Limit theorems for non-degenerate U-statistics of block maxima for time series

Axel Bücher[1] and **Torben Staud**[1]

[1]*Ruhr-Universität Bochum Fakultät für Mathematik*
e-mail: axel.buecher@rub.de; torben.staud@rub.de

Abstract: The block maxima method is a classical and widely applied statistical method for time series extremes. It has recently been found that respective estimators whose asymptotics are driven by empirical means can be improved by using sliding rather than disjoint block maxima. Similar results are derived for general non-degenerate U-statistics of arbitrary order, in the multivariate time series case. Details are worked out for selected examples: the empirical variance, the probability weighted moment estimator and Kendall's tau statistic. The results are also extended to the case where the underlying sample is piecewise stationary. The finite-sample properties are illustrated by a Monte Carlo simulation study.

1. Introduction

A common target parameter in various domains of application is the distribution of componentwise yearly or seasonal maxima calculated from some underlying multivariate time series [25, 1]. Statistical inference on the target distribution typically involves the assumption that the block maximum distribution is an extreme value distribution. The latter is justified by probabilistic results from extreme value theory: under broad conditions on the time series, the only possible limit distribution of affinely standardized componentwise maxima, as the block size goes to infinity, are extreme value distributions; see [28] for the univariate case and [24] for multivariate extensions.

The statistical literature on estimation and testing for extreme-value distributions is abundant, ranging from univariate estimators for the parameters of the generalized extreme value distribution [34, 23] to nonparametric estimators for extreme value copulas [20] and parametric estimators for max-stable process models [32].

---

arXiv: 2308.13761

Mathematically, statistical methods are typically validated under the additional assumption that the block maxima sample is serially independent. However, heuristically, both the independence assumption as well as the assumption that block maxima genuinely follow an extreme-value distribution should only be satisfied asymptotically, for the block size tending to infinity. [16, 19, 17] have shown that specific univariate estimators are consistent and asymptotically normal in a sampling scheme where the block size tends to infinity, while maintaining an i.i.d. assumption on the underlying time series. For specific univariate and multivariate estimators, [6, 7] also relax the i.i.d. assumption, and allow for more general stationary time series satisfying certain mixing conditions. It has moreover been found that estimators based on block maxima may be made more efficient by considering sliding rather than disjoint block maxima, both in the univariate [35, 8, 10] and in the multivariate case [41]. A simple but incomplete heuristic argument for this superiority consists of the fact that the sliding block maxima sample contains some observations that are not present in the disjoint block maxima sample; hence, it can be considered more informative. A formal argument for the superiority is provided in Lemma A.4 in [41].

In general, the field of asymptotic statistics is based on a number of fundamental theoretical tools like the central limit theorem, the delta-method, the empirical process or the concept of U-statistics [38]. While the efficiency gain of the sliding block maxima method over its disjoint blocks counterpart mentioned in the previous paragraph has been established for classical empirical means as well as empirical (copula) processes, it has not been studied yet for the case of U-statistics. The present paper aims at filling this gap by studying non-degenerate U-statistics of disjoint and sliding block maxima samples. The topic is related but different to [31], who study U-statistics in the univariate case where the kernel of order m is evaluated blockwise in the largest m order statistics of a (disjoint) block of observations.

In general, U-statistics comprise a number of important estimators like the empirical covariance, Wilcoxon's statistic or Kendall's tau statistic. Prominent examples from extremes are empirical probability weighted moments of order $k \geq 2$ in the univariate case (which give rise to the probability weighted moment estimator for the parameters of a Generalized Extreme Value distribution [23]), or sample versions of Kendall's tau and Spearman's rho in the bivariate case [1, Page 274–275]. Mathematical theory for i.i.d. random variables dates back to [22]; since then, several favorable statistical properties have been demonstrated [38, Chapter 12]. Asymptotic results on U-statistics have also been generalized to the time series context [37, 40, 14]; unbiasedness then only holds asymptotically.

The main result of this paper is Theorem 3.2, where we establish a central limit theorem on the estimation error of U-statistics for multivariate disjoint and sliding block maxima under mild assumptions on the serial dependence and the kernel function. As in the papers mentioned before, the disjoint blocks version is found to be at most as efficient as the sliding blocks version. In selected examples, it is in fact found to be less efficient. The results are extended to a sampling scheme involving piecewise stationarities which is used to capture certain applications from environmental extremes where maxima are calculated

based on, for example, summer days [10]. The model is interesting mathemati-
cally, because unlike the disjoint block maxima sample the sliding block maxima
sample is not stationary anymore.

The remaining parts of this paper are organized as follows: the underlying
model assumptions and the definition of respective U-Statistics for disjoint and
sliding block maxima are presented in Section 2. The main limit results are
discussed in Section 3, and illustrated for three selected examples in Section 4.
Extensions to piecewise stationary time series are presented in Section 5. Results
from a Monte Carlo simulation study illustrate the behavior in finite-sample
situations (Section 6). Finally, the proofs are deferred to Section 7. Additional
limit results under strong mixing assumptions and lengthy calculations of some
asymptotic variances are postponed to a supplement [11].

2. U-statistics of block maxima

Recall the Generalized Extreme Value (GEV) distribution with parameters μ
(location), σ (scale) and γ (shape), defined by its cumulative distribution func-
tion (cdf)

$$G_{(\mu,\sigma,\gamma)}(x) := \exp\left[-\left\{1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right\}^{-\frac{1}{\gamma}}\right], \qquad 1 + \gamma\,\frac{x-\mu}{\sigma} > 0.$$

If $\eta = (\mu, \sigma, \gamma)' = (0, 1, \gamma)'$, we will use the abbreviation $G_{(0,1,\gamma)} = G_\gamma$. The
support of G_γ is denoted by $S_\gamma = \{x \in \mathbb{R} : 1 + \gamma x > 0\}$.

An extension of the classical extremal types theorem to strictly stationary
time series [28] implies that, under suitable broad conditions, affinely standard-
ized maxima extracted from a stationary time series converge to the GEV-
distribution. This was generalized to the multivariate case in [24], where the
marginals are necessarily GEV-distributed. We make this an assumption, and
additionally require the scaling sequences to exhibit some common regularity
inspired by the max-domain of attraction condition in the i.i.d. case [12].

Condition 1 (Multivariate Max-domain of attraction). Let $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ denote a
stationary time series in $\mathbb{R}^d$ with continuous margins, where $d \in \mathbb{N} = \{1, 2, \dots\}$.
There exist sequences $(\boldsymbol{a}_r)_r = (a_r^{(1)}, \dots, a_r^{(d)}))_r \subset (0, \infty)^d, (\boldsymbol{b}_r)_r = (b_r^{(1)}, \dots,$
$b_r^{(d)})_r \subset \mathbb{R}^d$ and $\boldsymbol{\gamma} = (\gamma^{(1)}, \dots, \gamma^{(d)}) \in \mathbb{R}$, such that, for any $s > 0$ and $j \in$
$\{1, \dots, d\}$,

$$\lim_{r\to\infty} \frac{a_{\lfloor rs\rfloor}^{(j)}}{a_r^{(j)}} = s^{\gamma^{(j)}} \qquad \lim_{r\to\infty} \frac{b_{\lfloor rs\rfloor}^{(j)} - b_r^{(j)}}{a_r^{(j)}} = \frac{s^{\gamma^{(j)}} - 1}{\gamma^{(j)}}, \tag{2.1}$$

where the second limit is interpreted as $\log(s)$ if $\gamma^{(j)} = 0$. Moreover, for $r \to \infty$,

$$\boldsymbol{Z}_r = (Z_r^{(1)}, \dots, Z_r^{(d)}) \longrightarrow_d \boldsymbol{Z} \sim G \tag{2.2}$$

where G denotes a d-variate extreme-value distribution with marginal c.d.f.s $G_{\gamma^{(1)}}, \ldots, G_{\gamma^{(d)}}$ and where

$$Z_r^{(j)} = \frac{\max(X_1^{(j)}, \ldots, X_r^{(j)}) - b_r^{(j)}}{a_r^{(j)}}, \quad j \in \{1, \ldots, d\}.$$

In the case $d = 1$, we omit all upper indexes; e.g., we write $\gamma = \gamma^{(1)}$. In the case $d \geq 2$, let C denote the unique extreme value copula associated with $\boldsymbol{Z}$. As is well-known, C can be written as

$$C(\boldsymbol{u}) = \exp\left(-L(-\log u^{(1)}, \ldots, -\log u^{(d)})\right) \tag{2.3}$$

for some *stable dependence function* $L \colon [0, \infty]^d \to [0, \infty]$, which satisfies

(L1) L is homogeneous: $L(s\boldsymbol{x}) = sL(\boldsymbol{x})$ for all $s > 0$ and all $\boldsymbol{x} \in [0, \infty]^d$;
(L2) $L(\boldsymbol{e}_j) = 1$ for $j = 1, \ldots, d$, where $\boldsymbol{e}_j$ denotes the j-th unit vector in $\mathbb{R}^d$;
(L3) $\max\left(x^{(1)}, \ldots, x^{(d)}\right) \leq L(\boldsymbol{x}) \leq x^{(1)} + \ldots + x^{(d)}$ for all $\boldsymbol{x} \in [0, \infty]^d$;
(L4) L is convex;

see, e.g., [21, 33].

Note that (2.1) and (2.2) may for instance be deduced from Leadbetter's $D(u_n)$ condition, a domain-of-attraction condition on the associated i.i.d. sequence with stationary distribution equal to that of $\boldsymbol{X}_0$ and a weak requirement on the convergence of the c.d.f. of $\boldsymbol{Z}_r$, see Theorem 10.22 in [1].

From now on, we assume to observe $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, an excerpt from a strictly stationary d-dimensional time series $(\boldsymbol{X}_t)_t$ satisfying Condition 1 (some generalizations will be discussed in Section 5). For block size parameter $r \ll n$, define componentwise block maxima of size r by

$$\boldsymbol{M}_{r,i} := \left(M_{r,i}^{(1)}, \ldots, M_{r,i}^{(d)}\right), \qquad M_{r,i}^{(j)} := \max\left\{X_i^{(j)}, \ldots, X_{i+r-1}^{(j)}\right\},$$

where $i \in \{1, \ldots, n - r + 1\}$ denotes the first observation within each block.

The traditional block maxima method is based on applying statistical methods to the sample of disjoint block maxima. The latter is given by $\mathcal{M}_{n,r}^{(\mathrm{db})} = \left(\boldsymbol{M}_{r,i} \colon i \in I_n^{\mathrm{db}}\right)$, where $I_n^{\mathrm{db}} := \{(i-1)r + 1 \colon 1 \leq i \leq m\}$ with $m = m_n := \lfloor n/r \rfloor$. Note that m is the number of disjoint blocks of size r that fit into the sampling period. Under Condition 1, the sample of disjoint block maxima is stationary and approximately follows the multivariate extreme value distribution G.

Instead of partitioning the observation period into disjoints blocks, one may alternatively slide the blocks through the observation period, thereby taking successive maxima of only one to the right instead of r. The resulting sliding block maxima sample is given by $\mathcal{M}_{n,r}^{(\mathrm{sb})} = \left(\boldsymbol{M}_{r,i} \colon i \in I_n^{\mathrm{sb}}\right)$, where $I_n^{\mathrm{sb}} := \{1, \ldots, n - r + 1\}$. Under Condition 1, the sliding block maxima sample is stationary as well, with approximate c.d.f. G. Hence, statistical methods that are based on estimating unknown expectations by empirical means are meaningful.

The case of classical empirical means has been treated in varying generality in [8, 41, 10]. It was found that estimators based on sliding block maxima are

typically more efficient than their disjoint block maxima counterparts, despite the fact that the sample $\mathcal{M}_{n,r}^{(\mathrm{sb})}$ is heavily dependent over time, even if $(\boldsymbol{X}_t)_t$ is an i.i.d. sequence. In this paper we generalize these results to U-statistics of order $p \in \mathbb{N}$, with $p = 1$ corresponding to classical empirical means.

More precisely, let $h \colon (\mathbb{R}^d)^p \to \mathbb{R}$ be a known symmetric measurable function of p d-dimensional input variables, subsequently referred to as a kernel of order p. The main objects of interest in this paper are the associated U-statistic of order p given by, for $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$,

$$U_{n,r}^{\mathrm{mb}} := U_{n,r}^{\mathrm{mb}}(h) := \binom{n_{\mathrm{mb}}}{p}^{-1} \sum_{(i_1,\ldots,i_p)\in J_n^{\mathrm{mb}}} h\big(\boldsymbol{M}_{r,i_1},\ldots,\boldsymbol{M}_{r,i_p}\big), \qquad (2.4)$$

where $n_{\mathrm{mb}} = |I_n^{\mathrm{mb}}|$ denotes the length of the block maxima sample (i.e., $n_{\mathrm{db}} = m$ if $\mathrm{mb} = \mathrm{db}$ and $n_{\mathrm{sb}} = n - r + 1$ if $\mathrm{mb} = \mathrm{sb}$) and where

$$J_n^{\mathrm{mb}} := J_n^{\mathrm{mb}}(p) := \{(i_1,\ldots,i_p) \in (I_n^{\mathrm{mb}})^p \colon i_1 < \cdots < i_p\}.$$

A standard heuristic argument suggests that, for the majority of summands in (2.4), the underlying block maxima can be considered as asymptotically independent. As a consequence, U_n^{mb} should be considered as an estimator for

$$\theta_r = \theta_r(h) := \int \cdots \int h(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_p)\,\mathrm{d}\mathbb{P}_{\boldsymbol{M}_{r,1}}(x_1)\ldots\mathrm{d}\mathbb{P}_{\boldsymbol{M}_{r,1}}(x_p)$$
$$= \mathbb{E}[h(\tilde{\boldsymbol{M}}_{r,1}^{(1)},\ldots,\tilde{\boldsymbol{M}}_{r,1}^{(p)})], \qquad (2.5)$$

where $\tilde{\boldsymbol{M}}_{r,1}^{(1)},\ldots,\tilde{\boldsymbol{M}}_{r,1}^{(p)}$ are i.i.d. copies of $\boldsymbol{M}_{r,1}$. We are interested in obtaining asymptotic results for the estimation error

$$U_{n,r}^{\mathrm{mb}}(h) - \theta_r(h)$$

in an asymptotic framework where $r = r_n \to \infty$ such that $r = o(n)$ for $n \to \infty$.

3. Limit theorems for U-statistics of block maxima

We start by introducing further conditions and notations. First, throughout the proofs we will use traditional blocking techniques relying on mixing coefficients. The latter are well-known to control the serial dependence of the underlying time series. A similar condition has been imposed in [6], among others.

Condition 2 (Block size and serial dependence). For the block size sequence $(r_n)_n$ it holds that, as $n \to \infty$,
   (a) $r_n \to \infty$ and $r_n = o(n)$.
   (b) There exists a sequence $(\ell_n)_n \subset \mathbb{N}$ such that $\ell_n \to \infty$, $\ell_n = o(r_n)$ and $\frac{r_n}{\ell_n}\alpha(\ell_n) = o(1)$ and $\frac{n}{r_n}\alpha(\ell_n) = o(1)$.
   (c) $\left(\frac{n}{r_n}\right)^{1+\omega}\beta(r_n) = o(1)$ for some $\omega > 0$.
Here, α and β denote the alpha- and beta-mixing coefficients, see [5] for exact definitions and basic properties. Subsequently, we often write $r = r_n$ and $\ell = \ell_n$.

The expectation and higher order moments of $h(\boldsymbol{M}_{r,i_1}, \ldots, \boldsymbol{M}_{r,i_p})$ in (2.4) will be controlled by uniform integrablity and by relying on the convergence of rescaled block maxima from Condition 1. For that purpose, we need the kernel function h to behave well under location-scale transformations; see also [36], Chapter 5, and [31] for a similar, slightly more restrictive assumption.

Condition 3 (Location-scale property of the kernel function). There exist functions $f\colon (\mathbb{R}^d)^p \to (0, \infty)$, $\ell\colon (\mathbb{R}^d)^p \to \mathbb{R}$ such that, for all $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p, \boldsymbol{b} \in \mathbb{R}^d$ and $\boldsymbol{a} \in (0, \infty)^d$,

$$h\left(\frac{\boldsymbol{x}_1 - \boldsymbol{b}}{\boldsymbol{a}}, \ldots, \frac{\boldsymbol{x}_p - \boldsymbol{b}}{\boldsymbol{a}}\right) = \frac{h(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)}{f(\boldsymbol{a}, \boldsymbol{b})} - \ell(\boldsymbol{a}, \boldsymbol{b}), \qquad (3.1)$$

where $\boldsymbol{x}/\boldsymbol{y} := \left(x^{(1)}/y^{(1)}, \ldots, x^{(d)}/y^{(d)}\right)$ for $\boldsymbol{x} \in \mathbb{R}^d, \boldsymbol{y} \in (0, \infty)^d$.

Example 3.1. Condition 3 is met for the following kernel functions. Note that the kernels in (5) to (7) may be used to construct tests for stochastic independence; see, for instance, [29]. In the current case, this corresponds to testing asymptotic independence of the coordinates of $\boldsymbol{X}_1$.
(1) The mean kernel: $h(x) = x$ with $d = 1$, $p = 1$, $f(a, b) = a$, $\ell(a, b) = b$.
(2) The variance kernel: $h(x, y) = (x - y)^2/2$ with $d = 1$, $p = 2$, $f(a, b) = a^2$, $\ell \equiv 0$.
(3) Gini's mean difference kernel: $h(x, y) = |x - y|/2$ with $d = 1, p = 2, f(a, b) = a, \ell \equiv 0$.
(4) The modified probability weighted moment kernel of degree $k \in \mathbb{N}$ (see also Section 4.2): $h_k(x_1, \ldots, x_k) = \max\{x_1, \ldots, x_k\}/k$ with $d = 1$, $p = k$, $f(a, b) = a$, $l(a, b) = \frac{1}{k} \cdot \frac{b}{a}$.
(5) Kendall's τ kernel: $h(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{1}\{(x^{(1)} - y^{(1)})(x^{(2)} - y^{(2)}) > 0\}$ with $d = 2$, $p = 2$, $f \equiv 1$, $\ell \equiv 0$.
(6) Spearman's ρ kernel: $h(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) = 2^{-1} \sum_{\pi \in S_3} \operatorname{sgn}(x_{\pi_1}^{(1)} - x_{\pi_2}^{(1)}) \operatorname{sgn}(x_{\pi_1}^{(2)} - x_{\pi_3}^{(2)})$ with $d = 2$, $p = 3$, $f \equiv 1, \ell \equiv 0$ and where S_n denotes the symmetric group of order n.
(7) Hoeffding's D kernel and Bergsma and Dassio's t^* kernel: we refer to [29] for the kernel definition, which satisfy $d = 2, f \equiv 1, \ell \equiv 0$ and $p = 4$ and $p = 5$, respectively.

From now on, for the ease of notation, we only consider the case $p = 2$ (see also [14], among others). For $i \in \{1, \ldots, n - r + 1\}$, let

$$\boldsymbol{Z}_{r,i} := \left(Z_{r,i}^{(1)}, \ldots, Z_{r,i}^{(d)}\right), \qquad Z_{r,i}^{(j)} := (M_{r,i}^{(j)} - b_r^{(j)})/a_r^{(j)}.$$

with $\boldsymbol{a}_r$ and $\boldsymbol{b}_r$ from Condition 1. Note that $(\boldsymbol{Z}_{r,i})_i$ is stationary with $\boldsymbol{Z}_{r,1} \rightsquigarrow G$ as $n \to \infty$. Further, under Condition 3 one has

$$h\left(\boldsymbol{M}_{r,i}, \boldsymbol{M}_{r,j}\right) = f(\boldsymbol{a}_r, \boldsymbol{b}_r)\{h\left(\boldsymbol{Z}_{r,i}, \boldsymbol{Z}_{r,j}\right) + \ell(\boldsymbol{a}_r, \boldsymbol{b}_r)\}, \qquad (3.2)$$

which will ultimately allow to deduce asymptotic results on $U_{n,r}^{\mathrm{mb}}$ defined in (2.4)

from respective results on

$$
U_{n,r,Z}^{\mathrm{mb}} := U_{n,r,Z}^{\mathrm{mb}}(h) := \binom{n_{\mathrm{mb}}}{2}^{-1} \sum_{(i,j) \in J_n^{\mathrm{mb}}} h(\boldsymbol{Z}_{r,i}, \boldsymbol{Z}_{r,j}). \tag{3.3}
$$

Heuristically, the expectation of $U_{n,Z}^{\mathrm{mb}}$ is close to

$$
\vartheta_r = \vartheta_r(h) = \int \int h(\boldsymbol{x}, \boldsymbol{y}) \, \mathrm{d}\mathbb{P}_{\boldsymbol{Z}_{r,1}}(\boldsymbol{x}) \, \mathrm{d}\mathbb{P}_{\boldsymbol{Z}_{r,1}}(\boldsymbol{y}) = \mathbb{E}[h(\boldsymbol{Z}_{r,1}, \tilde{\boldsymbol{Z}}_{r,1})] \tag{3.4}
$$

with $\tilde{\boldsymbol{Z}}_{r,1}$ an independent copy of $\boldsymbol{Z}_{r,1}$. The sequence ϑ_r in turn converges to

$$
\vartheta := \mathbb{E}[h(\boldsymbol{Z}, \tilde{\boldsymbol{Z}})], \tag{3.5}
$$

under suitable integrability assumptions; here $\boldsymbol{Z}, \tilde{\boldsymbol{Z}} \sim G$ are independent (see Lemma 8.1 below). The necessary integrability condition, which will also ensure convergence of higher order moments, is as follows.

Condition 4 (Asymptotic integrability). There exists a $\nu > 2/\omega$ with ω from Condition 2 such that:

(a) $\limsup_{r\to\infty} \int \int |h(\boldsymbol{x}, \boldsymbol{y})|^{2+\nu} \, \mathrm{d}\mathbb{P}_{\boldsymbol{Z}_{r,1}}(\boldsymbol{x}) \, \mathrm{d}\mathbb{P}_{\boldsymbol{Z}_{r,1}}(\boldsymbol{y}) < \infty$,
(b) $\limsup_{r\to\infty} \sup_{s\in\mathbb{N}} \int |h(\boldsymbol{x}, \boldsymbol{y})|^{2+\nu} \, \mathrm{d}\mathbb{P}_{(\boldsymbol{Z}_{r,1}, \boldsymbol{Z}_{r,1+s})}(\boldsymbol{x}, \boldsymbol{y}) < \infty$.

Note that the two moment assumptions may be understood as an asymptotic formulation of uniform moments as used in [15]. In many situations, the conditions are easily satisfied, see, e.g., Section 4. Finally, for kernels of higher order than $p = 2$, more complicated versions of this condition will be needed, see [40].

Additional notation is needed to formulate the asymptotic limit results for $U_{n,r}^{\mathrm{mb}}$. Recall G from Condition 1. Let L denote the stable tail dependence function of G if $d \geq 2$, and the identity on $[0, \infty]$ if $d = 1$. For $\boldsymbol{u}, \boldsymbol{v} \in [0, 1]^d$ and $\xi \geq 0$, let

$$
C_\xi(\boldsymbol{u}, \boldsymbol{v}) = \exp\Big[-L_\xi\Big(-\log u^{(1)}, \ldots, -\log u^{(d)}, -\log v^{(1)}, \ldots, -\log v^{(d)} \Big) \Big], \tag{3.6}
$$

where, for $\boldsymbol{x}, \boldsymbol{y} \in [0, \infty]^d$,

$$
\begin{aligned}
L_\xi(\boldsymbol{x}, \boldsymbol{y}) := (\xi \wedge 1) \cdot \Big\{ L(x^{(1)}, \ldots, x^{(d)}) + L\Big(y^{(1)}, \ldots, y^{(d)}\Big) \Big\} \\
+ (1 - (\xi \wedge 1)) \cdot L\Big(x^{(1)} \vee y^{(1)}, \ldots, x^{(d)} \vee y^{(d)}\Big). \tag{3.7}
\end{aligned}
$$

As shown in Lemma 8.6 below, C_ξ defines a $2d$-variate extreme-value copula with stable tail dependence function L_ξ. Let G_ξ denote the $2d$-variate extreme value distribution with copula C_ξ and margins $G_{\gamma^{(1)}}, \ldots, G_{\gamma^{(d)}}, G_{\gamma^{(1)}}, \ldots, G_{\gamma^{(d)}}$, i.e.,

$$
G_\xi(\boldsymbol{x}, \boldsymbol{y}) = C_\xi\Big\{ G_{\gamma^{(1)}}(x^{(1)}), \ldots, G_{\gamma^{(d)}}(x^{(d)}), G_{\gamma^{(1)}}(y^{(1)}), \ldots, G_{\gamma^{(d)}}(y^{(d)}) \Big\} \tag{3.8}
$$

for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$. Note that $G_\xi(\boldsymbol{x}, \boldsymbol{y}) = G(\boldsymbol{x})G(\boldsymbol{y})$ for $\xi > 1$. Further, G_ξ is the multivariate analogue of $G_{\alpha_0, \xi}$ in Formula (5.1) in [8] and, in the case $d = 1$, also appeared in Formula (13) in [10].

Finally, for $(\boldsymbol{Z}_{1,\xi}, \boldsymbol{Z}_{2,\xi}) \sim G_\xi$, let

$$\sigma_{\mathrm{db}}^2 := 4 \operatorname{Var}(h_1(\boldsymbol{Z})), \qquad \sigma_{\mathrm{sb}}^2 := 8 \int_0^1 \operatorname{Cov}\big(h_1(\boldsymbol{Z}_{1,\xi}), h_1(\boldsymbol{Z}_{2,\xi})\big)\, \mathrm{d}\xi. \qquad (3.9)$$

where,

$$h_1 \colon \mathbb{R}^d \to \mathbb{R},\ h_1(\boldsymbol{z}) := \mathbb{E}[h(\boldsymbol{z}, \boldsymbol{Z})] - \vartheta \qquad (3.10)$$

with $\boldsymbol{Z} \sim G$ and ϑ from (3.5). The following result is the main result of this paper. Here and throughout, $\boldsymbol{\lambda}^{2d}$ denotes the Lebesgue measure on $\mathbb{R}^{2d}$.

Theorem 3.2. *Suppose Conditions 1, 2, 3 and 4 are met. Furthermore let h be $\boldsymbol{\lambda}^{2d}$-a.e. continuous and bounded on compact sets. Then, for* $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$,

$$\frac{\sqrt{m}}{f(\boldsymbol{a}_r, \boldsymbol{b}_r)} \cdot (U_{n,r}^{\mathrm{mb}} - \theta_r) \rightsquigarrow \mathcal{N}\big(0, \sigma_{\mathrm{mb}}^2\big),$$

with θ_r from (2.5) and σ_{mb}^2 from (3.9). Moreover, $\sigma_{\mathrm{sb}}^2 \le \sigma_{\mathrm{db}}^2$.

Note that, under Condition 3, $\theta_r = f(\boldsymbol{a}_r, \boldsymbol{b}_r)\{\vartheta_r + \ell(\boldsymbol{a}_r, \boldsymbol{b}_r)\}$ with ϑ_r from (3.4). In certain situations (in particular when $\ell = 0$ and $f \equiv \mathrm{const}$; see, e.g., Kendall's tau), one may be willing to regard $U_{n,r}^{\mathrm{mb}}$ as an estimator for the asymptotic analogue

$$\tilde{\vartheta}_r = f(\boldsymbol{a}_r, \boldsymbol{b}_r)\{\vartheta + \ell(\boldsymbol{a}_r, \boldsymbol{b}_r)\}. \qquad (3.11)$$

For instance, in case of the variance kernel (see also Section 4.1), $\tilde{\vartheta}_r$ is the variance of the $\mathrm{GEV}(b_r, a_r, \gamma)$-distribution, which is exactly the GEV-distribution approximating the distribution of $M_{r,1}$, see Assumption 1. Under an additional bias condition, we may deduce the following result on the estimation error.

Corollary 3.3. *Additionally to the assumptions made in Theorem 3.2, suppose that the limit $B = \lim_{n \to \infty} B_n$ exists, where*

$$B_n := \sqrt{m}(\vartheta_r - \vartheta). \qquad (3.12)$$

Then, for $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$,

$$\frac{\sqrt{m}}{f(\boldsymbol{a}_r, \boldsymbol{b}_r)} \cdot (U_{n,r}^{\mathrm{mb}} - \tilde{\vartheta}_r) \rightsquigarrow \mathcal{N}\big(B, \sigma_{\mathrm{mb}}^2\big),$$

with σ_{mb}^2 from (3.9) and $\tilde{\vartheta}_r$ from (3.11).

Remark 3.4 (Generalizations). Using the Cramér-Wold Theorem it is possible to generalize the limit theorems to the case of joint convergence involving a finite number of kernel functions. Moreover, as mentioned before and at the cost

of a more complicated notation, one might extend the results to higher kernel degrees $p \in \mathbb{N}$. Joint weak convergence then even holds for kernels of different degrees. These generalizations allow, for example, to handle the joint convergence of probability weighted moments estimators of different order, which would be needed to deduce the asymptotics of the PWM-estimator for the parameters of the GEV-distribution. Further generalizations concerning different model assumptions are worked out in Section 5 and Section B in the supplement.

Remark 3.5 (A bias-corrected version of the sliding blocks estimator). In view of Lemma 8.6 below, the block maxima $\boldsymbol{M}_{r,i}$ and $\boldsymbol{M}_{r,j}$ are asymptotically independent for $|i - j| \geq r$ and asymptotically dependent otherwise. As a consequence, the summands $h(\boldsymbol{M}_{r,i}, \boldsymbol{M}_{r,j})$ with $|i - j| < r$ induce a *dependency bias*, which suggests to replace $U_{n,r}^{\mathrm{sb}}$ by

$$\tilde{U}_{n,r}^{\mathrm{sb}} := \binom{\tilde{n}_{\mathrm{sb}}}{2}^{-1} \sum_{(i,j) \in \tilde{J}_n^{\mathrm{sb}}} h(\boldsymbol{M}_{r,i}, \boldsymbol{M}_{r,j}),$$

where $\tilde{J}_n^{\mathrm{sb}} = \{(i,j) \in (I_n^{\mathrm{sb}})^2 \colon j - i \geq r\}$; see also [10], Remark 3.1. Note that $|\tilde{J}_n^{\mathrm{sb}}| = \binom{n_{\mathrm{sb}} - r + 1}{2} = \binom{n - 2r + 2}{2}$ and $|J_n^{\mathrm{sb}} \setminus \tilde{J}_n^{\mathrm{sb}}| = O(nr)$, which can used to show that $\frac{\sqrt{m}}{f(a,b)}(\tilde{U}_{n,r}^{\mathrm{sb}} - U_{n,r}^{\mathrm{sb}}) = O_{\mathbb{P}}\big((1 + \ell(\boldsymbol{a}_r, \boldsymbol{b}_r))m^{-1/2}\big)$ with ℓ from Condition 3. Hence, the two estimators are typically asymptotically equivalent. However, in finite-sample situations, the bias-reduction may actually be superimposed by an increase in estimation variance (see Section D in the supplement), whence we cannot recommend its usage in general (this is akin to the recommendation in [10], Section E.3).

4. Examples

Details are worked out for specific kernel functions of interest.

4.1. Variance estimation

The variance is one of the most fundamental parameters to describe a distribution of interest, which, in our case, is $\sigma_r^2 := \mathrm{Var}(M_{r,1})$, where $(X_t)_t$ is a univariate time series. The respective empirical variance, based on either disjoint or sliding block maxima, is given by

$$\hat{\sigma}^2_{n,r,\mathrm{mb}} = \frac{1}{n_{\mathrm{mb}} - 1} \sum_{i \in I_n^{\mathrm{mb}}} \big(M_{r,i} - \overline{M}_r^{\mathrm{mb}}\big)^2, \qquad \mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\},$$

where $\overline{M}_r^{\mathrm{mb}} := n_{\mathrm{mb}}^{-1} \sum_{i \in I_n^{\mathrm{mb}}} M_{r,i}$. As is well-known, the empirical variance can be written as a U-statistic of order $p = 2$, that is,

$$\hat{\sigma}^2_{n,r,\mathrm{mb}} = U_{n,r}^{\mathrm{mb}}(h_{\mathrm{Var}}), \qquad h_{\mathrm{Var}}(x,y) = (x - y)^2/2.$$

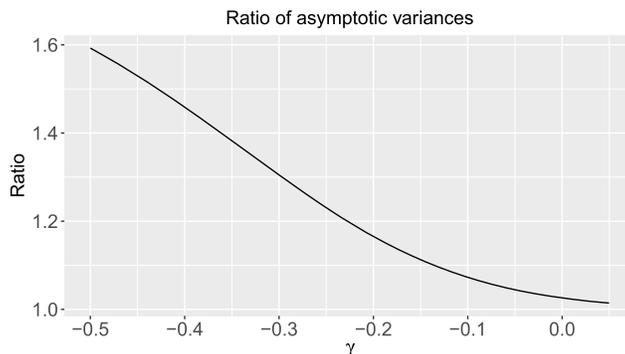The following result is a direct consequence of Theorem 3.2.

Ratio of asymptotic variances

FIG 1. *Graph of $\gamma \mapsto \sigma_{\mathrm{db}}^2/\sigma_{\mathrm{sb}}^2$ with σ_{mb}^2 as in (C.1) and (C.2).*

Corollary 4.1. *Let $d = 1$ and suppose that Conditions 1 and 2 are met with $\gamma < 1/4$. Moreover, assume that there exists a constant $\nu > 2/\omega$ such that $\limsup_r \mathbb{E}|Z_{r,1}|^{4+\nu} < \infty$. Then*

$$\frac{\sqrt{m}}{a_r^2}\left(\hat{\sigma}_{n,r,\mathrm{mb}}^2 - \sigma_r^2\right) \rightsquigarrow \mathcal{N}\left(0, \sigma_{\mathrm{mb}}^2\right),$$

where σ_{db}^2 and σ_{sb}^2 only depend on the tail index γ. Explicit formulas are provided in (C.1) and (C.2) in the supplement, respectively. Moreover, $\sigma_{\mathrm{sb}}^2 < \sigma_{\mathrm{db}}^2$.

The assumption $\gamma < 1/4$ is natural, as asymptotic normality results on empirical variances require finite fourth moments; in the case of the GEV-distribution, this exactly corresponds to $\gamma < 1/4$. Figure 1 shows the ratio of the asymptotic variances, $\sigma_{\mathrm{db}}^2/\sigma_{\mathrm{sb}}^2$ as a function of γ. We observe that the estimator based on sliding blocks has a significantly smaller variance for negative γ, say $\gamma < -.25$, while hardly any difference is visible for positive γ.

The previous results may be made more explicit when imposing a specific time series model. We exemplary work out details for a marginal transformed version of the ARMAX-model. The model is defined as follows: for an i.i.d. sequence $(W_t)_{t\in\mathbb{Z}}$ of Fréchet(1) distributed random variables and $\alpha \in (0,1]$, consider the ARMAX(1) recursion defined as

$$Y_t = \max\left\{\alpha Y_{t-1}, (1-\alpha)W_t\right\}, \quad t \in \mathbb{Z}. \tag{4.1}$$

The recursion has the stationary solution $Y_t := \max_{j\geq 0}(1-\alpha)\alpha^j W_{t-j}$, which has Fréchet(1) distributed marginals and extremal index $\theta = 1-\alpha$, see Example 10.5 in [1]. Define X_t as the transformed random variables $X_t := F_\gamma^{\leftarrow}(F_W(Y_t))$, where F_W is the c.d.f. of a Fréchet(1) distribution, F_γ is the c.d.f. of the Pareto family defined as

$$F_\gamma(x) := \begin{cases} \left(1 - (1+\gamma x)^{-1/\gamma}\right)\mathbf{1}\{x \geq 0\}, & \gamma > 0 \\ \left(1 - (1+\gamma x)^{-1/\gamma}\right)\mathbf{1}\{0 \leq x \leq -1/\gamma\}, & \gamma < 0, \\ \left(1 - \exp(-x)\right)\mathbf{1}\{x \geq 0\}, & \gamma = 0 \end{cases} \tag{4.2}$$

and where $F^{\leftarrow}$ is the left continuous generalized inverse of F. By [3] and [5] the untransformed time series $(Y_t)_t$ is exponentially β-mixing, which implies the same for $(X_t)_t$. This results in a large spectrum of choices for r_n and ℓ_n satisfying Condition 2, which can hence be regarded as non-restrictive. We will prove in Section 7.2 that, if $\gamma < 1/4$ and if $r = o(n), n = o(r^3)$, all assumptions from Corollary 4.1 are met, with $a_r = (r(1-\alpha))^\gamma$ and $b_r = \{(r(1-\alpha))^\gamma - 1\}/\gamma$. Hence,

$$\frac{\sqrt{m}}{(r(1-\alpha))^{2\gamma}}(\hat{\sigma}_{n,r,\mathrm{mb}}^2 - \sigma_r^2) \rightsquigarrow \mathcal{N}\big(0, \sigma_{\mathrm{mb}}^2\big) \tag{4.3}$$

as asserted. Moreover, one may show that the bias condition is met with $B = 0$, whence σ_r^2 may be replaced by $(r(1-\alpha))^{2\gamma}\tau^2(\gamma)$, where $\tau^2(\gamma) := \frac{1}{\gamma}\{\Gamma(1-2\gamma) - \Gamma(1-\gamma)^2\}\mathbf{1}(\gamma \in (-\infty, 1/2)\backslash\{0\}) + (\pi^2/6)\mathbf{1}(\gamma = 0)$ is the variance of the $\mathrm{GEV}(\gamma)$ distribution, with $\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}\,\mathrm{d}t$, $x > 0$, the Gamma function.

4.2. The probability weighted moment estimator

Let M be a random variable following a GEV distribution with parameter $\eta = (\mu, \sigma, \gamma)' \in \mathbb{R} \times (0, \infty) \times (-\infty, 1)$ and cdf G_η. For $k \in \mathbb{N}_0$, the kth probability weighted moment (PWM) of M is given by

$$\beta_{\eta,k} := \mathbb{E}[MG_\eta^k(M)]. \tag{4.4}$$

It is well-known that η is a one-to-one function of the first three probability weighted moments [23]. Replacing the moments in (4.4) by suitable estimators and plugging those into the one-to-one function results in (the) PWM estimator for η. One version, as proposed in [30], is given by

$$\tilde{\beta}_0 := \frac{1}{n}\sum_{i=1}^n M_i, \quad \tilde{\beta}_k := \frac{1}{n}\sum_{i=1}^n \frac{(i-1)\cdot\ldots\cdot(i-k)}{(n-1)\cdot\ldots\cdot(n-k)}M_{(i)}, \quad k \geq 1 \tag{4.5}$$

where $\mathcal{M} = (M_1, \ldots, M_n)$ is a sample of random variables distributed as M and $M_{(1)} \leq \ldots \leq M_{(n)}$ is the ordered sample. If $\mathcal{M}$ is an i.i.d. sample, then there are no ties with probability 1, whence $\tilde{\beta}_k = \hat{\beta}_k$, where, for $k \in \mathbb{N}$,

$$\hat{\beta}_{k-1} = \binom{n}{k}^{-1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} h_{\mathrm{pwm},k}(M_{i_1}, \ldots, M_{i_k}) \tag{4.6}$$

with the permutation invariant kernel function

$$h_{\mathrm{pwm},k}(x_1, \ldots, x_k) := \frac{1}{k}\sum_{j=1}^k \mathbf{1}\Big\{\max_{1 \leq i \leq k, i \neq j} x_i \leq x_j\Big\}x_j \tag{4.7}$$

Clearly, $\hat{\beta}_{k-1}$ is a U-statistic of order k that is unbiased for $\mathbb{E}[\hat{\beta}_{k-1}] = \beta_{\eta,k-1}$ in case the sample is i.i.d.

In this section, we apply Theorem 3.2 to derive limit results for the estimator

$$\hat{\beta}_{k-1}^{\mathrm{mb}} = U_{n,r}^{\mathrm{mb}}(h_{\mathrm{pwm},k})$$

with $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$. For simplicity, we restrict attention to the case $k = 2$, which yields a U-statistic of order 2. Since the function $h_{\mathrm{pwm},2}$ does not satisfy Condition 3, we will need the modified kernel function $\tilde{h}_{\mathrm{pwm},2}(x,y) := \max(x,y)/2$ from Example 3.1.

Proposition 4.2. *Let $d = 1$ and suppose that $(X_t)_{t\in\mathbb{Z}}$ satisfies Condition 1, does not contain ties with probability one, and that Condition 2 is met. If there exists $\nu > 2/\omega$ such that $\limsup_{r\to\infty} \mathbb{E}|Z_{r,1}|^{2+\nu} < \infty$, then, for $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$,*

$$\frac{\sqrt{m}}{a_r} \left\{ U_{n,r}^{\mathrm{mb}}(h_{\mathrm{pwm},2}) - U_{n,r}^{\mathrm{mb}}(\tilde{h}_{\mathrm{pwm},2}) \right\} \xrightarrow{L^2} 0.$$

If, moreover, the limit $B := \lim_n B_n$ exists, where $B_n = \sqrt{m}\,\mathbb{E}[F_r(Z_{r,1})Z_{r,1} - G_\gamma(Z)Z]$ with F_r the c.d.f. of $Z_{r,1}$ and $Z \sim G_\gamma$, then

$$\sqrt{m} \left(\frac{\hat{\beta}_1^{\mathrm{mb}} - \beta_{(b_r,a_r,\gamma),1}}{a_r} \right) \rightsquigarrow \mathcal{N}\left(B, \sigma_{\mathrm{mb}}^2\right),$$

with $0 < \sigma_{\mathrm{sb}}^2 < \sigma_{\mathrm{db}}^2$.

Note that similar asymptotics have also been worked out in [10] and [19]. In [10], who also provide explicit formulas for the asymptotic variances, the derivation was based on explicit expansions of the kernel function involving empirical cumulative distribution functions. Comparing our result with their Theorem 3.5, we observe that our result is slightly more restrictive, since we impose β-mixing rather than α-mixing. An extension to α-mixing is given in Section B in the supplement. [19] only consider the i.i.d. case and the disjoints blocks estimator. Under this setting and using an approach based on the quantile process of the block maxima sample, they were able to provide explicit expressions for the asymptotic bias under a mild and natural second order condition; see their Theorem 2.2. They also provide an alternative representation for the asymptotic normal distribution.

4.3. Estimation of Kendall's tau

Kendall's tau statistic is a well-known nonparametric distribution-free measure of rank correlation that quantifies the degree of association between two variables [26]. The population version $\tau = \tau(\boldsymbol{X})$ for a bivariate vector $\boldsymbol{X} = (X^{(1)}, X^{(2)})$ is defined as follows: for i.i.d. copies $\boldsymbol{X}_1, \boldsymbol{X}_2$ of $\boldsymbol{X}$, we have $\tau := \pi_c - \pi_d = 2\pi_c - 1$, where $\pi_c := \mathbb{P}((X_1^{(1)} - X_2^{(1)})(X_1^{(2)} - X_2^{(2)}) > 0)$ and $\pi_d := \mathbb{P}((X_1^{(1)} - X_2^{(1)})(X_1^{(2)} - X_2^{(2)}) < 0)$ denote the probabilities of concordance and discordance of $\boldsymbol{X}_1, \boldsymbol{X}_2$, respectively. Applied to bivariate extreme

value distributions, Kendall's tau provides a useful summary of extremal dependence; see [1, pp. 274–275] and the references therein.

For a bivariate sample $(\boldsymbol{X_1}, \ldots, \boldsymbol{X_n})$ Kendall's τ-statistic can be written as $\hat{\tau}_n = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \{2h_\tau(\boldsymbol{X}_i, \boldsymbol{X}_j) - 1\}$ with h_τ as in Example 3.1(5). For mb $\in \{\text{db}, \text{sb}\}$, let $\hat{\tau}_{n,r}^{\text{mb}}$ denote Kendall's τ-statistic applied to the sample of disjoint or sliding block maxima. An application of Theorem 3.2 yields the following result.

Proposition 4.3. *Let $d = 2$ and suppose Conditions 1 and 2 are met. Then, with $\tau_r := \tau(M_{r,1}^{(1)}, M_{r,1}^{(2)})$, we have, for mb $\in \{\text{db}, \text{sb}\}$,*

$$\sqrt{m}\big(\hat{\tau}_{n,r}^{\text{mb}} - \tau_r\big) \rightsquigarrow \mathcal{N}\big(0, \sigma_{\text{mb}}^2\big),$$

where the asymptotic variances can be represented as a function of the extreme-value copula C from (2.3) as follows:

$$\sigma_{\text{db}}^2 = 16 \bigg\{ \int_{[0,1]^2} \big\{C(\boldsymbol{u}) + \bar{C}(\boldsymbol{u})\big\}^2 \, \mathrm{d}C(\boldsymbol{u}) - 4\Big(\int_{[0,1]^2} C(\boldsymbol{u}) \, \mathrm{d}C(\boldsymbol{u}) \Big)^2 \bigg\},$$

$$\sigma_{\text{sb}}^2 = 32 \int_0^1 \bigg(\int_{[0,1]^2 \times [0,1]^2} \big\{C(\boldsymbol{u}) + \bar{C}(\boldsymbol{u})\big\}\big\{C(\boldsymbol{v}) + \bar{C}(\boldsymbol{v})\big\} \, \mathrm{d}C_\xi(\boldsymbol{u}, \boldsymbol{v})$$

$$- 4\Big(\int_{[0,1]^2} C(\boldsymbol{u}) \, \mathrm{d}C(\boldsymbol{u}) \Big)^2 \bigg) \, \mathrm{d}\xi,$$

where $\bar{C}(\boldsymbol{u}) = 1 - u^{(1)} - u^{(2)} + C(1 - u^{(1)}, 1 - u^{(2)})$ denotes the survival copula of C.

For the case $C = \Pi$, where Π denotes the independence copula, one can show that $\sigma_{\text{db}}^2 = 4/9, \sigma_{\text{sb}}^2 = 32(7/12 - 2\log(4/3))$ resulting in $\sigma_{\text{db}}^2/\sigma_{\text{sb}}^2 \approx 1.7428$, which has also been validated in a simulation experiment.

Remark 4.4 (Treating $\hat{\tau}_{n,r}^{\text{mb}}$ as an estimator $\tau = \tau(C)$). Proposition 4.3 quantifies the (asymptotic) estimation error when treating $\hat{\tau}_{n,r}^{\text{mb}}$ as an estimator for τ_r. In some situations one may rather be interested in treating $\hat{\tau}_{n,r}^{\text{mb}}$ as an estimator for $\tau = \tau(C)$, with C the max-attractor copula from Condition 1. In that case, a bias term may show up, which can be calculated more explicitly under some suitable second-oder conditions.

It is instructive to start with the i.i.d. case: let D denote the copula of $\boldsymbol{X}_1 = (X_1^{(1)}, X_1^{(2)})$, such that the copula C_r of the block maximum distribution with block size r satisfies $C_r(u^{(1)}, u^{(2)})^{1/r} = D((u^{(1)})^{1/r}, (u^{(2)})^{1/r})$. Condition 1 implies that, for all $\boldsymbol{u} = (u^{(1)}, u^{(2)}) \in [0,1]^2$, $\lim_{r \to \infty} C_r(\boldsymbol{u}) = C(\boldsymbol{u})$. A natural second order condition [9] reads as follows: suppose there exists a function $\varphi : (0, \infty) \to (0, \infty)$ with $\lim_{r \to \infty} \varphi(r) = 0$ and a non-null-function S on $[0,1]^2$ such that

$$\lim_{r \to \infty} \frac{C_r(\boldsymbol{u}) - C(\boldsymbol{u})}{\varphi(r)} = S(\boldsymbol{u}), \quad \boldsymbol{u} \in [0,1]^2, \tag{4.8}$$

where the convergence is uniform on $[\delta, 1]^2$, for all $\delta > 0$. As shown in [9], under some mild assumptions, the condition is equivalent to common second order conditions imposed on the stable tail dependence function L, and holds for selected copula families. Moreover, the convergence in (4.8) is necessarily uniform on $[0, 1]^2$ and the function φ is regularly varying of some order $\rho \le 0$.

Next, recall that $\tau_r = 4 \int C_r \, \mathrm{d}C_r - 1$. A standard argument shows that

$$\lim_{r \to \infty} \int \frac{C_r - C}{\varphi(r)} \, \mathrm{d}(C_r - C) = 0.$$

Further, $\int C \, \mathrm{d}C_r = \int C_r \, \mathrm{d}C$; this equality holding for all pairs of copulas. As a consequence,

$$\begin{aligned}
\frac{1}{4}(\tau_r - \tau) &= \int C_r \, \mathrm{d}C_r - \int C \, \mathrm{d}C \\
&= \int (C_r - C) \, \mathrm{d}(C_r - C) + \int (C_r - C) \, \mathrm{d}C + \int C \, \mathrm{d}(C_r - C) \\
&= o(\varphi(r)) + 2 \int (C_r - C) \, \mathrm{d}C.
\end{aligned}$$

Overall, we obtain that

$$\lim_{r \to \infty} \frac{\tau_r - \tau}{\varphi(r)} = 8 \int S(\boldsymbol{u}) \, \mathrm{d}C(\boldsymbol{u}).$$

Hence, if the block size $r = r_n$ is chosen in such a way that the limit $\lambda_0 := \lim_{n \to \infty} \sqrt{m} \varphi(r) \ge 0$ exists, we obtain, under the conditions of Proposition 4.3 and in the i.i.d. case,

$$\sqrt{m}(\hat{\tau}_{n,r}^{\mathrm{mb}} - \tau) = \sqrt{m}(\hat{\tau}_{n,r}^{\mathrm{mb}} - \tau_r) + \sqrt{m}\varphi(r)\frac{\tau_r - \tau}{\varphi(r)} \rightsquigarrow \mathcal{N}(\lambda_0 B, \sigma_{\mathrm{mb}}^2),$$

where $B = 8 \int S(\boldsymbol{u}) \, \mathrm{d}C(\boldsymbol{u})$.

Treating the block size r as a tuning parameter, the previous result allows to make statements on rate-optimal choices of r. Indeed, assuming that $\varphi(r) = r^\rho$ with $\rho < 0$ for simplicity (see [9] for examples), the previous equation can be restated as

$$\hat{\tau}_{n,r}^{\mathrm{mb}} - \tau \approx_d \mathcal{N}\left(r^\rho B, \frac{\sigma_{\mathrm{mb}}^2}{m}\right).$$

Minimizing the (asymptotic) MSE defined as $r^{2\rho}B^2 + \sigma_{\mathrm{mb}}^2/m$ with respect to r, we obtain the 'optimal choice' $r \propto n^{-2\rho/(1-2\rho)}$, with the (asymptotic) MSE then being of the order $n^{\rho/(1-2\rho)}$.

Finally, note that the previous argumentation remains true in the time series case, provided we require that the convergence in (4.8) holds uniformly on $[0, 1]^2$. However, it will typically be much harder to calculate C_r, let alone to obtain asymptotic expansions.

5. Extensions to piecewise stationarity

Environmental data typically involve different forms of non-stationarity. A particular source is seasonality, which may statistically be approached by restricting attention to seasons rather than years, bearing in mind that the inner-season variability should be approximately stationary. This idea may be approached mathematically by working with data satisfying the following assumption taken from [10].

Condition 5 (Piecewise stationary observation scheme). For sample size $n \in \mathbb{N}$, we have observations $\boldsymbol{X}_{n,1}, \ldots, \boldsymbol{X}_{n,n}$ taking values in $\mathbb{R}^d$. Moreover, for some block length sequence $(r_n)_n \subset \mathbb{N}$ diverging to infinity such that $r_n = o(n)$, we have

$$(\boldsymbol{X}_{n,1}, \ldots, \boldsymbol{X}_{n,n}) = (\boldsymbol{Y}_{1,1}, \ldots, \boldsymbol{Y}_{1,r_n}, \boldsymbol{Y}_{2,1}, \ldots, \boldsymbol{Y}_{2,r_n}, \ldots$$
$$\ldots, \boldsymbol{Y}_{n_{\mathrm{db}},1}, \ldots, \boldsymbol{Y}_{n_{\mathrm{db}},r_n}, \boldsymbol{Y}_{n_{\mathrm{db}}+1,1}, \ldots, \boldsymbol{Y}_{n_{\mathrm{db}}+1,n-n_{\mathrm{db}}r_n}),$$

where $n_{\mathrm{db}} = \lfloor n/r_n \rfloor$ and where $(\boldsymbol{Y}_{1,t})_t, (\boldsymbol{Y}_{2,t})_t, \ldots$ denote i.i.d. copies from a stationary time series satisfying Condition 1 with continuous marginal c.d.f. F. Note that $\boldsymbol{Y}_{j,t}$ should be regarded as the t-th observation in the j-th season.

We refer to [10] for further discussions of Condition 5, see in particular Remark 2.3. For the rest of this section, we tacitly assume Condition 5 and write $\boldsymbol{X}_j := \boldsymbol{X}_{n,j}$ for simplicity. Note that the triangular array $(\boldsymbol{X}_n)_n$ is r_n-dependent, which in fact simplifies the analysis of the disjoint block maxima method. For the sliding block maxima method however, mathematical challenges arise from the fact that the sliding block maxima sample is typically non-stationary. Indeed, for $x \in \mathbb{R}^d$, generally

$$\mathbb{P}\left(\boldsymbol{M}_{r,1} \leq x\right) \neq \mathbb{P}\left(\boldsymbol{X}_2, \ldots, \boldsymbol{X}_r \leq \boldsymbol{x}\right) \cdot \mathbb{P}\left(\boldsymbol{X}_{r+1} \leq \boldsymbol{x}\right) = \mathbb{P}\left(\boldsymbol{M}_{r,2} \leq \boldsymbol{x}\right).$$

In [10], Lemma 2.4, it is shown that this non-stationarity disappears asymptotically, which suggests that statistical methodology derived under stationarity assumptions (as in Section 3) may also be applicable under Condition 5. For deriving respective limit results, some modifications of the previous conditions are necessary. First of all, the integrability conditions from Condition 4 take the following, slightly more involved form.

Condition 6. There exists a $\nu > 2/\omega$ with ω from Condition 2 such that
  (a) $\limsup_{r \to \infty} \sup_{1 \leq i \leq j \leq r} \int \int |h(\boldsymbol{x}, \boldsymbol{y})|^{2+\nu} \, d\mathbb{P}_{\boldsymbol{Z}_{r,i}}(\boldsymbol{x}) \, d\mathbb{P}_{\boldsymbol{Z}_{r,j}}(\boldsymbol{y}) < \infty,$
  (b) $\limsup_{r \to \infty} \sup_{1 \leq i \leq j \leq r} \mathbb{E}[|h(\boldsymbol{Z}_{r,i}, \boldsymbol{Z}_{r,j})|^{2+\nu}] < \infty.$

It is worth noting that, if there exist monotone functions g_1, g_2 such that $|h(x,y)| \leq |g_1(x)| + |g_2(y)|$, the inner supremum may be omitted; examples can be found in Section 4.

Next, we quantify the average non-stationarity for the sliding block maxima. For $i, j \in \{1, \ldots, r\}$, let

$$\vartheta_{r,i,j} := \mathbb{E}[h(\boldsymbol{Z}_{r,i}, \tilde{\boldsymbol{Z}}_{r,j})], \qquad \bar{\vartheta}_r := \frac{1}{r^2} \sum_{1 \leq i,j \leq r} \vartheta_{r,i,j}, \qquad (5.1)$$

where $(\tilde{\boldsymbol{Z}}_{r,j})_{j=1,\ldots,r}$ is an independent copy of $(\boldsymbol{Z}_{r,j})_{j=1,\ldots,r}$. Note that $\vartheta_{r,1,1} = \vartheta_r$ with ϑ_r from (3.4), while $\vartheta_{r,i,j} \neq \vartheta_r$ in general. We do however have $\bar{\vartheta}_r = \vartheta_r + o(1)$ under the previous conditions (see also Lemma B.5 and B.6 in [10] for similar results).

Lemma 5.1. *Suppose Conditions 3, 5 and 6(a),(b) are met and that h is $\boldsymbol{\lambda}^{2d}$-a.e. continuous. Then, for $n \to \infty$,*

$$\mathbb{E}[U_{n,r,Z}^{\text{sb}}] = \bar{\vartheta}_r + O(m^{-1}), \quad \bar{\vartheta}_r = \vartheta_r + o(1).$$

This result suggests that the non-stationarity of the sliding block maxima method under Condition 5 may show up in the asymptotic bias of the U-statistic $U_{n,r}^{\text{sb}}$. The following assumption requires r to be sufficiently large to make this bias negligible.

Condition 7 (Negligibility of the bias due to non-stationarity). The limit $D := \lim_{n\to\infty} D_n$ exists, where

$$D_n = \sqrt{m}\big(\bar{\vartheta}_r - \vartheta_r\big). \tag{5.2}$$

Theorem 5.2. *Within the setting of Condition 5, suppose that the block size and the underlying time series $(\boldsymbol{Y}_{j,t})_t$ satisfy Condition 2(a),(b) and that the kernel satisfies Condition 3. Additionally, for $\text{mb} = \text{db}$, suppose that Condition 4(a) is met, and for $\text{mb} = \text{sb}$, suppose that Condition 6 and 7 are met. Then, if h is $\boldsymbol{\lambda}^{2d}$-a.e. continuous and bounded on compact sets, we have, for $n \to \infty$,*

$$\frac{\sqrt{m}}{f(\boldsymbol{a}_r, \boldsymbol{b}_r)} \left(U_{n,r}^{\text{mb}} - \theta_r\right) \rightsquigarrow \begin{cases} \mathcal{N}\big(0, \sigma_{\text{db}}^2\big), & \text{mb} = \text{db} \\ \mathcal{N}\big(D, \sigma_{\text{sb}}^2\big), & \text{mb} = \text{sb} \end{cases}$$

with σ_{mb}^2 from (3.9) satisfying $\sigma_{\text{sb}}^2 \leq \sigma_{\text{db}}^2$. If, additionally, the limit $B = \lim_{n\to\infty} B_n$ with B_n from (3.12) exists, then, again for $n \to \infty$,

$$\frac{\sqrt{m}}{f(\boldsymbol{a}_r, \boldsymbol{b}_r)} \left(U_{n,r}^{\text{mb}} - \tilde{\vartheta}_r\right) \rightsquigarrow \begin{cases} \mathcal{N}\big(B, \sigma_{\text{db}}^2\big), & \text{mb} = \text{db}, \\ \mathcal{N}\big(B + D, \sigma_{\text{sb}}^2\big), & \text{mb} = \text{sb} \end{cases}$$

with $\tilde{\vartheta}_r$ from (3.11).

6. Simulation study

A Monte Carlo simulation study was conducted to evaluate the finite sample performance of two selected estimators based on U-statistics: the empirical variance (univariate) as well as Kendall's τ statistic (bivariate). The study mainly aims at comparing the disjoint and sliding block maxima method for various extreme value indices and time series models. The discussion is divided into two subsections, depending on the nature of the target parameter: it can either be a parameter of the block maximum distribution with some fixed r (e.g., $r = 365$ or $r = 90$), or a parameter of the max-attractor distribution from Condition 1; in that case, n should be considered fixed while r may be treated as a tuning parameter.

6.1. Estimating parameters of the block maximum distribution

6.1.1. Estimating the block maxima variance

In Section 4.1, the empirical variance based on sliding block maxima, $\hat{\sigma}^2_{n,r,\mathrm{sb}}$, was found to be an asymptotically more efficient estimator of $\sigma^2_r := \mathrm{Var}(M_{r,1})$ than its disjoint blocks counterpart, $\hat{\sigma}^2_{n,r,\mathrm{db}}$. We assess the performance in finite-sample situations for data-generating processes made up from the following marginal and temporal models:

Stationary distribution of X_t We consider the generalized Pareto distribution $\mathrm{GPD}(0,1,\gamma)$ with shape parameter $\gamma \in \{-0.4, -0.2, 0, 0.1\}$, see (4.2). Note that the largest value of $\gamma = 0.1$ is close to the non-integrability point 0.25 for the variance estimator.

Time series models In addition to the i.i.d. case, two time series models were considered, each with three parameter choices. The first model is the (transformed) ARMAX(1) model, see Section 4.1, with time series parameter $\alpha \in \{0.25, 0.5, 0.75\}$; note that the extremal index is $\theta = 1 - \alpha$. As the second model we chose the Cauchy-AR model, defined as the stationary solution $(Y_t)_t$ of the Cauchy-AR recursion

$$Y_t = \phi Y_{t-1} + W_t, \quad t \in \mathbb{Z}, \quad (W_t)_t \overset{i.i.d.}{\sim} \mathrm{Cauchy}(0,1),$$

with time series parameter $\phi \in \{0.25, 0.5, 0.75\}$. This corresponds to the extremal index $\theta = 1 - \phi$, see, e.g., Problem 7.9 in [27]. Realizations from the model were transformed to the $\mathrm{GPD}(0,1,\gamma)$ distribution by setting $X_t = F_\gamma^\leftarrow(F_Y(Y_t))$, where F_Y and F_γ denote the c.d.f. of the Cauchy(0,1) and the GPD(0,1, γ)-distribution, respectively.

Combining each marginal model with each time series models results in a total of $4 \times 7 = 28$ different models. Throughout, we chose to fix the block size to $r = 90$, which roughly corresponds to the number of days in the summer months and which is a common block length in environmental applications. The number of blocks, denoted as m, ranged from 10 to 100, resulting in corresponding sample sizes ranging from $n = 900$ to $n = 9{,}000$ observations. The performance of the estimators was assessed based on approximating the MSE, the squared bias and the variance of the estimators based on $N = 10{,}000$ simulation repetitions. For assessing the bias, the true variance σ^2_{90} was determined in a preliminary simulation experiment involving a huge sample of size 10^6 drawn from the distribution of $M_{r,1}$; with one such sample for each of the 28 models.

The results for the i.i.d. and the ARMAX-models are illustrated in Figure 2, where we depict the ratio $\mathrm{MSE}(\hat{\sigma}^2_{n,r,\mathrm{db}})/\mathrm{MSE}(\hat{\sigma}^2_{n,r,\mathrm{sb}})$ as a function of the number of seasons (results for the Cauchy-AR-model are omitted because they are qualitatively similar). Across all considered numbers of seasons, tail indices and time series parameter, the sliding blocks estimator consistently outperforms its disjoint blocks counterpart. Notably, the depicted ratio is significantly larger
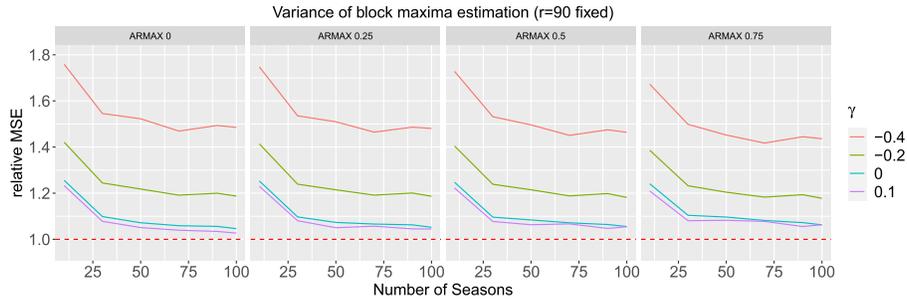
Variance of block maxima estimation (r=90 fixed)

FIG 2. *For the estimation of $\sigma_r^2 = \mathrm{Var}(M_{r,1})$, the ratio $\mathrm{MSE}(\hat{\sigma}_{n,r,\mathrm{db}}^2)/\mathrm{MSE}(\hat{\sigma}_{n,r,\mathrm{sb}}^2)$ is depicted as a function of number of blocks m.*

than one for small tail indices and for small sample sizes. This particular observation is promising because obtaining large sample sizes is sometimes challenging in the area of extreme value statistics. Also, it should be noted that the serial dependence does not substantially influence the relative performance (as was to be expected from the asymptotic results). Finally, we would like to report that the estimation variance was found to be of much larger order than the bias, whence the MSE-ratio is nearly the same as the respective variance ratio. This is different when the target variable is a parameter of the max-attractor distribution G from Condition 1, as will be discussed in Section 6.2 below.

6.1.2. Estimating Kendall's tau

We investigate the finite-sample performance in the bivariate case for the estimation of Kendall's $\tau = \tau_r = \tau(M_{r,1}^{(1)}, M_{r,1}^{(2)})$ based on the estimators $\hat{\tau}_{n,r}^{\mathrm{db}}$ and $\hat{\tau}_{n,r}^{\mathrm{sb}}$ from Section 4.3. Note that both Kendall's τ and its estimators do not depend on the marginal distributions (in case they are continuous). The data generating processes are as follows:

Time series models Three types of time series models were considered: bivariate versions of the ARMAX(1) and Cauchy-AR(1) model from the previous section as well as i.i.d. observations. The bivariate ARMAX(1) model is defined as the stationary bivariate solution to the recursion equation:

$$X_t^{(j)} = \max\{\alpha X_{t-1}^{(j)}, (1-\alpha)W_t^{(j)}\}, \quad t \in \mathbb{Z}, \quad j \in \{1, 2\},$$

where $\alpha \in (0, 1]$ and where $(\boldsymbol{W}_t)_t$ is an i.i.d. sequence with Fréchet(1)-distributed margins and with copula as specified below. Throughout, the value of α was fixed to 0.5; and the i.i.d. case is obtained for $\alpha = 0$. The bivariate Cauchy-AR(1) model is defined as the stationary solution of the bivariate Cauchy-AR(1) recursion

$$X_t^{(j)} = \phi X_{t-1}^{(j)} + W_t^{(j)}, \quad t \in \mathbb{Z}, \quad j \in \{1, 2\},$$
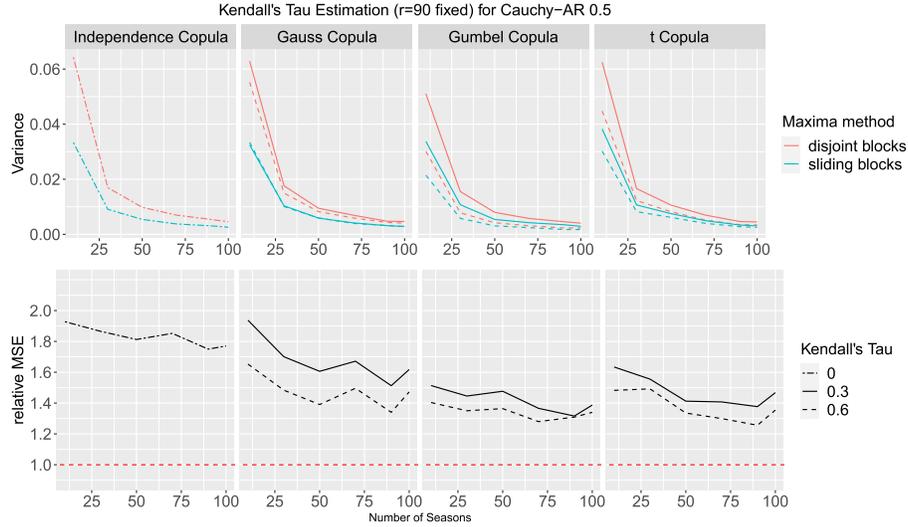
Kendall's Tau Estimation (r=90 fixed) for Cauchy−AR 0.5

FIG 3. *MSE of $\hat{\tau}_{n,r}^{\mathrm{mb}}$ (upper panel) and MSE ratio* $\mathrm{MSE}(\hat{\tau}_{n,r}^{\mathrm{db}})/\mathrm{MSE}(\hat{\tau}_{n,r}^{\mathrm{sb}})$ *(lower panel) plotted against the number of blocks m.*

where $\phi \in (0,1]$ and where $(\boldsymbol{W}_t)_t$ is an i.i.d. sequence with Cauchy(1) margins and with copula as specified below. Throughout, the value of ϕ was fixed to 0.5.

Copula of $\boldsymbol{W}_t$ Seven different copulas were considered: the independence copula, the Gaussian copula, the t_ν-copula with $\nu = 4$ degrees of freedom, and the Gumbel-Hougard copula, where the parameter of three last-named copulas was chosen in such a way that the associated value of Kendall's tau is in $\{0.3, 0.6\}$. Note that the Gaussian copula is tail independent, while the t- and Gumbel copula exhibit upper tail dependence. The upper tail dependence coefficients as a function of Kendall's tau are given by $2 \cdot t_5(-\sqrt{5(1-\sin(\pi\tau/2))/(1+\sin(\pi\tau/2))}) \in \{0.23, 0.5\}$ and $2 - 2^{1-\tau} \in \{0.375, 0.68\}$ for the t_4 and Gumbel-Hougard copula, respectively; see [18].

Overall, we obtain $3 \times 7 = 21$ different models. As in the previous section, we fix the block length to $r = 90$ and vary m between 10 and 100, resulting in sample sizes $n = mr$ ranging from 900 to 9,000 observations (results for other choices of fixed r were found to be qualitatively similar). The estimators are evaluated in terms of the mean squared error (MSE), the bias and the variance, based on $N = 1,000$ simulation repetitions. The true value of τ_r was assessed in a preliminary simulation involving a sample of size 100,000 from $\boldsymbol{M}_{r,1}$.

The results are presented in Figure 3, where we restrict attention to the Cauchy-AR(1) model, as the performance in the other two time series is nearly identical. As in the previous section, the bias was found to be of much smaller order than the variance, whence we further restrict attention to $\mathrm{MSE}(\hat{\tau}_{n,r}^{\mathrm{db}})$ and to the MSE ratio $\mathrm{MSE}(\hat{\tau}_{n,r}^{\mathrm{db}})/\mathrm{MSE}(\hat{\tau}_{n,r}^{\mathrm{sb}})$. We observe that the sliding blocks estimator consistently outperforms the disjoint blocks counterpart. The level of

dependence impacts the performance in that the estimation is more precise for higher dependence (for both estimators), and in that the advantage of the sliding blocks estimator over its disjoint blocks counterpart is highest for low levels of dependence/independence. Furthermore, as in the previous section, the sliding blocks estimator's advantage is slowly decreasing in the number of blocks.

6.2. Estimating parameters of the max-attractor distribution

If the target variable is a parameter of the max-attractor distribution, the total sample size n can be considered fixed, with the block size r serving as a tuning parameter to be chosen by the statistician. The arguments from Remark 4.4 then suggest that a bias-variance trade-off shows up: a large block size should correspond to small bias and large variance, and vice versa.

We restrict attention to the estimators $\hat{\tau}_{n,r}^{\mathrm{db}}$ and $\hat{\tau}_{n,r}^{\mathrm{sb}}$ from Section 4.3, now considered as estimators for $\tau = \tau(Z^{(1)}, Z^{(2)})$ with $(Z^{(1)}, Z^{(2)})$ from Condition 1. We chose to fix the sample size to $n = 1000$, and consider various choices for the block length between $r = 4$ and $r = 50$. For simplicity, we restrict attention to the iid case, where explicit calculations of $\tau = \tau(C)$ are possible such that we do not need to rely on preliminary Monte Carlo approximations. Regarding the copula D of $\boldsymbol{X}_i = (X_i^{(1)}, X_i^{(2)})$, we chose to work with to the outer power transform of the Clayton copula (see [6], Formula (4.5)) with parameter θ fixed to $\theta = 1$ and with parameter β chosen in such a way that Kendall's tau $\tau = \tau(C)$ of the extreme value attractor C (which is the Gumbel-Hougaard copula with parameter β) varies in $\{0.35, 0.4, 0.45, 0.5\}$.

The results, based on $N = 1000$ simulation runs each, are summarized in Figure 4. As expected, the squared bias is an increasing function of m, while the variance is decreasing. In fact, the curves agree with the theoretical results from Remark 4.4 and Section 3.1 in [9], where it is shown that φ from (4.8) can be chosen as r^{-1} for i.i.d. samples from the outer power Clayton copula. As a consequence, the squared bias is of the theoretical order $r^{-2} = (m/n)^2$, while the variance is of the order m^{-2}. The sum of the two rates corresponds to the MSE, which is a u-shaped function of m with minimal value attained at a point proportional to $m = n^{2/3}$, which in our case is $n = 100$. This is indeed close to the argmins of the MSE-curves depicted in Figure 4.

7. **Proofs**

Recall the definitions of θ_r, ϑ_r, σ_{mb}^2 from (2.5), (3.4) and (3.9), respectively.

7.1. Proofs for Section 3

Proof of Theorem 3.2. Recall the definition of $U_{n,r,\boldsymbol{Z}}^{(\mathrm{mb})}$ in (3.3). By Condition 3 we have
$$\frac{U_{n,r}^{\mathrm{mb}} - \theta_r}{f(\boldsymbol{a}_r, \boldsymbol{b}_r)} = U_{n,r,\boldsymbol{Z}}^{(\mathrm{mb})} - \vartheta_r.$$
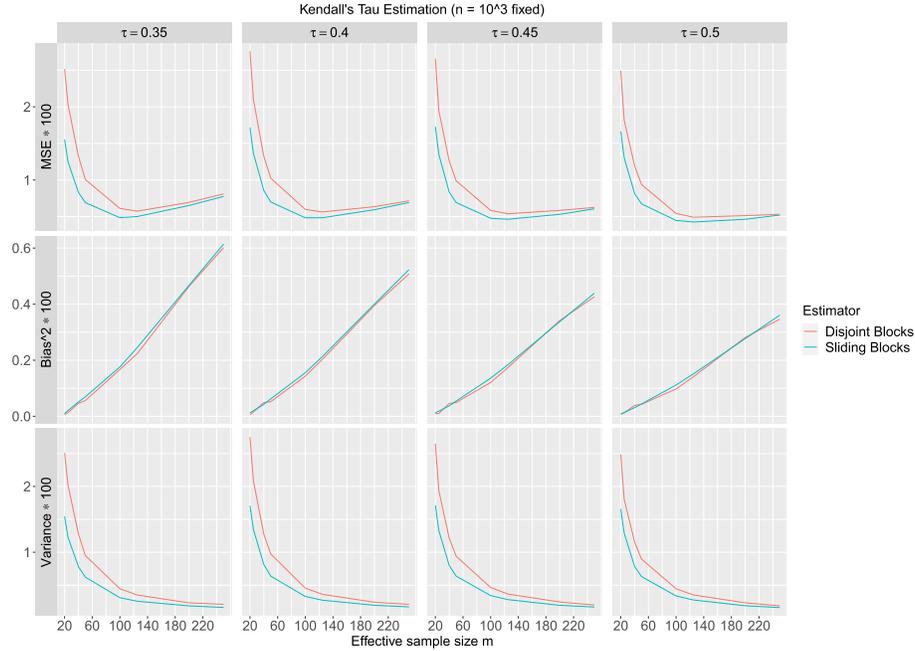
Fig 4. *Mean squared error (top row), squared bias (middle row) and estimation variance (bottom row) when considering $\hat{\tau}_{n,r}^{\mathrm{mb}}$ as an estimator for $\tau = \tau(C)$ with fixed sample size $n = 1000$. From left to right: Kendall's tau of the attractor copula $0.35, 0.4, 0.45, 0.5$.*

Hence it suffices to show that

$$\sqrt{m} \cdot \left(U_{n,r,\boldsymbol{Z}}^{\mathrm{mb}} - \vartheta_r\right) \rightsquigarrow \mathcal{N}\left(0, \sigma_{\mathrm{mb}}^2\right), \tag{7.1}$$

for $\mathrm{mb} \in \{\mathrm{db}, \mathrm{sb}\}$ and

$$\sigma_{\mathrm{sb}}^2 \leq \sigma_{\mathrm{db}}^2. \tag{7.2}$$

For the proof of (7.1) we will use a Hoeffding decomposition and verify weak convergence of the linear part to the normal limit and L^2-convergence to zero of the asymptotically degenerate part. For both parts we will employ common blocking techniques to deal with the serial dependence, (see e.g., [13], page 31). Define

$$h_{1,r} \colon \mathbb{R}^d \to \mathbb{R}, \quad \boldsymbol{x} \mapsto h_{1,r}(\boldsymbol{x}) := \mathbb{E}[h(x, \boldsymbol{Z}_{r,1})] - \vartheta_r \tag{7.3}$$

$$h_{2,r} \colon \mathbb{R}^{2d} \to \mathbb{R} \quad (\boldsymbol{x}, \boldsymbol{y}) \mapsto h_{2,r}(\boldsymbol{x}, \boldsymbol{y}) := h(\boldsymbol{x}, \boldsymbol{y}) - h_{1,r}(\boldsymbol{x}) - h_{1,r}(\boldsymbol{y}) - \vartheta_r$$

and notice the algebraic identity

$$U_{n,r}^{\mathrm{mb}} - \vartheta_r = \frac{2}{n_{\mathrm{mb}}} \sum_{i \in I_n^{\mathrm{mb}}} h_{1,r}(\boldsymbol{Z}_{r,i}) + \frac{2}{n_{\mathrm{mb}} \cdot (n_{\mathrm{mb}} - 1)} \sum_{(i,j) \in J_n^{\mathrm{mb}}} h_{2,r}(\boldsymbol{Z}_{r,i}, \boldsymbol{Z}_{r,j})$$

$$\equiv L_{n,r}^{\mathrm{mb}} + D_{n,r}^{\mathrm{mb}}. \tag{7.4}$$

Proof of (7.1) for mb = db: We start by proving

$$\sqrt{m} \cdot L_{n,r}^{\mathrm{db}} = \frac{2}{\sqrt{m}} \sum_{i \in I_n^{\mathrm{db}}} h_{1,r}(\boldsymbol{Z}_{r,i}) \rightsquigarrow \mathcal{N}\big(0, \sigma_{\mathrm{db}}^2\big). \qquad (7.5)$$

By Lemma 8.5, we may switch to i.i.d. copies of $\boldsymbol{Z}_{r,i}$. The assertion then follows from Ljapunov's central limit theorem, with the Ljapunov Condition being a straightforward consequence of Lemma 8.2 and Condition 4.

In the next part we show that the (asymptotically) degenerate part converges to zero in L^2, i.e., $\mathbb{E}[(\sqrt{m} \cdot L_{n,r}^{\mathrm{db}})^2] = o(1)$. For that purpose, it is sufficient to show that

$$\frac{m}{n_{\mathrm{db}}^4} \sum_{\substack{(i_1,i_2) \in J_n^{\mathrm{db}} \\ (j_1,j_2) \in J_n^{\mathrm{db}}}} \mathbb{E}[h_{2,r}(\boldsymbol{Z}_{r,i_1}, \boldsymbol{Z}_{r,i_2}) h_{2,r}(\boldsymbol{Z}_{r,j_1}, \boldsymbol{Z}_{r,j_2})] = o(1). \qquad (7.6)$$

The Cauchy-Schwarz inequality, standard inequalities for the expectation and Condition 4 imply that

$$\sup_{\substack{(i_1,i_2) \in J_n^{\mathrm{db}} \\ (j_1,j_2) \in J_n^{\mathrm{db}}}} \big| \mathbb{E}[h_{2,r}(\boldsymbol{Z}_{r,i_1}, \boldsymbol{Z}_{r,i_2}) h_{2,r}(\boldsymbol{Z}_{r,j_1}, \boldsymbol{Z}_{r,j_2})] \big| \leq \sup_{s \in \mathbb{N}} \mathbb{E}[h_{2,r}^2(\boldsymbol{Z}_{r,1}, \boldsymbol{Z}_{r,1+s})]$$

$$= O(1).$$

Consider a tuple $(\boldsymbol{i}, \boldsymbol{j}) = (i_1, i_2, j_1, j_2) \in \{(i_1, i_2) \in J_n^{\mathrm{db}}, (j_1, j_2) \in J_n^{\mathrm{db}}\}$ such that both the distance between the smallest, $\min(\boldsymbol{i}, \boldsymbol{j})$, and the second smallest index and the largest, $\max(\boldsymbol{i}, \boldsymbol{j})$, and the second largest index is at most $2r$. Clearly, the cardinality of the set of all those $(\boldsymbol{i}, \boldsymbol{j})$ is of the order $O(m^2)$, whence the expression in (7.6) with the sum restricted to those tuples is of the order $O(m^{-1})$. It is hence sufficient to consider the sum over those summands for which either the distance between the smallest index and all other indices is strictly larger than $2r$, or the distance between the largest index and all other indices is strictly larger than $2r$. We only consider the first case, as the other can be treated similarly. Without loss of generality, let i_1 be the smallest index, and let $\mathcal{J}_{n,\mathrm{db}} =$ denote the respective set of indices, that is, $\mathcal{J}_{n,\mathrm{db}} = \{(\boldsymbol{i}, \boldsymbol{j}) \in J_n^{\mathrm{db}} \times J_n^{\mathrm{db}} : i_2 - i_1 > 2r, j_1 - i_1 > 2r\}$.

For each tuple $(\boldsymbol{i}, \boldsymbol{j}) \in \mathcal{J}_{n,\mathrm{db}}$, we may use Berbee's coupling Lemma [2] to construct a random variable $\boldsymbol{Z}_{r,i_1}^*$ having the same distribution as $\boldsymbol{Z}_{r,i_1}$ that is independent of $(\boldsymbol{Z}_{r,i_2}, \boldsymbol{Z}_{r,j_1}, \boldsymbol{Z}_{r,j_2})$ and which satisfies $\mathbb{P}(\boldsymbol{Z}_{r,i_1} \neq \boldsymbol{Z}_{r,i_1}^*) \leq \beta(\sigma(\boldsymbol{Z}_{r,i_1}), \sigma(\boldsymbol{Z}_{r,i_2}, \boldsymbol{Z}_{r,j_1}, \boldsymbol{Z}_{r,j_2})) \leq \beta(r)$ where $\sigma(X)$ denotes the initial σ-field of X. Now decompose

$$\mathbb{E}[h_{2,r}(\boldsymbol{Z}_{r,i_1}, \boldsymbol{Z}_{r,i_2}) h_{2,r}(\boldsymbol{Z}_{r,j_1}, \boldsymbol{Z}_{r,j_2})]$$
$$= \mathbb{E}[\{h_{2,r}(\boldsymbol{Z}_{r,i_1}, \boldsymbol{Z}_{r,i_2}) - h_{2,r}(\boldsymbol{Z}_{r,i_1}^*, \boldsymbol{Z}_{r,i_2})\} h_{2,r}(\boldsymbol{Z}_{r,j_1}, \boldsymbol{Z}_{r,j_2})]$$
$$+ \mathbb{E}[h_{2,r}(\boldsymbol{Z}_{r,i_1}^*, \boldsymbol{Z}_{r,i_2}) h_{2,r}(\boldsymbol{Z}_{r,j_1}, \boldsymbol{Z}_{r,j_2})] =: I_{1,n}^{(\boldsymbol{i},\boldsymbol{j})} + I_{2,n}^{(\boldsymbol{i},\boldsymbol{j})}.$$

Using stationarity, basic properties of the conditional expectation and the properties of $\boldsymbol{Z}_{r,i_1}^*$ we obtain, via conditioning on $(\boldsymbol{Z}_{r,i_2}, \boldsymbol{Z}_{r,j_1}, \boldsymbol{Z}_{r,j_2})$, that $I_{2,n}^{(\boldsymbol{i},\boldsymbol{j})} \equiv 0$.

Next, repeated applications of Hölder's inequality imply that, uniformly in $(\boldsymbol{i}, \boldsymbol{j}) \in \mathcal{J}_{n,\mathrm{db}}$,

$$
|I_{1,n}^{(\boldsymbol{i},\boldsymbol{j})}| \leq \mathbb{E}\left[\mathbf{1}\big\{\boldsymbol{Z}_{r,i_1} \neq \boldsymbol{Z}_{r,i_1}^*\big\}\,\big|\big\{h_{2,r}(\boldsymbol{Z}_{r,i_1}, \boldsymbol{Z}_{r,i_2}) - h_{2,r}(\boldsymbol{Z}_{r,i_1}^*, \boldsymbol{Z}_{r,i_2})\big\} \\
\times h_{2,r}(\boldsymbol{Z}_{r,j_1}, \boldsymbol{Z}_{r,j_2})\big|\right]
$$

$$
\lesssim \beta(r)^{\nu/(2+\nu)},
$$

where we have used Condition 4. Overall,

$$
\frac{m}{n_{\mathrm{db}}^4} \sum_{(\boldsymbol{i},\boldsymbol{j}) \in \mathcal{J}_{n,\mathrm{db}}} |\mathbb{E}[h_{2,r}(\boldsymbol{Z}_{r,i_1}, \boldsymbol{Z}_{r,i_2}) h_{2,r}(\boldsymbol{Z}_{r,j_1}, \boldsymbol{Z}_{r,j_2})]| \lesssim m \cdot \sup_{(\boldsymbol{i},\boldsymbol{j}) \in I} |I_{1,n}^{(\boldsymbol{i},\boldsymbol{j})} + I_{2,n}^{(\boldsymbol{i},\boldsymbol{j})}|
$$

$$
\lesssim \left(m^{1+2/\nu}\beta(r)\right)^{\nu/(2+\nu)}
$$

which converges to zero by Condition 2 (c) as $2/\nu < \omega$. This implies (7.6), and in combination with (7.4) and (7.5) we obtain (7.1).

Proof of (7.1) for mb = sb: In order to show that the degenerate part of the rescaled sliding blocks U-statistic converges to zero, it is sufficient to show that

$$
\frac{m}{n_{\mathrm{sb}}^4} \sum_{\substack{(i_1,i_2) \in J_n^{\mathrm{sb}} \\ (j_1,j_2) \in J_n^{\mathrm{sb}}}} \mathbb{E}[h_{2,r}(\boldsymbol{Z}_{r,i_1}, \boldsymbol{Z}_{r,i_2}) h_{2,r}(\boldsymbol{Z}_{r,j_1}, \boldsymbol{Z}_{r,j_2})] \to 0.
$$

This can be worked out analogously to the disjoint case: again, we may restrict the sum in the upper display to tuples in $J_n^{\mathrm{sb}} = \{(\boldsymbol{i}, \boldsymbol{j}) \in J_n^{\mathrm{sb}} \times J_n^{\mathrm{sb}} : i_2 - i_1 > 2r, j_1 - i_1 > 2r\}$, as the set of the remaining tuples is of the order $O((nr)^2)$. We can then copy the disjoint blocks proof verbatim by replacing $\mathcal{J}_{n,\mathrm{db}}$ and n_{db} with $\mathcal{J}_{n,\mathrm{sb}}$ and n_{sb}.

It remains to show

$$
\frac{2\sqrt{m}}{n} \sum_{i \in I_n^{\mathrm{sb}}} h_{1,r}(\boldsymbol{Z}_{r,i}) \rightsquigarrow \mathcal{N}\big(0, \sigma_{\mathrm{sb}}^2\big).
$$

For this purpose use Theorem 8.7 with $f_{r,s} := h_{1,r}$, $f := h_1$ and note that all conditions are satisfied, where we use Lemma 8.6 and an easy adaptation of Lemma B.15 in [10] to obtain the weak convergence condition in (8.3).

Proof of (7.2): The inequality follows from Lemma A.10 in [41], where $X_{n,i} := h_{1,r}(\boldsymbol{Z}_{r,i})$ and the preconditions of Lemma A.10 can be deduced from Condition 2(a), (c) and 4(a). $\qquad\square$

Proof of Corollary 3.3. By Condition 3 and the assumption on B_n, we have

$$
\frac{\sqrt{m}}{f(\boldsymbol{a}_r, \boldsymbol{b}_r)}\big(\theta_r - f(\boldsymbol{a}_r, \boldsymbol{b}_r)(\vartheta + \ell(\boldsymbol{a}_r, \boldsymbol{b}_r))\big) = \sqrt{m}(\vartheta_r - \vartheta) = B_n = B + o(1).
$$

Hence, the assertion follows from Theorem 3.2 and Slutsky's theorem. $\qquad\square$

7.2. *Proofs for Section 4*

Proof of Corollary 4.1. Note that $|h_{\mathrm{Var}}(x,y)|^{2+\nu/2} \leq 2^{2+\nu/2}(|x|^{4+\nu} + |y|^{4+\nu})$. Hence, by assumption, Condition 4 is met for a $\nu > 0$ as in the formulation of Theorem 3.2. The statement follows by the continuity of h_{Var} and Example 3.1. $\qquad\square$

Proof of Equation (4.3). Fix $\gamma < 1/4$ and omit the lower index 1 everywhere; e.g., write Z_r instead of $Z_{r,1}$. We need to verify the conditions of Corollary 4.1.

We start by proving Condition 1, for which we restrict attention to the case $\gamma > 0$ since the other cases can be treated similarly. Using $F_W^{\leftarrow}(F_\gamma(t)) = -1/\log\{1 - (1 + \gamma t)^{-1/\gamma}\}$ for $t > 0$ and equation (10.5) from [1] we have

$$
\begin{aligned}
\mathbb{P}\left(\frac{M_r - b_r}{a_r} \leq x\right) &= \exp\left[-\frac{1 + (1-\alpha)(r-1)}{F_W^{\leftarrow}(F_\gamma(a_r x + b_r))}\right] \\
&= \exp\left[r(1-\alpha)\log\{1 - (1 + \gamma(a_r x + b_r))^{-1/\gamma}\}\right] + o(1) \\
&= \exp\left[-r(1-\alpha)(1 + \gamma(a_r x + b_r))^{-1/\gamma}\right] + o(1) \\
&= G_\gamma(x) + o(1),
\end{aligned}
$$

where we substituted $a_r = (r(1-\alpha))^\gamma, b_r = \{(r(1-\alpha))^\gamma - 1\}/\gamma$. Hence Condition 1 is satisfied.

Condition 2(a) holds by assumption. Conditions (b) and (c) hold since $r = o(n^3)$ and since there exists $c > 0$ with $\alpha(k) \leq \beta(k) \leq \exp(-ck)$ for $k \in \mathbb{N}$ by the discussion in Section 4.1. Note that (c) does hold for all $\omega > 0$.

In view of the latter statement, it remains to prove $\limsup_r \mathbb{E}[|Z_r|^{4+\nu}] < \infty$ for some $\nu > 0$. Note that this in turn is equivalent to $\limsup_r \mathbb{E}[|Z'_r|^{4+\nu}] < \infty$, where $Z'_r := (M_r - b'_r)/a'_r$ and where $b'_r \in \mathbb{R}, a'_r > 0$ are sequences with $Z'_r \rightsquigarrow \mathrm{GEV}(\mu, \sigma, \gamma)$ for some $\mu \in \mathbb{R}, \sigma > 0$.

Define $a'_r := r^\gamma$ and $b'_r := (r^\gamma - 1)/\gamma$, where the latter is defined by continuity as $b'_r = \log r$ if $\gamma = 0$. The p.d.f. of Z'_r is then given by

$$
f_{Z'_r}(t) = (1 - \alpha + \alpha/r) \cdot
\begin{cases}
(1 + \gamma t)^{-(1+1/\gamma)}\left(1 - \frac{(1+t\gamma)^{-1/\gamma}}{r}\right)^{r(1-\alpha)+\alpha-1}, & \gamma \neq 0 \\
e^{-t}\left(1 - \frac{e^{-t}}{r}\right)^{r(1-\alpha)+\alpha-1}, & \gamma = 0
\end{cases}
$$

for $t \in \mathrm{supp}(\tilde{Z}_r)$. We will only present the case $\gamma > 0$ as the other cases use similar ideas. Substituting $1 + t\gamma$, we obtain

$$
\begin{aligned}
&\mathbb{E}[|\tilde{Z}_r|^{4+\nu}] \\
&= \frac{1 - \alpha + \alpha/r}{\gamma}\int_{1/r^\gamma}^\infty \left(\frac{|t-1|}{\gamma}\right)^{4+\nu} t^{-1-1/\gamma}\left(1 - \frac{t^{-1/\gamma}}{r}\right)^{r(1-\alpha)+\alpha-1} \mathrm{d}t \\
&\leq \frac{1 - \alpha + \alpha/r}{\gamma^{5+\nu}}\left\{\int_{1/r^\gamma}^{1/2} t^{-1-1/\gamma}\left(1 - \frac{t^{-1/\gamma}}{r}\right)^{r(1-\alpha)+\alpha-1} \mathrm{d}t\right.
\end{aligned}
$$

$$+ \int_{1/2}^{\infty} t^{3+\nu-1/\gamma} \left(1 - \frac{t^{-1/\gamma}}{r}\right)^{r(1-\alpha)+\alpha-1} \mathrm{d}t \Bigg\} =: I_{r1} + I_{r2}$$

By the monotone convergence theorem the first integral converges to $\int_0^{1/2} t^{-1-1/\gamma} \exp(-(1-\alpha)t^{-1/\gamma}) \, \mathrm{d}t < \infty$; hence $\lim_{r\to\infty} I_{r1} < \infty$. Finally, let $\nu = 1/(2\gamma) - 2$ and invoke the monotone convergence theorem again to obtain

$$\lim_{r\to\infty} I_{r2} = \frac{1-\alpha}{\gamma^{3+\gamma/2}} \int_{1/2}^{\infty} t^{1-1/(2\gamma)} \exp(-(1-\alpha)t^{-1/\gamma}) \, \mathrm{d}t < \infty$$

as $1 - 1/(2\gamma) < -1$. Overall, we have shown that $\limsup_r \mathbb{E}[|Z_r|^{4+\nu}] < \infty$ as asserted.

Using similar ideas as before, one can show that $n = o(r^3)$ implies $\lim_{n\to\infty} B_n = 0$. $\square$

Proof of Proposition 4.2. Write $h_{\mathrm{pwm},2} = h_{\mathrm{pwm}}$ and $\tilde{h}_{\mathrm{pwm},2} = \tilde{h}_{\mathrm{pwm}}$. First of all, we have

$$\frac{\sqrt{m}}{a_r} \left\{U_{n,r}^{\mathrm{mb}}(h_{\mathrm{pwm}}) - U_{n,r}^{\mathrm{mb}}(\tilde{h}_{\mathrm{pwm}})\right\} = S_n^{\mathrm{mb}} + R_n^{\mathrm{mb}}$$

where

$$S_n^{\mathrm{mb}} = \sqrt{m} \binom{n_{\mathrm{mb}}}{2}^{-1} \sum_{\substack{(i,j)\in J_n^{\mathrm{mb}} \\ j-i>2r}} \left\{h_{\mathrm{pwm}}(Z_{r,i}, Z_{r,j}) - \tilde{h}_{\mathrm{pwm}}(Z_{r,i}, Z_{r,j})\right\}$$

$$R_n^{\mathrm{mb}} = \sqrt{m} \binom{n_{\mathrm{mb}}}{2}^{-1} \sum_{\substack{(i,j)\in J_n^{\mathrm{mb}} \\ j-i\le 2r}} \left\{h_{\mathrm{pwm}}(Z_{r,i}, Z_{r,j}) - \tilde{h}_{\mathrm{pwm}}(Z_{r,i}, Z_{r,j})\right\}.$$

The number of summands in R_n^{mb} is of the order $O(nr)$ for mb = sb and of the order $O(m)$ for mb = db, whence $R_n^{\mathrm{mb}} = O_{L^2}(m^{-1/2}) = o_{L^2}(1)$ by the integrability assumption.

Next, we have

$$S_n^{\mathrm{mb}} = \sqrt{m} \binom{n_{\mathrm{mb}}}{2}^{-1} \sum_{\substack{(i,j)\in J_n^{\mathrm{mb}} \\ j-i>2r}} \frac{1}{2}\mathbf{1}(Z_{r,i} = Z_{r,j})Z_{r,i}$$

which is zero with probability one by the no-ties assumption; note that all indices in the sum refer to blocks that do not overlap.

The second statement follows from Corollary 3.3, applied to $U_{n,r}^{\mathrm{mb}}(\tilde{h}_{\mathrm{pwm}})$. Finally, the inequality for the asymptotic variances can be found in [10]. $\square$

Proof of Proposition 4.3. Recall Example 3.1(5) and apply Theorem 3.2. A short calculation yields the formulas for the asymptotic variances. $\square$

7.3. Proofs for Section 5

Proof of Lemma 5.1. For $\xi \in (0, 1)$, let $\xi_r = 1 + \lfloor r\xi \rfloor$. Then,

$$\bar{\vartheta}_r = \int_0^1 \int_0^1 \mathbb{E}[h(\boldsymbol{Z}_{r,\xi_r}, \tilde{\boldsymbol{Z}}_{r,\xi'_r})] \, \mathrm{d}\xi \, \mathrm{d}\xi'$$

By Lemma 8.9, we have $\boldsymbol{Z}_{r,\xi_r} \rightsquigarrow \boldsymbol{Z} \sim G$ for any $\xi \geq 0$. Hence, by independence and the continuous mapping theorem, $h(\boldsymbol{Z}_{r,\xi_r}, \tilde{\boldsymbol{Z}}_{r,\xi'_r}) \rightsquigarrow h(\boldsymbol{Z}, \tilde{\boldsymbol{Z}})$. Therefore, by the previous display, dominated convergence (use Condition 6(a)) and Example 2.21 in [38], we obtain $\bar{\vartheta}_r = \vartheta + o(1)$. This implies the second statement, since $\vartheta_r = \vartheta + o(1)$ by Lemma 8.1.

For the first convergence assume that $n = mr$ for simplicity. We have

$$\mathbb{E}[U_{n,r,Z}^{\mathrm{sb}}] = \frac{1}{n_{\mathrm{sb}}(n_{\mathrm{sb}} - 1)} \sum_{\substack{1 \leq i \neq j \leq n_{\mathrm{sb}} \\ j-i>2r}} \mathbb{E}[h(\boldsymbol{Z}_{r,i}, \boldsymbol{Z}_{r,j})] + O(m^{-1}),$$

where the O-term is due to leaving nearby summands out. Next, by independence, piecewise stationarity and including nearby summands again,

$$\sum_{\substack{1 \leq i \neq j \leq n_{\mathrm{sb}} \\ j-i>2r}} \mathbb{E}[h(\boldsymbol{Z}_{r,i}, \boldsymbol{Z}_{r,j})] = \sum_{\substack{1 \leq i \neq j \leq n_{\mathrm{sb}} \\ j-i>2r}} \mathbb{E}[h(\boldsymbol{Z}_{r,i}, \tilde{\boldsymbol{Z}}_{r,j})] = m^2 \sum_{1 \leq i,j \leq r} \vartheta_{r,i,j} + O(rn).$$

Overall,

$$\mathbb{E}[U_{n,r,Z}^{\mathrm{sb}}] = \frac{m^2 r^2}{n_{\mathrm{sb}}(n_{\mathrm{sb}} - 1)} \bar{\vartheta}_r + O(m^{-1}) = \bar{\vartheta}_r + O(m^{-1}). \qquad \square$$

Proof of Theorem 5.2. For mb = db, $(\boldsymbol{Z}_{r,i})_{i \in I_n^{\mathrm{db}}}$ is an i.i.d. sample. Thus the proof essentially is an easier version of the proof of Theorem 3.2.

For mb = sb note that, by Lemma 5.1, Conditions 7 and 3, it is sufficient to show that $\sqrt{m}(U_{n,r,Z}^{\mathrm{sb}} - \mathbb{E}[U_{n,r,Z}^{\mathrm{sb}}]) \rightsquigarrow \mathcal{N}(0, \sigma_{\mathrm{sb}}^2)$. Note that we might replace $n - r + 1$ with n, since $(n - r + 1)/n = 1 + O(m^{-1})$. Unlike in the situation from Theorem 3.2, the sliding block maxima sample is not stationary anymore, which requires a different version of the Hoeffding decomposition. For $1 \leq i, j \leq n$, introduce functions $h_{1,r,i} \colon \mathbb{R}^d \to \mathbb{R}$ and $h_{2,r,i,j} \colon \mathbb{R}^{2d} \to \mathbb{R}$ by

$$h_{1,r,i}(\boldsymbol{x}) := \mathbb{E}[h(\boldsymbol{x}, \boldsymbol{Z}_{r,i})] - \vartheta_{r,i,i}$$
$$h_{2,r,i,j}(\boldsymbol{x}, \boldsymbol{y}) := h(\boldsymbol{x}, \boldsymbol{y}) - h_{1,r,i}(\boldsymbol{x}) - h_{1,r,j}(\boldsymbol{y}) - \vartheta_{r,i,j},$$

with $\vartheta_{r,i,j}$ from (5.1). Note, that by Lemma 5.1

$$U_{n,r,Z}^{\mathrm{sb}} - \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \mathbb{E}[h(\boldsymbol{Z}_{r,i}, \boldsymbol{Z}_{r,j})]$$

$$= \frac{2}{n} \sum_{i=1}^n h_{1,r,i}(\boldsymbol{Z}_{r,i}) + \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h_{2,r,i,j}(\boldsymbol{Z}_{r,i}, \boldsymbol{Z}_{r,j}) + O(m^{-1}) \quad (7.7)$$

$$\equiv L_{n,r}^{\mathrm{sb}} + D_{n,r}^{\mathrm{sb}} + O(m^{-1})$$

The asymptotic normality of $\sqrt{m}L_{n,r}^{\mathrm{sb}}$ follows from Theorem 8.8 with $f_{r,i} := 2h_{1,r,i}$, where the preconditions are met since the time series is piecewise stationary and by assumption; moreover, (8.3) is a consequence of Lemma 8.11. We omit the proof of $\sqrt{m}D_{n,r}^{\mathrm{sb}} = o_{\mathbb{P}}(1)$ as the proof is similar to the proof of the respective statement in the proof of Theorem 3.2.   □

8. Auxiliary results

8.1. Disjoint blocks – stationary case

Lemma 8.1 (Convergence of ϑ_r). *Assume Condition 1 is met. Furthermore suppose that h is $\boldsymbol{\lambda}^{2d}$-a.e. continuous and that there exists $\nu > 0$ with $\limsup_{r\to\infty} \int \int |h(\boldsymbol{x},\boldsymbol{y})|^{1+\nu} \, \mathrm{d}\mathbb{P}_{\boldsymbol{Z}_{r,1}}(\boldsymbol{x}) \, \mathrm{d}\mathbb{P}_{\boldsymbol{Z}_{r,1}}(\boldsymbol{y}) < \infty$. Then $\lim_{r\to\infty} \vartheta_r = \vartheta$ with ϑ_r and ϑ from (3.4) and (3.5), respectively.*

Proof. We have $h(\boldsymbol{Z}_{r,1}, \tilde{\boldsymbol{Z}}_{r,1}) \rightsquigarrow h(\boldsymbol{Z}, \tilde{\boldsymbol{Z}})$ by independence and the continuous mapping theorem. The assertion then follows from Example 2.21 in [38] and the integrability assumption.   □

Recall the definition of $h_{1,r}$ from (7.3).

Lemma 8.2 (Weak convergence of $h_{1,r}(\boldsymbol{Z}_{r,1})$). *Suppose Conditions 1, 4(a) hold and that h is $\boldsymbol{\lambda}^{2d}$-a.e. continuous and bounded on compact sets. Then, for $r \to \infty$,*

$$h_{1,r}(\boldsymbol{Z}_{r,1}) \rightsquigarrow h_1(\boldsymbol{Z}).$$

Moreover, for any $p < 2 + \nu$ with $p \in \mathbb{N}$, we have $\lim_{r\to\infty} \mathbb{E}[h_{1,r}^p(\boldsymbol{Z}_{r,1})] = \mathbb{E}[h_1^p(\boldsymbol{Z})]$.

Proof. Since $\vartheta_r \to \vartheta$ by Lemma 8.1, we may assume $\vartheta_r \equiv 0$. We will use Wichura's Theorem [4, Theorem 4.2]. Note that

$$T_r := h_{1,r}(\boldsymbol{Z}_{r,1}) = \int h(\boldsymbol{Z}_{r,1}, \boldsymbol{y}) \, \mathrm{d}\mathbb{P}_{\boldsymbol{Z}_{r,1}}(\boldsymbol{y}), \quad T := h_1(\boldsymbol{Z}) = \int h(\boldsymbol{Z}, \boldsymbol{y}) \, \mathrm{d}\mathbb{P}_{\boldsymbol{Z}}(\boldsymbol{y})$$

and define, for $B := B(b) := [-b, b]^d$ with $b \in \mathbb{N}$,

$$T_r(b) := \int_B h(\boldsymbol{Z}_{r,1}, \boldsymbol{y}) \, \mathrm{d}\mathbb{P}_{\boldsymbol{Z}_{r,1}}(\boldsymbol{y}), \quad T(b) := \int_B h(\boldsymbol{Z}, \boldsymbol{y}) \, \mathrm{d}\mathbb{P}_{\boldsymbol{Z}}(\boldsymbol{y}).$$

In order to show weak convergence of $T_r(b)$ to $T(b)$ we use the extended continuous mapping theorem (Theorem 1.11.1 in [39]). Let $\boldsymbol{x}_r \to \boldsymbol{x} \in \mathbb{R}^d$ and note that the map $(\boldsymbol{x}, \boldsymbol{y}) \mapsto h(\boldsymbol{x}, \boldsymbol{y})\mathbf{1}\{\boldsymbol{y} \in B\}$ is $\mathbb{P}_{\boldsymbol{Z}}^{\otimes 2}$-a.e. continuous. By the ordinary continuous mapping theorem we obtain weak convergence of $h(\boldsymbol{x}_r, \boldsymbol{Z}_{r,1})\mathbf{1}\{\boldsymbol{Z}_{r,1} \in\}$ to $h(\boldsymbol{x}, \boldsymbol{Z})\mathbf{1}\{\boldsymbol{Z} \in B\}$. Next, since there exists a compact set A containing $(\boldsymbol{x}_r)_r$, we have

$$\limsup_{r\to\infty} \mathbb{E}[h^2(\boldsymbol{x}_r, \boldsymbol{Z}_{r,1}) \cdot \mathbf{1}\{\boldsymbol{Z}_{r,1} \in B\}] \leq \sup_{\boldsymbol{x}\in A, \boldsymbol{z}\in B} h^2(\boldsymbol{x}, \boldsymbol{z}) < \infty,$$

which in turn implies moment convergence of $h(\boldsymbol{x}_r, \boldsymbol{Z}_{r,1})\mathbf{1}\{\boldsymbol{Z}_{r,1} \in B\}$. This shows continuous convergence of the mapping sequence $\boldsymbol{x} \mapsto \int_B h(\boldsymbol{x}, \boldsymbol{y}) \, d\mathbb{P}_{\boldsymbol{Z}_{r,1}}(\boldsymbol{y})$, and the extended continuous mapping theorem finally implies weak convergence of $T_r(b)$ to $T(b)$ as asserted.

Next, we have weak convergence of $T(b)$ to T, for $b \to \infty$. Indeed, with $\tilde{\boldsymbol{Z}}$ an independent copy of $\boldsymbol{Z}$, we have

$$\mathbb{E} \left| T - T(b) \right| \leq \mathbb{E} \left[|h(\boldsymbol{Z}, \tilde{\boldsymbol{Z}})|\mathbf{1}\{\tilde{\boldsymbol{Z}} \in B(b)\} \right] \leq \|h(\boldsymbol{Z}, \tilde{\boldsymbol{Z}})\|_{L^2(\mathbb{P})} \cdot \mathbb{P}(\tilde{\boldsymbol{Z}} \notin B(b))^{1/2}$$
$$= o(1)$$

as $b \to \infty$.

We finally verify $\lim_{b\to\infty} \limsup_{r\to\infty} \mathbb{P}\left(|T_r(b) - T_r| > \varepsilon\right) = 0$ for any fixed $\varepsilon > 0$. Let $\tilde{\boldsymbol{Z}}_{r,1}$ be an independent copy of $\boldsymbol{Z}_{r,1}$. Applying the Markov inequality, we have

$$\mathbb{P}\left(|T_r(b) - T_r| > \epsilon\right) \leq \mathbb{P}\left(\int_{B^c} |h(\boldsymbol{Z}_{r,1}, \boldsymbol{y})| \, d\mathbb{P}_{\boldsymbol{Z}_{r,1}}(\boldsymbol{y}) > \varepsilon\right)$$
$$\leq \varepsilon^{-1} \mathbb{E}[|h(\boldsymbol{Z}_{r,1}, \tilde{\boldsymbol{Z}}_{r,1})|\mathbf{1}\{\tilde{\boldsymbol{Z}}_{r,1} \in B(b)^c\}].$$

Applying the Cauchy-Schwarz inequality and taking the limit over r results in the upper bound $C \cdot \mathbb{P}\left(\boldsymbol{Z} \in B(b)^c\right)^{1/2}$ by the Portmanteau Theorem and Condition 4 (a), for a constant C not depending on b. The bound goes to 0 for $b \to \infty$ since $B(b)^c \downarrow \emptyset$.

Wichura's theorem implies weak convergence of $h_{1,r}(\boldsymbol{Z}_{r,1})$ to $h_1(\boldsymbol{Z})$, for $r \to \infty$. The stated convergence of moments follows by Example 2.21 in [38] using the Jensen inequality and Condition 4 (a). □

Let $\ell = \ell_n \in \mathbb{N}$ denote the sequence from Condition 2(b). We may assume that $1 < \ell < r$. For $j \in \mathbb{N}$, recall that

$$\boldsymbol{M}_{r-\ell,j} = \max(\boldsymbol{X}_j, \ldots \boldsymbol{X}_{j+r-\ell-1}), \quad \boldsymbol{Z}_{r-\ell,j} = (\boldsymbol{M}_{r-\ell,j} - \boldsymbol{b}_{r-\ell})/\boldsymbol{a}_{r-\ell}. \quad (8.1)$$

Lemma 8.3 (Weak convergence of clipped blocks)**.** *Suppose Conditions 1 and 2(a) and (b) are met. Then, as $n \to \infty$,*

$$(\boldsymbol{Z}_{r,1}, \boldsymbol{Z}_{r-\ell,1}) \rightsquigarrow (\boldsymbol{Z}, \boldsymbol{Z}).$$

Proof. Since $(\boldsymbol{Z}_{r,1}, \boldsymbol{Z}_{r-\ell,1}) = (\boldsymbol{Z}_{r,1}, \boldsymbol{Z}_{r,1}) - (0, \boldsymbol{Z}_{r-\ell,1} - \boldsymbol{Z}_{r,1})$ and since $\boldsymbol{Z}_{r,1}$ converges weakly to $\boldsymbol{Z}$ by assumption, it suffices to show that $\boldsymbol{Z}_{r-\ell,1} - \boldsymbol{Z}_{r,1} = o_{\mathbb{P}}(1)$. In particular, we may assume $d = 1$ and note $(M_{r-\ell,1} - b_{r-\ell})/a_{r-\ell}$ converges weakly to $\boldsymbol{Z}$.

Condition 1 yields local uniform convergence, see the proof of Lemma B.15 in [10], hence $a_r/a_{r-\ell} = 1 + o(1)$ and $(b_{r-\ell} - b_r)/a_r = o(1)$. By Lemma B.15 from [10] we have, for any $\varepsilon > 0$,

$$\mathbb{P}\left(|Z_{r-\ell,1} - Z_{r,1}| \geq \varepsilon\right) = \mathbb{P}\left(|Z_{r-\ell,1} - Z_{r,1}| \geq \varepsilon, M_{r-\ell,1} = M_{r,1}\right) + o(1).$$

Using the convergence of the rescaling sequences and that $Z_{r-\ell}$ is stochastically bounded we have

$$\frac{M_{r-\ell,1} - b_{r-\ell}}{a_{r-\ell}} - \frac{M_{r-\ell,1} - b_r}{a_r} = Z_{r-\ell,1}\left(1 - \frac{a_{r-\ell}}{a_r}\right) - \frac{b_{r-\ell} - b_r}{a_r}$$
$$= O_{\mathbb{P}}(1)o(1) + o(1) = o_{\mathbb{P}}(1).$$

This implies $Z_{r-\ell,1} - Z_{r,1} = o_{\mathbb{P}}(1)$. $\square$

For the next results, let $\Delta_{r,\ell}(j) := h_{1,r}(\boldsymbol{Z}_{r,j}) - h_{1,r-\ell}(\boldsymbol{Z}_{r-\ell,j})$ for $j \in I_n^{\mathrm{db}}$. Furthermore let $(\tilde{\boldsymbol{X}}_j, \ldots, \tilde{\boldsymbol{X}}_{j+r-1})_{j \in I_n^{\mathrm{db}}}$ be i.i.d. copies of $(\boldsymbol{X}_j, \ldots, \boldsymbol{X}_{j+r-1})_{j \in I_n^{\mathrm{db}}}$ and define $\tilde{\boldsymbol{M}}_{r,j} = \max(\tilde{\boldsymbol{X}}_j, \ldots, \tilde{\boldsymbol{X}}_{j+r-1})$ and $\tilde{\boldsymbol{Z}}_{r,j} = (\tilde{\boldsymbol{M}}_{r,j} - \boldsymbol{b}_r)/\boldsymbol{a}_r$ and $\tilde{\boldsymbol{M}}_{r-\ell,j}, \tilde{\boldsymbol{Z}}_{r-\ell,j}$ analogously to (8.1).

Lemma 8.4. *Suppose Conditions 1, 2(a), (b) and 4(a) are met and that h is $\boldsymbol{\lambda}^{2d}$-a.e. continuous. Then*

$$\lim_{n \to \infty} \mathbb{E}\left[\left\{\frac{1}{\sqrt{m}} \sum_{j \in I_n^{\mathrm{db}}} \Delta_{r,\ell}(j)\right\}^2\right] = 0, \qquad \lim_{n \to \infty} \mathbb{E}\left[\left|\Delta_{r,\ell}(1)\right|^p\right] = 0$$

for all $p \in \mathbb{N}$ with $p < 2 + \nu$.

Proof. We start by showing the second convergence. Let $\|\cdot\|_p$ denote the L_p-norm. Writing $\Delta_{r,\ell}(1) = \int h(\boldsymbol{Z}_{r,1}, y_1) - h(\boldsymbol{Z}_{r-\ell,1}, y_2)\,\mathrm{d}\mathbb{P}_{\tilde{\boldsymbol{Z}}_{r,1}} \otimes \mathbb{P}_{\tilde{\boldsymbol{Z}}_{r-\ell,1}}(y_1, y_2) + \vartheta_{r-\ell} - \vartheta_r$, we obtain

$$\|\Delta_{r,\ell}(1)\|_p \le \|h(\boldsymbol{Z}_{r,1}, \tilde{\boldsymbol{Z}}_{r,1}) - h(\boldsymbol{Z}_{r-\ell,1}, \tilde{\boldsymbol{Z}}_{r-\ell,1})\|_p + |\vartheta_r - \vartheta_{r-\ell}| =: r_{1,n} + r_{2,n},$$

by Jensen's inequality. Since $\lim_{r \to \infty} \vartheta_r = \vartheta$, we have $r_{2,n} = o(1)$ for $n \to \infty$. By Lemma 8.3 and independence, the vector $(\boldsymbol{Z}_{r,1}, \boldsymbol{Z}_{r-\ell,1}, \tilde{\boldsymbol{Z}}_{r,1}, \tilde{\boldsymbol{Z}}_{r-\ell,1})$ converges weakly to $(\boldsymbol{Z}, \boldsymbol{Z}, \tilde{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}})$, where $\boldsymbol{Z}, \tilde{\boldsymbol{Z}}$ are i.i.d. with cdf G. Therefore, the continuous mapping theorem yields $|h(\boldsymbol{Z}_{r,1}, \tilde{\boldsymbol{Z}}_{r,1}) - h(\boldsymbol{Z}_{r-\ell,1}, \tilde{\boldsymbol{Z}}_{r-\ell,1})| = o_{\mathbb{P}}(1)$. By Condition 4(a) we have asymptotic uniform integrability of $|h(\boldsymbol{Z}_{r,1}, \tilde{\boldsymbol{Z}}_{r,1})|^{2+\nu}$ so that $r_{1,n} = o(1)$ by Example 2.21 in [38] and stationarity.

Using stationarity and observing that the $\Delta_{r,\ell}(j)$ are centered we have

$$\mathbb{E}\left[\left\{\frac{1}{\sqrt{m}} \sum_{j \in I_n^{\mathrm{db}}} \Delta_{r,\ell}(j)\right\}^2\right]$$
$$\le 3\operatorname{Var}(\Delta_{r,\ell}(1)) + 2\sum_{s=2}^{m-1}(1 - \frac{s}{m})|\operatorname{Cov}(\Delta_{r,\ell}(1), \Delta_{r,\ell}(1+rs))|.$$

By Lemma 3.11 in [13], Condition 4(a) and 2 (c), there exists a constant $C > 0$ that is independent of $s \ge 2$ and n such that $|\operatorname{Cov}(\Delta_{r,\ell}(1), \Delta_{r,\ell}(1+rs))| \le C\alpha(r)^{\nu/(2+\nu)}$. Now $\alpha(r) \le \beta(r)$ and Condition 2 (c) imply that the sum in the upper display converges to 0. Hence, an application of the second claim of this lemma implies the first claim and the proof is finished. $\square$

Lemma 8.5 (Restriction to independent blocks)**.** *Suppose Conditions 1, 2(a), (b), and 4(a) are met, and that h is $\boldsymbol{\lambda}^{2d}$-a.e. continuous. Then $m^{-1/2} \sum_{j \in I_n^{\mathrm{db}}} h_{1,r}(\boldsymbol{Z}_{r,j})$ converges weakly if and only if $m^{-1/2} \sum_{j \in I_n^{\mathrm{db}}} h_{1,r}(\tilde{\boldsymbol{Z}}_{r,j})$ converges weakly. In that case the weak limits coincide.*

Proof. The result follows from a standard argument involving characteristic functions and Lemma 8.4; see, for instance, the proof of Theorem 3.6 in [7]. $\square$

8.2. Sliding blocks – stationary case

Recall the definitions of G_ξ, L_ξ and C_ξ from (3.8), (3.7) and (3.6), respectively. Recall the convention that $L := \mathrm{id}_{[0,\infty]}$ if $d = 1$, which implies $C = \mathrm{id}_{[0,1]}$. The following is a generalization of Lemma B.3 in (the supplementary material to) [10] for dimensions $d \geq 1$.

Lemma 8.6. *Suppose that Conditions 1, 2(a) and (b) are met. Then, for any $\xi \geq 0$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$,*

$$\lim_{n \to \infty} \mathbb{P}(\boldsymbol{Z}_{r,1} \leq \boldsymbol{x}, \boldsymbol{Z}_{r,\xi_r} \leq \boldsymbol{y}) = G_\xi(\boldsymbol{x}, \boldsymbol{y}),$$

where $\xi_r = 1 + \lfloor r\xi \rfloor$. Furthermore, G_ξ is the cdf of a 2d-variate extreme value distribution with copula C_ξ and stable tail dependence function L_ξ.

Proof. We only consider the case $\xi \in [0,1]$; the case $\xi > 1$ can be treated similarly. By the same arguments as in the proof of Lemma B.3 in [10], we obtain that

$$\lim_{n \to \infty} \mathbb{P}(\boldsymbol{Z}_{r,1} \leq \boldsymbol{x}, \boldsymbol{Z}_{r,\xi_r} \leq \boldsymbol{y})$$

$$= G\left(\xi^{-\gamma^{(1)}} x^{(1)} + \frac{\xi^{-\gamma^{(1)}} - 1}{\gamma^{(1)}}, \ldots, \xi^{-\gamma^{(d)}} x^{(d)} + \frac{\xi^{-\gamma^{(d)}} - 1}{\gamma^{(d)}} \right)$$

$$\times G\left(\xi^{-\gamma^{(1)}} y^{(1)} + \frac{\xi^{-\gamma^{(1)}} - 1}{\gamma^{(1)}}, \ldots, \xi^{-\gamma^{(d)}} y^{(d)} + \frac{\xi^{-\gamma^{(d)}} - 1}{\gamma^{(d)}} \right)$$

$$\times G\bigg((1-\xi)^{-\gamma^{(1)}} \left(x^{(1)} \wedge y^{(1)} \right) + \frac{(1-\xi)^{-\gamma^{(1)}} - 1}{\gamma^{(1)}}, \ldots$$

$$\ldots, (1-\xi)^{-\gamma^{(d)}} \left(x^{(d)} \wedge y^{(d)} \right) + \frac{(1-\xi)^{-\gamma^{(d)}} - 1}{\gamma^{(d)}} \bigg). \quad (8.2)$$

Since $-\log G_\gamma(x) = (1 + \gamma x)^{-1/\gamma}$, we may write

$$G(\boldsymbol{x}) = \exp\left\{ - L\left(-(1 + \gamma^{(1)} x^{(1)})^{-1/\gamma^{(1)}}, \ldots, (1 + \gamma^{(d)} x^{(d)})^{-1/\gamma^{(d)}} \right) \right\},$$

A straightforward calculation then shows that the expression on the right-hand side of (8.2) can be written as $G_\xi(\boldsymbol{x}, \boldsymbol{y})$. In particular, C_ξ is a copula, which

can easily be seen to be max-stable, i.e., $C_\xi(\boldsymbol{u}^s, \boldsymbol{v}^s) = C_\xi(\boldsymbol{u}, \boldsymbol{v})^s$ for all $s > 0$ and $\boldsymbol{u}, \boldsymbol{v} \in [0,1]^d$. It is hence an extreme-value copula with the given stable tail-dependence function L_ξ and G_ξ is the cdf of an extreme-value distribution. □

Theorem 8.7 (CLT for sliding blocks)**.** *Suppose that Conditions 1, 2(a), (b) are satisfied and that there exists an $\omega > 0$ with $m_n^{1+\omega} \alpha(r_n) \to 0$. For each $r = r_n$ let $\mathcal{F}_r = \{f_{r,i} \colon \mathbb{R}^d \to \mathbb{R} \mid i \in \mathbb{N}\}$ be a family of deterministic maps with the following properties:*

(i) *$f_{r,r+s} = f_{r,s}$ for all $s \in \mathbb{N}$ and $r = r_n$ with $n \in \mathbb{N}$;*
(ii) *The random variables $f_{r,i}(\boldsymbol{Z}_{r,i})$ are centered for all $i \in \mathbb{N}$ and $r = r_n$ with $n \in \mathbb{N}$;*
(iii) *There exists a $\nu > 2/\omega$ with $\limsup_{n\to\infty} \sup_{i\in\mathbb{N}} \mathbb{E}[|f_{r,i}(\boldsymbol{Z}_{r,i})|^{2+\nu}] < \infty$.*
(iv) *There exists a map $f \colon \mathbb{R}^d \to \mathbb{R}$ such that for, all $\xi \in [0,1], \xi' \in [0,2]$, as $n \to \infty$,*

$$\left(f_{r,\xi_r}(\boldsymbol{Z}_{r,\xi_r}), f_{r,\xi_r'}(\boldsymbol{Z}_{r,\xi_r'})\right) \rightsquigarrow \left(f(\boldsymbol{Z}_{1,|\xi-\xi'|}), f(\boldsymbol{Z}_{2,|\xi-\xi'|})\right), \qquad (8.3)$$

where $\xi_r = 1 + \lfloor r\xi \rfloor, \xi_r' = 1 + \lfloor r\xi' \rfloor$ and $(\boldsymbol{Z}_{1,|\xi-\xi'|}, \boldsymbol{Z}_{2,|\xi-\xi'|}) \sim G_{|\xi-\xi'|}$.

Then, for $n \to \infty$,

$$\frac{\sqrt{m}}{n} \sum_{i=1}^n f_{r,i}(\boldsymbol{Z}_{r,i}) \rightsquigarrow \mathcal{N}(0, \sigma_f^2), \quad \sigma_f^2 := 2 \int_0^1 \mathrm{Cov}(f(\boldsymbol{Z}_{1,\xi}), f(\boldsymbol{Z}_{2,\xi})) \, \mathrm{d}\xi.$$

Proof. The proof is very similar to the one of Theorem 2.6 in [8]. For completeness, it is carried out in the supplement. □

8.3. Sliding blocks – *non-stationary case*

The following theorem is an adaptation of Theorem 8.7 to the non-stationary setting of Section 5.

Theorem 8.8. *Suppose that the sampling scheme from Condition 5 is met and that the underlying time-series $(\boldsymbol{Y}_t)_t$ satisfies Conditions 2(a), (b) and $m_n^{1+\omega} \alpha(r_n) \to 0$ for some $\omega > 0$. For each $r = r_n$, let $\mathcal{F}_r = \{f_{r,i} \colon \mathbb{R}^d \to \mathbb{R} \mid i \in \mathbb{N}\}$ be a family of deterministic maps satisfying Conditions (i) – (iv) of Theorem 8.7. Then, for $n \to \infty$,*

$$\frac{\sqrt{m}}{n} \sum_{i=1}^n f_{r,i}(\boldsymbol{Z}_{r,i}) \rightsquigarrow \mathcal{N}(0, \sigma_f^2), \quad \sigma_f^2 := 2 \int_0^1 \mathrm{Cov}(f(\boldsymbol{Z}_{1,\xi}), f(\boldsymbol{Z}_{2,\xi})) \, \mathrm{d}\xi.$$

Proof. The proof is essentially the same as for Theorem 8.7, with the following simple adaptation: independence of $S_{n,1}^+, S_{n,2}^+, \ldots$ is a direct consequence of the imposed sampling scheme. □

The following result is an extension of Lemma 2.4 from [10] to multivariate time series.

Lemma 8.9. *Suppose the sampling scheme from Condition 5 is met. Then, for every $\xi \geq 0$ and $\boldsymbol{x} \in \mathbb{R}^d$,*

$$\lim_{n \to \infty} \mathbb{P}(\boldsymbol{Z}_{r,\xi_r} \leq \boldsymbol{x}) = G(\boldsymbol{x}),$$

with G from Condition 1 and with $\xi_r := 1 + \lfloor r\xi \rfloor$.

Proof. Note first that for univariate $x \in \mathbb{R}$ and $s > 0, \gamma \in \mathbb{R}$ we have $G_\gamma(x/s^\gamma + (s^{-\gamma} - 1)/\gamma) = G_\gamma(x)^s$. This implies, for $\boldsymbol{x} \in \mathbb{R}^d, \boldsymbol{\gamma} \in \mathbb{R}^d$,

$$G\Big[\Big(\tfrac{x^{(i)}}{s^{\gamma^{(i)}}} + \tfrac{s^{-\gamma^{(i)}}-1}{\gamma^{(i)}}\Big)_{i=1,\ldots,d}\Big] = C\Big[\Big(G_{\gamma^{(i)}}\Big(\tfrac{x^{(i)}}{s^{\gamma^{(i)}}} + \tfrac{s^{-\gamma^{(i)}}-1}{\gamma^{(i)}}\Big)\Big)_{i=1,\ldots,d}\Big]$$
$$= C^s\Big[\Big(G_{\gamma^{(i)}}\big(x^{(i)}\big)\Big)_{i=1,\ldots d}\Big] = G^s(\boldsymbol{x}),$$

by (2.3) and (L1) from Condition 1.

By piecewise stationarity and Condition 1 it suffices to show the result for $\xi \in (0,1)$. Analogous to the proof of Lemma 2.4 from [10] we have

$$\lim_{r \to \infty} \mathbb{P}(\boldsymbol{Z}_{r,\xi_r} \leq \boldsymbol{x})$$
$$= G\Big[\Big(\tfrac{x^{(i)}}{(1-\xi)^{\gamma^{(i)}}} + \tfrac{(1-\xi)^{-\gamma^{(i)}}-1}{\gamma^{(i)}}\Big)_{i=1,\ldots,d}\Big] \cdot G\Big[\Big(\tfrac{x^{(i)}}{\xi^{\gamma^{(i)}}} + \tfrac{\xi^{-\gamma^{(i)}}-1}{\gamma^{(i)}}\Big)_{i=1,\ldots,d}\Big]$$
$$= G(\boldsymbol{x}),$$

where the last equality follows from the identity in the previous display. $\square$

Lemma 8.10. *Suppose the sampling scheme from Condition 5 is met and that the underlying time series $(\boldsymbol{Y}_t)_t$ satisfies Conditions 2(a) and (b). Then, for any $\xi, \xi' \geq 0$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$,*

$$\lim_{n \to \infty} \mathbb{P}(\boldsymbol{Z}_{r,\xi_r} \leq \boldsymbol{x}, \boldsymbol{Z}_{r,\xi'_r} \leq \boldsymbol{y}) = G_{|\xi-\xi'|}(\boldsymbol{x}, \boldsymbol{y}).$$

Proof. This is a slight adaption of the proof of Lemma 8.6 using standard clipping techniques and Lemma 8.9. $\square$

Lemma 8.11. *Suppose the sampling scheme from Condition 5 is met and that the underlying time series $(\boldsymbol{Y}_t)_t$ satisfies Conditions 2(a) and (b). Furthermore, let h be $\boldsymbol{\lambda}^{2d}$-a.e. continuous and bounded on compact sets and suppose that Condition 6(a) is met. Then, for $\xi, \xi' \in [0, \infty)$*

$$\big(h_{1,r,\xi_r}(\boldsymbol{Z}_{r,\xi_r}), h_{1,r,\xi'_r}(\boldsymbol{Z}_{r,\xi'_r})\big) \rightsquigarrow \big(h_1(\boldsymbol{Z}_{1,|\xi-\xi'|}), h_1(\boldsymbol{Z}_{2,|\xi-\xi'|})\big)$$

and marginal moments of order $p < 2 + \nu$, with $p \in \mathbb{N}$, converge.

Proof. We proceed similar as in the proof of Lemma 8.2 and employ the Cramér-Wold Theorem and Wichura's Theorem. Fix $\boldsymbol{a} = (a^{(1)}, a^{(2)}) \in \mathbb{R}^2 \setminus \{\boldsymbol{0}\}$ and let

$$T_n := a^{(1)} h_{1,r,\xi_r}(\boldsymbol{Z}_{r,\xi_r}) + a^{(2)} h_{1,r,\xi'_r}(\boldsymbol{Z}_{r,\xi'_r})$$
$$= \int a^{(1)} h(\boldsymbol{Z}_{r,\xi_r}, \boldsymbol{y}_1) + a^{(2)} h(\boldsymbol{Z}_{r,\xi'_r}, \boldsymbol{y}_2) \, \mathrm{d}\mathbb{P}_{\boldsymbol{Z}_{r,\xi_r}} \otimes \mathbb{P}_{\boldsymbol{Z}_{r,\xi'_r}}(\boldsymbol{y}_1, \boldsymbol{y}_2),$$

$$T := a^{(1)}h_1(\boldsymbol{Z}_{1,|\xi-\xi'|}) + a^{(2)}h_1(\boldsymbol{Z}_{2,|\xi-\xi'|})$$
$$= \int a^{(1)}h(\boldsymbol{Z}_{1,|\xi-\xi'|}, \boldsymbol{y}_1) + a^{(2)}h(\boldsymbol{Z}_{2,|\xi-\xi'|}, \boldsymbol{y}_2) \; \mathrm{d}\mathbb{P}_{\boldsymbol{Z}}^{\otimes 2}(\boldsymbol{y}_1, \boldsymbol{y}_2)$$

and define, for $B := B(b) := [-b,b]^d$ with $b \in \mathbb{N}$,

$$T_n(b) := \int_{B \times B} a^{(1)}h(\boldsymbol{Z}_{r,\xi_r}, \boldsymbol{y}_1) + a^{(2)}h(\boldsymbol{Z}_{r,\xi'_r}, \boldsymbol{y}_2) \; \mathrm{d}\mathbb{P}_{\boldsymbol{Z}_{r,\xi_r}} \otimes \mathbb{P}_{\boldsymbol{Z}_{r,\xi'_r}}(\boldsymbol{y}_1, \boldsymbol{y}_2),$$
$$T(b) := \int_{B \times B} a^{(1)}h(\boldsymbol{Z}_{1,|\xi-\xi'|}, \boldsymbol{y}_1) + a^{(2)}h(\boldsymbol{Z}_{2,|\xi-\xi'|}, \boldsymbol{y}_2) \; \mathrm{d}\mathbb{P}_{\boldsymbol{Z}}^{\otimes 2}(\boldsymbol{y}_1, \boldsymbol{y}_2).$$

We may now proceed analogous to the proof of Lemma 8.2, where we use the extended continuous mapping theorem and the weak convergence from Lemma 8.10 to show that $T_n(b) \rightsquigarrow T(b)$ for $n \to \infty$. □

Acknowledgments

Funding

Supplementary Material

Supplement to "Limit theorems for non-degenerate U-statistics of block maxima for time series"

(doi: 10.1214/24-EJS2269SUPP; .pdf). We provide the proof of Theorem 8.7, an extension of Theorem 3.2 to strong mixing, explicit formulas for some asymptotic variances from Section 4 and additional simulation results on the bias-corrected sliding blocks estimator.

References

[1] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. (2004). *Statistics of Extremes: Theory and Applications. Wiley Series in Probability and Statistics.* John Wiley & Sons Ltd., Chichester. MR2108013
[2] BERBEE, H. C. (1979). Random walks with stationary increments and renewal theory. *MC Tracts.* MR0547109

[3] BERGHAUS, B. and BÜCHER, A. (2018). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Ann. Statist.* **46** 2307–2335. https://doi.org/10.1214/17-AOS1621. MR3845019.

[4] BILLINGSLEY, P. (2013). *Convergence of Probability Measures.* John Wiley & Sons. MR0233396

[5] BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. Update of, and a supplement to, the 1986 original. https://doi.org/10.1214/154957805100000104. MR2178042

[6] BÜCHER, A. and SEGERS, J. (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes* **17** 495–528. MR3252823

[7] BÜCHER, A. and SEGERS, J. (2018). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli* **24** 1427–1462. https://doi.org/10.3150/16-BEJ903. MR3706798

[8] BÜCHER, A. and SEGERS, J. (2018). Inference for heavy tailed stationary time series based on sliding blocks. *Electron. J. Stat.* **12** 1098–1125. https://doi.org/10.1214/18-EJS1415. MR3780041.

[9] BÜCHER, A., VOLGUSHEV, S. and ZOU, N. (2019). On second order conditions in the multivariate block maxima and peak over threshold method. *J. Multivariate Anal.* **173** 604–619. https://doi.org/10.1016/j.jmva.2019.04.011. MR3959186

[10] BÜCHER, A. and ZANGER, L. (2023). On the disjoint and sliding block maxima method for piecewise stationary time series. *Ann. Statist.* **51** 573–598. https://doi.org/10.1214/23-aos2260. MR4600993.

[11] BÜCHER, A. and STAUD, T. (2024). Supplement to "Limit theorems for non-degenerate U-statistics of block maxima for time series". DOI: https://doi.org/10.1214/24-EJS2269.

[12] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory. Springer Series in Operations Research and Financial Engineering.* Springer, New York. An introduction. https://doi.org/10.1007/0-387-34471-3. MR2234156.

[13] DEHLING, H. and PHILIPP, W. (2002). Empirical process techniques for dependent data. In *Empirical Process Techniques for Dependent Data* 3–113. Birkhäuser Boston, Boston, MA. https://doi.org/10.1007/978-1-4612-0099-4_1. MR1958777

[14] DEHLING, H. and WENDLER, M. (2010). Central limit theorem and the bootstrap for U-statistics of strongly mixing data. *Journal of Multivariate Analysis* **101** 126–137. MR2557623

[15] DEHLING, H. and WENDLER, M. (2010). Law of the iterated logarithm for *U*-statistics of weakly dependent observations. In *Dependence in Probability, Analysis and Number Theory* 177–194. Kendrick Press, Heber City, UT. MR2731072.

[16] DOMBRY, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block max-

ima framework. *Bernoulli* **21** 420–436. https://doi.org/10.3150/13-BEJ573. MR3322325

[17] Dombry, C. and Ferreira, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli* **25** 1690–1723. https://doi.org/10.3150/18-BEJ1032. MR3961227

[18] Embrechts, P., Lindskog, F. and McNeil, A. (2001). Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich* **14** 1–50.

[19] Ferreira, A. and de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* **43** 276–298. https://doi.org/10.1214/14-AOS1280. MR3285607

[20] Genest, C. and Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist.* **37** 2990–3022. https://doi.org/10.1214/08-AOS672. MR2541453 (2010i:62095).

[21] Gudendorf, G. and Segers, J. (2010). Extreme-value copulas. In *Copula Theory and Its Applications. Lect. Notes Stat. Proc.* **198** 127–145. Springer, Heidelberg. https://doi.org/10.1007/978-3-642-12465-5_6. MR3051266.

[22] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19** 293–325. https://doi.org/10.1214/aoms/1177730196. MR0026294

[23] Hosking, J. R. M., Wallis, J. R. and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27** 251–261. MR0797563

[24] Hsing, T. (1989). Extreme value theory for multivariate stationary sequences. *J. Multivariate Anal.* **29** 274–291. https://doi.org/10.1016/0047-259X(89)90028-6. MR1004339.

[25] Katz, R. W., Parlange, M. B. and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources* **25** 1287–1304. https://doi.org/10.1016/S0309-1708(02)00056-8.

[26] Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika* **30** 81–93. https://doi.org/10.1093/biomet/30.1-2.81. MR0138175

[27] Kulik, R. and Soulier, P. (2020). *Heavy-Tailed Time Series. Springer Series in Operations Research and Financial Engineering.* Springer, New York. https://doi.org/10.1007/978-1-0716-0737-4. MR4174389.

[28] Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **65** 291–306. https://doi.org/10.1007/BF00532484. MR722133 (85b:60033).

[29] Leung, D. and Drton, M. (2018). Testing independence in high dimensions with sums of rank correlations. *The Annals of Statistics* **46** 280–307. https://doi.org/10.1214/17-AOS1550. MR3766953

[30] Maciunas Landwehr, J., Matalas, N. and Wallis, J. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research* **15** 1055–1064.

[31] Oorschot, J., Segers, J. and Zhou, C. (2023). Tail inference using

extreme U-statistics. *Electron. J. Stat.* **17** 1113–1159. `https://doi.org/10.1214/23-ejs2129`. MR4575028.

[32] PADOAN, S. A., RIBATET, M. and SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.* **105** 263–277. `https://doi.org/10.1198/jasa.2009.tm08577`. MR2757202.

[33] PICKANDS, J. III (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd Session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)* **49** 859–878, 894–902. With a discussion. MR820979.

[34] PRESCOTT, P. and WALDEN, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* **67** 723–724. `https://doi.org/10.1093/biomet/67.3.723`. MR601119 (81m:62046).

[35] ROBERT, C. Y., SEGERS, J. and FERRO, C. A. T. (2009). A sliding blocks estimator for the extremal index. *Electron. J. Stat.* **3** 993–1020. `https://doi.org/10.1214/08-EJS345`. MR2540849.

[36] SEGERS, J. (2001). Extreme of a random sample: limit theorems and statistical applications, PhD thesis, Katholieke Universiteit Leuven, Belgium.

[37] SEN, P. K. (1972). Limiting behavior of regular functionals of empirical distributions for stationary *-mixing processes. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **25** 71–82. `https://doi.org/10.1007/BF00533337`. MR0329003

[38] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge University Press, Cambridge. MR1652247 (2000c:62003).

[39] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics.* Springer, New York. MR1385671

[40] YOSHIHARA, K.-I. (1976). Limiting behavior of *U*-statistics for stationary, absolutely regular processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **35** 237–252. `https://doi.org/10.1007/BF00532676`. MR418179.

[41] ZOU, N., VOLGUSHEV, S. and BÜCHER, A. (2021). Multiple block sizes and overlapping blocks for multivariate time series extremes. *Ann. Statist.* **49** 295–320. `https://doi.org/10.1214/20-AOS1957`. MR4206679