty, \tag{B.1}$$

or that there exists $H \in (1/2, 1)$ and a function ℓ slowly varying at infinity such that

$$\lim_{n \to \infty} \frac{\gamma_n(i, j)}{n^{2H-2}\ell(n)} = b_{i,j}, \tag{B.2}$$

and the $b_{i,j}$s are not identically zero. Denote then $\gamma_n = n^{2H-2}\ell(n)$. Then, we have the following inequality due to Arcones (1994).

For any function G such that $\mathbb{E}[G^2(\mathbf{X}_0)] < \infty$ and with Hermite rank q with respect to $\mathbf{X}_0$,

$$\operatorname{var}\left(n^{-1} \sum_{r=1}^{n} G(\mathbf{X}_j)\right) \leq C(\ell^q(n) n^{2q(H-1)}) \vee n^{-1} \operatorname{var}(G(\mathbf{X}_0)). \tag{B.3}$$

where the constant C depends only on the Gaussian process $\{\mathbf{X}_n\}$ and not on the function G. This bound summarizes Eqs. 2.18, 3.10 and 2.40 in Arcones (1994). The rate obtained is n^{-1} in the weakly dependent case where Eq. B.1 holds and in the case where Eq. B.2 holds and G has Hermite rank q such that $q(1 - H) > 1$. Otherwise, the rate is $\ell^q(n) n^{2q(H-1)}$.

Appendix C: A criterion for tightness

We state a criterion for the tightness of a sequence of random processes with path in $\mathcal{D}(\mathbb{R}^d)$, which adapts to the present context Bickel and Wichura (1971, Theorem 3) and the remarks thereafter.

Let T be a rectangle $T = T_1 \times T_d \subset \mathbb{R}^d$. A block B in T is a subset of T of the form $\prod_{i=1}^d (s_i, t_i]$ with $s_i < t_i$, $1 \leq i \leq d$. Disjoint blocks $B = \prod_{i=1}^d (s_i, t_i]$ and $B' = \prod_{i=1}^d (s_i', t_i']$ are neighbours if there exists $p \in \{1, \ldots, d\}$ such that $s_p' = t_p$ or $s_p = t_p'$ and $s_i = s_i'$ and $t_i = t_i'$ for $i \neq p$. (In the terminology of Bickel and Wichura

(1971) the blocks B and B' are said to share a common face.) Let X be a random process indexed by T. The increment of the process X over a block $B = \prod_{i=1}^{d}(s_i, t_i]$ is defined by

$$X(B) = \sum_{(\epsilon_1,\dots,\epsilon_d)\in\{0,1\}^d} (-1)^{d-\sum_{i=1}^{d}\epsilon_i} X(s_1 + \epsilon_1(t_1 - s_1), \dots, s_d + \epsilon_d(t_d - s_d)).$$

(This is the usual d-dimensional increment of a random process X. If for instance $d = 2$, then $X(B) = X(t_1, t_2) - X(t_1, s_2) - X(s_1, t_2) + X(s_1, s_2)$). If X is an indicator, i.e. $X(\mathbf{y}) = \mathbf{1}_{\{\mathbf{Y}\le\mathbf{y}\}}$ for some T valued random variable $\mathbf{Y}$, then $X(B) = \mathbf{1}_{\{\mathbf{Y}\in B\}}$.

Lemma C.1 *Let $\{\zeta_n\}$ be sequence of stochastic processes indexed by a compact rectangle $T \subset \mathbb{R}^d$. Assume that the finite dimensional marginal distributions of ζ_n converges weakly to those of a process ζ which is continuous on the upper boundary of T. Assume moreover that there exist $\gamma \ge 0$ and $\delta > 1$ such that*

$$\mathbb{P}(|\zeta_n(B)| \wedge |\zeta_n(B')| \ge \lambda) \le C\lambda^{-\gamma}\mathbb{E}[\mu_n^{\delta}(B \cup B')] \tag{C.1}$$

for some sequence of random probability measures μ_n which converges weakly in probability to a (possibly random) probability measure μ with (almost surely) continuous marginals. Then the sequence of processes $\{\zeta_n\}$ is tight in $\mathcal{D}(T, \mathbb{R})$.

Sketch of proof For f defined on $T = T_1 \times \cdots \times T_d$, $i \in \{1, \dots, d\}$ and $t \in T_i$, define $f_t^{(i)}$ on $T_1 \times \cdots \times T_{i-1} \times T_{i+1} \times \cdots \times T_d$ by

$$f_t^{(i)}(t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_d) = f(t_1, \dots, t_{i-1}, t, t_{i+1}, \dots, t_d)$$

and define, for $s < t \in T_i$ and $\delta > 0$,

$$w_i''(f, s, t) = \sup_{s<u<v<w<t} \|f_u^{(i)} - f_v^{(i)}\|_\infty \wedge \|f_v^{(i)} - f_w^{(i)}\|_\infty,$$
$$w_i''(f, \delta) = \sup_{u<v<w<u+\delta} \|f_u^{(i)} - f_v^{(i)}\|_\infty \wedge \|f_v^{(i)} - f_w^{(i)}\|_\infty.$$

By the corollary of Bickel and Wichura (1971), a sequence of processes $\{X_n\}$ defined on T converges weakly in $\mathcal{D}(T)$ to a process X which is continuous at the upper boundary of T with probability one, if the finite-dimensional marginal distributions of X_n converges to those of X and if, for all $\delta, \lambda > 0$, and al $i = 1, \dots, d$,

$$\mathbb{P}(w_i''(X_n, \delta) > \lambda) \to 0. \tag{C.2}$$

For any measure μ on T, define its i-th marginal $\mu^{(i)}$ by

$$\mu^{(i)}((s, t]) = \mu(T_1 \times \cdots \times T_{i-1} \times (s, t] \times T_{i+1} \times \cdots \times T_d), s, t \in T_i.$$

As mentioned in the remarks after the proof of Bickel and Wichura (1971, Theorem 3), an easy adaptation of the proof of Billingsley (1968, Theorem 15.6) shows that Eq. C.2 is implied by

$$\mathbb{P}(w_i''(X_n, s, t) > \lambda) \leq C\lambda^{-\gamma}\mathbb{E}[\{\mu_n^{(i)}(s, t])\}^{\delta}], \tag{C.3}$$

where μ_n satisfies the assumptions of the Lemma. So we must show that Eq. C.1 implies Eq. C.3. The proof is by induction, so the first step is to prove it in the one-dimensional case, where Eq. C.1 becomes, for $u < v < w \in T$,

$$\mathbb{P}(|\zeta_n(v) - \zeta_n(u)| \wedge |\zeta_n(w) - \zeta_n(v)| \geq \lambda) \leq C\lambda^{-\gamma}\mathbb{E}[\mu_n^{\delta}((u, w])]. \tag{C.4}$$

The proof of Eq. C.3 under the assumption C.4 follows the lines of the proof of Billingsley (1968, (15.26)) under the assumption (Billingsley 1968, (15.21)). The key ingredient is the maximal inequality (Billingsley 1968, Theorem 12.5), which can be easily adapted as follows in the present context. Let $S_0, \dots, S_n$ be random variables. Assume that there exists nonnegative random variables $u_1, \dots, u_n$ such that

$$\mathbb{P}(|S_i - S_j| \wedge |S_k - S_j| > \lambda) \leq \lambda^{-\gamma}\mathbb{E}[(u_i + \cdots + u_k)^{\delta}]$$

for some $\delta > 1$ and $\gamma \geq 0$ and all $1 \leq i \leq j \leq k \leq n$ and, then there exists a constant C that depends only on δ and γ such that

$$\mathbb{P}\left(\max_{1\leq i\leq j\leq k\leq n} |S_i - S_j| \wedge |S_k - S_j| > \lambda\right) \leq C\lambda^{-\gamma}\mathbb{E}[(u_1 + \cdots + u_n)^{\delta}].$$

Proving by induction that Eq. C.1 implies Eq. C.3 in the d-dimensional case can be done exactly along the lines of Step 5 of the proof of Bickel and Wichura (1971, Theorem 1). □

In order to apply this criterion to the context of empirical processes, we need the following lemma which slightly extends the bound Billingsley (1968, (13.18)).

Lemma C.2 *Let $\{(B_i, B_i')\}$ be a sequence of m-dependent vectors, where B_i and B_i' are Bernoulli random variables, with parameters p_i and q_i, respectively, and such that $B_iB_i' = 0$ a.s. Denote $S_n = \sum_{r=1}^n (B_j - p_j)$ and $S_n' = \sum_{r=1}^n (B_j' - q_j)$. Then, there exists a constant C which depends only on m, such that*

$$\mathbb{E}[S_n^2 S_n'^2] \leq C\left(\sum_{i=1}^n p_i\right)\left(\sum_{i=1}^n q_i\right) \leq C\left(\sum_{i=1}^n p_i \vee q_i\right)^2. \tag{C.5}$$

Proof We start by assuming that the pairs (B_i, B_i') are i.i.d. and we prove Eq. C.5 by induction. For any integrable random variable X, denote $\bar{X} = X - \mathbb{E}[X]$. For $n = 1$, since $B_1B_1' = 0$, we obtain $\mathbb{E}[\bar{B}_i\bar{B}_i'] = -p_iq_i$ and

$$\begin{aligned}\mathbb{E}[\bar{B}_1^2\bar{B}'^2_1] &= \mathbb{E}[(B_1 - 2p_1B_1 + p_1^2)(B_1' - 2q_1B_1' + q_1^2)] \\ &= p_1q^2 + p_1^2q_1 - 3p_1^2q^2 = p_1q_1(p_1 + q_1 - 3p_1q_1) \leq p_1q_1.\end{aligned}$$

The last inequality comes from the fact that $B_1 B_1' = 0$ a.s. implies that $p_i + q_i \leq 1$, and $0 \leq p + q - 3pq \leq p + q \leq 1$ for all $p, q \geq 0$ such that $p + q \leq 1$. Assume now that Eq. C.5 holds with $C = 3$ for some $n \geq 1$. Then, denoting $s_n = \sum_{r=1}^n p_j$ and $s_n' = \sum_{r=1}^n q_j$, we have

$$\begin{aligned}\mathbb{E}[S_{n+1}^2 {S'_{n+1}}^2] &= \mathbb{E}[S_n^2 {S'_n}^2] + \mathbb{E}[S_n^2]\mathbb{E}[{\bar{B}'}^2_{n+1}] + \mathbb{E}[{S'_n}^2]\mathbb{E}[\bar{B}_{n+1}^2] \\ &\quad + 4\mathbb{E}[S_n S_n']\mathbb{E}[B_{n+1} B_{n+1}'] + \mathbb{E}[\bar{B}_{n+1}^2 {\bar{B}'}^2_{n+1}] \\ &\leq 3 s_n s_n' + s_n q_{n+1} + s_n' p_{n+1} + 4 p_{n+1} q_{n+1} \sum_{i=1}^n p_i q_i + p_{n+1} q_{n+1} \\ &\leq 3 s_n s_n' + 3 s_n q_{n+1} + 3 s_n' p_{n+1} + p_{n+1} q_{n+1} \leq 3 s_{n+1} s_{n+1}'.\end{aligned}$$

This proves that Eq. C.5 holds for al $n \geq 1$.

We now consider the case of m-dependence. Let a_i, $1 \leq i \leq n$ be a sequence of real numbers and set $a_i = 0$ if $i > n$. Then

$$\left(\sum_{i=1}^n a_i\right)^2 = \left(\sum_{q=1}^m \sum_{r=1}^{\lceil n/m \rceil} a_{(j-1)m+q}\right)^2 \leq m \sum_{q=1}^m \left(\sum_{r=1}^{\lceil n/m \rceil} a_{(j-1)m+q}\right)^2 .$$

Applying this and the bound for the independent case (extending all sequences by zero after the index n) yields

$$\mathbb{E}[S_n^2 {S'_n}^2] \leq 3m^2 \sum_{q=1}^m \sum_{q'=1}^m \sum_{r=1}^{\lceil n/m \rceil} \sum_{j'=1}^{\lceil n/m \rceil} p_{(j-1)m+q} p_{(j'-1)m+q'} = 3m^2 s_n s_n'.$$

□

Let us apply this criterion in the context of Section 3. Fix a cone $\mathbf{j}$ and a relatively compact subset $A \in \mathbf{j}$. Recall that $E_{n,1}$ and $E_{n,2}$ are defined in Eqs. 46 and 47.

Lemma C.3 *Under the assumptions of Theorem 5 or 7, for any fixed $B \in \mathbb{R}^{h'+1}$, $E_{n,1}(B, \cdot)$ is tight in $\mathcal{D}([a, b])$, and if moreover $\Psi_{A,m,h}$ is continuous, then $E_{n,1}$ is tight in $\mathcal{D}(\mathcal{K} \times [a, b])$ for any $0 < a < b$ and any compact set $\mathcal{K}$ of $\mathbb{R}^{h'+1}$.*

Proof Since A is a cone, if $s < t$, then $tA \subset sA$. Thus, a sequence of random measures $\hat{\mu}_n$ on $\mathbb{R}^d \times (0, \infty)$ can be defined by

$$\begin{aligned}\hat{\mu}_n((-\infty, \mathbf{y}] \times (s, \infty)) &= \frac{1}{n} \sum_{r=1}^n \frac{\mathbb{P}(\mathbf{Y}_{j,h} \in s u_n A \mid \mathcal{X})}{g(k/n)} \mathbb{P}(\mathbf{Y}_{r+m,r+m+h'} \leq \mathbf{y} \mid \mathcal{X}) \\ &= \frac{1}{n} \sum_{r=1}^n G_n(A, B, s, \mathbf{X}_{j,h} \mathbf{X}_{r+m,r+m+h'}, \mathbf{y}),\end{aligned}$$

where G_n is defined in Eq. 25. Then $\hat{\mu}_n$ converges vaguely in probability to the measure μ defined by

$$\mu((-\infty, \mathbf{y}] \times (s, \infty)) = \mu(A)T(s)\Psi_{A,m,h}(\mathbf{y}).$$

Then, by conditional m-dependence, for any neighbouring relatively compact blocs D, D' of $\mathbb{R}^d \times (0, \infty]$, applying Lemma C.2 yields

$$\mathbb{E}[E_{n,1}^2(D)E_{n,2}^2(D') \mid \mathcal{X}] \le C\hat{\mu}_n(D)\hat{\mu}_n(D').$$

Taking unconditional expectations then yields

$$\mathbb{E}[E_{n,2}^2(D)E_{n,2}^2(D')] \le C\hat{\mathbb{E}}[\mu_n(D)\hat{\mu}_n(D')] \le \mathbb{E}[\hat{\mu}_n^2(D \cup D')].$$

Thus Eq. C.1 holds with $\delta = \gamma = 2$. In the context of Theorem 5, for any fixed B, this implies that for each B, the sequence of processes $E_{n,1}(B, \cdot)$ is tight, since the limiting distribution is proportional to $T(s)$ which is continuous. If the distribution function Ψ is assumed to be continuous, then Lemma C.1 applies and the process $E_{n,1}$ is tight with respect to both variables. □

Lemma C.4 *Under the assumptions of Theorem* 5*, for any fixed* $B \in \mathbb{R}^{h'+1}$*,* $E_{n,2}(B, \dot{)}$ *converges uniformly to zero on compact sets of* $(0, \infty]$*. Under the assumption of Corollary* 6*,* $E_{n,2}$ *converges uniformly to zero on compact sets of* $\mathbb{R}^{h'+1} \times (0, \infty]$.

Proof We only need to prove the tightness. By the variance inequality B.3 and Hölder's inequality, we have, for any relatively compact neighbouring blocks D, D' of $\mathbb{R}^d \times (0, \infty)$,

$$\begin{aligned}\mathbb{P}(|E_{2,n}(D)| \wedge |E_{2,n}(D')| \ge \lambda) &\le \lambda^{-2}\sqrt{\mathbb{E}[E_{2,n}^2(D)]\mathbb{E}[E_{2,n}^2(D')]} \\ &\le \lambda^{-2}\mathbb{E}[E_{2,n}^2(D \cup D')] \\ &\le C\lambda^{-2}n^{-1}\mathbb{E}[\tilde{\mu}_n^2(D \cup D')]\end{aligned}$$

where $\tilde{\mu}_n$ is the random measure defined by

$$\tilde{\mu}_n(\mathbf{y}, s) = \frac{\mathbb{P}(\mathbf{Y}_{1,h} \in su_nA \mid \mathcal{X})}{g(k/n)}\mathbb{P}(\mathbf{Y}_{m,m+h'} \le \mathbf{y} \mid \mathcal{X}).$$

The sequence $\tilde{\mu}_n$ converges vaguely on $\mathbb{R}^d \times (0, \infty]$, in probability and in the mean square to the measure $\hat{\mu}$ defined by

$$\hat{\mu}((-\infty, \mathbf{y}] \times (s, \infty]) = \frac{\nu_{\mathbf{j}}(\boldsymbol{\sigma}(\mathbf{X}_{1,h})^{-1} \cdot A)}{(\mathbb{E}[\nu_C(\boldsymbol{\sigma}(\mathbf{X}_{1,h})^{-1} \cdot A])^{\delta}}T(s)\mathbb{P}(\mathbf{Y}_{m,m+h'} \le \mathbf{y} \mid \mathcal{X}).$$

The measure $\hat{\mu}$ has continuous marginals if we consider the case of a fixed B (which takes care of Theorem 7). The marginals of $\hat{\mu}$ are almost surely continuous if F_Z is continuous, so Lemma C.1 applies. □

References

Arcones, M.A.: Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. Ann. Probab. **22**(4), 2242–2274 (1994)

Bickel, P.J., Wichura, M.J.: Convergence criteria for multiparameter stochastic processes and some applications. Ann. Math. Stat. **42**, 1656–1670 (1971)

Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)

Bingham, N.H., Goldie, C.M., Teugels, J.L.: Regular variation. In: Encyclopedia of Mathematics and its Applications, vol. 27. Cambridge University Press, Cambridge (1989)

Breidt, F.J., Crato, N., de Lima, P.: The detection and estimation of long memory in stochastic volatility. J. Econom. **83**(1–2), 325–348 (1998)

Breiman, L.: On some limit theorems similar to the arc-sine law. Theory Probab. Appl. **10**, 323–331 (1965)

Das, B., Resnick, S.I.: Conditioning on an extreme component: model consistency with regular variation on cones. Bernoulli **17**(1), 226–252 (2011)

Davis, R.A., Mikosch, T.: The extremogram: a correlogram for extreme events. Bernoulli **38A**, 977–1009 (2009). Probability, statistics and seismology

Drees, H.: Optimal rates of convergence for estimates of the extreme value index. Ann. Stat. **26**(1), 434–448 (1998)

Harvey, A.C.: Long memory in stochastic volatility. In: Knight, J., Satchell, S. (eds.) Forecasting Volatility in Financial Markets. Butterworth-Heinemann, London (1998)

Hurvich, C.M., Moulines, E., Soulier, P.: Estimating long memory in volatility. Econometrica **73**(4), 1283–1328 (2005)

Kulik, R., Soulier, P.: The tail empirical process for long memory stochastic volatility sequences. Stoch. Process. Their Appl. **121**(1), 109–134 (2011)

Mitra, A., Resnick, S.I.: Hidden regular variation and detection of hidden risks. Stoch. Models **27**, 591–614 (2011)

Omey, E., Willekens, E.: Second-order behaviour of distributions subordinate to a distribution with finite mean. Commun. Stat. Stoch. Models **3**(3), 311–342 (1987)

Orey, S.: A central limit theorem for m-dependent random variables. Duke Math. J. **25**, 543–546 (1958)

Resnick, S.I.: Heavy-tail phenomena. In: Springer Series in Operations Research and Financial Engineering. Probabilistic and statistical modeling. Springer, New York (2007)

Resnick, S.I.: Multivariate regular variation on cones: application to extreme values, hidden regular variation and conditioned limit laws. Stochastics **80**(2–3), 269–298 (2008)