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

Rafał Kulik\textsuperscript{*} Philipp Soulier\textsuperscript{†}

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).

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 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

\[
Y_j = \sigma(X_j)Z_j
\]

(1)

where \( \sigma \) 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

\[
\sum_{n=1}^{\infty} |\gamma_n| < \infty ,
\]

(2)

\textsuperscript{*}University of Ottawa
\textsuperscript{†}Université de Paris-Ouest
or that it has long memory with Hurst index \( H \in (1/2, 1) \), i.e.
\[
\gamma_n = \text{cov}(X_0, X_n) = n^{2H-2}\ell(n) \tag{3}
\]
where \( \ell \) 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 \( \alpha \) degrees of freedom, and the Pareto distribution with index \( \alpha \).

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 \( \alpha \) 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 \( a(t) = F_Z^{-1}(1-1/t) \) and \( F_Z^{-1} \) 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
\[
\lim_{t \to \infty} \mathbb{P}(Y_m \leq y \mid Y_0 > t) = \frac{\mathbb{E}[\sigma^\alpha(X_0)F_Z(y/\sigma(X_m))] \mathbb{E}[\sigma^\alpha(X_0)]}{\mathbb{E}[\sigma^\alpha(X_0)]}. \tag{8}
\]
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 Figure 1 estimates of the standard distribution function and of the conditional distribution for a simulated SV process. Clearly, the two estimated distributions are different, as suggested by (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 Equation (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:

$$
\rho(A, B, m) = \lim_{t \to \infty} \mathbb{P}((Y_{m}, \ldots, Y_{m+h'}) \in B \mid (Y_{1}, \ldots, Y_{h}) \in tA).
$$

(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 $B$:

$$
\lim_{t \to \infty} \mathbb{P}((Y_{m}, \ldots, Y_{m+h'}) \in tB \mid (Y_{1}, \ldots, Y_{h}) \in tA) = 0.
$$
The general aim of this paper is to investigate the existence of these limiting conditional distributions appearing in (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 →p, weak convergences of sequences of random variables or vectors by →d and weak convergence in the Skorokhod space D(\(\mathbb{R}^q\)) of cadlag functions defined on \(\mathbb{R}^q\) endowed with the \(J_1\) topology by ⇒.

Boldface letters denote vectors. Product of vectors and inequalities between vectors are taken componentwise: \(\mathbf{u} \cdot \mathbf{v} = (u_1v_1, \ldots, u_qv_q)\); \(\mathbf{x} \leq \mathbf{y}\) 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 \(h \leq h'\), let \(\xi_{h,h'}\) denote the \((h' - h + 1)\)-dimensional vector \((\xi_h, \ldots, \xi_{h'})\).

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

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 \(\sigma\)-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 \(\nu\) on the Borel sets of \(\mathbb{R}^h \setminus \{0\}\) 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 \(0\). A sequence of Radon measures \(\nu_n\) on \(\mathbb{R}^h\) is said to converge vaguely to a Radon measure \(\nu\), which will be denoted by \(\nu_n \rightarrow \nu\) if \(\nu_n(A) \rightarrow \nu(A)\) for all relatively compact set \(A\) of \(\mathbb{R}^k \setminus \{0\}\). Recall that \(a(t) = \frac{F_Z}{2}(1 - 1/t)\).
Proposition 1. The finite dimensional distributions of the process \( \{Y_j, j \in \mathbb{Z}\} \) are multivariate regularly varying and for each fixed integer \( h \)
\[
\lim_{t \to \infty} t \mathbb{P}( (Y_1, \ldots, Y_h) \in a(t) \cdot ) \to \alpha \mathbb{E}[\sigma^\alpha(X_1)] \nu(\cdot)
\]
where the measure \( \nu \) is characterized by
\[
\nu([x,y]^C) = (1-\beta) \sum_{i=1}^{h} (-x_i)^{-\alpha} + \beta \sum_{i=1}^{h} y_i^{-\alpha},
\]
for \( x = (x_1, \ldots, x_h) \in [-\infty,0)^h \) and \( y = (y_1, \ldots, y_h) \in (0,\infty]^h \), and \( \beta \) is defined in (4).

The special form of the measure \( \nu \) 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 \( C \) of \([0,\infty]^h\) such that \( tx \in C \) for all \( x \in C \) and \( t > 0 \). Hence, our discussion in this section is related to the concept of regular variation on cones or hidden regular variation (see Resnick (2008), Das and Resnick (2011), Mitra and Resnick (2011)). 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 \( C \) such that there exists an integer \( \beta_C \) and a Radon measure \( \nu_C \) on \( C \) such that, for all relatively compact subsets \( A \) of \( C \) with \( \nu_C(\partial A) = 0, \)
\[
\lim_{t \to \infty} \mathbb{P}( (Z_1, \ldots, Z_h) \in tA) / (F_Z(t))^{\beta_C} = \nu_C(A).
\]

Intuitively, the number \( \beta_C \) corresponds to the number of components of a point of a relatively compact subset \( A \) of the cone \( C \) 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 \( P_1, \ldots, P_k \) be nonempty subsets of \( \{1, \ldots, h\} \) 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 \( C \) be the cone defined by
\[
C = \left\{ z \in [0,\infty]^h \mid \prod_{u=1}^{k} \left( \sum_{i \in P_u} z_i \right) > 0 \right\}
\]

In words, a vector \( z = (z_1, \ldots, z_h) \) belongs to \( C \) 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 \( C \) is easily characterized: a subset \( A \) is relatively compact in \( C \) if and only if there exists \( \eta > 0 \) such that \( \sum_{i \in P_u} z_i > \eta \) for all \( u = 1, \ldots, k \). Examples of such cones are \( C_1 = \{ z_1 > 0, z_2 > 0, z_3+z_4 > 0 \} \)
in $[0, \infty]^4$ and $C_2 = \{z_1 + z_2 > 0, z_2 + z_3 > 0, z_3 + z_4 > 0, z_4 + z_5 > 0\}$ in $[0, \infty]^5$; for $a, b, c > 0$, \{z_1 > a, z_2 > b, z_3 > c\} is relatively compact in $C_1$ and \{z_2 > a, z_4 > b\} is relatively compact in $C_2$. More detailed examples will be given in Section 2.1.

**Proposition 2.** Assume that there exists $\epsilon > 0$ such that
\[
E[\sigma^{2h\alpha+}(X_0)] < \infty .
\] (13)

Let $C$ be one of the cones defined in (12). Then there exists an integer $\beta_C$ and a Radon measure $\nu_C$ on $C$ such that (11) holds and for all relatively compact sets $A \subset C$ with $\nu_C(\partial A) = 0$, for $m > h$ and $h' \geq 0$, and for any Borel measurable set $B \subset \mathbb{R}^{h'+1}$, we have
\[
\lim_{t \to +\infty} \frac{\mathbb{P}(Y_{1,h} \in tA, Y_{m,m+h'} \in B)}{(F_Z(t))^{\beta_C}} = \mathbb{E} \left[ \nu_C(\sigma(X_{1,h})^{-1} \cdot A) \mathbb{P}(Y_{m,m+h'} \in B \mid \mathcal{X}) \right] .
\] (14)

Furthermore, for $r = 1, \ldots, h$, there exist functions $\mathcal{L}_r$ such that for all $s, s' \geq 1$, $u, v \in (0, \infty)^h$,
\[
\lim_{t \to +\infty} \frac{\mathbb{P}(u \cdot Z_{1,h} \in tsA, v \cdot Z_{r,r+h-1} \in ts'A)}{(F_Z(t))^{\beta_C}} = \mathcal{L}_r(A, u, v, s, s') .
\] (15)

Some comments are in order. Note first that we assume that $h < m$. Otherwise, if $m < h$, then vectors $Y_{m,m+h'}$ and $Y_{1,h}$ may be asymptotically dependent. For example, if $\{Z_t\}$ is i.i.d with the tail distribution as in (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 primarily interested in estimating distribution of future vector $Y_{m,m+h'}$ based on the past observations $Y_{1,h}$. It is easily seen that the coefficient $\beta_C$ is the smallest integer $\ell$ for which there exists $i_1, \ldots, i_\ell \in \{1, \ldots, h\}$ such that $z_{i_1} > 0, \ldots, z_{i_\ell} > 0$ implies $\prod_{u=1}^\ell (\sum_{i \in P_u} z_i) > 0$. The measure $\nu_C$ is cumbersome to write precisely in general, but is easily obtained in each example. See Eq. (33) in the proof of Proposition 2. The condition (13) obviously holds if $\sigma(x) = e^x$ or if $\sigma$ is a polynomial. For $B = \mathbb{R}^{h'+1}$, (14) specializes to
\[
\lim_{t \to +\infty} \frac{\mathbb{P}(\sigma(X_{1,h}) \cdot Z_{1,h} \in tA)}{(F_Z(t))^{\beta_C}} = \mathbb{E}[\nu_C(\sigma(X_{1,h})^{-1} \cdot A)] .
\]

If $\nu_C(A) > 0$, then $\mathbb{E}[\nu_C(\sigma(X_{1,h})^{-1} \cdot A)] > 0$ and (14) implies that the extremogram defined in (9) can be expressed as
\[
\rho(A, B, m) = \frac{\mathbb{E}[\nu_C(\sigma(X_{1,h})^{-1} A) \mathbb{P}(Y_{m,m+h'} \in B \mid \mathcal{X})]}{\mathbb{E}[\nu_C(\sigma(X_{1,h})^{-1} \cdot A)]} .
\] (16)

It may happen that $\mathcal{L}_r(A, \cdot) \equiv 0$ for $r = 2, \ldots, h$. Intuitively, this happens if $u \cdot Z_{1,h}$ and $v \cdot Z_{k,k+h-1}$ belonging simultaneously to $tA$ implies that at least $\beta_C + 1$ coordinates of $Z_{1,h+k-1}$ 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 $C$. 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 $[0, \infty]^h \setminus \{0\}$. If there exists two cones $C$ and $C'$ such that (11) hold, $A$ is relatively compact in both $C$ and $C'$, $\nu_C(A) > 0$ and $\nu_{C'}(A) > 0$, then $\beta_C = \beta_{C'}$ and for all $u \in (0, \infty)^h$, $\nu_C(u \cdot A) = \nu_{C'}(u \cdot A)$. 

6
2.1 Examples

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

\[
\nu_{C}(dz_{1}, \ldots, dz_{h}) = \alpha^{h} \prod_{i=1}^{h} z_{i}^{-\alpha - 1} dz_{i}.
\]

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 (13) holds, then for \( m > h \), and \( B \in \mathbb{R}^{h + 1} \), Proposition 2 yields

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

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

\[
\Psi_{h}(y) = \lim_{t \to \infty} \mathbb{P}(Y_{m} \leq y \mid Y_{1} > t, \ldots, Y_{h} > t) = \frac{\mathbb{E}[\prod_{i=1}^{h} \sigma^{\alpha}(X_{i})F_{Z}(y/\sigma(X_{m}))]}{\mathbb{E}[\prod_{i=1}^{h} \sigma^{\alpha}(X_{i})]}.
\] (17)

Finally, note that the function \( \mathcal{L}_{r} \) defined in (15) vanish for \( r = 2, \ldots, h \).

Example 2. Consider now \( C = (0, \infty] \). 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 \( m > 1 \),

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

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

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

\[
\nu_{C}(dz) = \alpha \{z_{1}^{-\alpha - 1} d_{1}\delta_{0}(dz_{2}) + \delta_{0}(dz_{1})z_{2}^{-\alpha - 1} d_{z_{2}}\},
\]

where \( \delta_{0} \) 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

\[
\lim_{t \to \infty} \mathbb{P}(Y_{m,m+h'} \in B \mid Y_{1} + Y_{2} > t) = \frac{\mathbb{E} \left[ \mathbb{P}(Y_{m,m+h'} \in B \mid \mathcal{X})(\sigma^{\alpha}(X_{1}) + \sigma^{\alpha}(X_{2})) \right]}{\mathbb{E}[\sigma^{\alpha}(X_{1})] + \mathbb{E}[\sigma^{\alpha}(X_{2})]}.
\]

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

\[
\Lambda(y) = \lim_{t \to \infty} \mathbb{P}(Y_{m} \leq y \mid Y_{1} + Y_{2} > t) = \frac{\mathbb{E}[\{\sigma^{\alpha}(X_{1}) + \sigma^{\alpha}(X_{2})\}F_{Z}(y/\sigma(X_{m}))]}{\mathbb{E}[\sigma^{\alpha}(X_{1})] + \mathbb{E}[\sigma^{\alpha}(X_{2})]}.
\]

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 $C = ([0, \infty)^2 \setminus \{0\}) \times ([0, \infty)^2 \setminus \{0\}) \times (0, \infty)$. Then $\beta_C = 2$ and
\[
\nu_C(dz) = \alpha^3 \{z_1^{\alpha-1}d_1 \delta_0(d_2) + \delta_0(d_1)z_2^{\alpha-1}d_2\} \\
\times \{z_3^{\alpha-1}d_3 \delta_0(d_4) + \delta_0(d_3)z_4^{\alpha-1}d_4\}z_5^{\alpha-1}d_5.
\]
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 $\mathbb{E}[\sigma^{3+t}(X_0)] < \infty$ for some $\epsilon > 0$, then we obtain, for $m > 3$,
\[
\lim_{t \to \infty} \mathbb{P}(\mathbf{Y}_{m,m+h'} \in B \mid Y_1 + Y_2 > t, Y_3 + Y_4 > t, Y_5 > t) = \frac{\mathbb{E}[\mathbb{P}(\mathbf{Y}_{m,m+h'} \in B \mid \mathbf{X}) \{\sigma^\alpha(X_1) + \sigma^\alpha(X_2)\} \{\sigma^\alpha(X_3) + \sigma^\alpha(X_4)\} \sigma^\alpha(X_5)\]}{\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 $L_r$ vanish for $r \geq 2$.

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 $C_1 = ([0, \infty)^2 \setminus \{0\}) \times ([0, \infty)^2 \setminus \{0\})$, $C_2 = (0, \infty) \times ([0, \infty)^3 \setminus \{0\})$ and $C_3 = [0, \infty)^4 \setminus \{0\}$. Then it is easily seen that $\beta_{C_1} = \beta_{C_2} = 2$ and $\beta_{C_3} = 1$, and for all $u \in ([0, \infty)^4$, $\nu_{C_1}(u^{-1} \cdot A) = \nu_{C_2}(u^{-1} \cdot A) = u_1^\alpha (u_3^a + u_4^a) a^{-\alpha} b^{-\alpha}$ and $\nu_{C_3}(A) = 0$.

## 3 Estimation

Let $C$ be a cone defined in (12) and let $A$ be a relatively compact subset of $C$ such that $\nu_C(A) > 0$. To simplify the notation, assume that we observe $Y_1, \ldots, Y_{n+m+h'}$. An estimator $\hat{\rho}_n(A, B, m)$ is naturally defined by
\[
\hat{\rho}_n(A, B, m) = \frac{\sum_{r=1}^n 1_{Y_{r+h} \in Y_{(n+m-k)}} 1_{Y_{m+m+h'} \in B}}{\sum_{r=1}^n 1_{Y_{r+h} \in Y_{(n+m-k)}}},
\]
where $k$ is a user chosen threshold and $Y_{(n,1)} \leq \cdots \leq Y_{(n,n)}$ are the increasing order statistics of the observations $Y_1, \ldots, Y_n$. We will also consider the case $B = (-\infty, y]$, i.e. the case of the limiting conditional distribution of $\mathbf{Y}_{m,m+h'}$ given $Y_{1,h} \in tA$, that means
\[
\Psi_{A,m,h'}(\mathbf{y}) = \lim_{t \to \infty} \mathbb{P}(\mathbf{Y}_{m,m+h'} \leq \mathbf{y} \mid Y_{1,h} \in tA) = \rho(A, \infty, \mathbf{y}, m) = \frac{\mathbb{E}[\nu_C(\sigma(X_{1+h})^{-1} \cdot A) \prod_{i=1}^{h'} F_{Y_i/\sigma(X_{m+i})}] \mathbb{E}[\nu_C(\sigma(X_{1,h})^{-1} \cdot A)]}{\mathbb{E}[\nu_C(\sigma(X_{1,h})^{-1} \cdot A)]}.
\]  
(18)

An estimator $\hat{\Psi}_{n,A,m,h'}$ of $\Psi_{A,m,h'}$ is defined on $\mathbb{R}^{h'+1}$ by
\[
\hat{\Psi}_{n,A,m,h'}(\mathbf{y}) = \frac{\sum_{r=1}^n 1_{Y_{r+h} \in Y_{(n+m-k)}} 1_{Y_{m+m+h'} \leq \mathbf{y}}}{\sum_{r=1}^n 1_{Y_{r+h} \in Y_{(n+m-k)}}}.
\]  
(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 $u_n = (1/F_Y)^{\alpha}(n/k)$. Consider the measure defined on $C$ by
\[
\mu_C(A) = \frac{\mathbb{E}[\nu_C(\sigma(X_{1,h})^{-1} \cdot A)]}{(\mathbb{E}[\sigma^\alpha(X)])^\beta_C}.
\]  
(20)
We introduce a rate of convergence:
\[
\nu_n(A) = \mathbb{E} \left[ \sup_{s \geq 1} \frac{\mathbb{P}(Y_{1,h} \in u_n s A \mid X)}{(k/n)^{3c}} - s^{-n^{3c} \mu_c(A)} \right].
\]  
\tag{21}

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

We need also the following quantities, which are well defined when (13) holds. For \( r = 2, \ldots, h \) and measurable subsets \( B, B' \) of \( \mathbb{R}^{h'+1} \), define
\[
\mathcal{R}_r(A, B, B') = \frac{\mathbb{E} \left[ L(A, \sigma(X_{1,h})), \sigma(X_{k,k+h-1}), 1, 1 \right] \times \mathbb{P}(Y_{m,m+h'} \in B, Y_{m+k-1,m+h'+k-1} \in B' \mid X)}{\mu_c(A)}
+ \frac{\mathbb{E} \left[ L(A, \sigma(X_{1,h})), \sigma(X_{k,k+h-1}), 1, 1 \right] \times \mathbb{P}(Y_{m,m+h'} \in B', Y_{m+k-1,m+h'+k-1} \in B \mid X)}{\mu_c(A)}.
\]
\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 summa-

bility (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 assumptions is satisfied by all reasonable examples.

**Theorem 5.** Let (2), (4), (13) hold. Assume moreover that \( A \) is a relatively compact subcone of \( \mathcal{C} \) such that \( \mu_c(A) > 0 \), that \( k/n \to 0 \), \( n(k/n)^{3c} \to \infty \) and
\[
\lim_{n \to \infty} n(k/n)^{3c} \nu_n(A) = 0.
\]  
\tag{23}

Then
\[
\sqrt{n(k/n)^{3c} \mu_c(A)} \{ \rho_n(A, B, m) - \rho(A, B, m) \}
\]
converges weakly to a centered Gaussian distribution with variance
\[
\rho(A, B, m) \{ 1 - \rho(A, B, m) \}
+ \sum_{r=2}^{h \wedge (m-h)} \{ \mathcal{R}_r(A, B) - 2 \rho(A, B, m) \mathcal{R}_r(A, B, \mathbb{R}^{h'+1}) + \rho^2(A, B, m) \mathcal{R}_r(A, \mathbb{R}^{h'+1}) \}. \tag{24}
\]

If \( h = 1 \) or if the functions \( L_r \) defined in (15) are identically zero for \( r \geq 2 \), then the limiting covariance in (24) is simply \( \rho(A, B, m) \{ 1 - \rho(A, B, m) \} \). Otherwise, the additional terms can be canceled by modifying the estimator of \( \rho_n(A, B, m) \). Assuming we have \( nh + m + h' + 1 \) observations, we can define
\[
\tilde{\rho}_n(A, B, m) = \frac{\sum_{r=1}^{n} 1_{\{Y_{(r-1)h+1,rh} \in Y(n,n-k)A\}} 1_{\{Y_{(r-1)h+1,m+(r-1)h+k+h'} \in B\}}}{\sum_{r=1}^{n} 1_{\{Y_{(r-1)h+1,rh} \in Y(n,n-k)A\}}}
\]
Noting that the events \( \{Y_{r,r+h-1} \in A\} \) are \( h \)-dependent conditionally on \( X \), the proof of Theorem 5 can be easily adapted to show that the limiting variance of \( \sqrt{n(k/n)^{3C}\{\tilde{\rho}_n(A,B,m) - \rho(A,B,m)\}} \) is the same as in the case where \( L_r \equiv 0 \) for \( r = 2, \ldots, h \). 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 (19) and (18).

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

\[
\sqrt{n(k/n)^{3C}} \{\tilde{\Psi}_{n,A,m,h'} - \Psi_{A,m,h'}\}
\]

converges in \( D(\mathbb{R}^{h'+1}) \) to a Gaussian process. If \( h = 1 \) or if the functions \( L_r \) 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.

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 \( (x, x') \in \mathbb{R}^h \times \mathbb{R}^{h'+1} \) and \( s \geq 1 \) by

\[
G_n(A,B,s,x,x') = \frac{\mathbb{P}(\sigma(x) \cdot Z_{1,h} \in u_n s A)}{(k/n)^{3C}} \mathbb{P}(\sigma(x') \cdot Z_{m,m+h'} \in B)
\]

\[
G(A,B,x,x') = \lim_{n \to \infty} G_n(A,B,1,x,x') = \frac{\nu_C(\sigma(x)^{-1} A)}{\mathbb{E}[\sigma^\alpha(X_1)]^{3C}} \mathbb{P}(\sigma(x') \cdot Z_{m,m+h'} \in B).
\]

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

**Assumption 1.** For large \( n \), \( \inf_x \tau_n(A,B,s) = \tau(A,B) \) 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 \( \sigma \) 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 (3). Assume that \( A \) is a relatively compact subcone of \( C \) such that \( \nu_C(A) > 0 \). Let Assumption 1 and (13) hold, and \( k/n \to 0 \), \( n(k/n)^{3C} \to \infty \) and

\[
\lim_{n \to \infty} \left\{ n(k/n)^{3C} \wedge \gamma_n^{-(\tau(A,B)/2)} \right\} v_n(A) = 0.
\]
1. If \( n(k/n)^\beta \gamma_n^{\tau(A,B)} \rightarrow 0 \), then
\[
\sqrt{n(k/n)^\beta \mu(A)} \{ \hat{\rho}_n(A, B, m) - \rho(A, B, m) \}
\]
converges to a centered Gaussian distribution with variance given in (24).

2. If \( n(k/n)^\beta \gamma_n^{\tau(A,B)} \rightarrow \infty \), then \( \frac{\gamma_n^{\tau(A,B)}}{2} \{ \hat{\rho}_n(A, B, m) - \rho(A, B, m) \} \)
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 \( \nu \) 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 \( \tau^*(A) = \inf_{y \in \mathbb{R}^h+1} \tau(A, (\infty, y)) \).

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

- If \( n(k/n)^\beta \gamma_n^{\tau^*(A)} \rightarrow 0 \), then
\[
\sqrt{n(k/n)^\beta \mu(A)} \{ \hat{\Psi}_{n,A,m,h'} - \Psi_{A,m,h'} \}
\]
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 \( n(k/n)^\beta \gamma_n^{\tau^*(A)} \rightarrow \infty \), then \( \frac{\gamma_n^{\tau^*(A)}}{2} \{ \hat{\Psi}_{n,A,m,h'} - \Psi_{A,m,h'} \} \)
converges in \( \mathcal{D}((\infty, +\infty)^{h'+1}) \) to a process which can be expressed as \( J_{A,m,h'} \cdot \kappa \) where \( J_{A,m,h'} \) is a deterministic function and \( \kappa \) 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 \( \kappa \) 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 rate of convergence (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 \( \eta^* \) 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 $\eta$ such that for $z > 0$,

$$
\mathbb{P}(Z > z) = cz^{-\alpha} \exp \left( \int_1^z \frac{\eta(s)}{s} \, ds \right),
$$

$$
\exists C > 0 , \forall s \geq 0 , |\eta(s)| \leq C\eta^*(s).
$$

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 $E[\sigma^\alpha(X)] < \infty$. 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 $\sigma$ 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, (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 the two different possibility for the 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

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

with a user chosen $k$.

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

$$
v_n(A) = O(\eta^*(u_n)). \quad (28)
$$

**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

$$
\lim_{n \to \infty} (n(k/n)^h)^{1/2} \eta^*(u_n) = 0. \quad (29)
$$

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

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

weakly in $D((\infty, \infty))$, where $\mathbb{B}$ is the standard Brownian bridge.
3.3.2 Example 3 continued

Consider the estimation of

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

An estimator if defined by

\[ \hat{\Lambda}_n(y) = \frac{\sum_{r=1}^n 1\{Y_r + Y_{r+1} > Y_{(n:n-k)}\} 1\{Y_r \leq y\}}{\sum_{r=1}^n 1\{Y_r + Y_{r+1} > Y_{(n:n-k)}\}}. \]

As argued before, if (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 \( v_n(A) \):

\[ v_n(A) = O \left( \eta^*(u_n) + u_n^{-1} \int_0^{u_n} \hat{F}_Z(s) \, ds \right). \tag{30} \]

**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

\[ \lim_{n \to \infty} k^{1/2} \left( \eta^*(u_n) + u_n^{-1} \int_0^{u_n} \hat{F}_Z(s) \, ds \right) = 0. \]

In the weakly dependent case (2) or in the long memory case (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 \( \mathcal{D}((-, \infty)) \), where \( \mathbb{W} \) is a Gaussian process with covariance

\[ \text{cov}(\mathbb{W}(y), \mathbb{W}(y')) = \Lambda(y \wedge y') - 2\Lambda(y)\Lambda(y') \]

\[ + \frac{\mathbb{E}[\sigma^\alpha(X_2)\{F_Z(y/\sigma(X_m))F_Z(y'/\sigma(X_{m+1})) + F_Z(y/\sigma(X_m))F_Z(y'/\sigma(X_{m+1}))\}]}{\mathbb{E}[\sigma^\alpha(X_1) + \sigma^\alpha(X_2)]}. \]

**Remark.** If the estimator if modified by taking only every other observation,

\[ \hat{\Lambda}_n(y) = \frac{\sum_{r=1}^n 1\{Y_{2(r-1)+1} + Y_{2r} > Y_{(n:n-k)}\} 1\{Y_r \leq y\}}{\sum_{r=1}^n 1\{Y_{2(r-1)+1} + Y_{2r} > Y_{(n:n-k)}\}}. \]

then \( \sqrt{k}(\hat{\Lambda}_n - \Lambda) \) converges weakly to \( 2\mathbb{B} \circ \Lambda \) where \( \mathbb{B} \) is the standard Brownian bridge. Indeed, 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 (8)

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

\[
\lim_{t \to \infty} P(Y_m \leq y \mid Y_0 > t) = \frac{E[\sigma(0)F_Z(y/\sigma(Z_m))]}{E[\sigma(0)]}.
\] (31)

Proof. Conditioning on the sigma-field \( \mathcal{X} \) yields

\[
P(Y_m \leq y, Y_0 > t) = E[P(\sigma(0)Z_0 \leq y, \sigma(Z_m) > t) \mid \mathcal{X}] = E[P(\sigma(0)Z_0 \leq y \mid \mathcal{X})P(\sigma(Z_m) > t \mid \mathcal{X})]
\]

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

\[
\frac{F_Z(y/\sigma(0))F(t/\sigma(0))}{F(t)} \leq C(\sigma(0) + 1)^{\alpha + \epsilon}.
\]

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

\[
\lim_{t \to \infty} \frac{P(Y_m \leq y, Y_0 > t)}{F(t)} = E\left[\lim_{t \to \infty} \frac{F(y/\sigma(0))F(t/\sigma(0))}{F(t)}\right] = E[\sigma(0)F_Z(y/\sigma(Z_m))].
\]

Finally, noting that by (6) we have \( P(Y_0 > t) \sim E[\sigma(0)]F(t) \) as \( t \to \infty \) yields (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

\[
\lim_{t \to \infty} tP(a^{-1}(t)(u_1Z_1, \ldots, u_hZ_h) \in [x, y]^c)
\]

\[
= (1 - \beta) \sum_{i=1}^h u_i^\alpha |x_i|^{-\alpha} + \beta \sum_{i=1}^h u_i^\alpha y_i^{-\alpha} = \sum_{i=1}^h u_i^\alpha \nu_{\alpha,\beta}([x_i, y_i]^c),
\]

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

\[
\nu_{\alpha,\beta}(dx) = \alpha((1 - \beta)(-x)^{-\alpha - 1}1_{x<0}) + \beta x^{-\alpha - 1}1_{x>0}) \, dx.
\]

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

\[
tP(a^{-1}(t)(u_1Z_1, \ldots, u_kZ_k) \in [x, y]^c) \leq C \sum_{i=1}^k (u_i \vee 1)^{\alpha + \epsilon}.
\] (32)

Thus, we can apply the bounded convergence Theorem and obtain

\[
\lim_{t \to \infty} tP(a^{-1}(t)(\sigma(X_1)Z_1, \ldots, \sigma(X_k)Z_k) \in [x, y]^c)
\]

\[
= E[\lim_{t \to \infty} tP(a^{-1}(t)(\sigma(x_1)Z_1, \ldots, \sigma(X_k)Z_k) \in [x, y]^c \mid \mathcal{X})] = E[\sigma(0)] \sum_{i=1}^k \nu_{\alpha,\beta}([x_i, y_i]^c).
\]

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

$$\frac{\mathbb{P}(Z_{1,h} \in tA \mid Z \geq \eta) - \mathbb{P}(Z_{1,h} \in tA \mid Z \leq \eta)}{(\mathbb{F}(Z(t)))^{\beta_c}} \sim \frac{\mathbb{P}(Z_{1,h} \in tA \mid Z_{i_1} > \eta, Z_{i_2} > \eta, \ldots, Z_{i_{\beta_C}} > \eta) - \mathbb{P}(Z_{1,h} \in tA \mid Z_{i_1} < \eta, Z_{i_2} < \eta, \ldots, Z_{i_{\beta_C}} < \eta)}{(\mathbb{F}(Z(t)))^{\beta_c}}$$

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

$$\lim_{t \to \infty} \frac{\mathbb{P}(Z_{i_1} > \eta, Z_{i_2} > \eta, \ldots, Z_{i_{\beta_C}} > \eta, Z_{i_1'} > \eta, Z_{i_2'} > \eta, \ldots, Z_{i_{\beta_C'}} > \eta)}{(\mathbb{F}(Z(t)))^{\beta_c}} = 0.$$ 

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

$$\mathbb{P}(Z_{1,h} \in tA \mid Z_1 > \eta, \ldots, Z_{\beta_C} > \eta) \sim \mathbb{P}(Z_{1,h} \in tA \mid Z_1 > \eta, \ldots, Z_{\beta_C} > \eta, Z_{\beta_C + 1} > \epsilon, \ldots, Z_h > \epsilon)$$

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

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

Thus, we obtain

$$\lim_{t \to \infty} \mathbb{P}(Z_{1,h} \in tA \mid Z_1 > \eta, \ldots, Z_{\beta_C} > \eta) \leq \lim_{t \to \infty} \frac{\mathbb{P}((Z_1, \ldots, Z_{\beta_C}) \in tA_\zeta) - \mathbb{P}((Z_1, \ldots, Z_{\beta_C}) \in tA_0)}{(\mathbb{F}(Z(t)))^{\beta_c}} = \alpha^{\beta_c} \int_{A_\zeta} \prod_{i=1}^{\beta_C} z_i^{-\alpha-1} dz_i.$$ 

Moreover, $\lim_{\zeta \to 0} \int_{A_\zeta} \prod_{i=1}^{\beta_C} z_i^{-\alpha-1} dz_i = \int_{A_0} \prod_{i=1}^{\beta_C} z_i^{-\alpha-1} dz_i$, thus it actually holds that

$$\lim_{t \to \infty} \frac{\mathbb{P}(Z_{1,h} \in tA \mid Z_1 > \eta, \ldots, Z_{\beta_C} > \eta)}{(\mathbb{F}(Z(t)))^{\beta_c}} \leq \alpha^{\beta_c} \int_{A_0} \prod_{i=1}^{\beta_C} z_i^{-\alpha-1} dz_i.$$ 

Conversely, for the lower bound, it obviously holds that

$$\frac{\mathbb{P}(Z_{1,h} \in tA \mid Z_1 > \eta, \ldots, Z_{\beta_C} > \eta)}{(\mathbb{F}(Z(t)))^{\beta_c}} \geq \frac{\mathbb{P}((Z_1, \ldots, Z_{\beta_C}) \in tA_0)}{(\mathbb{F}(Z(t)))^{\beta_c}} \to \alpha^{\beta_c} \int_{A_0} \prod_{i=1}^{\beta_C} z_i^{-\alpha-1} dz_i.$$
Comparing the lower bound and the upper bound and summing over \( i \in P^* \) yields

\[
\lim_{t \to \infty} \frac{\mathbb{P}(Z_{1,h} \in tA)}{(F_Z(t))^{\beta_C}} = \alpha^{\beta_C} \sum_{i \in P^*} \int_{A_0(i)} \prod_{q=1}^{\beta_C} z_{iq}^{-\alpha_i - 1} \, dz_{iq},
\]

where \( i = (i_1, \ldots, i_{\beta_C}) \), \( A_0(i) = \{ (z_{i_1}, \ldots, z_{i_{\beta_C}}) \mid \sum_{i \in P \cap (i_1, \ldots, i_{\beta_C})} z_i \geq a_u, u = 1, \ldots, k \} \). This proves that the measure \( \nu_C \) has the following expression

\[
\nu_C(dz) = \alpha^{\beta_C} \sum_{i \in P^*} \prod_{q=1}^{\beta_C} z_{iq}^{-\alpha_i - 1} \prod_{i \notin \{i_1, \ldots, i_{\beta_C}\}} \delta_0(dz_i).
\]  

(33)

where \( \delta_0 \) denotes the Dirac point measure at 0.

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

\[
\lim_{t \to \infty} \frac{\mathbb{P}(u \cdot Z_{1,h} \in tA)}{(F_Z(t))^{\beta_C}} = \nu_C(u^{-1} \cdot A).
\]  

(34)

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

\[
\frac{\mathbb{P}(u \cdot Z_{1,h} \in tA)}{(F_Z(t))^{\beta_C}} \leq \frac{\mathbb{P}(\exists i_1, \ldots, i_{\beta_C} \in \{1, \ldots, h\}, u_{i_j}Z_{i_j} > \eta)}{(F_Z(t))^{\beta_C}} \leq C \sum_{1 \leq i_1 < \cdots < i_{\beta_C} \leq h} \prod_{q=1}^{\beta_C} (u_{i_q} \lor 1)^{\alpha_i + \epsilon}.
\]

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

\[
\sup_{t \geq 1} \frac{\mathbb{P}(u \cdot Z_{1,h} \in tA)}{(F_Z(t))^{\beta_C}} \leq CM(u).
\]  

(35)

Assumption (13) implies that \( \mathbb{E}[M(\sigma(X_{1,h}))] < \infty \). Then (34), (35) and bounded convergence yield (14). We now prove (15). For \( r \geq 2 \), let \( C_r \) be the subcone of \( [0, \infty)^{h+r-1} \) defined by

\[
(z_1, \ldots, z_{h+r-1}) \in C_r \iff (z_1, \ldots, z_h) \in C, (z_r, \ldots, z_{h+r-1}) \in C.
\]

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

\[
(z_1, \ldots, z_{h+r-1}) \in C_r \iff \prod_{u=1}^{k} \left( \sum_{i \in P_u} z_i \right) \prod_{u=1}^{k} \left( \sum_{i \in P^r_u} z_i \right) > 0.
\]

The sum over the sets \( P^r_u \) which include one of the sets \( P_u \) can be removed from the second product, and thus we see that \( C_r \) is of the form (12) and (11) holds. Necessarily, it holds
that $\beta_{C_r} \geq \beta_C$. Indeed, if there exists only $\ell < \beta_C$ indices $i_1, \ldots, i_\ell$ such that $z_i > 0$, then the first product above is zero, hence $(z_1, \ldots, z_{h+r-1}) \notin C_r$. 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 $C$, then $A_r(s, s')$ is also relatively compact in $C_r$ and thus, it holds that

$$
\lim_{t \to \infty} \frac{\mathbb{P}(Z_{1,h} \in tsA, Z_{r,r+h-1} \in ts'\lambda)}{(F_Z(t))^{\beta_{C_r}}} = \lim_{t \to \infty} \frac{\mathbb{P}(Z_{1,r+h-1} \in tA_r(s, s'))}{(F_Z(t))^{\beta_{C_r}}} = \nu_{C_r}(A_r(s, s')) ,
$$

and (15) follows straightforwardly, with $\mathcal{L}_r \equiv 0$ if $\beta_{C_r} > \beta_C$.

**Proof of Lemma 3.** By assumption, we have

$$
\lim_{t \to \infty} \frac{\mathbb{P}((Z_1, \ldots, Z_h) \in tA)}{(F_Z(t))^{\beta_C}} = \nu_C(A) , \quad \lim_{t \to \infty} \frac{\mathbb{P}((Z_1, \ldots, Z_h) \in tA)}{(F_Z(t))^{\beta_{C'}}} = \nu_{C'}(A) ,
$$

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

We now prove the results of Section 3. For clarity of notation, denote $\sigma_i = \sigma(X_i)$, $\nu_C = \nu$, $\mu_C = \mu$ and define $g(t) = t^{\beta_C}$ and $T(s) = s^{-\alpha\beta_C}$. Recall that $F_Y$ denotes the distribution function of $Y$ and $u_n = (1/F_Y)^\nu(n/k)$. By (4), (13), Breiman’s Lemma applies and thus it holds that $F_Y(u_n) \sim \mathbb{E}[\sigma_0^\alpha] F_Z(u_n)$ and

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

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

$$
W_{r,n}(s) = 1_{\{Y_{r+h-1} \in u_n sA\}} , \quad V_r(B) = 1_{\{Y_{r+m+r+m+h} \in B\}} .
$$

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

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

(37)

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

(38)

Recall the definition (25) of the function $G_n$:

$$
G_n(A, B, s, x, x') = \frac{\mathbb{P}(\sigma(x) \cdot Z_{1,h} \in u_n sA)}{g(k/n)} \mathbb{P}(\sigma(x') \cdot Z_{m,m+h} \in B) .
$$

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

$$
L_n(s, x) = \frac{\mathbb{P}(\sigma(x) \cdot Z_{1,h} \in u_n sA)}{g(k/n)} ,
$$

(39)
With these notations, we have,
\[ L_n(s, X_{r,r+h-1}) = \frac{\mathbb{E}[W_{r,n}(s) \mid X]}{g(k/n)}, \]
\[ G_n(A, B, s, X_{r,r+h-1}, X_{r+m,r+m+h}) = \frac{\mathbb{E}[W_{r,n}(s) V_r(B) \mid X]}{g(k/n)}. \]

For \( x \in \mathbb{R}^h \), denote
\[ L(x) = \frac{\nu(\sigma(x)^{-1} \cdot A)}{(\mathbb{E}[\sigma(X)])^\delta}, \]
so that \( \mathbb{E}[L(X_{1,h})] = \mu(A) \).

**Proof of Lemma 4.** Write
\[
L_n(s, x) - T(s)L(x) = \left\{ \frac{g(\hat{F}_Z(u_n s))}{g(k/n)} - (\mathbb{E}[\sigma(X)])^{-\delta} T(s) \right\} \frac{\mathbb{P}(\sigma(x) \cdot Z_{1,h} \in u_n s A)}{g(\hat{F}_Z(u_n s))} \\
+ (\mathbb{E}[\sigma(X)])^{-\delta} \mathbb{E} \left[ \sup_{s \geq 1} \left| \frac{\mathbb{P}(\sigma(X_{1,h}) \cdot Z_{1,h} \in u_n s A \mid X)}{g(\hat{F}_Z(u_n s))} - (\mathbb{E}[\sigma(X)])^\delta L(x) \right| \right].
\]

Thus, recalling the definition of \( v_n \) from (21), we have
\[
v_n(A) \leq \sup_{s \geq 1} \left| \frac{g(\hat{F}_Z(u_n s))}{g(k/n)} - (\mathbb{E}[\sigma(X)])^{-\delta} T(s) \right| \mathbb{E}[M(\sigma(X_{1,h}))] \\
+ (\mathbb{E}[\sigma(X)])^{-\delta} \mathbb{E} \left[ \sup_{s \geq 1} \left| \frac{\mathbb{P}(\sigma(X_{1,h}) \cdot Z_{1,h} \in u_n s A \mid X)}{g(\hat{F}_Z(u_n s))} - (\mathbb{E}[\sigma(X)])^\delta L(X_{1,h}) \right| \right].
\]

For all \( x \in \mathbb{R}^h \), we have
\[
\lim_{n \to \infty} \sup_{s \geq 1} \left| \frac{\mathbb{P}(\sigma(x) \cdot Z_{1,h} \in u_n s A)}{g(\hat{F}_Z(u_n s))} - (\mathbb{E}[\sigma(X)])^\delta L(x) \right| = 0.
\]

Moreover, by (35),
\[
\sup_{s \geq 1} \left| \frac{\mathbb{P}(\sigma(x) \cdot Z_{1,h} \in u_n s A)}{g(\hat{F}_Z(u_n s))} - (\mathbb{E}[\sigma(X)])^\delta L(x) \right| \leq CM(\sigma(x)).
\]

Thus, by (13) and bounded convergence,
\[
\lim_{n \to \infty} \mathbb{E} \left[ \sup_{s \geq 1} \left| \frac{\mathbb{P}(\sigma(X_{1,h}) \cdot Z_{1,h} \in u_n s A \mid X)}{g(\hat{F}_Z(u_n s))} - (\mathbb{E}[\sigma(X)])^\delta L(X_{1,h}) \right|^2 \right] = 0.
\]

Since \( g \circ \hat{F} \) is regularly varying at infinity with negative index, by (Bingham et al., 1989, Theorem 1.5.2), the convergence of \( g(\hat{F}_Z(u_n s))/g(k/n) \) to \( (\mathbb{E}[\sigma(X)])^{-\delta} T(s) \) is uniform on \([1, \infty)\). Thus we have proved that \( v_n(A) \to 0 \). \( \square \)
Proof of Theorem 5. For \( s \geq 1 \), define

\[
K(B, s) = T(s)\mu(A)\rho(A, B, m), \quad \hat{K}_n(B, s) = \frac{1}{ng(k/n)} \sum_{r=1}^{n} W_{r,n}(s)V_r(B),
\]

\[
\hat{e}_n(s) = \hat{K}_n(\mathbb{R}^{k+1}, s) = \frac{1}{ng(k/n)} \sum_{r=1}^{n} W_{r,n}(s), \quad \xi_n = \frac{Y_{(n:n-k)}}{u_n}.
\]

With this notation, we have

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

Equations (38) and (37) imply, respectively, that

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

With this in mind, we split

\[
\hat{\rho}_n(A, B, m) - \rho(A, B, m) = \frac{\hat{K}_n(B, \xi_n) - K(B, \xi_n)}{\hat{e}_n(\xi_n)} - \frac{\rho(A, B, m)}{\hat{e}_n(\xi_n)} \{\hat{e}_n(\xi_n) - \mu(A)T(\xi_n)\}.
\]

Thus, we only need to find the correct norming sequence \( w_n \) and asymptotic distribution in \( D([a, b]) \) for any \( 0 < a < b \) of the sequence of processes \( w_n\{\hat{K}_n(B, \cdot) - K(B, \cdot)\} \). To do this, define further

\[
K_n(B, s) = \mathbb{E}[\hat{K}_n(B, s)]
\]

Then

\[
\hat{K}_n(B, s) - K(B, s) = \hat{K}_n(B, s) - K_n(B, s) + K_n(B, s) - K(B, s).
\]

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 \( \hat{K}_n - K_n = (ng(k/n))^{-1/2}E_{n,1} + E_{n,2} \) with

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

\[
E_{n,2}(B, s) = \frac{1}{ng(k/n)} \sum_{r=1}^{n} \mathbb{E}[W_{r,n}(s)V_r(B) | \mathcal{X}] - K_n(B, s)
\]

\[
= \frac{1}{n} \sum_{r=1}^{n} \{G_n(A, B, s, X_{r,r+h-1}, X_{r+m,r+m+h'}) - K_n(B, s)\}.
\]

The term in (43) will be called the i.i.d. term. It is a sum of conditionally independent random variables. The term in (44) will be called the dependent term. It is a function of the dependent vectors \( (X_{r,r+h-1}, X_{r+m,r+m+h'}) \).

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

$$
(\mathbb{E}[\sigma^a(X_1))]^d \text{cov}(W(B, s), W(B', s'))
= \mathbb{E}\left[\mathcal{L}_1(A, \sigma(X_{1,k}),\sigma(X_{1,k}), s, s') \times \mathbb{P}(Y_{m,m+h'} \in B, Y_{m,m+h'} \in B' \mid \mathcal{X})\right]
+ \sum_{r=2}^{h \wedge (m-h)} \mathbb{E}\left[\mathcal{L}_r(A, \sigma(X_{1,k}), \sigma(X_{r,r+h-1}), s, s')
\times \{\mathbb{P}(Y_{m,m+h'} \in B, Y_{m+r-1,m+h'+r-1} \in B' \mid \mathcal{X})
+ \mathbb{P}(Y_{m,m+h'} \in B', Y_{m+r-1,m+h'+r-1} \in B \mid \mathcal{X})\}\right],
$$

where the functions $L_r$ are defined in (15).

Claim 2. For each fixed $B$, $E_{n,1}(B, \cdot)$ is tight in $D([a, b])$ 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 $E_{n,2}(B, \cdot) = O_P(\sqrt{m})$, 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 $a > 0$, $\sup_{s \geq a} \sup_B |K_n(B, s) - K(B, s)| = O_P(v_n(A))$.

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

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

This convergence and the decomposition (41) imply

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

This distribution is Gaussian. Applying (45) and the fact that $\rho(A, \mathbb{R}^{h'+1}, m) = 1$, it is easily checked that its variance is given by (24). This concludes the proof of Theorem 5. $\square$

We now prove the claims.

Proof of Claim 1. For $r = 1, \ldots, n$, 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 \cite{Orey1958}. Let \( C(B, B', s, s') \) denote the quantity in the right hand side of (45). We need to check that
\[
\text{cov} \left( \sum_{r=1}^{n} \zeta_{n,r}(B, s), \sum_{r=1}^{n} \zeta_{n,r}(B', s') \mid \mathcal{X} \right) \overset{p}{\rightarrow} C(B, B', s, s') ,
\]
\[
\sum_{r=1}^{n} \mathbb{E}[\zeta_{n,r}^4(B, s) \mid \mathcal{X}] \rightarrow p 0 .
\]
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 (46) and (47).

For \( u \geq 1, x, x' \in \mathbb{R}^h \), denote
\[
\mathcal{L}_{n,u}(A, x, x', s, s') = \frac{\mathbb{P}(\sigma(x) \cdot Z_{1,h} \in uA, \sigma(x') \cdot Z_{u,u+h-1} \in uA')}{g(F_Z(u_n))}.
\]
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 (15). For \( u > h \), \( Z_{1,h} \) and \( Z_{u,u+h-1} \) 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,
\[
\text{cov} \left( \sum_{r=1}^{n} \zeta_{n,r}(B, s), \sum_{r=1}^{n} \zeta_{n,r}(B', s') \mid \mathcal{X} \right) = \sum_{r=1}^{n} \text{cov} \left( \zeta_{n,r}(B, s), \zeta_{n,r}(B', s') \mid \mathcal{X} \right)
\]
\[+ \sum_{r=1}^{n} \sum_{u=1}^{m+h'} \text{cov} \left( \zeta_{n,r}(B, s), \zeta_{n,j+u}(B', s') \mid \mathcal{X} \right)
\]
\[+ \sum_{r=1}^{n} \sum_{u=1}^{m+h'} \text{cov} \left( \zeta_{n,j+u}(B, s), \zeta_{n,r}(B', s') \mid \mathcal{X} \right) .
\]

For \( u = 1, \ldots, h \wedge (m - h) \) it is easily seen that
\[
\sum_{r=1}^{n} \text{cov} \left( \zeta_{n,r}(B, s), \zeta_{n,r+u}(B', s') \mid \mathcal{X} \right)
\]
\[\sim \frac{g(F_Z(u_n))}{ng(k/n)} \sum_{r=1}^{n} \mathcal{L}_{n,u}(\mathbf{X}_{r,r+h-1}, \mathbf{X}_{r+u,r+u+h-1}, s, s')
\]
\[\times \mathbb{P}(\mathbf{Y}_{r+m,r+m+h} \in B, \mathbf{Y}_{r+u+m,r+u+m+h'} \in B' \mid \mathcal{X})
\]
\[\rightarrow_p \frac{\mathbb{E} \left[ \mathcal{L}_u(A, \mathbf{X}_{1,h}, \mathbf{X}_{u,u+h-1}, s, s') \mathbb{P}(\mathbf{Y}_{m,m+h} \in B, \mathbf{Y}_{u,u+m+h'} \in B' \mid \mathcal{X}) \right]}{(\mathbb{E}[\sigma^\alpha(\mathbf{X})])^\delta} .
\]

This yields the right-hand side of (45), so we must prove that the terms in (48) and (49) are negligible. If \( h > m - h \), then for large \( n \) and \( m - h < u \leq h \), we have \( (uA') \cap B = 0 \), so, for all \( r = 1, \ldots, n \),
\[
\mathbb{P}(\mathbf{Y}_{r,r+h-1} \in uA, \mathbf{Y}_{r+u,r+u+h-1} \in uA', \mathbf{Y}_{r+m,r+m+h} \in B, \mathbf{Y}_{r+u+m,r+u+m+h'} \in B' \mid \mathcal{X}) = 0 .
\]
For \( u > h \), then as mentioned above, \( L_u(A, \cdot, \cdot, s, s') \) converges to 0 in \( L^1(X_{1,h}, X_{u,u+h-1}) \) so
\[
\sum_{r=1}^{n} \text{cov}(\zeta_{n,r}(B, s), \zeta_{n,r+u}(B', s') | \mathcal{X}) \to P 0 .
\]
This proves (46). Next, since \( \zeta_{n,r} \) are indicators and applying (38)
\[
\sum_{r=1}^{n} \mathbb{E}[\zeta_{n,r}^4(B, s)] \leq C \frac{\mathbb{E}[W_1,n(s,A)V_1(B)]}{ng(k/n)} \to 0 .
\]
This proves (47) and the weak convergence of finite dimensional distributions. \(\square\)

Proof of Claim 3. By definition of the functions \( L_n \) and \( G_n \) (cf. (39) and (25)), it clearly holds that
\[
|G_n(A, B, s, X_{r,r+h-1}, X_{r+m,r+m+h'})| \leq L_n(s, X_{r,r+h-1}) .
\]
We apply the variance inequality (B.3) in the weak dependence case to get
\[
\text{var}(E_{n,2}(B, s)) \leq C \frac{\text{var}(G_n(A, B, s, X_{1,h}, X_{1+m,1+m+h'}))}{n} \leq \frac{1}{n} \mathbb{E}[L_n^2(s, X_{1,h})] .
\]
By (35), \( L_n(s, x) \leq C \mathbb{M}(\sigma(x)) \). Thus, by (13), the right hand side is uniformly bounded, thus \( \text{var}(E_{n,2}(B, s)) = O(1/n) \) 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 \( (0, \infty] \). \(\square\)

Proof of Claim 5. Consider now the bias term \( K_n - K \). Recall that (see (42) and (38))
\[
K_n(B, s) = \mathbb{E}[\tilde{K}_n(B, s)] \to T(s)\mu(A)\rho(A, B, m) = K(B, s)
\]
Therefore, \( K_n(B, s) \) converges pointwise to \( K(B, s) \). The goal here is to show that this convergence is uniform. Using the definition of \( K_n \), (39) and (25) we have
\[
K_n(B, s) = \mathbb{E}[G_n(A, B, s, X_{1,h}, X_{m, m+h'})] = \mathbb{E}[L_n(s, X_{1,h})\mathbb{P}(\sigma(X_{m, m+h'}) \cdot Z_{m, m+h'} \in B | \mathcal{X})] .
\]
Using this definition and recalling the formula for \( \rho(A, B, m) \) (see (16))
\[
K(B, s) = T(s)\mathbb{E}[L(X_{1,h})\mathbb{P}(\sigma(X_{m, m+h'}) \cdot Z_{m, m+h'} \in B | \mathcal{X})] .
\]
Therefore, recalling the definition (21) of \( v_n(A) \), we obtain that
\[
|K_n(B, s) - K(B, s)| \leq \mathbb{E} \left[ \sup_{s \geq 1} |L_n(s, X_{1,h}) - T(s)L(X_{1,h})| \right] = v_n(A) .
\]
\(\square\)

Proof of Corollary 6. In the following, \( y \) stands for the set \( (-\infty, y] \) in the previous notation. For \( y \in \mathbb{R}^{h+1} \), rewrite the decomposition (41) in the present context to get
\[
\tilde{\Psi}_n(y) - \Psi(y) = \frac{\tilde{K}_n(y, \xi_n) - K(y, \xi_n)}{\tilde{e}_n(\xi_n)} - \frac{\Psi(y)}{\tilde{e}_n(\xi_n)} \{ \tilde{e}_n(\xi_n) - \mu(A) T(\xi_n) \} .
\]
Thus we need only prove that the sequence of suitably normalized processes \( \tilde{K}_n(s, y) - K_n(y, s) \) 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. \(\square\)
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 2σ(\(A, B)(1 - H) < 1\), then \(\gamma_n^{-\tau(A, B)/2}E_n,A(B, \cdot)\) 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 \(d \in \mathbb{N}^*, q \in \mathbb{N}^d\) and \(x \in \mathbb{R}^d\), denote

\[
H_q(x) = \prod_{i=1}^d H_{q_i}(x_i).
\]

Define \(X_j = (X_{j+1}, \ldots, X_{j+h}, X_{j+m}, \ldots, X_{j+m+h'})\). The Hermite coefficients of \(G_n(A, B, s, \cdot)\) and \(G\) with respect to \(X_0\) can be expressed, for \(q \in \mathbb{N}^{h+h'+1}\), as

\[
J_n(q, s) = E[H_q(X_0)G_n(A, B, s, X_0)] ,
\]

\[
J(q) = E[H_q(X_0)G(X_0)].
\]

Since \(G_n(A, B, s, \cdot)\) converges to \(T(s)G(\cdot)\) in \(L^p(X_0)\) for some \(p > 1\), \(J_n(q, s)\) converges to \(s^{-\alpha\delta}J(q)\). 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 \(X_0\). Define \(J^*_n(q, s) = E[H_q(UX_0)G_n(A, B, s, UX_0)]\) and \(J^*(q) = E[H_q(UX_0)G(X_0)]\). Under Assumption 1, the function \(G_n\) can be expanded for \(x \in \mathbb{R}^{h+h'+1}\) as

\[
G_n(A, B, s, x) = E[G_n(A, B, s, X_0)] = \sum_{|q| = \tau(A, B)} \frac{J^*_n(q, s)}{q!} H_q(Ux) + r_n(s, x),
\]

where \(r_n\) is implicitly defined and has Hermite rank at least \(\tau(A, B) + 1\) with respect to \(UX_0\). Denote \(R_n(s) = n^{-1}\sum_{r=1}^n r_n(s, X_r)\). Applying (B.3), we have

\[
\text{var}(R_n(s)) \leq C \left( \gamma_n^{-\tau(A, B)+1} \sqrt{\frac{1}{n}} \right) \text{var}(G_n(A, B, s, X_0)) \leq C \left( \gamma_n^{-\tau(A, B)+1} \sqrt{\frac{1}{n}} \right) E[L_n^2(s, X_{1:n})].
\]

By Assumption (13), \(E[L_n^2(s, X_{1:n})]\) is uniformly bounded, thus \(\text{var}(R_n(s)) = o(\gamma_n^{-\tau(A, B)})\) 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

\[
Z_n(s) = \sum_{|q| = \tau(A, B)} \frac{J^*_n(q, s)n^{-1}}{q!} \gamma_n^{-\tau(A, B)/2} \sum_{r=1}^n H_q(UX_r).
\]

By (Arcones, 1994, Theorem 6), there exist random variables \(R^*(q)\) such that \(Z_n(s)\) converges to

\[
T(s) \sum_{|q| = \tau(A, B)} \frac{J^*(q)}{q!} R^*(q)
\]

for each \(s \geq 0\). To prove that the convergence is uniform, we only need to prove that \(J^*_n(q, s)\) converges uniformly to \(T(s)J^*(q)\) for each \(q\) such that \(|q| = \tau(A)\). 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 \( a > 0 \),

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

As already shown in the proof of Lemma 4, this last quantity converges to 0 for \( p = 2 \). \( \square \)

**Appendix**

**A Second order regular variation of convolutions**

Denote \( A \propto B \) if there exists positive constant \( c_1 \) and \( c_2 \) such that \( c_1 A \leq B \leq c_2 B \).

**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

\[
\left| \mathbb{P}(u_1 Z_1 + u_2 Z_2 > t) - \bar{F}(t/u_1) - \bar{F}(t/u_2) \right| \leq Cu_1^{a+\epsilon} u_2^{a+\epsilon} t^{-1} \bar{F}(t) \int_0^t \bar{F}(s) ds.
\]

**Proof.** Obviously, we have

\[
\mathbb{P}(u_1 Z_1 + u_2 Z_2 > t) = \bar{F}(t/u_1) + \bar{F}(t/u_2) - \bar{F}(t/u_1) \bar{F}(t/u_2)
\]

\[
+ \mathbb{P}(t/2 < u_1 Z_1 \leq t) \mathbb{P}(t/2 < u_2 Z_2 \leq t)
\]

\[
+ \mathbb{P}(u_1 Z_1 \leq t/2, u_2 Z_2 \leq t, u_1 Z_1 + u_2 Z_2 > t)
\]

\[
+ \mathbb{P}(u_2 Z_2 \leq t/2, u_1 Z_1 \leq t, u_1 Z_1 + u_2 Z_2 > t).
\]

Consider for instance the second last term. It may be written as

\[
I_1 := \mathbb{E} \left[ 1_{\{u_1 Z_1 \leq t/2\}} \left\{ \frac{\bar{F}(t(1 - u_1 Z_1/t)/u_2)}{\bar{F}(t/u_2)} - 1 \right\} \right].
\]

Since \( F \) satisfies Assumption 2, we have, for \( u \in [1/2, 1] \),

\[
0 \leq \frac{\bar{F}(ut)}{\bar{F}(t)} - 1 = u^{-\alpha} e^{\int_0^1 \frac{\eta(t_s)}{s} ds} - 1 = \{u^{-\alpha} - 1\} e^{\int_0^1 \frac{\eta(t_s)}{s} ds} + e^{\int_0^1 \frac{\eta(t_s)}{s} ds} \int_0^1 \frac{\eta^*(ts)}{s} ds.
\]

Since \( \eta^*(t) \) is decreasing, we have, for all \( u \in [1/2, 1] \),

\[
0 \leq \frac{\bar{F}(ut)}{\bar{F}(t)} - 1 \leq C\{ |u^{-\alpha} - 1| + \log(u) \} \leq C(1 - u).
\]

Applying this inequality with \( 1 - u = u_1 Z_1/t \) on the event \( u_1 Z_1 \leq t/2 \) yields

\[
I_1 \leq Cu_1 t^{-1} \mathbb{E} \left[ Z_1 1_{\{u_1 Z_1 \leq t/2\}} \right] \leq Ct^{-1} \int_0^{t/u_1} \bar{F}(s) ds = Ct^{-1} u_1^{-1} \int_0^{t/u_1} \bar{F}(s/u_1) ds.
\]
By Potter’s bounds, for any $\epsilon > 0$, there exists a constant $C$ such for any $s, t > 0,$
\[ \frac{\tilde{F}(s/u)}{F(s)} \leq C(u_1^{-1} \wedge 1)^{-\alpha - \epsilon}. \]

Applying this bound we obtain
\[ I_1 \leq C(u_1 \vee 1)^{\alpha + \epsilon} (u_2 \vee 1)^{\alpha + \epsilon} t^{-1} \tilde{F}(t) \int_0^t \tilde{F}(s) \, ds. \]

To conclude, note that $\tilde{F}^2(t) = O(t^{-1} \tilde{F}(t) \int_0^t \tilde{F}(s) \, ds)$ if $\alpha < 1$ and $\tilde{F}^2(t) = o(t^{-1} \tilde{F}(t) \int_0^t \tilde{F}(s) \, ds)$ if $\alpha \geq 1$.

**Remark.** By induction, we can obtain the bound
\[ \left| \mathbb{P}(Z_1 + \cdots + Z_n > t) - n \tilde{F}(t) \right| \leq C t^{-1} \tilde{F}(t) \int_0^t \tilde{F}(s) \, ds, \]
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 $\mathbb{E}[Z_1] < \infty$,
\[ \lim_{t \to \infty} t \left\{ \frac{\mathbb{P}(Z_1 + \cdots + Z_n > t)}{\mathbb{P}(Z_1 > t)} - n \right\} = \frac{n(n - 1)}{2} \mathbb{E}[Z_1]. \]

**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 $X = (X^{(1)}, \ldots, X^{(k)})$ be a $k$-dimensional centered Gaussian vector with covariance matrix $\Gamma$. The Hermite coefficients of $G$ with respect to $X$ are defined as
\[ J(G, X, q) = \mathbb{E} \left[ G(X) \prod_{j=1}^k H_{q_j}(X^{(j)}) \right], \]
where $q = (q_1, \ldots, q_k) \in \mathbb{N}^k$. If $\Gamma$ is the $k \times k$ identity matrix (denoted by $I_k$), i.e. the component of $X$ are i.i.d. standard Gaussian, then the corresponding Hermite coefficients are denoted by $J^*(G, q)$. The Hermite rank of $G$ with respect to $X$, is the smallest integer $\tau$ such that
\[ J(G, X, q) = 0 \quad \text{for all} \quad q \text{ such that} \quad 0 < |q_1 + \cdots + q_k| < \tau. \]
B.2 Variance inequalities

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

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

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

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

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 \( \mathbb{E}[G^2(X_0)] < \infty \) and with Hermite rank \( q \) with respect to \( X_0 \),

\[
\text{var} \left( n^{-1} \sum_{r=1}^{n} G(X_j) \right) \leq C(\ell^q(n)n^{2q(H-1)}) \lor n^{-1} \text{var}(G(X_0)). \quad (B.3)
\]

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

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 \( \prod_{i=1}^{\ell}(s_i, t_i) \) with \( s_i \leq t_i, 1 \leq i \leq d \). 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

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

(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. \( X(y) = 1_{y \leq y} \) for some \( T \) valued random variable \( Y \), then \( X(B) = 1_{y \in B} \).

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 \( \zeta_n \) converges weakly to
those of a process $\zeta$ which is continuous on the upper boundary of $T$. Assume moreover that there exist $\gamma \geq 0$ and $\delta > 1$ such that

$$\mathbb{P}(|\zeta_n(B)| \wedge |\zeta_n(B')| \geq \lambda) \leq C\lambda^{-\gamma}E[\mu_\delta^n(B \cup B')]$$

(C.1)

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

**Sketch of proof.** For $f$ defined on $T = T_1 \times \cdots \times T_d$, $i \in \{1, \ldots, d\}$ 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

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

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

$$w_i^{\text{PrimePrime}}(f, s, t) = \sup_{s<u<v<w<t} \|f_t^{(i)}(t) - f_u^{(i)}(t)\|_\infty \wedge \|f_t^{(i)}(t) - f_v^{(i)}(t)\|_\infty,$$

$$w_i^{\text{PrimePrime}}(f, \delta) = \sup_{u<v<w<u+\delta} \|f_t^{(i)}(t) - f_u^{(i)}(t)\|_\infty \wedge \|f_t^{(i)}(t) - f_v^{(i)}(t)\|_\infty.$$

By the Corollary of Bickel and Wichura (1971), a sequence of processes $\{X_n\}$ defined on $T$ converges weakly in $D(T)$ 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 all $i = 1, \ldots, d$,

$$\mathbb{P}(w_i^{\text{PrimePrime}}(X_n, \delta) > \lambda) \to 0.$$  

(C.2)

For any measure $\mu$ 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 (C.2) is implied by

$$\mathbb{P}(w_i^{\text{PrimePrime}}(X_n, s, t) > \lambda) \leq C\lambda^{-\gamma}E[\mu^{(i)}(s, t)\delta],$$

(C.3)

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

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

(C.4)

The proof of (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 $S_0, \ldots, S_n$ be random variables. Assume that there exists nonnegative random variables $u_1, \ldots, u_n$ such that

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

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

Proving by induction that (C.1) implies (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

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

**Proof.** We start by assuming that the pairs $(B_i, B'_i)$ are i.i.d. and we prove (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 $\mathbb{E}[B_1B'_1] = -p_1q_1$ and

$$
\mathbb{E}[B_1^2B'_1^2] = \mathbb{E}[(B_1 - 2p_1B_1 + p_1^2)(B'_1 - 2q_1B'_1 + q_1^2)]
= p_1q_1^2 + p_1^2q_1 - 3p_1^2q_1^2 = p_1q_1(p_1 + q_1 - 3p_1q_1) \leq p_1q_1.
$$

The last inequality comes from the fact that $B_1B'_1 = 0$ 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 (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

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

This proves that (C.5) holds for all $n \geq 1$.

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

$$
\left( \sum_{i=1}^n a_i \right)^2 \leq m \sum_{q=1}^{[n/m]} \left( \sum_{r=1}^{[n/m]} a_{(j-1)m+q} \right)^2 \leq m \sum_{q=1}^{[n/m]} \left( \sum_{r=1}^{[n/m]} a_{(j-1)m+q} \right)^2.
$$
Applying this and the bound for the independent case (extending all sequences by zero after the index $n$) yields
\[
\mathbb{E}[s_n^2 s_n'{}^2] \leq 3m^2 \sum_{q=1}^{m} \sum_{q'=1}^{m} \sum_{r=1}^{\lfloor n/m \rfloor} \sum_{r'=1}^{\lfloor n/m \rfloor} P(j-1)m+qP(j'-1)m+q' = 3m^2s_n s_n'.
\]

Let us apply this criterion in the context of section 3. Fix a cone $J$ and a relatively compact subset $A \in J$. Recall that $E_{n,1}$ and $E_{n,2}$ are defined in (43) and (44).

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

\[\text{Proof.} \quad \text{Since } A \text{ is a cone, if } s < t, \text{ then } tA \subset sA. \text{ Thus, a sequence of random measures } \hat{\mu}_n \text{ on } \mathbb{R}^d \times (0, \infty) \text{ can be defined by}
\]
\[
\hat{\mu}_n((-\infty, y] \times (s, \infty)) = \frac{1}{n} \sum_{r=1}^{n} \mathbb{P}(Y_{j,h} \in su_nA \mid \mathcal{A}) \mathbb{P}(Y_{r+m,r+m+h'} \leq y \mid \mathcal{A})
\]
\[
= \frac{1}{n} \sum_{r=1}^{n} G_n(A, B, s, X_{j,h}X_{r+m,r+m+h'}, y),
\]
where $G_n$ is defined in (25). Then $\hat{\mu}_n$ converges vaguely in probability to the measure $\mu$ defined by
\[
\mu((-\infty, y] \times (s, \infty)) = \mu(A)T(s)\Psi_{A,m,h}(y).
\]

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

Taking unconditional expectations then yields
\[
\mathbb{E}[E_{n,2}^2(D)E_{n,2}^2(D')] \leq C\mathbb{E}[,\hat{\mu}_n(D)\hat{\mu}_n(D')] \leq \mathbb{E}[\hat{\mu}_n(D \cup D')].
\]

Thus (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 $\Psi$ is assumed to be continuous, then Lemma C.1 applies and the process $E_{n,1}$ is tight with respect to both variables.

\[\text{Lemma C.4.} \quad \text{Under the assumptions of Theorem 5, for any fixed } B \in \mathbb{R}^{h' + 1}, E_{n,2}(B, \cdot) \text{ converges uniformly to zero on compact sets of } (0, \infty). \text{ Under the assumption of Corollary 6, } E_{n,2} \text{ converges uniformly to zero on compact sets of } \mathbb{R}^{h' + 1} \times (0, \infty).\]
Proof. We only need to prove the tightness. By the variance inequality (B.3) and H"older’s inequality, we have, for any relatively compact neighbouring blocks $D, D'$ of $\mathbb{R}^d \times (0, \infty)$,

$$\mathbb{P}(|E_{2,n}(D)| \land |E_{2,n}(D')| \geq \lambda) \leq \lambda^{-2} \sqrt{\mathbb{E}[E_{2,n}^2(D)]\mathbb{E}[E_{2,n}^2(D')]} \leq \lambda^{-2} \mathbb{E}[E_{2,n}^2(D \cup D')] \leq C\lambda^{-2}n^{-1}\mathbb{E}[\tilde{\mu}_n^2(D \cup D')]$$

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

$$\tilde{\mu}_n(y, s) = \frac{\mathbb{P}(Y_{1,h} \in su_n A \mid X) - g(k/n)}{\mathbb{P}(Y_{1,h} \in su_n A \mid X)} \mathbb{P}(Y_{m,m+h'} \leq y \mid X).$$

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

$$\hat{\mu}((\infty, y] \times (s, \infty]) = \frac{\nu_3(\sigma(X_{1,h})^{-1} \cdot A)}{(\mathbb{E}[\nu_3(\sigma(X_{1,h})^{-1} \cdot A)]^3)T(s)\mathbb{P}(Y_{m,m+h'} \leq y \mid X)}.$$

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 $F_\mathcal{Z}$ is continuous, so Lemma C.1 applies.

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

References

Miguel A. Arcones. Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *The Annals of Probability*, 22(4):2242–2274, 1994.

Peter J. Bickel and Michael J. Wichura. Convergence criteria for multiparameter stochastic processes and some applications. *Annals of Mathematical Statistics*, 42:1656–1670, 1971.

Patrick Billingsley. *Convergence of probability measures*. New York, Wiley, 1968.

Nicholas H. Bingham, Charles M. Goldie, and Jan L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989.

F. Jay Breidt, Nuno Crato, and Pedro de Lima. The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83(1-2):325–348, 1998.

Leo Breiman. On some limit theorems similar to the arc-sine law. *Theory of Probability and Applications*, 10:323–331, 1965.

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

Richard A. Davis and Thomas Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 38A:977–1009, 2009. Probability, statistics and seismology.
Holger Drees. Optimal rates of convergence for estimates of the extreme value index. *The Annals of Statistics*, 26(1):434–448, 1998.

Andrew C. Harvey. Long memory in stochastic volatility. In J. Knight and S. Satchell (eds), *Forecasting volatility in financial markets*. Butterworth-Heinemann, London, 1998.

Clifford M. Hurvich, Eric Moulines, and Philippe Soulier. Estimating long memory in volatility. *Econometrica*, 73(4):1283–1328, 2005.

Rafał Kulik and Philippe Soulier. The tail empirical process for long memory stochastic volatility sequences. *Stochastic Processes and their Applications*, 121(1):109 – 134, 2011.

Abhimanyu Mitra and Sidney I. Resnick. Hidden regular variation and detection of hidden risks. *Stochastic Models*, 27:591–614, 2011.

Edward Omey and Eric Willekens. Second-order behaviour of distributions subordinate to a distribution with finite mean. *Communications in Statistics. Stochastic Models*, 3(3):311–342, 1987.

Steven Orey. A central limit theorem for $m$-dependent random variables. *Duke Mathematical Journal*, 25:543–546, 1958.

Sidney I. Resnick. *Heavy-tail phenomena*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2007. Probabilistic and statistical modeling.

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