The Annals of Statistics
2021, Vol. 49, No. 1, 295–320
https://doi.org/10.1214/20-AOS1957

MULTIPLE BLOCK SIZES AND OVERLAPPING BLOCKS FOR MULTIVARIATE TIME SERIES EXTREMES

BY NAN ZOU[1,*], STANISLAV VOLGUSHEV[1,†] AND AXEL BÜCHER[2]

[1] *Department of Statistical Sciences, University of Toronto, [*]nan.zou@utoronto.ca; [†]stanislav.volgushev@utoronto.ca*
[2] *Mathematisches Institut, Heinrich-Heine-Universität Düsseldorf, axel.buecher@hhu.de*

Block maxima methods constitute a fundamental part of the statistical toolbox in extreme value analysis. However, most of the corresponding theory is derived under the simplifying assumption that block maxima are independent observations from a genuine extreme value distribution. In practice, however, block sizes are finite and observations from different blocks are dependent. Theory respecting the latter complications is not well developed, and, in the multivariate case, has only recently been established for disjoint blocks of a single block size. We show that using overlapping blocks instead of disjoint blocks leads to a uniform improvement in the asymptotic variance of the multivariate empirical distribution function of rescaled block maxima and any smooth functionals thereof (such as the empirical copula), without any sacrifice in the asymptotic bias. We further derive functional central limit theorems for multivariate empirical distribution functions and empirical copulas that are uniform in the block size parameter, which seems to be the first result of this kind for estimators based on block maxima in general. The theory allows for various aggregation schemes over multiple block sizes, leading to substantial improvements over the single block length case and opens the door to further methodology developments. In particular, we consider bias correction procedures that can improve the convergence rates of extreme-value estimators and shed some new light on estimation of the second-order parameter when the main purpose is bias correction.

1. Introduction. Extreme-value theory provides a central statistical ingredient in various fields like hydrology, meteorology and financial risk management, which all have to deal with highly unlikely but important events; see, for example, Beirlant et al. (2004) for an overview. Mathematically, the properties of such events can be understood by studying the (multivariate) tail of probability distributions and the potential temporal dependence of tail events. Respective statistical methodology typically relies on some version of one of two fundamental approaches: the peaks-over-threshold (POT) method which considers only observations that exceed a certain high threshold, or the block maxima (BM) method which is based on taking maxima of observed values over consecutive blocks of observations and treating those maxima as (approximate) data from an extreme value distribution.

While historically the BM approach was the first to be invented (Gumbel (1958)), the mathematical interest soon shifted toward the POT approach. POT methods are by now well understood, and there is a rich and mature literature on various theoretical and practical aspects of such methods; see de Haan and Ferreira (2006) for a review of many classical results and Drees and Rootzén (2010), Can et al. (2015), Fougères, de Haan and Mercadier (2015), Einmahl, de Haan and Zhou (2016) for recent developments. In the last couple of years, there has been an increased interest in the theoretical aspects of the BM approach for univariate

Received March 2019; revised February 2020.
MSC2020 subject classifications. Primary 62G32, 62E20; secondary 60G70, 62G20.
Key words and phrases. Extreme-value copula, second-order condition, bias correction, mixing coefficients, block maxima, empirical process.

observations, and recent work in this direction includes Dombry (2015), Dombry and Ferreira (2019), Ferreira and de Haan (2015), Bücher and Segers (2018b, 2018a). The case of multivariate observations has received much less attention, and the only theoretical analysis of (componentwise) block maxima in the multivariate setting that we are aware of is due to Bücher and Segers (2014). The present paper is motivated by this apparent imbalance of theoretical developments for BM methods as compared to POT methods in the multivariate case.

It is well known that the analysis of multivariate distributions can be decomposed into two distinct parts: the analysis of marginal distributions and the analysis of the dependence structure as described by the associated copula. Classical results from extreme-value theory further show that the possible dependence structures of extremes have to satisfy certain constraints, but do not constitute a parametric family. In fact, the possible dependence structures may be described in various equivalent ways (see, e.g., Resnick (1987), Beirlant et al. (2004), de Haan and Ferreira (2006)): by the exponent measure μ (Balkema and Resnick (1977)), by the spectral measure Φ (de Haan and Resnick (1977)), by the Pickands dependence function A (Pickands (1981)), by the stable tail dependence function L (Huang (1992)), by the tail copula Λ (Schmidt and Stadtmüller (2006)), by the madogram ν (Naveau et al. (2009)), by the extreme-value copula C_∞ (see Gudendorf and Segers (2010) for an overview), or by other less popular objects.

Since statistical theory for estimators of, for example, the Pickands dependence function, the stable tail dependence function, or the madogram may be derived from corresponding results for the empirical copula process (see, e.g., Genest and Segers (2009)), we focus on constructing estimators for the extreme-value copula C_∞, which can in turn serve as a fundamental building block for subsequent developments. This approach was also taken in the above-mentioned reference Bücher and Segers (2014), who analyze the empirical copula process based on (disjoint) block maxima, and then apply the results to obtain the asymptotic behavior of estimators for the Pickands dependence function.

The basic observational setting that we consider is the same as in Bücher and Segers (2014): data are assumed to come from a strictly stationary multivariate time series, and we assume that the copula of the random vector of componentwise block-maxima converges, as the block length tends to infinity, to a copula C_∞ which is our main object of interest. However, in contrast to Bücher and Segers (2014), we base our estimators on overlapping instead of disjoint blocks. While the corresponding theoretical analysis is more involved due to the additional dependence introduced by overlaps in the blocks, we show that this always leads to a reduction in the asymptotic variance of the resulting empirical copula process and smooth functionals thereof. Another major difference with Bücher and Segers (2014) is that we consider functional central limit theorems which explicitly involve the block size as a parameter. This generalization is crucial for various applications, some of which are considered in Section 3.

As a first simple but useful application, we consider estimators for C_∞ which are based on aggregating over various block length parameters, thereby providing estimators which are less sensitive to the choice of a single block length parameter. The corresponding asymptotic theory is a straightforward consequence of the asymptotic theory mentioned before. A Monte Carlo simulation study reveals the superiority of the aggregated estimators over their nonaggregated versions in typical finite-sample situations.

A second more involved application concerns the construction of bias-reduced estimators for C_∞ (see Fougères, de Haan and Mercadier (2015), Beirlant et al. (2016) for recent proposals in the multivariate POT approach for i.i.d. observations). As is typically done when tackling the problem of bias reduction in extreme value statistics, the estimators are obtained by explicitly taking into account the second-order structure of the extreme value model in the

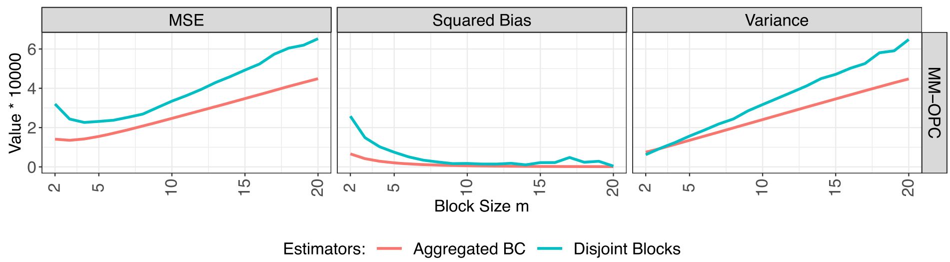

FIG. 1. $10^4\times$*average MSE, squared bias and variance of the disjoint blocks estimator from* Bücher and Segers (2014) *and the aggregated bias corrected estimator proposed in this paper. Data generating process and estimators are as described in Section* 4, *Model* (*MM-OPC*).

estimation step. We are not aware of any results on bias-reduced estimators within the block maxima framework in general. In fact, even for POT methods such results do not seem to exist in the multivariate time series setting (some results on the univariate time series case can be found in de Haan, Mercadier and Zhou (2016)). As a necessary intermediate step for bias correction, we need to consider estimation of a second-order parameter which naturally shows up in the second-order condition. We show that special care needs to be taken when estimating this parameter for its use in bias correction, and propose a penalized estimator which explicitly takes this specific aim into account.

The improvement in both variance and bias of one of the estimators for C_∞ proposed in this paper over the disjoint blocks estimator from Bücher and Segers (2014) is illustrated in Figure 1.

The idea of using sliding/overlapping block maxima for statistical inference appears to be quite new to the extreme value community, whence similar results in the literature actually are rare, even in univariate situations. To the best of our knowledge, the idea first appeared in the context of estimating the extremal index of a univariate stationary time series; see Berghaus and Bücher (2018), Northrop (2015), Robert, Segers and Ferro (2009). The only paper we are aware of in the classical univariate case is Bücher and Segers (2018a), which is restricted to the heavy tailed case. The idea of basing inference on multiple block sizes seems to be new, and is possibly transferable to the univariate case as well.

We further remark that there is a rich and mature literature that deals with estimation of extreme-value copulas and related objects when observations from an extreme-value copula are available (see, among many others, Pickands (1981), Capéraà, Fougères and Genest (1997) for early contributions and Genest and Segers (2009), Gudendorf and Segers (2010) for rank-based methods). However, the setting in that literature is different from ours since we do not assume that data from the extreme value copula are available directly.

Finally, we would like to stress that this paper's main contribution consists of providing a sound theoretical and methodological foundation for further developments in the field, for example, for constructing and validating new application-oriented methods. In fact, many practically relevant statistics such as the madogram (Naveau et al. (2009)) can be written as functionals of the empirical copula, whence a rather straightforward application of our results might concern a sliding blocks version of the madogram.

The remaining parts of this paper are organized as follows: the sliding block maxima (empirical) copula process, including the block length as an argument of the process, is considered in Section 2. The applications on aggregated estimators, bias-reduced estimators and estimators of second-order parameters are worked out in Section 3. Some theoretical examples, as well as a detailed Monte Carlo simulation study are presented in Section 4. All proofs are deferred to a Supplementary Material (Zou, Volgushev and Bücher (2020)).

Throughout, for $\xi \in \mathbb{R}$, let $\lceil \xi \rceil$ be the smallest integer greater or equal to ξ. Let $\langle \xi \rangle$ be the largest integer smaller or equal to ξ if $\xi \geq 0$ and the smallest integer greater or equal to ξ if $\xi < 0$. For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$, write $\boldsymbol{u} \leq \boldsymbol{v}$ if $u_j \leq v_j$ for all j, and $\boldsymbol{u} \nleq \boldsymbol{v}$ if there exists j such that $u_j > v_j$. Let $\boldsymbol{u} \wedge \boldsymbol{v} = (\min(u_1, v_1), \dots, \min(u_d, v_d))$. All convergences will be for $n \to \infty$, if not mentioned otherwise. The arrow $\Rightarrow$ denotes weak convergence in the sense of Hoffman–Jørgensen; see van der Vaart and Wellner (1996).

2. Functional weak convergence of empirical copula processes based on sliding block maxima. Suppose $(\boldsymbol{X}_t)_{t \in \mathbb{Z}} = (X_{t,1}, \dots, X_{t,d})_{t \in \mathbb{Z}}$ is a multivariate strictly stationary process, and that $(\boldsymbol{X}_t)_{t=1}^n$ is observable data. Let $m \in \{1, \dots, n\}$ be a block size parameter and, for $i = 1, \dots, n-m+1$ and $j = 1, \dots, d$, let $M_{m,i,j} = \max\{X_{t,j} : t \in [i, i+m) \cap \mathbb{Z}\}$ be the maximum of the ith sliding block of observations in the jth coordinate. For $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, let $\boldsymbol{M}_{m,i} = (M_{m,i,1}, \dots, M_{m,i,d})$ and

$$F_{m,j}(x) = \mathbb{P}(M_{m,1,j} \leq x), \qquad F_m(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{M}_{m,1} \leq \boldsymbol{x}),$$
$$U_{m,i,j} = F_{m,j}(M_{m,i,j}), \qquad \boldsymbol{F}_m(\boldsymbol{x}) = \big(F_{m,1}(x_1), \dots, F_{m,d}(x_d)\big),$$
$$\boldsymbol{U}_{m,i} = (U_{m,i,1}, \dots, U_{m,i,d}), \qquad \boldsymbol{F}_m^{\leftarrow}(\boldsymbol{x}) = \big(F_{m,1}^{\leftarrow}(x_1), \dots, F_{m,d}^{\leftarrow}(x_d)\big),$$

where $G^{\leftarrow}$ denotes the left-continuous generalized inverse of a c.d.f. G. Subsequently, we assume that the marginal c.d.f.'s of $X_{1,1}, \dots, X_{1,d}$ are continuous. In that case, the marginal c.d.f.'s of $\boldsymbol{M}_{m,1}$ are continuous as well and

$$C_m(\boldsymbol{u}) = \mathbb{P}(\boldsymbol{U}_{m,1} \leq \boldsymbol{u}), \quad \boldsymbol{u} \in [0,1]^d,$$

is the unique copula associated with $\boldsymbol{M}_{m,1}$. Throughout, we shall work under the following fundamental domain-of-attraction condition.

ASSUMPTION 2.1. There exists a copula C_∞ such that

$$\lim_{m \to \infty} C_m(\boldsymbol{u}) = C_\infty(\boldsymbol{u}), \quad \boldsymbol{u} \in [0,1]^d.$$

Note that the convergence is necessarily uniform, by Lipschitz continuity of C_m and C_∞. Typically, the limit C_∞ will be an extreme value copula (Hsing (1989), Hüsler (1990)), that is, $C_\infty(\boldsymbol{u}^{1/s})^s = C_\infty(\boldsymbol{u})$ for all $s > 0$ and $\boldsymbol{u} \in [0,1]^d$ and

$$C_\infty(\boldsymbol{u}) = \exp\{-L(-\log u_1, \dots, -\log u_d)\}, \quad \boldsymbol{u} \in [0,1]^d,$$

for some *stable tail dependence function* $L : [0,\infty]^d \to [0,\infty]$ satisfying:

(i) L is homogeneous: $L(s\cdot) = sL(\cdot)$ for all $s > 0$;
(ii) $L(\boldsymbol{e}_j) = 1$ for $j = 1, \dots, d$, where $\boldsymbol{e}_j$ denotes the jth unit vector;
(iii) $\max(x_1, \dots, x_d) \leq L(\boldsymbol{x}) \leq x_1 + \dots + x_d$ for all $\boldsymbol{x} \in [0,\infty)^d$;
(iv) L is convex;

see, for example, Beirlant et al. (2004). By Theorem 4.2 in Hsing (1989), this is for instance the case if the time series $(\boldsymbol{X}_t)_t$ is beta-mixing. However, C_∞ is in general different from the extreme value attractor, say $C_\infty^{\text{i.i.d.}}$, in case the observations are i.i.d. from the stationary distribution of the time series; see, for instance, Section 4.1 in Bücher and Segers (2014). In fact, (block) maxima calculated from time series naturally incorporate information about the serial dependence (as, e.g., measured by the multivariate extremal index, see Section 10.5.2. in Beirlant et al. (2004)), whence the BM approach is typically more suitable when it comes to, for example, assessing return levels or periods. In the i.i.d. case, Assumption 2.1 is equivalent to the existence of a stable tail dependence function L such that

$\lim_{t\to\infty} t\{1 - C_1(\mathbf{1} - \boldsymbol{x}/t)\} = L(\boldsymbol{x})$ for $x \in [0,\infty)^d$, where the copula C_1 is extended to a c.d.f. on $\mathbb{R}^d$.

Assumption 2.1 does not contain any information about the rate of convergence of C_m to C_∞. In many cases, more precise statements about this rate can be made, and it is even possible to write down higher order expansions for the difference $C_m - C_\infty$. For some of the material in the paper, we will assume the validity of such expansions. Recall that a function φ defined on the integers is regularly varying if $t \mapsto \varphi(\langle t\rangle)$ is regularly varying as a function $(0,\infty) \to \mathbb{R}$.

ASSUMPTION 2.2 (Second-order condition). There exists a regularly varying function $\varphi : \mathbb{N} \to (0,\infty)$ with coefficient of regular variation $\rho_\varphi < 0$ and a (necessarily continuous) nonnull function S on $[0,1]^d$ such that

$$C_m(\boldsymbol{u}) - C_\infty(\boldsymbol{u}) = \varphi(m)S(\boldsymbol{u}) + o\big(\varphi(m)\big) \quad (m \to \infty),$$

uniformly in $\boldsymbol{u} \in [0,1]^d$.

We refer to the accompanying paper Bücher, Volgushev and Zou (2019) for a detailed account on second-order conditions in the i.i.d. case. In particular, the latter paper shows that the block maxima second-order condition above follows from the more common second-order condition imposed on a POT-type convergence to L under fairly general assumptions; see also equation (6) in Fougères, de Haan and Mercadier (2015). It was further shown in Bücher, Volgushev and Zou (2019) that, in the i.i.d. case, the function φ in the condition above must be regularly varying (the part can hence be removed from the assumption), that the function S has certain homogeneity properties and that local uniform convergence on $[\delta, 1]^d$ is sufficient for uniform convergence on $[0,1]^d$. Specific examples in the i.i.d. and time series case are discussed in more detail in Section 4.1.

2.1. *Estimation in the case of known marginal distributions.* We begin by estimating C_∞ in the case of known marginal c.d.f.'s $F_{1,1},\dots,F_{1,d}$, which, on the level of proofs, is a necessary intermediate step when considering the realistic case of unknown marginal c.d.f.'s in the subsequent section. For block size $m' \in \{1,\dots,n\}$, let

$$\hat{C}^{\circ}_{n,m'}(\boldsymbol{u}) = \frac{1}{n-m'+1}\sum_{i=1}^{n-m'+1} \mathbb{1}(\boldsymbol{U}_{m',i} \le \boldsymbol{u}), \quad \boldsymbol{u} \in [0,1]^d, \tag{2.1}$$

denote the empirical c.d.f. of the sample of standardized sliding block maxima $\boldsymbol{U}_{m',1},\dots,\boldsymbol{U}_{m',n-m'+1}$. Subsequently, we will consider block sizes of the form $m' = \langle ma\rangle$ with scaling parameter $a > 0$. The respective centred empirical process we are interested in is

$$\begin{aligned}\mathbb{C}^{\Diamond}_{n,m}(\boldsymbol{u},a) &= \sqrt{n/m}\{\hat{C}^{\circ}_{n,\langle ma\rangle}(\boldsymbol{u}) - C_{\langle ma\rangle}(\boldsymbol{u})\} \\ &= \sqrt{n/m}\frac{1}{b_a}\sum_{i=1}^{b_a}\{\mathbb{1}(\boldsymbol{U}_{\langle ma\rangle,i} \le \boldsymbol{u}) - \mathbb{P}(\boldsymbol{U}_{\langle ma\rangle,i} \le \boldsymbol{u})\},\end{aligned}$$

where $b_a = n - \langle ma\rangle + 1$. For the functional weak convergence results to follow, we consider $\mathbb{C}^{\Diamond}_{n,m}$ as an element of $(\ell^\infty([0,1]^d \times A), \|\cdot\|_\infty)$, the space of bounded function on $[0,1]^d \times A$ equipped with the supremum norm, where $A = [a_\wedge, a_\vee] \subset (0,\infty)$ is a fixed interval, the case $a_\wedge = a_\vee$ being explicitly allowed. We impose the following assumptions on the block length parameter $m = m_n$ and the serial dependence of the time series.

ASSUMPTION 2.3. Denote by $\alpha(\cdot)$ and $\beta(\cdot)$ the α and β mixing coefficients of the process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$, respectively. Assume:

(i) $m = m_n \to \infty, n/m \to \infty$,
(ii) $\alpha(h) = o(h^{-(1+\varrho)})$ as $h \to \infty$, for some $\varrho > 0$,
(iii) $\beta(m)(n/m)^{1/2} \to 0$,
(iv) $\alpha(m)(n/m)^{1/2+\zeta} \to 0$, for some$\zeta \in (0, 1/2)$.

Condition (i) is a typical condition in extreme value statistics, and in fact a necessary condition to allow for consistent estimation of C_∞. Condition (ii) is a short-range dependence condition that we introduce merely for technical reasons associated with our method of proof. At the cost of more sophisticated proofs, the condition may possibly be relaxed. However, since the condition is known to be satisfied for many common time series model, we feel that such a relaxation is not necessarily needed. Assumptions (iii) and (iv) relate the block length parameter to the serial dependence and allow for obtaining central limit theorems (alpha-mixing) and proofs of tightness based on coupling arguments (beta-mixing).

THEOREM 2.4. *Under Assumptions 2.1 and 2.3,*

$$\mathbb{C}_{n,m}^{\Diamond} \Rightarrow \mathbb{C}^{\Diamond} \quad \text{in } \ell^\infty([0,1]^d \times A),$$

where $\mathbb{C}^{\Diamond}$ denotes a tight centred Gaussian process on $[0,1]^d \times A$ with continuous sample paths and covariance function

$$\begin{aligned}\gamma(\boldsymbol{v}, \boldsymbol{u}, c, a) :=& \operatorname{Cov}(\mathbb{C}^{\Diamond}(\boldsymbol{u}, a), \mathbb{C}^{\Diamond}(\boldsymbol{v}, c)) \\ =& \int_{-a}^{0} (C_\infty(\boldsymbol{u}^{1/a}))^{-\xi} (C_\infty(\boldsymbol{v}^{1/c} \wedge \boldsymbol{u}^{1/a}))^{\xi+a} (C_\infty(\boldsymbol{v}^{1/c}))^{c-\xi-a} \, \mathrm{d}\xi \\ &+ \int_{0}^{c-a} (C_\infty(\boldsymbol{v}^{1/c}))^{c-a} (C_\infty(\boldsymbol{v}^{1/c} \wedge \boldsymbol{u}^{1/a}))^{a} \, \mathrm{d}\xi \\ &+ \int_{c-a}^{c} (C_\infty(\boldsymbol{v}^{1/c}))^{\xi} (C_\infty(\boldsymbol{v}^{1/c} \wedge \boldsymbol{u}^{1/a}))^{c-\xi} (C_\infty(\boldsymbol{u}^{1/a}))^{\xi+a-c} \, \mathrm{d}\xi \\ &- (c+a) C_\infty(\boldsymbol{v}) C_\infty(\boldsymbol{u}) \quad (a_\wedge \le a \le c \le a_\vee, \boldsymbol{u}, \boldsymbol{v} \in [0,1]^d).\end{aligned}$$

Perhaps surprisingly, the limiting covariance does not depend on the serial dependence of the original time series, except through C_∞ itself. In the univariate case, this was also observed in Bücher and Segers (2018a).

REMARK 2.5. Under a slightly weaker version of Assumption 2.3, Bücher and Segers (2014), Theorem 3.1, investigated the corresponding empirical process based on disjoint block maxima with $a = c = 1$, that is, the process in $\ell^\infty([0,1]^d)$ defined by

$$\boldsymbol{u} \mapsto \sqrt{n/m} \left\{ \frac{1}{\langle m/n \rangle} \sum_{i=1}^{\langle m/n \rangle} \mathbb{1}(\boldsymbol{U}_{m,1+m(i-1)} \le \boldsymbol{u}) - C_m(\boldsymbol{u}) \right\},$$

and with tight centred Gaussian limit denoted by $\mathbb{C}^D(\boldsymbol{u})$. The covariance function of the limiting process is given by

$$\gamma^D(\boldsymbol{u}, \boldsymbol{v}) = \operatorname{Cov}(\mathbb{C}^D(\boldsymbol{u}), \mathbb{C}^D(\boldsymbol{v})) = C_\infty(\boldsymbol{u} \wedge \boldsymbol{v}) - C_\infty(\boldsymbol{u}) C_\infty(\boldsymbol{v}).$$

A comparison between the covariance functionals γ and γ^D is worked out in Section 2.3 below; cf. Section A.4 in the Supplementary Material (Zou, Volgushev and Bücher (2020)) for an alternative expression for γ.

PROOF OF THEOREM 2.4. Recall $b_a = n - \langle ma \rangle + 1$, let $b = b_1 = n - m + 1$ and define

$$\mathbb{C}_{n,m}^{\Diamond,b}(\boldsymbol{u}, a) = \sqrt{n/m}\frac{1}{b}\sum_{i=1}^{b}\big(\mathbb{1}(\boldsymbol{U}_{\langle ma \rangle,i} \le \boldsymbol{u}) - \mathbb{P}(\boldsymbol{U}_{\langle ma \rangle,i} \le \boldsymbol{u})\big).$$

The proof consists of several steps, which are explicitly taken care of in the Supplementary Material (Zou, Volgushev and Bücher (2020)):

(i) In Lemma A.1, we prove that $\|\mathbb{C}_{n,m}^{\Diamond} - \mathbb{C}_{n,m}^{\Diamond,b}\|_\infty \xrightarrow{p} 0$. Hence it suffices to prove weak convergence of $\mathbb{C}_{n,m}^{\Diamond,b}$.

(ii) In Lemma A.2, we show that $\mathbb{C}_{n,m}^{\Diamond,b}$ is asymptotically uniformly equicontinuous in probability with respect to the $\|\cdot\|_\infty$-norm on $[0, 1]^d \times A$.

(iii) In Lemma A.5, we prove that the finite-dimensional distributions of $\mathbb{C}_{n,m}^{\Diamond,b}$ converge weakly to those of $\mathbb{C}^{\Diamond}$.

Weak convergence of $\mathbb{C}_{n,m}^{\Diamond}$ then follows by combining (i)–(iii). □

The proofs of Step (ii) and Step (iii) are quite lengthy and technical, but it is instructive to present the main ideas within the next two remarks.

REMARK 2.6 (Proving fidi-convergence). The main steps for proving weak convergence of the finite-dimensional distributions are as follows:

(i) *Calculation of the limiting covariance functional* γ. This is treated in Lemma A.4, and bears similarities with common long run variance calculations in classical time series analysis. The integrals in γ are due to the fact that some of the sliding blocks are overlapping, with the integration variable ξ controlling the relative position of two overlapping blocks, and with each of the three integrals corresponding to one of three possibilities for two blocks to overlap: (1) a block of length a starts before a block of length c and ends inside, (2) a block of length a lies completely within a block of length c, or (3) a block of length a starts inside a block of length c and ends outside. Consider for instance the latter case, which would correspond to $0 < c - a < \xi < c$ and amounts to consideration of the event $\{\boldsymbol{M}_{1:\langle mc \rangle} \le \boldsymbol{x}, \boldsymbol{M}_{\langle m\xi \rangle+1:\langle m\xi \rangle+\langle ma \rangle} \le \boldsymbol{y}\}$. The main idea consist of rewriting this event as

$$\{\boldsymbol{M}_{1:\langle m\xi \rangle} \le \boldsymbol{x}\} \cap \{\boldsymbol{M}_{\langle m\xi \rangle+1:\langle mc \rangle} \le \boldsymbol{x} \wedge \boldsymbol{y}\} \cap \{\boldsymbol{M}_{\langle mc \rangle+1:\langle m\xi \rangle+\langle ma \rangle} \le \boldsymbol{y}\}.$$

We then use alpha mixing to show that the three events are asymptotically independent; this eventually gives rise to the three-fold product in the third integral in the definition of γ with each of the factors corresponding to the probability of one of the events above.

(ii) *Big-blocks-small-blocks technique.* The summands of the estimator of interest are collected in successive blocks of (block maxima) observations, with a "big block" followed by a "small block" followed by a "big block," etc. The small blocks are then shown to be negligible, while the big blocks are shown to be asymptotically independent (via alpha mixing). Weak convergence of the sum corresponding to big blocks can finally be shown by an application of the Lyapunov central limit theorem.

REMARK 2.7 (Proving asymptotic tightness). The main steps for proving the tightness part (see Lemma A.2) are as follows:

(i) *Getting rid of serial dependence.* Based on a coupling lemma for beta mixing sequences by Berbee (1979) and a blocking argument, proving tightness of $\mathbb{C}_{n,m}^{\Diamond,b}$ may be reduced to proving tightness of two empirical processes based on row-wise i.i.d. observations.

In contrast to classical time series settings where blocks are based on the original observations, we consider blocks of collections of block maxima corresponding to all block sizes considered. Blocking vectors of block maxima is needed to deal with the additional block length parameter in our setting.

(ii) *Proving tightness via a moment bound.* After the reduction in step (i), we now deal with row-wise i.i.d. observations and the results in van der Vaart and Wellner (1996) can be applied. Here, each "observation" corresponds to a block of collections of block maxima mentioned in the previous step. The moment bound in Theorem 2.14.2 in the latter book allows to deduce tightness of the corresponding processes from controlling the bracketing numbers of certain function classes which map collections of block maxima to pieces in the sum defining $\mathbb{C}^{\Diamond}_{n,m}(\boldsymbol{u}, a)$.

(iii) *Bounding a certain bracketing number.* The last step is based on some explicit lengthy calculations, which take the precise definition of the triangular arrays into account, and in particular the fact that the "observations" are block maxima (with arguments similar to the one given in Remark 2.6 for the calculation of the limiting covariance).

2.2. *Estimation in the case of unknown marginal c.d.f.'s.* The results in Section 2.1 are based on the assumption that the marginal c.d.f.'s are known. In practice, this is not realistic and marginals are typically standardized by taking componentwise ranks of observed block maxima. For $x \in \mathbb{R}$, $j = 1, \dots, d$ and block size m', let $\hat{F}_{n,m',j}(x) = \frac{1}{n-m'+1} \sum_{i=1}^{n-m'+1} \mathbb{1}(M_{m',i,j} \le x)$ and consider observable pseudo-observations from $C_{m'}$ defined as

$$\hat{\boldsymbol{U}}_{n,m',i} = (\hat{U}_{n,m',i,1}, \dots, \hat{U}_{n,m',i,d}), \quad \hat{U}_{n,m',i,j} = \hat{F}_{n,m',j}(M_{m',i,j})$$

The observable analog of the estimator $\hat{C}^{\circ}_{n,m'}$ in (2.1) is then given by

$$\hat{C}_{n,m'}(\boldsymbol{u}) = \frac{1}{n-m'+1} \sum_{i=1}^{n-m'+1} \mathbb{1}(\hat{\boldsymbol{U}}_{n,m',i} \le \boldsymbol{u}),$$

and we are interested in the asymptotic behavior of the associated empirical copula process, indexed by $\boldsymbol{u} \in [0, 1]^d$ and block length scaling parameter $a \in A$, defined as

$$\widehat{\mathbb{C}}^{\Diamond}_{n,m}(\boldsymbol{u}, a) = \sqrt{n/m}\{\hat{C}_{n,\langle ma\rangle}(\boldsymbol{u}) - C_{\langle ma\rangle}(\boldsymbol{u})\}.$$

Subsequently, the process will be called *extended empirical copula process based on sliding block maxima*. Additional assumptions are needed for a corresponding weak convergence result.

ASSUMPTION 2.8. For any $j = 1, \dots, d$, the first-order partial derivative $\dot{C}_{\infty,j}(\boldsymbol{u}) = \frac{\partial}{\partial u_j} C_\infty(\boldsymbol{u})$ exists and is continuous on $\{\boldsymbol{u} \in [0, 1]^d : u_j \in (0, 1)\}$.

Recall that such an assumption is even needed for weak convergence of the classical empirical copula process based on i.i.d. observations from C_∞ (Segers (2012)). For completeness, define $\dot{C}_{\infty,j}(\boldsymbol{u}) = 0$ if $u_j \in \{0, 1\}$. Following Bücher and Segers (2014), we do not need differentiability of C_m for finite m. Instead, we will work with the functions

$$\dot{C}_{m,j}(\boldsymbol{v}) := \limsup_{h \downarrow 0} h^{-1}\{C_m(\boldsymbol{v} + h\boldsymbol{e}_j) - C_m(\boldsymbol{v})\},$$

where $j = 1, \dots, d, m \in \mathbb{N}, \boldsymbol{v} \in [0, 1]^d$ and $\boldsymbol{e}_j$ denotes the jth canonical unit vector in $\mathbb{R}^d$. Note that $\dot{C}_{m,j}$ is always defined and satisfies $0 \le \dot{C}_{m,j} \le 1$.

For the upcoming main theorem of this paper, we will need an additional assumption on the quality of convergence of C_m to C_∞, which will eventually allow us to move from the known margins to the unknown margins case. Any of the following three conditions will be sufficient; the first two assumptions have also been considered in Bücher and Segers (2014) (with $k_n = m_n$ in Part (a)), while the third part (a more refined version of (a)) is included specifically for the bias corrections worked out in Section 3.2, where (a) is typically not met.

ASSUMPTION 2.9 (Quality of convergence of C_m to C_∞).

(a) A sequence $(k_n)_{n\in\mathbb{N}}$ of natural numbers with $k_n \to \infty$ is said to satisfy $\mathrm{SC}_1(k_n)$ if $\sqrt{n/k_n}(C_{k_n} - C_\infty)$ is relatively compact in $\mathcal{C}([0,1]^d)$ (the space of continuous, real-valued functions on $[0,1]^d$).

(b) For every $\delta \in (0, 1/2)$, letting $S_{j,\delta} := [0,1]^{j-1} \times [\delta, 1-\delta] \times [0,1]^{d-j}$,

$$\lim_{m\to\infty} \max_{j=1,\dots,d} \sup_{\boldsymbol{u}\in S_{j,\delta}} \left|\dot{C}_{m,j}(\boldsymbol{u}) - \dot{C}_{\infty,j}(\boldsymbol{u})\right| = 0.$$

(c) A sequence $(k_n)_{n\in\mathbb{N}}$ of natural numbers with $k_n \to \infty$ is said to satisfy $\mathrm{SC}_2(k_n)$ if Assumption 2.2 holds, S is uniformly Hölder-continuous of order $\delta \in (0,1]$, $(n/k_n)^{(1-\delta)/2} \times \varphi(k_n) = o(1)$ as $n \to \infty$ and $n \mapsto \sqrt{n/k_n}\{C_{k_n} - C_\infty - \varphi(k_n)S\}$ is relatively compact in $\mathcal{C}([0,1]^d)$.

THEOREM 2.10 (Functional weak convergence of the extended empirical copula process based on sliding block maxima). *Let Assumptions 2.1, 2.3 and 2.8 hold. If either* $\mathrm{SC}_1(\langle m_n a_n\rangle)$ *from Assumption 2.9*(a) *holds for every converging sequence* $(a_n)_{n\in\mathbb{N}}$ *in* A, *or if Assumption 2.9*(b) *holds, or if* $\mathrm{SC}_2(\langle m_n a_n\rangle)$ *from Assumption 2.9*(c) *holds for every converging sequence* $(a_n)_{n\in\mathbb{N}}$ *in* A, *then*

$$\widehat{\mathbb{C}}_{n,m}^{\Diamond} \Rightarrow \widehat{\mathbb{C}}^{\Diamond} \quad \text{in } \ell^\infty([0,1]^d \times A),$$

where, letting $\boldsymbol{u}^{(j)} = (1, \dots, 1, u_j, 1, \dots, 1)$ *with* u_j *at the* jth *coordinate*,

$$\widehat{\mathbb{C}}^{\Diamond}(\boldsymbol{u}, a) = \mathbb{C}^{\Diamond}(\boldsymbol{u}, a) - \sum_{j=1}^{d} \dot{C}_{\infty,j}(\boldsymbol{u})\mathbb{C}^{\Diamond}(\boldsymbol{u}^{(j)}, a). \tag{2.2}$$

If additionally Assumption 2.2 is met, then Theorem 2.10 shows that the uniform convergence rate of $\hat{C}_{n,m}$ to C_∞ is given by $O_{\mathbb{P}}(\sqrt{m/n} + \varphi(m))$, where $\sqrt{m/n}$ corresponds to the stochastic part, while $\varphi(m)$ is due to the deterministic difference between C_m and C_∞. Assuming for simplicity that Assumption 2.2 holds with $\varphi(m) = m^{\rho_\varphi}$, we find that the best possible convergence rate of $\hat{C}_{n,m}$ is obtained by setting $m \asymp n^{1/(1-2\rho_\varphi)}$. In Section 3, we will show that this rate can in fact be improved by combining estimators $\hat{C}_{n,\langle ma\rangle}$ for several values of a. Establishing the asymptotic properties of those estimators will require the full power of Theorem 2.10, including the process convergence uniformly over the block length parameter a.

REMARK 2.11. If Assumption 2.2 is met and if $\sqrt{n/m}\varphi(m) = O(1)$, then it is easy to show (using regular variation of φ) that $\mathrm{SC}_1(\langle m_n a_n\rangle)$ from Assumption 2.9(a) holds for every converging sequence $(a_n)_{n\in\mathbb{N}}$ in A. Similarly, under Assumption 3.1 below and if $\sqrt{n/m}\varphi(m)\psi(m) = O(1)$, then $\mathrm{SC}_2(\langle m_n a_n\rangle)$ holds for every converging sequence $(a_n)_{n\in\mathbb{N}}$ in A.

2.3. *A comparison of the asymptotic variances based on disjoint and sliding block maxima.* The asymptotic variance of the sliding blocks version of the empirical copula with known and estimated margins will be shown to be less than or equal to the asymptotic variance of the corresponding disjoint blocks versions. Since the asymptotic bias of both approaches is the same, this suggests that the sliding blocks estimator, when available, should always be used instead of the disjoint blocks estimator.

THEOREM 2.12. *Suppose C_∞ is an extreme value copula satisfying Assumption 2.8. Let $\widehat{\mathbb{C}}^\Diamond(\boldsymbol{u}, 1)$ denote the weak limit of the empirical copula process based on sliding block maxima defined in (2.2). Similarly, recall $\mathbb{C}^D(\boldsymbol{u})$ as defined in Remark 2.5 and let*

$$\widehat{\mathbb{C}}^D(\boldsymbol{u}) = \mathbb{C}^D(\boldsymbol{u}) - \sum_{j=1}^{d} \dot{C}_{\infty,j}(\boldsymbol{u})\mathbb{C}^D(\boldsymbol{u}^{(j)})$$

denote the weak limit of the corresponding disjoint blocks version (Theorem 3.1 in Bücher and Segers (2014)). Then, for any $\boldsymbol{u}_1, \dots, \boldsymbol{u}_k \in [0,1]^d$, $k \in \mathbb{N}$,

$$\operatorname{Cov}(\widehat{\mathbb{C}}^\Diamond(\boldsymbol{u}_1, 1), \dots, \widehat{\mathbb{C}}^\Diamond(\boldsymbol{u}_k, 1)) \le_L \operatorname{Cov}(\widehat{\mathbb{C}}^D(\boldsymbol{u}_1), \dots, \widehat{\mathbb{C}}^D(\boldsymbol{u}_k))$$

and

$$\operatorname{Cov}(\mathbb{C}^\Diamond(\boldsymbol{u}_1, 1), \dots, \mathbb{C}^\Diamond(\boldsymbol{u}_k, 1)) \le_L \operatorname{Cov}(\mathbb{C}^D(\boldsymbol{u}_1), \dots, \mathbb{C}^D(\boldsymbol{u}_k)), \tag{2.3}$$

where $\le_L$ denotes the Loewner-ordering between symmetric matrices.

The proof is given in Section A.3. In Figure 2, we depict $\operatorname{Var}(\widehat{\mathbb{C}}^\Diamond(\boldsymbol{u}, 1))$ and $\operatorname{Var}(\widehat{\mathbb{C}}^D(\boldsymbol{u}))$, for $\boldsymbol{u} = (u, u)$ with $u \in [0, 1]$, for the Gumbel–Hougaard copula in (4.1) with shape parameter $\beta = 1$ and $\beta = \ln 2/\ln(3/2)$ (see Section A.4 in the Supplementary Material (Zou, Volgushev and Bücher (2020)) for analytical expressions). Note that when $\beta = 1$, the Gumbel–Hougaard copula degenerates to the independence copula, that is, $C_\infty(u_1, u_2) = u_1u_2$ while $\beta = \ln 2/\ln(3/2)$ results in a tail dependence coefficient of $1/2$. The difference between $\operatorname{Var}(\widehat{\mathbb{C}}^\Diamond(\boldsymbol{u}, 1))$ and $\operatorname{Var}(\widehat{\mathbb{C}}^D(\boldsymbol{u}))$ is seen to be substantial, in particular for small values of u.

As a consequence of the previous result, whenever T is a continuous and linear (real-valued) functional on the space of continuous functions on $[0, 1]^d$ (e.g., the Hadamard derivative of a functional $\Phi : \ell^\infty(T) \to \mathbb{R}$ at C_∞, tangentially to the subspace of continuous functions), then

$$\operatorname{Var}(T(\widehat{\mathbb{C}}^\Diamond(\cdot, 1))) \le \operatorname{Var}(T(\widehat{\mathbb{C}}^D)).$$

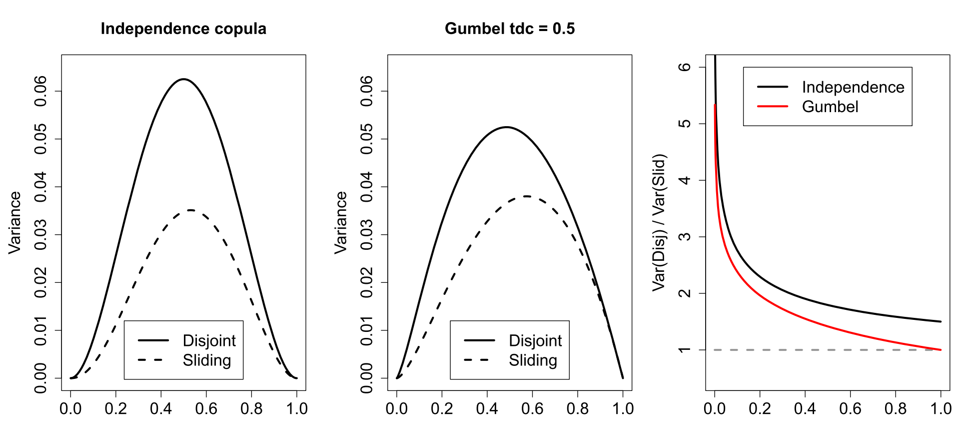

FIG. 2. *Left plot*: $\operatorname{Var}(\widehat{\mathbb{C}}^\Diamond(u, u, 1))$ *(dashed line) and* $\operatorname{Var}(\widehat{\mathbb{C}}^D(u, u))$ *(solid line) as a function of* $u \in [0, 1]$ *for* $C_\infty(u, v) = uv$. *Middle plot*: *same with Gumbel–Hougaard copula with tail dependence coefficient* $1/2$. *Right plot*: $u \mapsto \operatorname{Var}(\widehat{\mathbb{C}}^D(u, u))/\operatorname{Var}(\widehat{\mathbb{C}}^\Diamond(u, u, 1))$.

Indeed, by the Riesz representation theorem (Dudley (2002), Theorem 7.4.1), $T(\mathbb{C}) = \int_{[0,1]^d} \mathbb{C}\,\mathrm{d}\mu$ for some finite signed Borel measure μ on $[0,1]^d$, whence $\operatorname{Var}(T(\mathbb{C})) = \int_{[0,1]^d}\int_{[0,1]^d} \operatorname{Cov}(\mathbb{C}(\boldsymbol{u}), \mathbb{C}(\boldsymbol{v}))\,\mathrm{d}\mu(\boldsymbol{u})\,\mathrm{d}\mu(\boldsymbol{v})$. The claim then follows by measure-theoretic induction. Examples of interesting functionals T can for instance be found in Genest and Segers (2010), Section 3, which comprise Blomqvist's beta, Spearman's footrule, Spearman's rho and Gini's gamma.

3. Applications of the functional weak convergence. The functional weak convergence result in Theorem 2.10 can be applied to large variety of statistical problems. Classical applications include the derivation of the asymptotic behavior of estimators for the Pickands dependence function; see, for example, Section 3.3 in Bücher and Segers (2014). Throughout this section, we discuss applications that explicitly make use of the fact that we allow for various block sizes, allowing one to aggregate over those block sizes, to derive bias reduced estimators or to even estimate second-order characteristics.

Despite not being necessary for the bias correction to *work*, many of the results in this section can be formulated in a convenient explicit way under the assumption of a third-order condition.

ASSUMPTION 3.1 (Third-order condition). Assumption 2.2 holds and there exists a regularly varying function $\psi : \mathbb{N} \to (0,\infty)$ with coefficient of regular variation $\rho_\psi < 0$ and a (necessarily continuous) nonnull function T on $[0,1]^d$, not a multiple of S, such that, uniformly in $\boldsymbol{u} \in [0,1]^d$,

$$\lim_{k\to\infty} \frac{1}{\psi(k)}\left\{\frac{C_k(\boldsymbol{u}) - C_\infty(\boldsymbol{u})}{\varphi(k)} - S(\boldsymbol{u})\right\} = T(\boldsymbol{u}). \tag{3.1}$$

Under the additional assumption that $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence, it can be proved that ψ in the above condition must be regularly varying under mild additional assumptions (it can hence be removed from the assumption).

LEMMA 3.2. *Assume that the time series $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence. If Assumption 2.2 holds and additionally there exists a function $\psi : \mathbb{N} \to (0,\infty)$ with $\psi(k) = o(1) (k \to \infty)$ and a nonnull function T such that (3.1) holds uniformly in $\boldsymbol{u} \in [0,1]^d$, and if the functions S, S^2/C_∞ and T are linearly independent, then ψ is regularly varying of order $\rho_\psi \le 0$.*

Next, we discuss an additional property of the function φ from Assumption 2.2 which allows to quantify the speed of convergence of

$$r_x(k) = \left(\frac{\langle xk\rangle}{k}\right)^{\rho_\varphi} - \frac{\varphi(\langle xk\rangle)}{\varphi(k)} \tag{3.2}$$

(note that convergence to zero of this difference follows from regular variation of φ). This difference will be important in later parts of the manuscript as it will appear in several bounds that are related to bias correction.

LEMMA 3.3. *Assume that $\mathcal{X} \subset (0,\infty)$ is compact and that there exists a nonnegative function $\delta : \mathbb{N} \to [0,\infty)$ with $\lim_{k\to\infty} \delta(k) = 0$ such that, uniformly in $x \in \mathcal{X}$,*

$$C_{\langle xk\rangle}(\boldsymbol{u}) = C_k\big(\boldsymbol{u}^{1/x_k}\big)^{x_k} + O\big(\delta(k)\big) \quad (k\to\infty) \tag{3.3}$$

for any $\boldsymbol{u} \in (0,1)^d$, where $x_k := \langle xk\rangle/k$. Under Assumption 3.1 we have, uniformly in $x \in \mathcal{X}$,

$$r_x(k) = \left(\frac{\langle xk\rangle}{k}\right)^{\rho_\varphi} - \frac{\varphi(\langle xk\rangle)}{\varphi(k)} = O\big(\varphi(k) + \psi(k) + \delta(k)/\varphi(k)\big) \quad (k\to\infty).$$

In the i.i.d. case, equation (3.3) obviously holds with $\delta \equiv 0$. The next result provides a bound on the difference in (3.3) under mixing conditions.

LEMMA 3.4. *Let Assumption 2.1 hold with an extreme-value copula C_∞. Further, let $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ be α-mixing with mixing coefficients $\alpha(k) = O(k^{-(1+\varrho)})$ for some $\varrho > 0$. Then (3.3) holds with $\delta(k) = O(k^{-(1+\varrho)/(2+\varrho)} \log k)$.*

3.1. *Improved estimation by aggregation over block lengths.* Since the functional weak convergence result in Theorem 2.10 involves a scaling parameter for the block length, we may easily analyse estimators for C_∞ which are based on aggregating over several blocks. More formally, we consider the following general construction: for a set $M = M_n \subset \{1, \dots, n\}$ of block length parameters and a set $w = \{w_{n,k} : k \in M_n\}$ of weights satisfying $\sum_{k\in M} w_{n,k} = 1$ for all $n \in \mathbb{N}$, let

$$\hat{C}^{\mathrm{agg}}_{n,(M,w)}(\boldsymbol{u}) = \sum_{k\in M_n} w_{n,k}\hat{C}_{n,k}(\boldsymbol{u}), \quad \boldsymbol{u} \in [0, 1]^d.$$

To derive the asymptotic distribution of this weighted aggregated estimator, we make the following assumption on the tuple (M, w).

ASSUMPTION 3.5. Let $m = m_n$ denote the sequence from Assumption 2.3. For some closed interval $A = [a_\wedge, a_\vee] \subset (0, \infty)$ of positive length, we have

$$M = M_n = \{k \in \mathbb{N} : k/m \in A\}$$

and the weights $w_{n,k}$ satisfy $\lim_{n\to\infty} m w_{n,\langle ma\rangle} = f(a)$ uniformly over A for some continuous f on A with $\int_A f(a)\,\mathrm{d}a = 1$.

For instance, given a continuous function f on A that integrates to unity, we may choose the weights $w_{n,k} = f(k/m)/\{\sum_{\ell\in M_n} f(\ell/m)\}$.

PROPOSITION 3.6. *Let any of the sufficient conditions in Theorem 2.10 be met and assume that additionally Assumption 3.5 is true. Then, in $\ell^\infty([0, 1]^d)$,*

$$\sqrt{\frac{n}{m}}\Big(\hat{C}^{\mathrm{agg}}_{n,(M,w)}(\cdot) - C_\infty(\cdot) - \sum_{k\in M_n} w_{n,k}\big\{C_k(\cdot) - C_\infty(\cdot)\big\}\Big) \Rightarrow \int_A f(a)\widehat{\mathbb{C}}^\diamond(\cdot, a)\,\mathrm{d}a.$$

Note that the asymptotic results in Theorem 2.10 imply that the asymptotic variance of $\hat{C}_{n,\langle ma\rangle}(\boldsymbol{u})$ is proportional to ma/n. For simplicity ignoring the dependence between $\hat{C}_{n,k}(\boldsymbol{u})$ for different k, this motivates the choice $w_{n,k} = k^{-1}/(\sum_{\ell\in M_n} \ell^{-1})$, which is in fact the solution to the minimization problem "minimize $\sum_k (k/n)w^2_{n,k}$ over $w_{n,k}$ with $\sum_k w_{n,k} = 1$." The corresponding function f is $f(a) = c/a$, with c a normalizing constant such that the integral over f is one. Despite this being a crude approximation since $\hat{C}_{n,k}(\boldsymbol{u})$ will be strongly dependent for different values of k, it performs reasonably well in simulations where we will see that in many cases it leads to an improvement in MSE. An alternative approach to choosing $w_{n,k}$ would consist of estimating the entire variance-covariance matrix of $\{\hat{C}_{n,k}(\boldsymbol{u}) : k \in M_n\}$ and minimize a corresponding quadratic form of $w_{n,k}$. We leave a detailed investigation of this question to future research.

Finally, note that if the second-order condition from Assumption 2.2 holds, then the deterministic bias term (see also the discussion in the next section) in Proposition 3.6 can be further decomposed as

$$B^{\mathrm{agg}}_{n,(M,w)}(\boldsymbol{u}) \equiv \sum_{k\in M_n} w_{n,k}\{C_k(\boldsymbol{u}) - C_\infty(\boldsymbol{u})\}$$
$$= \varphi(m)S(\boldsymbol{u}) \sum_{k\in M_n} w_{n,m(k/m)}\{(k/m)^{\rho_\varphi} + o(1)\}$$
$$= \varphi(m)S(\boldsymbol{u}) \int_A f(a)a^{\rho_\varphi}\,\mathrm{d}a + o(\varphi(m)).$$

Note in particular that the asymptotic bias vanishes if $\varphi(m)\sqrt{n/m} = o(1)$.

3.2. *Bias correction.* We begin by commenting on the notion of bias of $\hat{C}_{n,\langle ma\rangle}(\boldsymbol{u})$ as an estimator for the attractor copula $C_\infty(\boldsymbol{u})$. The difference $\hat{C}_{n,\langle ma\rangle}(\boldsymbol{u}) - C_\infty(\boldsymbol{u})$ can be naturally decomposed into two terms

$$D^\diamondsuit_{n,m}(\boldsymbol{u}, a) = \hat{C}_{n,\langle ma\rangle}(\boldsymbol{u}) - C_{\langle ma\rangle}(\boldsymbol{u}), \qquad B^\diamondsuit_{n,m}(\boldsymbol{u}, a) = C_{\langle ma\rangle}(\boldsymbol{u}) - C_\infty(\boldsymbol{u}).$$

The first term captures the stochastic part of $\hat{C}_{n,\langle ma\rangle}(\boldsymbol{u}) - C_\infty(\boldsymbol{u})$ and may be rewritten as

$$D^\diamondsuit_{n,m}(\boldsymbol{u}, a) = \sqrt{\frac{m}{n}}\,\widehat{\mathbb{C}}^\diamondsuit_{n,m}(\boldsymbol{u}, a) = O_{\mathbb{P}}\left(\sqrt{\frac{m}{n}}\right).$$

Recall that $\widehat{\mathbb{C}}^\diamondsuit_{n,m}$ converges to a *centered* Gaussian process. For this reason, throughout the remaining part of this paper, when discussing the bias of an estimator, we mostly concentrate on (versions of) the deterministic sequence $B^\diamondsuit_{n,m}$, which might in fact be of larger order than $O((m/n)^{1/2})$ and which we will call the approximation part of the bias. Note that this is a slight abuse of terminology as we never prove results about $\mathbb{E}[D^\diamondsuit_{n,m}(\boldsymbol{u}, a)]$; however, a similar approach has also been taken in Fougères, de Haan and Mercadier (2015).

Regarding the approximation part of the bias, note that the fundamental Assumption 2.1 only guarantees that $B^\diamondsuit_{n,m} = o(1)$. Under the second-order condition from Assumption 2.2, however, we obtain a hold on both the size and the direction of the bias:

$$\begin{aligned} B^\diamondsuit_{n,m}(\boldsymbol{u}, a) &= \varphi(\langle ma\rangle)S(\boldsymbol{u}) + o(\varphi(\langle ma\rangle)) = \varphi(m)a^{\rho_\varphi}S(\boldsymbol{u}) + o(\varphi(m)) \\ &= O(\varphi(m)). \end{aligned} \tag{3.4}$$

In this section, we exploit the generality of Theorem 2.10 to construct estimators for C_∞ with a smaller order approximation bias.

More precisely, in the current Section 3.2, we present three approaches on how to reduce the bias under either the preliminary assumption that the second-order coefficient ρ_φ is known, or that an estimate $\hat{\rho}_\varphi$ is available. In the next section, we will then discuss how to obtain such an estimate. For the remaining parts of Section 3, suppose that the third-order condition from Assumption 3.1 is met, which implies the expansion

$$C_m(\boldsymbol{u}) - C_\infty(\boldsymbol{u}) = \varphi(m)S(\boldsymbol{u}) + \varphi(m)\psi(m)T(\boldsymbol{u}) + o(\varphi(m)\psi(m)), \tag{3.5}$$

$m \to \infty$, for the approximation part of the bias of $\hat{C}_{n,m} - C_\infty$.

3.2.1. *Naive bias-corrected estimator.* The expansion in (3.5) implies that, assuming $ma \in \mathbb{N}$ for simplicity for the moment,

$$C_{ma}(\boldsymbol{u}) - C_m(\boldsymbol{u}) = \{\varphi(ma) - \varphi(m)\}S(\boldsymbol{u}) + O(\varphi(m)\psi(m))$$
$$= (a^{\rho_\varphi} - 1)\varphi(m)S(\boldsymbol{u}) + O(\varphi(m)\psi(m)).$$

This suggests that the leading bias term $\varphi(m)S(\boldsymbol{u})$ in Expansion (3.5) can be estimated by the plug-in version $\{\hat{C}_{m',n}(\boldsymbol{u}) - \hat{C}_{m,n}(\boldsymbol{u})\}/\{(m'/m)^{\rho_\varphi} - 1\}$ where $m' \neq m$ is an integer and we set $a = m'/m$ in the expansion above. Subtracting this estimated bias from the estimator $\hat{C}_{n,m}$ naturally leads to the following *naive bias-corrected estimator*

$$\hat{C}^{\mathrm{bc,nai}}_{n,(m,m')}(\boldsymbol{u}) = \hat{C}_{n,m}(\boldsymbol{u}) - \frac{\hat{C}_{n,m'}(\boldsymbol{u}) - \hat{C}_{n,m}(\boldsymbol{u})}{(m'/m)^{\rho_\varphi} - 1}.$$

Note that this estimator is infeasible in practice since ρ_φ is unknown. A feasible estimator denoted by $\check{C}^{\mathrm{bc,nai}}_{n,(m,m')}$, can be obtained by replacing ρ_φ with an estimator $\hat{\rho}_\varphi$. In the result below, we quantify the impact of such a replacement under the mild condition $\hat{\rho}_\varphi = \rho_\varphi + o_{\mathbb{P}}(1)$, estimators satisfying this assumption will be presented in Section 3.3 below. Furthermore, it is worthwhile to mention that $\hat{C}^{\mathrm{bc,nai}}_{n,(m,m')} = \hat{C}^{\mathrm{bc,nai}}_{n,(m',m)}$ as can be verified by a simple calculation.

Assuming that $m' = \langle ma \rangle$ for some fixed value $a \in (0, \infty)$, $a \neq 1$, the asymptotic distribution of this estimator is as follows.

PROPOSITION 3.7. *Let any of the sufficient conditions in Theorem 2.10 be met. Additionally, suppose that Assumption 3.1 is met and assume that $m' = \langle ma \rangle$ for some fixed constant $0 < a \neq 1$. Then, in $\ell^\infty([0, 1]^d)$,*

$$\sqrt{\frac{n}{m}}\big(\hat{C}^{\mathrm{bc,nai}}_{n,(m,m')}(\cdot) - C_\infty(\cdot) - B^{\mathrm{bc,nai}}_{n,(m,m')}(\cdot)\big)$$
$$\Rightarrow \widehat{\mathbb{C}}^{\Diamond}_{\mathrm{bc,nai}}(\cdot, a) := \widehat{\mathbb{C}}^{\Diamond}(\cdot, 1) - \frac{\widehat{\mathbb{C}}^{\Diamond}(\cdot, a) - \widehat{\mathbb{C}}^{\Diamond}(\cdot, 1)}{a^{\rho_\varphi} - 1},$$

where the bias term $B^{\mathrm{bc,nai}}_{n,(m,m')}$ admits the expansion

$$B^{\mathrm{bc,nai}}_{n,(m,m')}(\boldsymbol{u}) = \left\{\varphi(m)r_a(m)\frac{S(\boldsymbol{u})}{a^{\rho_\varphi} - 1} + \varphi(m)\psi(m)\frac{1 - a^{\rho_\psi}}{1 - a^{-\rho_\varphi}}T(\boldsymbol{u})\right\}$$
$$+ \varphi(m)o\big(\psi(m) + |r_a(m)|\big),$$

with $r_a(m) = (\langle ma \rangle/m)^{\rho_\varphi} - \varphi(\langle ma \rangle)/\varphi(m) = o(1)$ as in (3.2). In particular, we have

$$\sup_{\boldsymbol{u}\in[0,1]^d} \big|B^{\mathrm{bc,nai}}_{n,(m,m')}(\boldsymbol{u})\big| = \varphi(m)O\big(\psi(m) + |r_a(m)|\big).$$

If moreover $\hat{\rho}_\varphi$ satisfies $\hat{\rho}_\varphi = \rho_\varphi + o_{\mathbb{P}}(1)$, then uniformly in $\boldsymbol{u} \in [0, 1]^d$

$$\check{C}^{\mathrm{bc,nai}}_{n,(m,m')}(\boldsymbol{u}) = \hat{C}^{\mathrm{bc,nai}}_{n,(m,m')}(\boldsymbol{u}) + O_{\mathbb{P}}\big(|\hat{\rho}_\varphi - \rho_\varphi|\{\varphi(m) + \sqrt{m/n}\}\big).$$

Note that the bias term $B^{\mathrm{bc,nai}}_{n,(m,m')}$ is of smaller order than the bias term $B^{\Diamond}_{n,m}$ of the plain empirical copula based on sliding block maxima; see (3.4). Moreover, in the i.i.d. case, we can further bound $|r_a(m)|$ by $O(\varphi(m) + \psi(m))$; see Lemma 3.3 and Remark 3.4.

3.2.2. *Improving the naive bias-corrected estimator by aggregation.* The naive bias-corrected estimator is fairly simple since it only considers two block length parameters m and $m' = \langle am \rangle$. One way to improve this estimator is to consider aggregation over different block lengths; an approach that was shown to work well in Fougères, de Haan and Mercadier (2015) for estimating the stable tail dependence function. Many kinds of aggregation are possible, but for the sake of brevity we will restrict our attention to the following version inspired by Section 3.1 (which works well in finite-sample settings as demonstrated in Section 4)

$$\hat{C}^{\mathrm{bc,agg}}_{n,(m,M,w)}(\boldsymbol{u}) = \sum_{k\in M_n} w_{n,k}\hat{C}^{\mathrm{bc,nai}}_{n,(m,k)}(\boldsymbol{u}).$$

Here, $M = M_n \subset \{1, \dots n\} \setminus \{m_n\}$ and $\{w_{n,k} : k \in M_n\}$ are assumed to satisfy Assumption 3.5. Similar to the discussion in Section 3.2.1, let $\check{C}^{\mathrm{bc,agg}}_{n,(m,M,w)}$ denote a feasible version of $\hat{C}^{\mathrm{bc,agg}}_{n,(m,M,w)}$, with ρ_φ replaced by $\hat{\rho}_\varphi$.

PROPOSITION 3.8. *Let any of the sufficient conditions in Theorem 2.10 be met. Additionally, suppose that Assumption 3.1 is met and that* $(M_n, \{w_{n,k} : k \in M_n\})$ *satisfies Assumption 3.5 and* $1 \notin A$. *Then, in* $\ell^\infty([0,1]^d)$,

$$\sqrt{\frac{n}{m}}\big(\hat{C}^{\mathrm{bc,agg}}_{n,(m,M,w)}(\cdot) - C_\infty(\cdot) - B^{\mathrm{bc,agg}}_{n,(m,M,w)}(\cdot)\big) \Rightarrow \int_A f(a)\widehat{\mathbb{C}}^{\diamondsuit}_{\mathrm{bc,nai}}(\cdot, a)\,\mathrm{d}a,$$

where the bias term $B^{\mathrm{bc,agg}}_{n,(m,M,w)}$ *satisfies*

$$B^{\mathrm{bc,agg}}_{n,(m,M,w)}(\boldsymbol{u}) = \int_A f(a)\Big\{\varphi(m)r_a(m)\frac{S(\boldsymbol{u})}{a^{\rho_\varphi}-1} + \varphi(m)\psi(m)\frac{(1-a^{\rho_\psi})T(\boldsymbol{u})}{1-a^{-\rho_\varphi}}\Big\}\,\mathrm{d}a + o\big(r(m)\big),$$

where, recalling $r_a(m)$ *from* (3.2),

$$r(m) = \varphi(m)\Big(\psi(m) + \sup_{a\in A}\big|r_a(m)\big|\Big). \tag{3.6}$$

In particular, $\sup_{\boldsymbol{u}\in[0,1]^d} |B^{\mathrm{bc,agg}}_{n,(m,M,w)}(\boldsymbol{u})| = O(r(m))$. *If moreover* $\hat{\rho}_\varphi = \rho_\varphi + o_{\mathbb{P}}(1)$ *then we have, uniformly in* $\boldsymbol{u} \in [0,1]^d$,

$$\check{C}^{\mathrm{bc,agg}}_{n,(m,M,w)}(\boldsymbol{u}) = \hat{C}^{\mathrm{bc,agg}}_{n,(m,M,w)}(\boldsymbol{u}) + O_{\mathbb{P}}\big(|\hat{\rho}_\varphi - \rho_\varphi|\{\varphi(m) + \sqrt{m/n}\}\big).$$

3.2.3. *Regression-based bias correction.* A more sophisticated, regression-based estimator (inspired by Beirlant et al. (2016), where the POT-case is tackled) can be motivated by the following consequence of the expansion in (3.5) and the regular variation of $\varphi(\cdot)$:

$$C_{\langle ma\rangle}(\boldsymbol{u}) = C_\infty(\boldsymbol{u}) + a^{\rho_\varphi}\varphi(m)S(\boldsymbol{u}) + r_m(\boldsymbol{u}), \quad m\to\infty, \tag{3.7}$$

for all $a > 0$, where $r_m(\boldsymbol{u}) = o(\varphi(m))$. Letting $y_{i,n} := \hat{C}_{n,k_i}(\boldsymbol{u})$ for suitable values k_i (to be determined below) we find that

$$y_{i,n} = C_\infty(\boldsymbol{u}) + (k_i/m)^{\rho_\varphi}\varphi(m)S(\boldsymbol{u}) + \varepsilon_{i,n}, \tag{3.8}$$

where the remainder $\varepsilon_{i,n}$ contains both the stochastic error $\hat{C}_{n,k_i}(\boldsymbol{u}) - C_{k_i}(\boldsymbol{u})$ and the deterministic error from expansion (3.7). This motivates the following weighted least square estimator for $C_\infty(\boldsymbol{u})$ and $B_m(\boldsymbol{u}) = \varphi(m)S(\boldsymbol{u})$:

$$\begin{aligned}&\big(\hat{C}^{\mathrm{bc,reg}}_{n,(M,w)}(\boldsymbol{u}), \hat{B}^{\mathrm{bc,reg}}_{n,(m,M,w)}(\boldsymbol{u})\big)\\ &\quad\in \operatorname*{arg\,min}_{(b,c)\in\mathbb{R}^2} \sum_{k\in M_n} w_{n,k}\big\{\hat{C}_{n,k}(\boldsymbol{u}) - b - (k/m)^{\rho_\varphi}c\big\}^2,\end{aligned} \tag{3.9}$$

where $w_{n,k}$ and $M = M_n \subset \{1,\dots,n\}$ are as in Section 3.1 with the additional assumption that the weights $w_{n,k}$ are nonnegative. Note that, since the parameter ρ_φ is fixed in the above minimization problem, the value of $\hat{C}^{\mathrm{bc,reg}}_{n,(M,w)}(\boldsymbol{u})$ does in fact not depend on m, and hence we do not need to consider m as an index in $\hat{C}^{\mathrm{bc,reg}}_{n,(M,w)}(\boldsymbol{u})$. Similar to the discussion in Section 3.2.1, let $(\check{C}^{\mathrm{bc,reg}}_{n,(M,w)}(\boldsymbol{u}), \check{B}^{\mathrm{bc,reg}}_{n,(m,M,w)}(\boldsymbol{u}))$ denote a feasible version of $(\hat{C}^{\mathrm{bc,reg}}_{n,(M,w)}(\boldsymbol{u}), \hat{B}^{\mathrm{bc,reg}}_{n,(m,M,w)}(\boldsymbol{u}))$, where ρ_φ is replaced by $\hat{\rho}_\varphi$.

Assuming that M_n contains sufficiently many elements so that the inverse matrix in the next display exists, the minimization problem above has the unique closed-form solution

$$\begin{pmatrix} \hat{C}^{\mathrm{bc,reg}}_{n,(M,w)}(\boldsymbol{u}) \\ \hat{B}^{\mathrm{bc,reg}}_{n,(m,M,w)}(\boldsymbol{u}) \end{pmatrix} = \begin{pmatrix} \mu_{0,n} & \mu_{1,n} \\ \mu_{1,n} & \mu_{2,n} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{k\in M_n} w_{n,k}\hat{C}_{n,k}(\boldsymbol{u}) \\ \sum_{k\in M_n} w_{n,k}(k/m)^{\rho_\varphi}\hat{C}_{n,k}(\boldsymbol{u}) \end{pmatrix}$$

where we defined $\mu_{v,n} := \sum_{k\in M_n} w_{n,k}(k/m)^{v\rho_\varphi}$, $v = 0, 1, 2$. To state the asymptotics of this estimator, define

$$\kappa_v := \int_A f(a)a^{v\rho_\varphi}\,\mathrm{d}a, \qquad T_v(\boldsymbol{u}) := \int_A f(a)a^{v\rho_\varphi}\widehat{\mathbb{C}}^{\Diamond}(\boldsymbol{u}, a)\,\mathrm{d}a,$$

and

$$\mathcal{T}_{m,v}(\boldsymbol{u}) := \int_A f(a)a^{v\rho_\varphi}\{a^{\rho_\varphi+\rho_\psi}\varphi(m)\psi(m)T(\boldsymbol{u}) - \varphi(m)r_a(m)S(\boldsymbol{u})\}\,\mathrm{d}a.$$

PROPOSITION 3.9. *Let any of the sufficient conditions in Theorem 2.10 be met. Additionally, suppose that Assumption 3.1 is met and that $(M_n, \{w_{n,k} : k \in M_n\})$ satisfies Assumption 3.5. Then, in $\ell^\infty([0,1]^d)$,*

$$\sqrt{\frac{n}{m}}\big(\hat{C}^{\mathrm{bc,reg}}_{n,(M,w)}(\cdot) - C_\infty(\cdot) - B^{\mathrm{bc,reg}}_{n,(m,M,w)}(\cdot)\big) \Rightarrow \frac{\kappa_2 T_0(\cdot) - \kappa_1 T_1(\cdot)}{\kappa_2\kappa_0 - \kappa_1^2},$$

where the bias term $B^{\mathrm{bc,reg}}_{n,(m,M,w)}$ satisfies

$$B^{\mathrm{bc,reg}}_{n,(m,M,w)}(\boldsymbol{u}) = \frac{\kappa_2\mathcal{T}_{m,0}(\boldsymbol{u}) - \kappa_1\mathcal{T}_{m,1}(\boldsymbol{u})}{\kappa_2\kappa_0 - \kappa_1^2} + o\big(r(m)\big) = O\big(r(m)\big),$$

with $r(m)$ as defined in (3.6). Moreover,

$$\sqrt{\frac{n}{m}}\big(\hat{B}^{\mathrm{bc,reg}}_{n,(m,M,w)}(\cdot) - \varphi(m)S(\cdot) - \Gamma^B_{n,(m,M,w)}(\cdot)\big) \Rightarrow \frac{\kappa_0 T_1(\cdot) - \kappa_1 T_0(\cdot)}{\kappa_2\kappa_0 - \kappa_1^2}$$

in $\ell^\infty([0,1]^d)$, where the bias term $\Gamma^B_{n,(m,M,w)}$ satisfies

$$\Gamma^B_{n,(m,M,w)}(\boldsymbol{u}) = \frac{\kappa_0\mathcal{T}_{m,1}(\boldsymbol{u}) - \kappa_1\mathcal{T}_{m,0}(\boldsymbol{u})}{\kappa_2\kappa_0 - \kappa_1^2} + o\big(r(m)\big) = O\big(r(m)\big),$$

and the processes involving $\hat{C}^{\mathrm{bc,reg}}_{n,(M,w)}$, $\hat{B}^{\mathrm{bc,reg}}_{n,(m,M,w)}$ converge jointly. If moreover $\hat{\rho}_\varphi = \rho_\varphi + o_{\mathbb{P}}(1)$, then we have, uniformly in $\boldsymbol{u} \in [0,1]^d$

$$\check{C}^{\mathrm{bc,reg}}_{n,(M,w)}(\boldsymbol{u}) = \hat{C}^{\mathrm{bc,reg}}_{n,(M,w)}(\boldsymbol{u}) + O_{\mathbb{P}}\big(r(m) + |\hat{\rho}_\varphi - \rho_\varphi|\{\varphi(m) + \sqrt{m/n}\}\big).$$

3.3. *Estimating the second-order parameter.* Estimators for ρ_φ can be obtained by considering the expansion in (3.7). A simple estimator can be based on the observation that, for any $\boldsymbol{u}$ with $S(\boldsymbol{u}) \neq 0$ and any $a \neq 1$,

$$\frac{C_{\langle ma^2\rangle}(\boldsymbol{u}) - C_m(\boldsymbol{u})}{C_{\langle ma\rangle}(\boldsymbol{u}) - C_m(\boldsymbol{u})} = \frac{a^{2\rho_\varphi} - 1}{a^{\rho_\varphi} - 1} + o(1) = a^{\rho_\varphi} + 1 + o(1), \quad m \to \infty.$$

Letting $m_\rho = m_\rho(n)$ denote a block length parameter (typically chosen of smaller order than the block length m used for estimating C_∞, whence the different notation here), this suggests the following naive estimator for ρ_φ:

$$\hat{\rho}^{\mathrm{nai}}_\varphi(a, \boldsymbol{u}) = \log_a\bigg(\frac{\hat{C}_{n,\langle m_\rho a^2\rangle}(\boldsymbol{u}) - \hat{C}_{n,m_\rho}(\boldsymbol{u})}{\hat{C}_{n,\langle m_\rho a\rangle}(\boldsymbol{u}) - \hat{C}_{n,m_\rho}(\boldsymbol{u})} - 1\bigg).$$

PROPOSITION 3.10. *Let Assumption 3.1 be met and let $m_\rho = m_\rho(n)$ be an increasing sequence of integers such that any of the sufficient conditions in Theorem 2.10 is met for that sequence. Further assume that $(m_\rho/n)^{1/2} = o(\varphi(m_\rho))$. Then, for any $\boldsymbol{u} \in [0,1]^d$ with $S(\boldsymbol{u}) \neq 0$ and any $a \neq 1$, we have*

$$\varphi(m_\rho)\sqrt{\frac{n}{m_\rho}}\big(\hat\rho_\varphi^{\mathrm{nai}}(a,\boldsymbol{u}) - \rho_\varphi - \Gamma_{n,m_\rho}^{\rho,\mathrm{nai}}(\boldsymbol{u},a)\big)$$
$$\Rightarrow \frac{\widehat{\mathbb{C}}^\diamond(\boldsymbol{u},a^2) - \widehat{\mathbb{C}}^\diamond(\boldsymbol{u},1) - (a^{\rho_\varphi}+1)\{\widehat{\mathbb{C}}^\diamond(\boldsymbol{u},a) - \widehat{\mathbb{C}}^\diamond(\boldsymbol{u},1)\}}{S(\boldsymbol{u})a^{\rho_\varphi}(a^{\rho_\varphi}-1)\log a},$$

where

$$\Gamma_{n,m_\rho}^{\rho,\mathrm{nai}}(\boldsymbol{u},a) = \psi(m_\rho)\frac{T(\boldsymbol{u})}{S(\boldsymbol{u})}\frac{(a^{\rho_\varphi+\rho_\psi}-1)(a^{\rho_\psi}-1)}{(a^{\rho_\varphi}-1)\log a}$$
$$+ O\big(r_{a^2}(m_\rho) + r_a(m_\rho) + m_\rho^{-1}\big) + o\big(\psi(m_\rho)\big).$$

In particular, we have

$$\hat\rho_\varphi^{\mathrm{nai}}(a,\boldsymbol{u}) - \rho_\varphi = O_{\mathbb{P}}\bigg(\frac{1}{\varphi(m_\rho)}\sqrt{\frac{m_\rho}{n}}\bigg) + O\big(m_\rho^{-1} + r_{a^2}(m_\rho) + r_a(m_\rho) + \psi(m_\rho)\big).$$

While the estimator $\hat\rho_\varphi^{\mathrm{nai}}(a,\boldsymbol{u})$ defined above is easy to motivate and analyze theoretically, we found in simulations that it does not work well when the sample size n is small or even moderate (up to $n = 5000$). This motivated us to consider alternative estimators by treating ρ_φ in equation (3.8) as unknown. Specifically, we considered estimators of the form

$$(\hat b_0, \hat b_1, \hat\rho_\varphi^{\mathrm{reg}}) \in \operatorname*{arg\,min}_{b_0,b_1,\rho<0} \sum_{k\in M_n} w_{n,k}\big(\hat C_{n,k}(\boldsymbol{u}) - b_0 - b_1(k/m_\rho)^\rho\big)^2, \tag{3.10}$$

where $w_{n,k}$ and $M = M_n \subset \{1,\dots,n\}$ are as in Section 3.1 with the additional assumption that the weights $w_{n,k}$ are nonnegative. This lead to some improvement in performance compared to using $\hat\rho_\varphi^{\mathrm{nai}}$, but still did not lead to very satisfactory results, prompting us to refine the estimator even further.

To gain an intuitive understanding of the shortcomings of $\hat\rho_\varphi^{\mathrm{nai}}$, $\hat\rho_\varphi^{\mathrm{reg}}$ as plug-in estimators for bias correction, we take a closer look at the properties of the quantity

$$\widetilde C_{n,(m,\langle ma\rangle)}^{\mathrm{nai}}(\boldsymbol{u};\gamma) := \hat C_{n,m}(\boldsymbol{u}) - \frac{\hat C_{n,\langle ma\rangle}(\boldsymbol{u}) - \hat C_{n,m}(\boldsymbol{u})}{(\langle ma\rangle/m)^\gamma - 1},$$

which is simply the naive bias-corrected estimator from Section 3.2.1 but with $\gamma < 0$ plugged in instead of the true ρ_φ. We next take a close look at the bias and variance of this "estimator" as a function of γ under the third-order condition from Assumption 3.1. The leading part of the bias is approximately given by

$$\varphi(m)S(\boldsymbol{u})\bigg(1 - \frac{a^{\rho_\varphi}-1}{a^\gamma - 1}\bigg) = \varphi(m)S(\boldsymbol{u})\frac{a^\gamma - a^{\rho_\varphi}}{a^\gamma - 1}.$$

A close analysis reveals that $\gamma \mapsto g(\gamma) := |a^\gamma - a^{\rho_\varphi}|/|a^\gamma - 1|$ is decreasing on $(-\infty,\rho_\varphi)$ with $\lim_{\gamma\to-\infty} g(\gamma) = a^{\rho_\varphi}$ if $a > 1$ and $\lim_{\gamma\to-\infty} g(\gamma) = 1$ if $a < 1$ and increasing on $(\rho_\varphi, 0)$ with $\lim_{\gamma\uparrow 0} g(\gamma) = \infty$ for $a \in (0,\infty)\setminus\{1\}$; see Figure 3 for a picture of the graph for two specific choices of a, ρ_φ. Hence the leading bias will never be increased compared to the original estimator if γ is smaller than ρ_φ, but can increase dramatically if $\gamma > \rho_\varphi$, especially if γ gets close to zero. Similarly, the asymptotic variance of the "bias correction part"

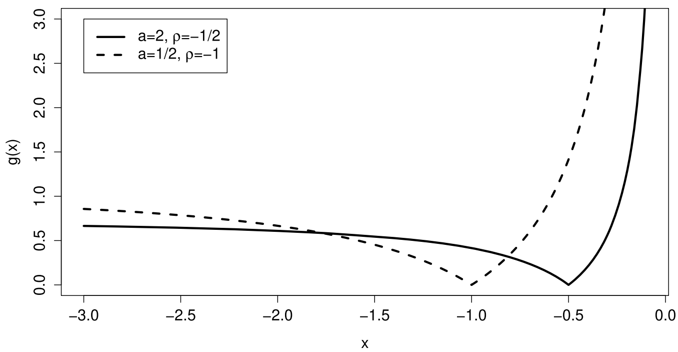

FIG. 3. *Function g for two choices of (a, ρ_φ).*

$\{\hat{C}_{n,\langle ma\rangle}(\boldsymbol{u}) - \hat{C}_{n,m}(\boldsymbol{u})\}/\{(\langle ma\rangle/m)^\gamma - 1\}$ can be found to be a strictly increasing function of γ.

In summary, the above findings suggest a very asymmetric behavior in the performance of the naive bias corrected estimator with respect to values of γ that are too large or too small relative to the true parameter ρ_φ. This apparent asymmetry is not taken into account in the minimization problem (3.10). It thus seems natural to introduce an additional penalty term which discourages the estimator of ρ_φ from being too close to 0. We hence consider the estimator

$$\big(\hat{b}_0(\boldsymbol{u}), \hat{b}_1(\boldsymbol{u}), \hat{\rho}_\varphi^{\text{pen}}(\boldsymbol{u})\big) \in \mathop{\arg\min}_{\rho\in[K',K''],b_0,b_1\in\mathbb{R}} \widehat{\text{RSS}}_\eta(b_0, b_1, \rho; \boldsymbol{u}),$$

where $K' < K'' < 0$ are fixed constants (in the simulations, we choose $K' = -2$ and $K'' = -0.1$), $\eta \ge 0$ denotes a penalty parameter, and

$$\widehat{\text{RSS}}_\eta(b_0, b_1, \rho; \boldsymbol{u}) = \widetilde{\text{RSS}}(b_0, b_1, \rho; \boldsymbol{u}) + \frac{\eta}{|\rho|} \min_{a_0,a_1\in\mathbb{R},K'\le\kappa\le K''} \widetilde{\text{RSS}}(a_0, a_1, \kappa; \boldsymbol{u}),$$

$$\widetilde{\text{RSS}}(b_0, b_1, \rho; \boldsymbol{u}) = \sum_{k\in M_m} w_{n,k}\big\{\hat{C}_{n,k}(\boldsymbol{u}) - b_0 - b_1(k/m_\rho)^\rho\big\}^2.$$

To motivate the factor $\min_{a_0,a_1\in\mathbb{R},K'\le\kappa\le K''} \widetilde{\text{RSS}}(a_0, a_1, \kappa; \boldsymbol{u})$ in the penalty, note that, provided this factor is nonzero, an equivalent representation for the corresponding minimization problem is to minimize

$$\frac{\widetilde{\text{RSS}}(b_0, b_1, \rho; \boldsymbol{u})}{\min_{a_0,a_1\in\mathbb{R},K'\le\kappa\le K''} \widetilde{\text{RSS}}(a_0, a_1, \kappa; \boldsymbol{u})} + \frac{\eta}{|\rho|}.$$

Since the minimal achievable value of the ratio equals 1, this automatically provides a scaling for the penalty part $\frac{\eta}{|\rho|}$ and makes this choice attractive in practice. Finally, observe that the procedure described above produces an estimator of ρ_φ for each value of $\boldsymbol{u}$. We hence propose to further aggregate estimators $\hat{\rho}_\varphi^{\text{pen}}(\boldsymbol{u})$ across different values of $\boldsymbol{u} \in U$ for some finite set $U \subset (0, 1)^d$ to obtain the aggregated estimator

$$\hat{\rho}_{\varphi,U}^{\text{pen,agg}} := \frac{1}{|U|} \sum_{\boldsymbol{u}\in U} \hat{\rho}_\varphi^{\text{pen}}(\boldsymbol{u}).$$

Next, we prove consistency of the estimators defined above.

PROPOSITION 3.11. *Suppose that Assumption 2.2 is met with $\rho_\varphi \in [K', K'']$ and let $m_\rho = m_\rho(n)$ be an increasing sequence of integers such that any of the sufficient conditions* *in Theorem 2.10 is met for that sequence. Further, assume that* $\sqrt{n/m_\rho}\varphi(m_\rho) \to \infty$, *that Assumption 3.5 is met with* m_ρ *instead of* m, *and that* $w_{n,k} > 0$ *for all* k, n. *Then, for any compact* $U \subset \{\boldsymbol{u} \in [0,1]^d : S(\boldsymbol{u}) \neq 0\}$ *and any fixed* $\eta \geq 0$,

$$\sup_{\boldsymbol{u} \in U} \left| \rho_\varphi^{\text{pen}}(\boldsymbol{u}) - \rho_\varphi \right| = o_{\mathbb{P}}(1).$$

Also, $\hat{\rho}_{\varphi,U}^{\text{pen,agg}} = \rho_\varphi + o_{\mathbb{P}}(1)$ *for any finite set* $U \subset \{\boldsymbol{u} \in [0,1]^d : S(\boldsymbol{u}) \neq 0\}$.

4. Examples and finite-sample properties. The proposed estimators will be compared in a simulation study. We begin by providing some details on several examples that will be used in the simulations. Throughout this section, copulas will be denoted by C and D, where D typically refers to the copula of i.i.d. innovations involved in a time series model whose stationary distribution has copula $C = C_1$. For the sake of simplicity, we only consider the case $d = 2$ below. For a generic $d \geq 2$, see Examples D.1 and D.2 in the Supplementary Material (Zou, Volgushev and Bücher (2020)).

4.1. *Examples.*

EXAMPLE 4.1 (t-copula, i.i.d. case). For degrees of freedom $\nu \in \mathbb{N}$ and correlation $\theta \in (-1, 1)$, the t-copula is defined, for $(u, v) \in [0, 1]^2$, as

$$D(u, v; \nu, \theta) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})\pi\nu|P|^{1/2}} \left(1 + \frac{\boldsymbol{x}' P^{-1} \boldsymbol{x}}{\nu}\right)^{-\frac{\nu+2}{2}} \mathrm{d}x_2 \, \mathrm{d}x_1,$$

where $\boldsymbol{x} = (x_1, x_2)'$, P is a 2×2 correlation matrix with off-diagonal element θ, and t_ν is the cumulative distribution function of a standard univariate t-distribution with degrees of freedom ν. Let L and M be the first-order and the second-order POT-type limits associated to D. More specifically,

$$L(x, y) = y t_{\nu+1}\left(\frac{(y/x)^{1/\nu} - \theta}{\sqrt{1-\theta^2}} \sqrt{\nu+1}\right) + x t_{\nu+1}\left(\frac{(x/y)^{1/\nu} - \theta}{\sqrt{1-\theta^2}} \sqrt{\nu+1}\right),$$

and $M = M(x, y)$ is defined in Section 4 and 4.1 of Fougères, de Haan and Mercadier (2015). Recall that $D_\infty(e^{-x}, e^{-y}) = e^{-L(x,y)}$. Let

$$\Gamma_2(x, y) = x^2 (\partial L/\partial x)(x, y) + y^2 (\partial L/\partial y)(x, y).$$

By Theorem 2.6 of Bücher, Volgushev and Zou (2019), Assumption 2.2 holds for $(D_m)_{m \in \mathbb{N}}$ with $D_m(u, v) = D(u^{1/m}, v^{1/m})^m$. Specifically, when $\nu = 1$, we have $\rho_\varphi = -1$, $\varphi(m) = (2m)^{-1}$, and

$$S(e^{-x}, e^{-y}) = D_\infty(e^{-x}, e^{-y})\big(\Gamma_2(x, y) - L^2(x, y)\big);$$

when $\nu = 2$, we have $\rho_\varphi = -1$, $\varphi(m) = (2m/3)^{-1}$, and

$$S(e^{-x}, e^{-y}) = D_\infty(e^{-x}, e^{-y})\big[(1/3)\big(\Gamma_2(x, y) - L^2(x, y)\big) - (2/3) M(x, y)\big];$$

when $\nu = 3, 4, \ldots$, we have $\rho_\varphi = -2\nu^{-1}$, $\varphi(m) = m^{\rho_\varphi}$, and

$$S(e^{-x}, e^{-y}) = -D_\infty(e^{-x}, e^{-y}) M(x, y).$$

EXAMPLE 4.2 (Outer-power transformation of Clayton copula, i.i.d. case). For $\theta > 0$ and $\beta \geq 1$, the outer-power transformation of a Clayton copula is defined as

$$D(u, v; \theta, \beta) = \big[1 + \big\{(u^{-\theta} - 1)^\beta + (v^{-\theta} - 1)^\beta\big\}^{1/\beta}\big]^{-1/\theta}, \quad (u, v) \in [0, 1]^2$$

which is to be interpreted as zero if $\min(u, v) = 0$. By Theorem 4.1 in Charpentier and Segers (2009), D is in the copula domain of attraction of the Gumbel–Hougaard copula with shape parameter β, defined by

$$D_\infty(u, v) = D(u, v; \beta) := \exp[-\{(-\log u)^\beta + (-\log v)^\beta\}^{1/\beta}], \tag{4.1}$$

which is again to be interpreted as zero if $\min(u, v) = 0$. Further, by Proposition 4.3 of Bücher and Segers (2014), for $D_m(u, v) = D(u^{1/m}, v^{1/m})^m$, Assumption 2.2 is met with $\rho_\varphi = -1$, $\varphi(m) = (2m)^{-1}$, and

$$S(u, v) = \theta \Lambda(u, v; \beta),$$

where, letting $x = -\log u$ and $y = -\log v$,

$$\Lambda(u, v; \beta) = D(u, v; \beta)\{(x^\beta + y^\beta)^{2/\beta} - (x^\beta + y^\beta)^{1/\beta - 1}(x^{\beta+1} + y^{\beta+1})\}.$$

EXAMPLE 4.3 (Moving-maximum-process). Let D denote a copula and let $(\boldsymbol{W}_t)_{t\in\mathbb{Z}}$ denote an i.i.d. sequence from D. Fix $p \in \mathbb{N}$ and let a_{ij} $(i = 0, \dots, p;\ j = 1, \dots, d)$ denote nonnegative constants satisfying $\sum_{i=0}^p a_{ij} = 1$ for $j = 1, \dots, d$. The moving maximum process $(\boldsymbol{U}_t)_{t\in\mathbb{Z}}$ of order p is defined as

$$U_{tj} = \max_{i=0,\dots,p} W_{t-i,j}^{1/a_{ij}} \quad (t \in \mathbb{Z};\ j = 1, \dots, d),$$

with the convention that $w^{1/0} = 0$ for $w \in (0, 1)$. As suggested by the notation, the random variables U_{tj} are uniformly distributed on $(0, 1)$, whence a model with arbitrary continuous margins can be obtained by considering $X_{tj} = \eta_j(U_{tj})$ for some strictly increasing (quantile) function $\eta_j : (0, 1) \to \mathbb{R}$.

Assume that the copula D is in the (i.i.d.) copula domain of attraction of an extreme-value copula D_∞, that is, for any $\boldsymbol{u} \in [0, 1]^d$,

$$D_m(\boldsymbol{u}) = \{D(\boldsymbol{u}^{1/m})\}^m \longrightarrow D_\infty(\boldsymbol{u}) \quad (m \to \infty). \tag{4.2}$$

Note that D_m is the copula of the componentwise block maximum of size m, based on the sequence $(\boldsymbol{W}_t)_{t\in\mathbb{N}}$.

As a consequence of Proposition 4.1 in Bücher and Segers (2014), if C_m denotes the copula of the componentwise block maximum of size m based on the sequence $(\boldsymbol{U}_t)_{t\in\mathbb{N}}$, then

$$\lim_{m\to\infty} C_m(\boldsymbol{u}) = D_\infty(u), \quad \boldsymbol{u} \in [0, 1]^d$$

as well, that is, Assumption 2.1 is met. We prove in the Appendix that if Assumption 2.2 is met for $(D_m)_m$ (denote the auxiliary function by φ_D and S_D), then it is also met for $(C_m)_m$ provided that $1/m = o(\varphi_D(m))$, with the same auxiliary functions. In case $1/m \neq o(\varphi_D(m))$, additional technical assumptions are needed and the functions φ_D, S_D and φ, S might differ. Details in the general case are omitted for the sake of brevity.

EXAMPLE 4.4 (Random-repetition-process). Let $\boldsymbol{X}_0, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots$ be a sequence of i.i.d. d-dimensional random vectors. Independently, let $I_1, I_2, \dots$ be a sequence of i.i.d. indicator random variables with $p := \mathbb{P}(I_t = 1) \in (0, 1]$. For $t = 1, 2, \dots$, define

$$\boldsymbol{X}_t = \begin{cases} \boldsymbol{\xi}_t & \text{if } I_t = 1, \\ \boldsymbol{X}_{t-1} & \text{if } I_t = 0. \end{cases}$$

The process $\boldsymbol{X}_0, \boldsymbol{X}_1, \dots$ is a simplification of the doubly stochastic model in Smith and Weissman (1994) and is a stationary process. By this stationarity and by Kolmogorov's extension theorem, we can embed the process $\boldsymbol{X}_0, \boldsymbol{X}_1, \ldots$ in a two-sided stationary process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$. By Bücher and Segers (2014), the process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is β-mixing with $\beta(h) = O((1-p)^h)$ as $h \to \infty$.

Assume that $\boldsymbol{\xi}_1$ has continuous marginal c.d.f.'s and copula D which is in the (i.i.d.) copula domain of attraction of an extreme value copula D_∞; see (4.2). Then, by Bücher and Segers (2014), if C_m denotes the copula of the componentwise block maximum of size m based on the sequence $(\boldsymbol{X}_t)_{t\in\mathbb{N}}$, $C_m \to D_\infty$ as $m \to \infty$ as well. In particular, unlike for the moving maximum process (see equation (8.3) in Bücher and Segers (2014)), the limit of C_m is equal to the i.i.d. attractor of the copula $C_1 = D$ of the stationary distribution.

4.2. *Finite-sample properties.* In this section, we compare the estimators for C_∞ introduced in the previous section by means of Monte Carlo simulations. We focus on the case $d = 2$ below; respective results in higher dimensions are quite similar and do not reveal additional deep insights; see the cases $d = 4, 8$ treated in Section D.2 in the Supplementary Material (Zou, Volgushev and Bücher (2020)). Results for all estimators are reported as follows: each estimator is computed for all values $\boldsymbol{u} \in \mathcal{U} := \{.1, .2, \ldots, .9\}^2$ and block size $m \in \{1, \ldots, 20\}$ (except for the aggregated versions, for which we specify the set of block length parameters below). Squared bias, variance and MSE of each estimator and in each point $\boldsymbol{u} \in \mathcal{U}$ for sample size $n = 1000$ was estimated based on 1000 Monte Carlo replications. For the sake of brevity, we only report summary results which correspond to taking averages of the squared bias, MSE and variance over all values $\boldsymbol{u} \in \mathcal{U}$. We present results on the following six models:

(IID-OPC) i.i.d. realizations from an outer power Clayton copula with $d = 2, \theta = 1, \beta = \log(2)/\log(2 - 0.25)$.

(MM-OPC) A moving maximum process based on an outer power Clayton copula with $d = 2, \theta = 1, \beta = \log(2)/\log(2 - 0.25)$, $a_{11} = 0.25, a_{12} = 0.5$.

(RR-OPC) A random repetition process based on an outer power Clayton copula with $d = 2, \theta = 1, \beta = \log(2)/\log(2 - 0.25)$, and $p = 0.5$.

(IID-T5) i.i.d. realizations from a t-copula with $d = 2, \nu = 5, \theta = 0.5$.

(MM-T5) A moving maximum process based on a t-copula with $d = 2, \nu = 5, \theta = 0.5$ and $a_{11} = 0.25, a_{12} = 0.5$.

(MM-T3) A moving maximum process based on a t-copula with $d = 2, \nu = 3, \theta = 0.25$ and $a_{11} = 0.25, a_{12} = 0.5$.

Additional results for models called (RR-T5), (IID-T3) and (RR-T3) can be found in Section D.1 of the Supplementary Material (Zou, Volgushev and Bücher (2020)). Note that we also investigated other parameter combinations, but chose to only present results for the above models as they provide, to a large extent, a representative subset of the results.

Following the heuristics after Proposition 3.6, weights $w = \{w_{n,k} : k \in M\}$ are always chosen as

$$w_{n,k} = k^{-1}\Big(\sum_{\ell\in M} \ell^{-1}\Big)^{-1}, \tag{4.3}$$

with block length sets $M = M_n$ as specified below, possibly depending on the specific estimator.

4.2.1. *Comparison of estimators without bias correction.* We first focus on the performance of three estimators that do not involve bias correction:

- the disjoint blocks estimator $\hat{C}^D_{n,m}$ from Bücher and Segers (2014); see also Section 2.3;

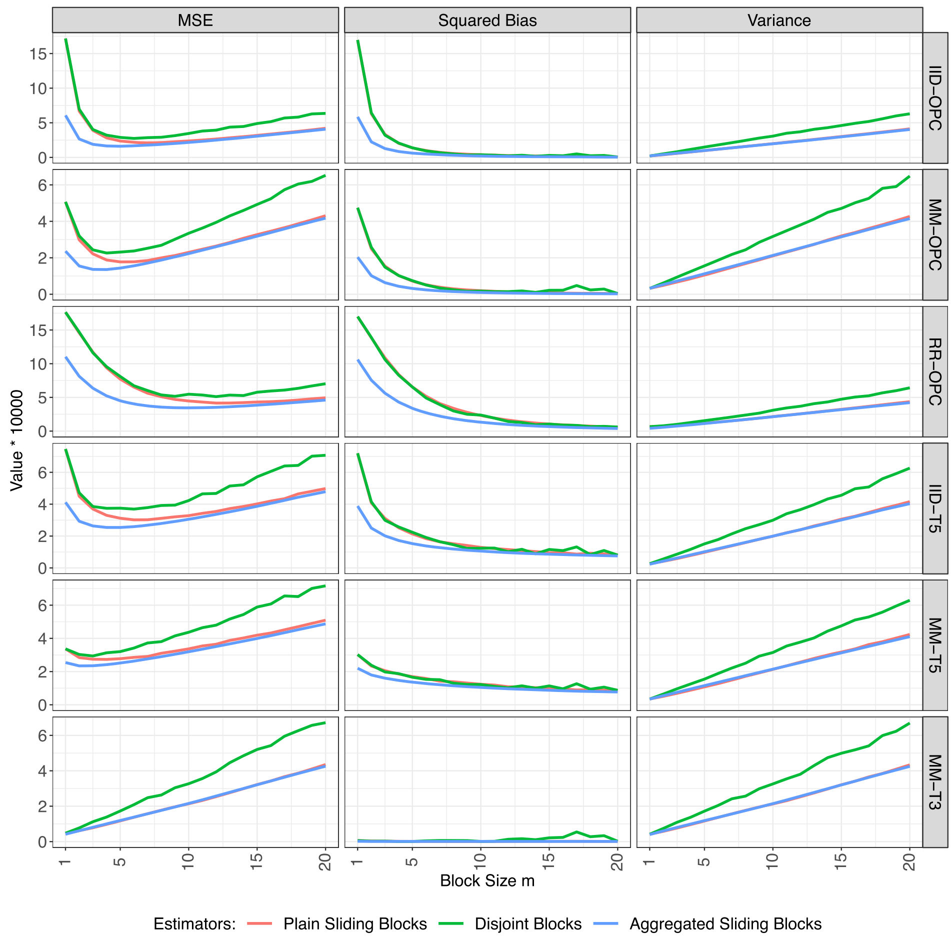

FIG. 4. $10^4 \times$ *average MSE, average squared bias and average variance of plain sliding blocks estimator, disjoint blocks estimator and aggregated sliding blocks estimator, for* $m \geq 1$.

- the plain sliding blocks estimator $\hat{C}_{n,m}$ from Section 2.2;
- the aggregated sliding blocks estimator $\hat{C}^{\text{agg}}_{n,(M,w)}$ from Section 3.1, with block length set $M = \{m, m+1, \ldots, m+9\}$ and weights as in (4.3).

The respective results are shown in Figure 4. As predicted by the theory, the variance curves are linear in m, with the disjoints blocks estimator always exhibiting the largest variance, while the variances of the aggregated and plain version of the sliding blocks estimator are both smaller and similar to each other. In terms of bias, the disjoint and plain sliding blocks estimators $\hat{C}^D_{n,m}$ and $\hat{C}_{n,m}$ show a very similar behavior, with only some smaller deviations (in particular visible for larger block sizes) which may possibly be explained by the fact that the disjoint blocks estimator does not make use of all observations in case the block length m is not a divisor of the sample size $n = 1000$. The aggregated sliding blocks estimator typically has the smallest bias among the three competitors. Finally, in terms of MSE, the aggregated sliding blocks estimator again shows the uniformly best performance. Except for Model (MM-T3), the global minimum of the MSE-curve for $\hat{C}^{\text{agg}}_{n,(M,w)}$ is substantially smaller than the minima for the other two estimators. Model (MM-T3), on the other hand, exhibits little to no bias for all block sizes under consideration, even for $m = 1$. As a consequence, at their minimal MSE, the three estimators yield comparably good results.

When comparing the i.i.d., the moving maximum, and the random repetition models, observant readers might note that the bias in the moving maximum Model (MM-OPC) seems to be the smallest, while the bias in the random repetition Model (RR-OPC) seems to be the largest. Intuitively, this can be explained as follows. First, realizations from moving maximum processes are already based on maxima, and thus it can be expected that their dependence structure is closer to that of a "limiting" max-stable model described by C_∞. Second, realizations from random repetition processes potentially have many repeated observations, and consequently a block size larger than that in the i.i.d. case is needed to bring the dependence structure of the block maxima close to its limit C_∞.

4.2.2. *Comparison of bias corrected estimators.* In this section, three bias corrected estimators for the plain sliding blocks estimator $\hat{C}_{n,m}$ are compared with $\hat{C}_{n,m}$ itself. In all cases, the second-order parameter ρ_φ is estimated through $\hat{\rho}_\varphi = \hat{\rho}^{\text{pen,agg}}_{\varphi,U}$, with the parameters of that estimator set to $K' = -2$, $K'' = -0.1$, $\eta = 1/2$, $U = \{(.1, .1), (.11, .11), \dots, (.5, .5)\}$, $M = \{2, \dots, 50\}$ and weights as in (4.3). We consider the following estimators:

- The naive bias corrected estimator $\check{C}^{\text{bc,nai}}_{n,(m,m')}$ with $m' = 1$ and $m \geq 2$.
- The aggregated naive bias corrected estimator $\check{C}^{\text{bc,agg}}_{n,(m',M,w)}$ with $(m', M) = (1, \{m, \dots, m+9\})$ (where $m \geq 2$ is on the x-axis) and with weights as in (4.3).
- The regression-based bias corrected estimator $\check{C}^{\text{bc,reg}}_{n,(M,w)}$ with $M = \{1, m, m+1, \dots, m+9\}$ (where $m \geq 2$ is on the x-axis) and with weights as in (4.3) (recall from the discussion right after (3.9) that $\check{C}^{\text{bc,reg}}_{n,(M,w)}$ does not depend on the parameter m in that equation).

The choice of small block sizes for the bias correction, in particular $m' = 1$, is motivated by the fact that this choice leads to the best performance in the simulations we tried. Similar observations were made in Fougères, de Haan and Mercadier (2015) who recommend using a very large value for the threshold k in the POT setting.

The results are presented in Figure 5. We observe that the naive bias corrected estimator exhibits, at each fixed block size, a slightly larger variance and a slightly smaller squared bias than the plain sliding blocks empirical copula. In terms of MSE, no universal statement regarding the ordering between the two estimators can be made. Their minimal MSEs (for each separate model, over all block length parameters) are however quite similar. We further find that aggregating the naive bias-corrected estimator leads to substantial improvements for small values of m and no major impact for larger values of m. This is similar to the findings in the previous section. Compared with the plain sliding block estimator, the aggregated bias corrected estimator shows much less sensitivity to the parameter m in the first four models, where there is a substantial bias. In Model (MM-T3), where the bias is negligible compared to the variance, attempts to correct the bias introduce a bit of variance leading to a slight increase in MSE for all block sizes. Finally, the aggregated naive and regression-based bias corrected estimators show very similar performance.

Based on the simulation results, we recommend using the aggregated bias corrected estimator among all bias corrected estimators since it leads to better results than the naive estimator, is reasonably fast to compute (see Section D.3 in the Supplementary Material (Zou, Volgushev and Bücher (2020))), and is simpler to implement than the regression-based estimator. At the same time, it is less sensitive to the choice of the block size parameter compared to the estimator without bias correction.

4.3. *Comments on the choice of m.* In practice, a choice must be made regarding the block length parameter m. This issue is delicate, and in fact comparable to the choice of k, the number of upper order statistics, in the POT-approach, which has no universal answer.

Our theoretical results show that increasing m will increase the (asymptotic) variance (which is proportional to m/n) but decrease the bias (which is proportional to $\varphi(m)$; see

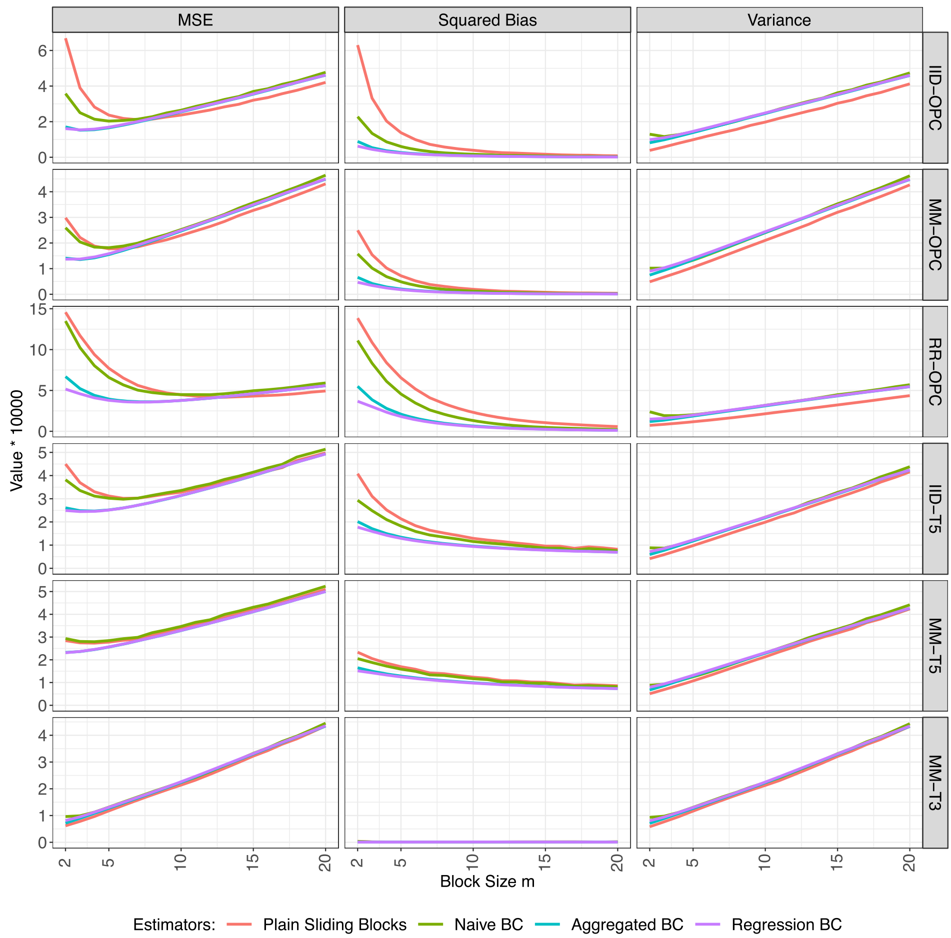

FIG. 5. $10^4 \times$ *average MSE, average squared bias and average variance of plain sliding blocks estimator, naive bias corrected estimator, aggregated naive bias corrected estimator and regression-based bias corrected estimator, for* $m \geq 2$.

also the discussion after Theorem 2.10). This theoretical prediction is clearly observed in the simulation results reported in the previous section. An (asymptotic) MSE-minimizing choice of m would need to balance the bias and variance and would depend on the parameters showing up in the second-order condition and on the asymptotic variance. Those parameters would need to be estimated to obtain a simple plug-in estimator for the optimal rate (our results on estimating ρ_φ provide a first step in that direction). However, within the similar POT-setting, such approaches are typically known to be quiet unreliable in finite samples. A very common and simple procedure consists of identifying stable regions within a plot of, in our case, m against the estimator (Drees, de Haan and Resnick (2000)). Moreover, as shown by our simulation results, applying aggregated or bias-reduced estimators may also make the choice of m less critical, as the dependence of the performance of the estimator on m is less pronounced.

Acknowledgments. The authors would like to thank Sebastian Engelke and Chen Zhou for fruitful discussions. We are also grateful to the associate editor and three anonymous referees for detailed feedback which helped to improve the presentation of our results.

S. Volgushev is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

A. Bücher is supported by the collaborative research center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG).

SUPPLEMENTARY MATERIAL

Supplement to "Multiple block sizes and overlapping blocks for multivariate time series extremes" (DOI: 10.1214/20-AOS1957SUPP; .pdf). The supplement contains the proofs for the results in this paper and additional simulation results.

REFERENCES

BALKEMA, A. A. and RESNICK, S. I. (1977). Max-infinite divisibility. *J. Appl. Probab.* **14** 309–319. MR0438425 https://doi.org/10.1017/s002190020010498x

BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes: Theory and Applications. Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2108013 https://doi.org/10.1002/0470012382

BEIRLANT, J., ESCOBAR-BACH, M., GOEGEBEUR, Y. and GUILLOU, A. (2016). Bias-corrected estimation of stable tail dependence function. *J. Multivariate Anal.* **143** 453–466. MR3431445 https://doi.org/10.1016/j.jmva.2015.10.006

BERBEE, H. C. P. (1979). *Random Walks with Stationary Increments and Renewal Theory. Mathematical Centre Tracts* **112**. Mathematisch Centrum, Amsterdam. MR0547109

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

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 https://doi.org/10.1007/s10687-014-0195-8

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

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

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. MR3959186 https://doi.org/10.1016/j.jmva.2019.04.011

CAN, S. U., EINMAHL, J. H. J., KHMALADZE, E. V. and LAEVEN, R. J. A. (2015). Asymptotically distribution-free goodness-of-fit testing for tail copulas. *Ann. Statist.* **43** 878–902. MR3325713 https://doi.org/10.1214/14-AOS1304

CAPÉRAÀ, P., FOUGÈRES, A.-L. and GENEST, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84** 567–577. MR1603985 https://doi.org/10.1093/biomet/84.3.567

CHARPENTIER, A. and SEGERS, J. (2009). Tails of multivariate Archimedean copulas. *J. Multivariate Anal.* **100** 1521–1537. MR2514145 https://doi.org/10.1016/j.jmva.2008.12.015

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

DE HAAN, L., MERCADIER, C. and ZHOU, C. (2016). Adapting extreme value statistics to financial time series: Dealing with bias and serial dependence. *Finance Stoch.* **20** 321–354. MR3479324 https://doi.org/10.1007/s00780-015-0287-6

DE HAAN, L. and RESNICK, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrsch. Verw. Gebiete* **40** 317–337. MR0478290 https://doi.org/10.1007/BF00533086

DOMBRY, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* **21** 420–436. MR3322325 https://doi.org/10.3150/13-BEJ573

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

DREES, H., DE HAAN, L. and RESNICK, S. (2000). How to make a Hill plot. *Ann. Statist.* **28** 254–274. MR1762911 https://doi.org/10.1214/aos/1016120372

DREES, H. and ROOTZÉN, H. (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Statist.* **38** 2145–2186. MR2676886 https://doi.org/10.1214/09-AOS788

DUDLEY, R. M. (2002). *Real Analysis and Probability. Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358 https://doi.org/10.1017/CBO9780511755347

EINMAHL, J. H. J., DE HAAN, L. and ZHOU, C. (2016). Statistics of heteroscedastic extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 31–51. MR3453645 https://doi.org/10.1111/rssb.12099

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

FOUGÈRES, A.-L., DE HAAN, L. and MERCADIER, C. (2015). Bias correction in multivariate extremes. *Ann. Statist.* **43** 903–934. MR3325714 https://doi.org/10.1214/14-AOS1305

GENEST, C. and SEGERS, J. (2009). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist.* **37** 2990–3022. MR2541453 https://doi.org/10.1214/08-AOS672

GENEST, C. and SEGERS, J. (2010). On the covariance of the asymptotic empirical copula process. *J. Multivariate Anal.* **101** 1837–1845. MR2651959 https://doi.org/10.1016/j.jmva.2010.03.018

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

GUMBEL, E. J. (1958). *Statistics of Extremes*. Columbia Univ. Press, New York. MR0096342

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

HUANG, X. (1992). Statistics of bivariate extreme values. Ph.D. thesis, Tinbergen Institute Research Series, The Netherlands.

HÜSLER, J. (1990). Multivariate extreme values in stationary random sequences. *Stochastic Process. Appl.* **35** 99–108. MR1062586 https://doi.org/10.1016/0304-4149(90)90125-C

NAVEAU, P., GUILLOU, A., COOLEY, D. and DIEBOLT, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika* **96** 1–17. MR2482131 https://doi.org/10.1093/biomet/asp001

NORTHROP, P. J. (2015). An efficient semiparametric maxima estimator of the extremal index. *Extremes* **18** 585–603. MR3418769 https://doi.org/10.1007/s10687-015-0221-5

PICKANDS, J. III (1981). Multivariate extreme value distributions. *Bull. Int. Stat. Inst.* **49** 859–878, 894–902. MR0820979

RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes. Applied Probability. A Series of the Applied Probability Trust* **4**. Springer, New York. MR0900810 https://doi.org/10.1007/978-0-387-75953-1

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. MR2540849 https://doi.org/10.1214/08-EJS345

SCHMIDT, R. and STADTMÜLLER, U. (2006). Non-parametric estimation of tail dependence. *Scand. J. Stat.* **33** 307–335. MR2279645 https://doi.org/10.1111/j.1467-9469.2005.00483.x

SEGERS, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli* **18** 764–782. MR2948900 https://doi.org/10.3150/11-BEJ387

SMITH, R. L. and WEISSMAN, I. (1994). Estimating the extremal index. *J. Roy. Statist. Soc. Ser. B* **56** 515–528. MR1278224

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics. Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

ZOU, N., VOLGUSHEV, S. and BÜCHER, A. (2021). Supplement to “Multiple block sizes and overlapping blocks for multivariate time series extremes.” https://doi.org/10.1214/20-AOS1957SUPP